AUDITORY PSYCHOPHYSICS IN BIRDS: THE EFFECTS OF UNIQUE NOISE ON SENSITIVITY

by

A. J. Taylor

Submitted for the degree of

Doctor of Philosophy
in
Psychology

at the Victoria University of Wellington

Wellington, New Zealand, 1984

I hereby declare that I am the sole author of this thesis.

I authorize Victoria University of Wellington to lend this thesis to other institutions or individuals for the purpose of scholarly research.

A. J. Taylor

I further authorize Victoria University of Wellington to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

A. J. Taylor

ABSTRACT

The performances of observers in auditory experiments are likely to be affected by extraneous noise from physiological or neurological sources and also by decision noise. Attempts have been made to measure the characteristics of this noise, in particular its level relative to that of masking noise provided by the experimenter. This study investigated an alternative approach, a method of analysis which seeks to reduce the effects of extraneous noise on measures derived from experimental data.

Group-Operating-Characteristic (GOC) analysis was described by Watson (1963) and investigated by Boven (1976). Boven distinguished between common and unique noise. GOC analysis seeks to reduce the effects of unique noise.

In the analysis, ratings of the same stimulus on different occasions are summed. The cumulative frequency distributions of the resulting variable define a GOC curve. This curve is analogous to an ROC curve, but since the effects of unique noise tend to be averaged out during the summation, the GOC is less influenced by extraneous noise. The amount of improvement depends on the relative variance of the unique and common noise ($\underline{k}$). Higher levels of unique noise lead to greater improvement.

In this study four frequency discrimination experiments were carried out with pigeons as observers, using a three-key operant procedure. In other experiments, computer-simulated observers were used.

The first two pigeon experiments, and the simulations, were based on known distributions of common noise. The ROCs for the constructed distributions provided a standard with which the GOC curve could be compared. In all cases the analysis led to improvements in the measures

of performance and increased the match of the experimental results and the ideal ROC.

The amount of improvement, as well as reflecting the level of unique noise, depended on the number of response categories. With smaller numbers of categories, improvement was reduced and k was underestimated. Since the pigeon observers made only "yes" or "no" responses, the results for the pigeon experiments were compared with the results of simulations with known distributions in order to obtain more accurate estimates of k.

The third and fourth pigeon experiments involved frequency discrimination tasks with a standard of 450 Hz and comparison frequencies of 500, 600, 700, 800 and 900 Hz, and 650 Hz, respectively. With the multiple comparison frequencies the results were very variable. This was due to the small number of trials for each frequency and the small number of replications. The results obtained with one comparison frequency were more orderly but, like those of the previous experiment, were impossible to distinguish from those which would be expected if there was no common noise.

A final set of experiments was based on a hardware simulation. Signals first used in the fourth pigeon experiment were processed by a system made up of a filter, a zero-axis crossing detector and a simulated observer. The results of these experiments were compatible with the possibility that the amount of unique noise in the pigeon experiments overwhelmed any evidence of common noise.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr John Whitmore and Mr Stuart Slater, for the help and support they have given me throughout this study. Dr Whitmore undertook most of the supervision, and gave me considerable help with acoustics, instrumentation and programming as each stage of the research was set up, in a new laboratory, often with new equipment. Among other things, he developed the systems and programs used for signal analysis and for the output of the digital signals. Later, he took the photographs included in the report. I have very much appreciated the moral support he has provided throughout. Mr Slater undertook an advisory role at a number of stages of the research and made extensive comments on two drafts of the report. He also eased administrative matters on a number of occasions and provided several useful utility programs.

Thanks also to Mr John Bowden, of the Department of Psychology workshop, who designed and built the apparatus used to control events in the pigeon chamber and to monitor the birds' responses. All the acoustic apparatus not given a maker's name in the report was produced in the Department's workshop.

I would also like to acknowledge the help and friendly cooperation of my fellow-student, Dianne Tomlinson, in the matter of obtaining and looking after the birds used in the research (and some that weren't used).

Thanks are also due to Dave Atmore of the Computer Services Centre for help with SYSPUB, used to produce this report, the Department of Psychology for equipment and computing support, and the Internal Research Committee for grants 18/79 and 133/79.

Finally, thanks to my wife Sue, who made it possible for me to do this work, and encouraged and helped me throughout, and to Beth, who takes a patient interest in what she first knew as "Dad's fesis".

Alan Taylor

CONTENTS

Appendix page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

UNIQUE NOISE AND GOC ANALYSIS

> On an operational level, internal noise is equivalent to the observation that the same physical stimulus may elicit different responses (Green, 1964, p. 397).

Internal Noise

Green and Luce (1974) suggest that "perhaps the single most pervasive characteristic of psychophysical data is the inconsistency of subjects when answering most questions we ask them about simple stimuli" (p. 373). The inconsistency may be attributed to noise within the observer. This internal noise degrades, or places a limit on, the observer's psychophysical performance.

Internal noise is sometimes equated with biological noise of the sort described by Soderquist and Lindsey (1972). This may be neural (e.g., random firings in the nervous system), physiological (the noise of activity of various systems, such as peristalsis, breathing and heartbeats) or aural (e.g., noise due to mechanical vibrations of the external and middle ear). These sorts of noises can affect the sensitivity of observers to external signals. Soderquist and Lindsey showed that in a signal detection task d' could vary from 0.5 to 1.5, depending on the relation between the signal presentation and the stage of the heartbeat cycle. Shaw and Piercy (1962) had earlier measured low frequency noise in an enclosure covering the ear. The overall noise level, and the levels in certain frequency bands, rose and fell in accordance with heartbeat, indicating the physiological origin of the noise.

The concept of internal noise has become particularly important in studies of binaural hearing phenomena. In research which led to an account of many earlier findings, Diercks and Jeffress (1962) showed that when tones at each ear were out of phase, the binaural threshold was lower than it was when the tones were identical. Diercks and Jeffress viewed the thresholds they obtained as being masked rather than absolute, with internal or "self"-noise as the masker. Their results, and those of the earlier experiments (e.g., Hirsh, 1948), could then be explained by assuming that the internal noise was made up of three components, one unique to each ear and the other common to both. Subsequent investigators of so-called masking level differences (MLDs) have incorporated this sort of internal noise into their theoretical accounts (e.g., McFadden, 1968; Osman, 1971). Wilbanks and Whitmore (1968), who investigated the detection of monaural signals as a function of interaural noise correlation, used the notion of internal noise in an attempt to explain the relatively small MLDs they obtained below 250 Hz. One possibility was that uncorrelated noise in each ear canal reduced the interaural noise correlation, an important determinant of MLDs. However, given the results of Shaw and Piercy (1962), body noise did not appear to provide a full explanation, and the writers suggested that neural noise also played some part.

Attempts have been made to measure the magnitude and other characteristics of internal noise in monaural experiments. In one method (Green, 1964; Spiegel & Green, 1981; Gilkey, Hanna & Robinson, 1981), the consistency of responding on trials based on identical signal-plus-noise or noise-alone stimuli is used to assess the relative variability of internal and external noise. If the effect of the external noise (the reproduced masker) is large compared with that of the internal noise, decisions will tend to be consistent over repetitions of the trials. If, on the other hand, the effect of external noise is relatively small, the decisions will be less consistent. Spiegel and Green, and Green, concluded that the ratio of the standard deviations of the internal and external noise (σ_I/σ_E) was about unity. Spiegel and Green also employed a method, first used by Siegel (1979), which estimated σ_I/σ_E from performance in a two-interval forced-choice task with identical or independent maskers in the two intervals. With independent maskers, both external and internal noise were assumed to contribute to the overall variability, whereas with identical maskers the variability was assumed to be due only to internal noise. Performance in the latter case was therefore expected to be superior. The smaller the difference between performance with identical and independent maskers the greater the internal noise variability. With this method the ratio of the standard deviations of the internal and external noise was again estimated to be around unity. Earlier work by Swets, Shipley, McKey and Green (1959) and Watson (1963) had also suggested a σ_I/σ_E value of around one. Swets et al. obtained this value even when different levels of masking were used, suggesting that internal noise was proportional to external noise.

In other studies, the concept of internal noise has been used in attempting to explain differences between theoretical predictions and obtained results, or incorporated in theories of monaural detection and descrimination (e.g., de Boer, 1966; Eijkman, Thijssen & Vendrik, 1966; Henning, 1967a, 1967b; Pfafflin & Mathews, 1962; Raab & Goldberg, 1975; Swets, 1961; Thijssen & Vendrik, 1968).

The possible dangers involved in using internal noise to account for experimental results in an ad hoc way were pointed out by Green (1960). He suggested that:

> If the concept is to have any importance, it must be made specific. This implies that we have to (1) state exactly what this noise is, i.e., that we have to characterize it

> mathematically, (2) specify in what way it interacts with the detection or discrimination process, and (3) evaluate specifically what effect it will have on performance (p. 1202).

Considering the diverse sources of such noise, these are very difficult requirements to meet.

The Group-Operating-Characteristic (GOC) technique (Watson 1963; Boven 1976), which is examined in this thesis, provides an approach to the problem of internal noise which is rather different from those described above. In other approaches the aim is to measure the characteristics of internal noise (especially its magnitude) or to incorporate them in theoretical accounts of hearing. The most important aim of GOC analysis, on the other hand, is to remove the effects of internal noise from measures of sensitivity. GOC analysis achieves this by producing a curve which is analogous to the traditional Receiver-Operating-Characteristic (ROC) curve but whose shape is less affected by the characteristics of internal noise. The technique is important because it provides a way of studying and testing theories about hearing which avoids the dangers involved in the ad hoc use of the concept of internal noise and also the difficulties of specifying its characteristics in the way outlined by Green.

Unique Noise

Internal noise has been defined in various ways, and identified with a number of different noise sources. In order to provide a firm basis for the discussion of GOC analysis, Boven distinguished amongst various sorts of noise as follows:

> Internal noise is a term which refers to noise which is coming from a source or sources within the observer and, in principle, could be specified. Individual noise is noise which is associated with an individual observer and which may or may not be internal and may or may not be unique to that observer. The individual noise includes any criterion variance of the observer and any effect of sequential dependencies. Unique noise ... is a statistical concept which

> refers to the idiosyncratic component of the total noise variance of an observer. The complement of unique noise is common noise. Since the common noise is of interest to the experimenter he will be interested in removing the effect of unique noise from his data (p. 5).

An important point of Boven's definition is that unique noise is a statistical concept and, unlike internal noise, is not identified with any particular source. Another important point is that the level of unique noise may be affected by decision noise arising from criterion variability, faulty memory (Swets, 1961) and inattention, as well as by biological noise.

Boven wanted to remove the effects of the unique noise associated with a group of observers. In the present research, the main aim was to remove the effects of unique noise on measures of individual performance. The situation for a single observer may be illustrated by considering an experiment consisting of signal-plus-noise and noise-alone trials. On each trial the observer makes a decision based on noise which is the sum of two independent noise sources (the two sources of noise are assumed here and throughout to be independent, so the variance of the overall noise is equal to the simple sum of the variances of the common and unique noise). The sources of the first noise are the signal and the masking noise, both provided by the experimenter, and any effects, such as sequential dependencies, which are correlated with these. The second noise arises from biological sources or is due to inconsistent decision-making (e.g., criterion variability, faulty memory or inattention), or from sources outside the observer, such as environmental noise which is not correlated with the experimental masking noise. If the first noise is repeated on a number of occasions it will be common to all the trials, while most of the components of the second noise will be unique to each trial. This situation is shown schematically in Figure 1. Exactly which components of the second noise are common and which are unique will depend on whether one or several observers are being considered. For example, the sensory system of an observer may respond to a masking noise in a way which is unique to him in relation to other observers, but common across repeated trials of that masking noise with that observer. The variability of an observer's responses over a number of repetitions of the experimental noise would be expected to tell us something about the relative variability of the unique and common noise. If the ratio of the variances of the two noises (σ_U^2/σ_C^2) denoted _k_, is equal to zero, the observer's judgements would be expected to be the same on each repetition of the trial. If, on the other hand, _k_ is very large (i.e., there is no common noise, or relatively little), the judgements on separate presentations should be independent, or nearly so. However, as Boven points out for the case of multiple observers, there may be some evidence of common noise even when independent maskers are used for each observer. This lack of independence may be due to sequential effects which are common to observers.

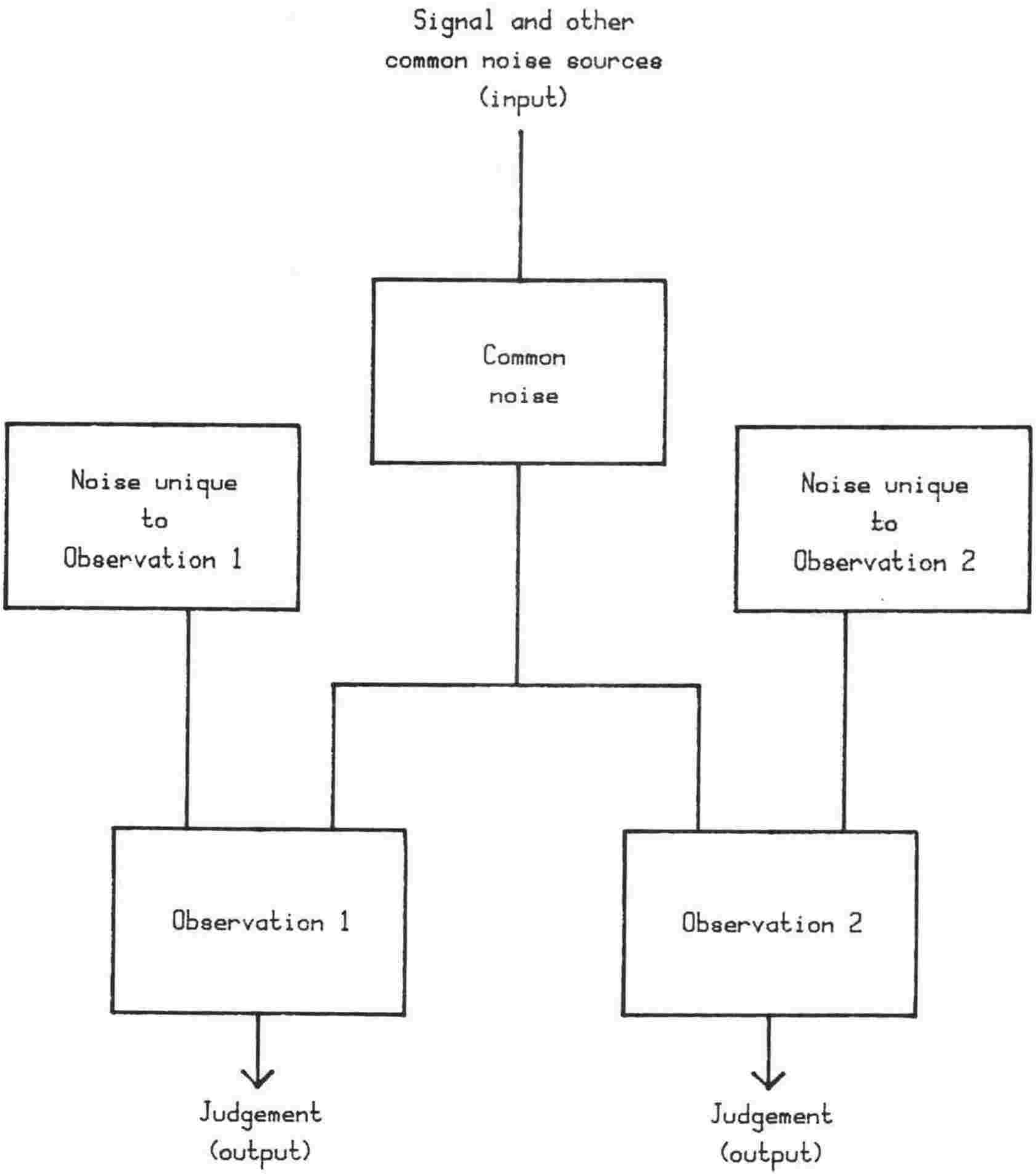


Figure 1. Block diagram showing how common and unique noise enter into judgements about the same signal on two different trials (after Boven, 1976).

GOC Analysis

The primary aim of GOC analysis is to reduce the effects of unique noise on measures of performance. With a single observer, this is achieved by combining the observer's judgements over one or more repetitions of a set of trials in order to average out the effects of the unique noise.

If the variance of the unique noise is relatively great (i.e., if k is large), GOC analysis should lead to significant improvement in measures of performance, but if k is relatively small, GOC analysis will produce little improvement. The relationship between k and the amount of improvement gives rise to a secondary aim of GOC analysis, which is to estimate the relative variances of the unique and common noise. The amount of improvement expected for different values of k will be discussed following a description of how GOC curves are produced.

In the Theory of Signal Detectability (TSD), the ROC curve shows the performance of a single observer. As utilised by Watson (1963), Jeffress and Gaston (c. 1967; Whitmore, personal communication) and Boven (1976), the GOC curve shows the performance of a group of observers. In the present research, most of the GOC curves were based on the judgements of one observer about the same signals on a number of occasions. The construction of the GOC curve will be described for this case.

For each observer a set of M reproducible signals is created. A replication consists of one presentation of each of these signals, making M trials. A number of replications, up to N, is run for each observer.

The results can be set out in a trial-by-replication matrix, as in the top part of Figure 2. The observer's responses, in this case "1" for "yes" and "2" for "no", are added across replications to create a new random variable, X_n. This variable is a sum of ratings, with a maximum value equal to the number of rating-scale categories times the number of replications. For example, if there are six replications, X_n could range between 6 and 12.

The next step is to create a stimulus-response (S-R) matrix in which the stimulus categories are values of X_n, as shown in the bottom part of Figure 2. Since this matrix is exactly analogous to a S-R matrix for a single observer, subsequent analysis follows conventional lines. The probabilities $P(x|SN)$ and $P(x|N)$ are calculated for each value of X, and then accumulated, starting from the smallest values of X. The resulting values of $P(x|SN)$ are plotted over the corresponding values of $P(x|N)$ in a conventional unit-square ROC space. The number of points defining the GOC curve is equal to the number of replications combined in the GOC analysis.

	Trial type	Replication number						Sum of ratings
		1	2	. . .	n	. . . . . .	N	
1	Std	1	1		2		1	5
2	Comp	2	2		2		2	8
.								.
.								.
.								.
.								.
.								.
m	Std	1	2		2		1	6
.								.
.								.
.								.
.								.
.								.
.								.
.								.
.								.
M	Comp	2	2		2		1	7

Trial type	N	.	.	N x number of response categories
Standard				
Comparison				

Figure 2. Trial-by-replication matrix showing how ratings are summed in GOC analysis (top), and a stimulus-response matrix for the resulting response categories (bottom).

The Theory of Multiple Observations

This section returns to the question of the amount of improvement which can be expected when information from different observations is combined. It looks at how the effects of unique noise are reduced during combination. The material is based mainly on Green and Swets (1966), Swets, Shipley, McKey and Green (1959) and Boven (1976).

Green and Swets (1966) considered two general ways in which multiple observations may be combined. In the first, an overall decision is a function solely of individual decisions made on individual observations. In the second, which is applicable to GOC analysis, information obtained from several observations is pooled and the accumulated evidence is used to reach a decision. In order to derive predictions from the second model, Green and Swets made three assumptions:

1. the observations are combined with no loss of information;
2. individual observations are independent (i.e., <u>k</u> approaches infinity), and
3. each observation is sampled from underlying distributions which are normal with equal variance.

The third assumption allowed Green and Swets to use d' as a measure of performance.

With these assumptions, the level of performance resulting from the combination of n observations, d'_n, is given by:

$$d'_n = \left[\sum_{i=1}^{n} (d'_i)^2\right]^{1/2} , \qquad (1)$$

where d'_i is d' for the ith observation.

Each replication combined in GOC analysis can be treated as if it were one observation (instead of a series of observations). Then the total number of observations is equal to N, the number of replications. If d' is constant over observations, equation 1 predicts that adding the results of N replications together as described in the previous section will increase the measure of performance to a level $(N)^{1/2}$ times that observed on individual replications.

In practice, one or more of the assumptions made by Green and Swets may not be met. Contrary to the first assumption, information will probably be lost, although this may be due to the nature of the observations rather than to the way they are combined. When an observer uses a rating scale with a finite number of categories, numbers of stimulus elements are mapped onto single response elements. In this kind of partial identification experiment (Bush, Galanter & Luce, 1963), knowledge of the judgements of an observer leads to only limited knowledge of the stimulus, even when the observer is performing perfectly. The judgements therefore involve information reduction (Coombs, Dawes & Tversky, 1970) and it is in this sense that information is lost when rating scales are used.

The assumption that the underlying distributions are normal is probably unjustified in many cases--for example, when the amplitude of narrow-band noise is the basis for the observer's performance (Jeffress, 1964).

Finally, there will usually be some dependence among observations because of the presence of common noise. The greater the dependence among observations, the smaller the improvement that would be expected to result from combining observations.

Swets et al. (1959) have shown the amount of improvement to be expected for a given ratio of unique to common noise variance (k). Their derivation was originally applied with observers who each made a number of observations of a signal-plus-noise or noise-alone. They found that when the masking noise on each of the five trials (presented successively) was independent, improvement in the performance over presentations was in accord with the predictions of equation 1 above. However, when an identical masking noise was used for each observation--that is, when the ratio of unique to common noise variance was reduced--the improvement was less marked, as would be expected from the relations given below.

The Swets et al. derivation will not be given in full here, but several important results will be discussed.[1] The first of these gives d'_n, the value of d' for n observations (or replications):

$$d'_n = \frac{nm}{(n^2\sigma_C^2 + n\sigma_U^2)^{1/2}} \qquad (2)$$

where n is the number of observations or replications,

m is the contribution of the signal to the evidence variable (the difference between the means of the signal-plus-noise and the noise-alone distributions),

σ_C^2 is the variance of the common noise, and

σ_U^2 is the variance of the unique noise.

This equation shows that as the number of observations increases, the relative contribution of σ_U^2, the unique noise variance, decreases. Thus, with $\sigma_C^2 = \sigma_U^2 = 1$, the initial ratio of $n\sigma_U^2/n^2\sigma_C^2$ is equal to 1, whereas after five observations have been combined the same ratio is equal to 0.20.

---

[1] Some of the terms in the Swets et al. equations have been altered to fit in with the terminology used in this thesis.

A second important result gives the amount of improvement for a particular value of k:

$$\frac{d'_n}{d'_1} = \frac{(1 + \underline{k})^{1/2}}{(1 + \underline{k}/n)^{1/2}} \qquad (3)$$

where d'_1 is the initial value of d'.

Equation 3 can be solved to provide a way of estimating k from d' values obtained before and after GOC analysis:

$$\underline{k} = \frac{n[(d'_n)^2 - (d'_1)^2]}{n(d'_1)^2 - (d'_n)^2} \quad . \qquad (4)$$

This result is analogous to that obtained by Watson (1963) for multiple observers.

Boven extended this derivation to cover the case where the normal underlying distributions of signal-plus-noise and noise-alone do not have equal variances. He used the index d_z (Jeffress, 1967), which is given by:

$$d_z = \frac{(2)^{1/2}m}{(\sigma_S^2 + \sigma_N^2)^{1/2}} \qquad (5)$$

where σ_S^2 and σ_N^2 are the variances of the signal-plus-noise and noise-alone distributions respectively, and

m is the difference between the means of the two distributions.

Boven showed that the ratio $d_{z(n)}/d_{z(1)}$, of the d_z value for n observers (or replications) to that for an individual observer (or replication), is equal to the ratio for the equivalent d' values (equation 3). Thus

d_z can be substituted for the d' values in equation 4 in order to estimate k.

Areas of Investigation

The overall aim of the work reported here was to evaluate the effectiveness of the GOC technique in reducing the effects of unique noise on measures of sensitivity. In describing the work carried out as part of this general aim, the report covers four main topics. The first is the effectiveness of GOC analysis when the underlying distributions--and the corresponding ROC curve--are relatively complex. The second topic concerns the estimation of the magnitudes of unique and common noise from the results of GOC analyses. The third topic is the application of the GOC technique to the results of a psychophysical experiment, while the fourth is an illustration of how GOC analysis and simulation might be used in the investigation of hearing. These areas will be described in more detail.

The first topic arises from Boven's finding that GOC analyses were less effective in recovering the ROC curve of known underlying distributions when the ROC was relatively complex. In his experiments, relatively simple ROC curves were obtained for distributions which were analogues of normal unequal variance distributions or which were rectangular. The distributions underlying the complex ROC were multimodal. Boven suggested that GOC analysis would be more successful in recovering complex ROC curves if greater numbers of trials per replication were used. The present study took up the question of whether an increase in the number of replications would be as effective as increasing the number of trials per replication. It also considered the relative effectiveness of different numbers of rating-scale categories in recovering complex ROC curves.

The number of response categories used by the observer is central to the second topic. As pointed out earlier, the Green and Swets (1966) and Swets et al. (1959) assumption that the observations are combined with no loss of information is unlikely to hold when observers use rating scales with finite numbers of categories. In general, the smaller the number of categories, the greater the loss of information. The effect of this loss of information is to reduce the change in d' as replications are added in GOC analysis, and so to underestimate the value of $\underline{k}$ when using equation 4. Boven demonstrated this point with a simulation in which two observers used four-category rating scales. This point was followed up in the research reported here and an alternative method of estimating $\underline{k}$ was developed.

In the third area of investigation, GOC analysis was applied to the results of a psychophysical experiment in auditory frequency discrimination. There are two main aims in this sort of analysis. One is to obtain operating characteristics and measures of sensitivity relatively unaffected by unique noise. The other is to find out whether there is in fact any common noise. If noise which is faithfully reproduced on different occasions does not evoke the same response in the sensory system of the observer on each occasion, all the noise in the experiment will be unique, and the addition of more and more observations will lead to perfect performance. Common noise could exist for one type of task and not for another (e.g., for amplitude discrimination but not for frequency discrimination). Similarly, common noise may exist for reproducible noise presented repeatedly to the same individual but not for the same noise presented to different subjects. In principle, the GOC technique could be used to discover whether common noise exists in a given situation. In practice, however, the results of experiments in which there is no common noise may be difficult to distinguish from those in which the amount of unique noise is so great as to mask the effects of common noise. This matter is considered in detail in Chapters 7 and 8.

The fourth topic concerns one way in which the results of GOC analysis may be followed up. Once well-defined GOC curves have been obtained, hardware models of the hearing system of the observers may be used (perhaps with the reproducible signals originally presented to the observers) in an attempt to produce results similar to those arising from the GOC analysis. Correspondence of the outcomes for the observers and the simulation is not proof that the model used in the simulation is an accurate representation of the observers' hearing system. However, the use of GOC analysis should, given the presence of common noise, provide a more valid criterion against which to evaluate models than would be available otherwise. As Boven points out, the presence of unique noise will lead to an ROC curve which is closer to a normal equal variance curve than the curve which would arise from the underlying distributions of common noise in the absence of unique noise. Thus in the absence of GOC analysis, important differences between underlying distributions may be overlooked.

Methods of Investigation

Several methods were used to investigate GOC analysis. In four experiments, pigeons were used as observers. The difficulties reported by various investigators who have attempted to train these birds to make aural discriminations (e.g., Heise, 1953; Krasnegor, 1971; Stebbins, 1970b; Hienz, Sinnott & Sachs, 1977) suggested that pigeons were particularly appropriate subjects for the investigation of unique noise. This is discussed in Chapter 3 along with the apparatus and procedures used with the birds.

Other experiments reported here were simulations based on computer models. In most cases the simulations were entirely computer-based, with the values of both the common and unique noise being generated by software. These simulations are introduced in Chapter 4. In the simulation described in Chapter 9, however, signals previously presented to the pigeons were monitored and processed by a hardware system via a microphone, and the computer made decisions based on information supplied by the equipment. This is referred to as a hardware simulation.

In the first two pigeon experiments, discrete distributions of common noise constructed from sinusoids of varying frequencies were used. As discussed in Chapter 2, these distributions provided ideal ROC curves with which the GOC curves could be compared. Discrete distributions of common noise were also used in the computer simulations, along with virtually continuous distributions of both common and unique noise. In the hardware simulation, the characteristics of the common noise distributions, but not those of unique noise, could be measured directly. Finally, in the last two experiments with pigeons, described in Chapters 7 and 8, the nature of both the common and unique noise distributions was unknown.

Summary

Performance in psychophysical experiments is degraded by noise which arises from various sources. GOC analysis provides a way of reducing the effects of this noise on measures of performance. Models described by Swets et al. (1959), Watson (1963) and Boven (1976) relate improvements in measures of performance to the relative variances of unique and common noise (*k*).

The research reported here covers four main topics. These concern the applicability of GOC analysis to complex distributions, the estimation of k from the results of GOC analysis, the use of the GOC technique to analyse the results of psychophysical experiments in hearing, and the use of simulations in conjunction with GOC analysis.

Some of the experiments used pigeons, while others were based on simulated observers. In some experiments the distributions underlying the psychophysical tasks were known, while in others, in particular the psychophysical experiments with pigeons, their characteristics were unknown.

CHAPTER 2

THE USE OF KNOWN UNDERLYING DISTRIBUTIONS

Two of the pigeon experiments, and all of the simulations, made use of known underlying distributions. This chapter briefly discusses the advantages of using such distributions in evaluating the GOC technique, and describes how the distributions were constructed.

Problems with Unknown Distributions

TSD assumes that an observer makes decisions about events, such as the presence or absence of a signal, on the basis of an evidence variable. In general, the way in which the evidence variable is distributed for the signal and non-signal events determines the shape of the ROC curve which summarises the observer's performance. In most psychophysical experiments the nature of the underlying distributions is unknown. Furthermore, their characteristics cannot be inferred from the ROC since different distributions can give rise to the same ROC (e.g., Egan, 1975). Establishing whether a GOC curve more clearly reflects the nature of the underlying distributions than does the conventional ROC curve is therefore impossible.

A way round this problem is to start with known underlying distributions. The ROC for such distributions represents the performance of an observer entirely unaffected by unique noise, and provides a clear standard against which to evaluate the effects of GOC analysis. As replications are added in the analysis, the GOC curve ought to approximate the ROC more and more closely. The test of GOC analysis can be made all the more clear-cut by constructing underlying distributions shaped very differently from those which might reasonably be assumed for unique noise.

Construction of Distributions

The concept of the constructed distributions used in the present research has its origins in unpublished work by Whitmore (c. 1970, personal communication). The discrete distributions in the first two pigeon experiments and in some simulations were constructed in the same way as those used by Boven (1976) and Podd (1975).

Discrete Distributions

In the pigeon experiments the evidence variable was sinusoidal frequency; the probabilities of different frequencies defined the shapes of the distributions, which can be seen as probability mass functions. The triangular probability functions used in the first pigeon experiment are shown in Figure 3. These are discrete analogues of distributions described by Egan (1975, Appendix D). Twenty-six different frequencies were used, ranging from 400 Hz to 1000 Hz in 24-Hz steps. The means of the standard and comparison distributions were 560.7 Hz and 839.3 Hz respectively; the standard deviation was 128.1 Hz. The discrete distributions used in the simulations were constructed out of the same number of elements as those used with the pigeons.

In the first pigeon experiment, the birds were trained to adopt a particular cutoff--to make one response when sinusoids of 712 Hz or above were presented and another when a lower frequency occurred. As with the triangular distributions themselves, this method was chosen for the initial experiment in order to make the task as easy as possible. At that stage, the birds' ability to maintain more difficult discriminations for long periods was uncertain. The effect of the initial training was to establish frequency as the decision axis.

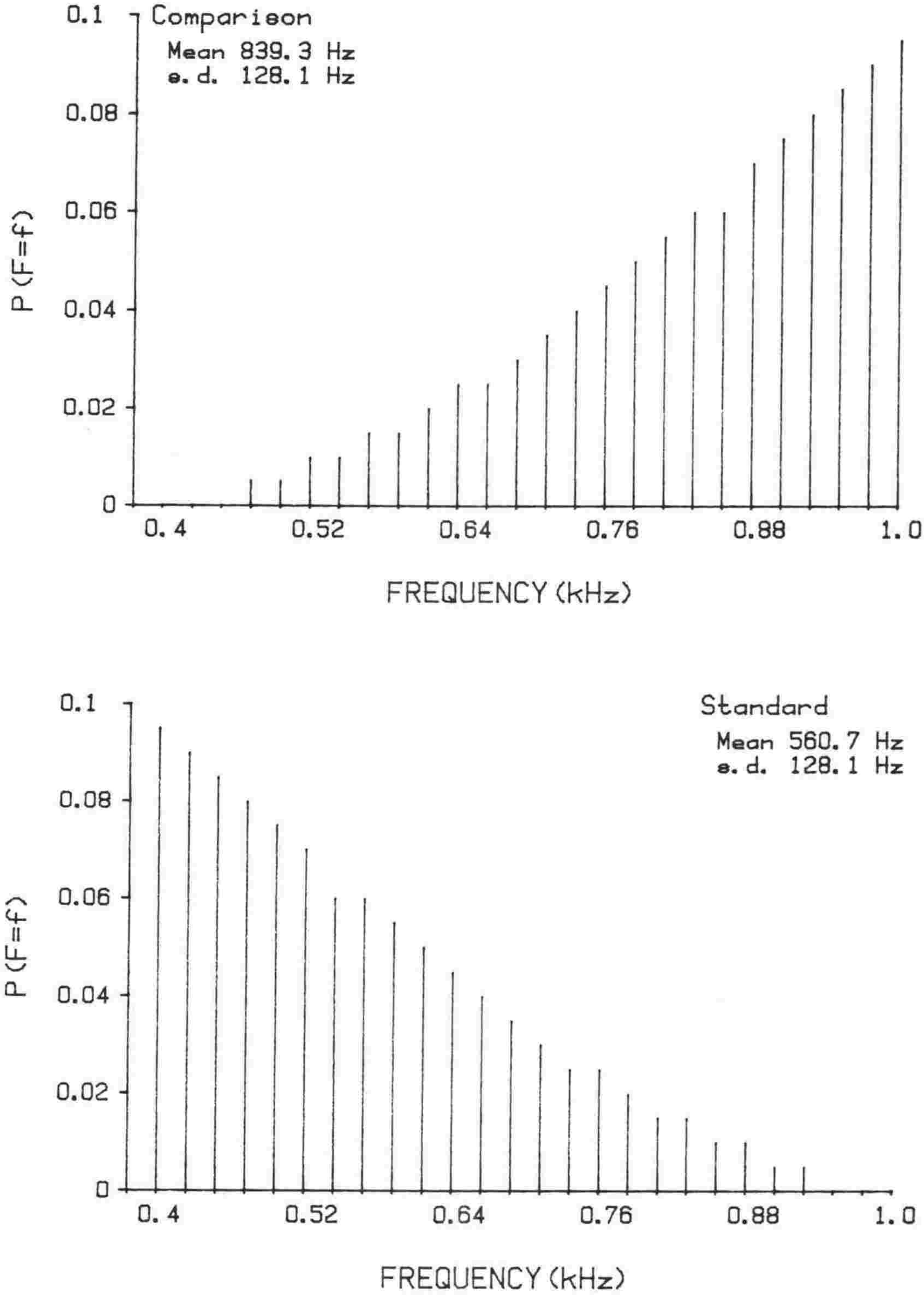


Figure 3. The triangular distributions used in the first pigeon experiment.

In the second experiment with pigeons, described in Chapter 6, different distributions were used, and the birds' responses were reinforced according to the distribution from which the current frequency had been sampled. This experiment was closer to a typical psychophysical experiment, in which the values of the evidence variable are not known to the experimenter, and the relation of responses to a precisely defined cutoff cannot be monitored.

Figure 4 shows the ROC for the triangular probability distributions, plotted on linear coordinates. As mentioned earlier, the performance of an observer unaffected by unique noise will fall somewhere on this curve. In examining the results for the pigeon and simulated observers, the observers' initial performances, as well as their GOCs, could be compared with these ideal ROCs.[2]

---

[2] Strictly, the ROC function consists of a finite number of points. Conventionally, successive points are connected by straight lines, and that custom is followed throughout the thesis.

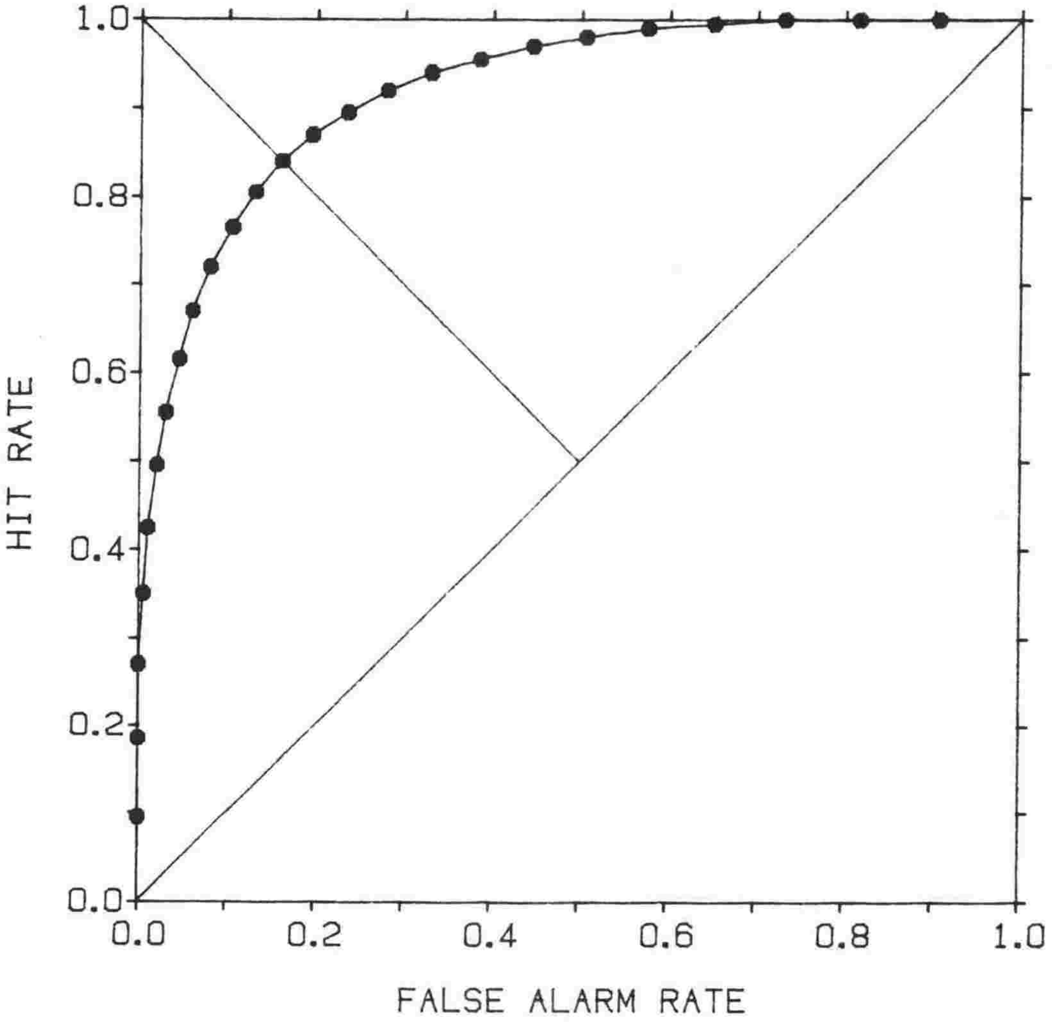


Figure 4. The ROC curve for the triangular distributions used in the first pigeon experiment and shown in Figure 3.

Continuous Distributions

Some of the distributions produced during the computer simulations are referred to as continuous since the only restriction on the value of the evidence variable was the floating point resolution of the computer.[3]

All the continuous distributions (both common and unique noise) used in the simulations were approximately normal, being generated by the Polar algorithm described by Knuth (1969). This algorithm, and the pseudo-random number generators used with it, are discussed in Chapter 4.

The means and standard deviations of the distributions were nominated for each simulation. The number of trials per replication and, in the case of unique noise, the number of replications, determined the number of sampled values making up each distribution.

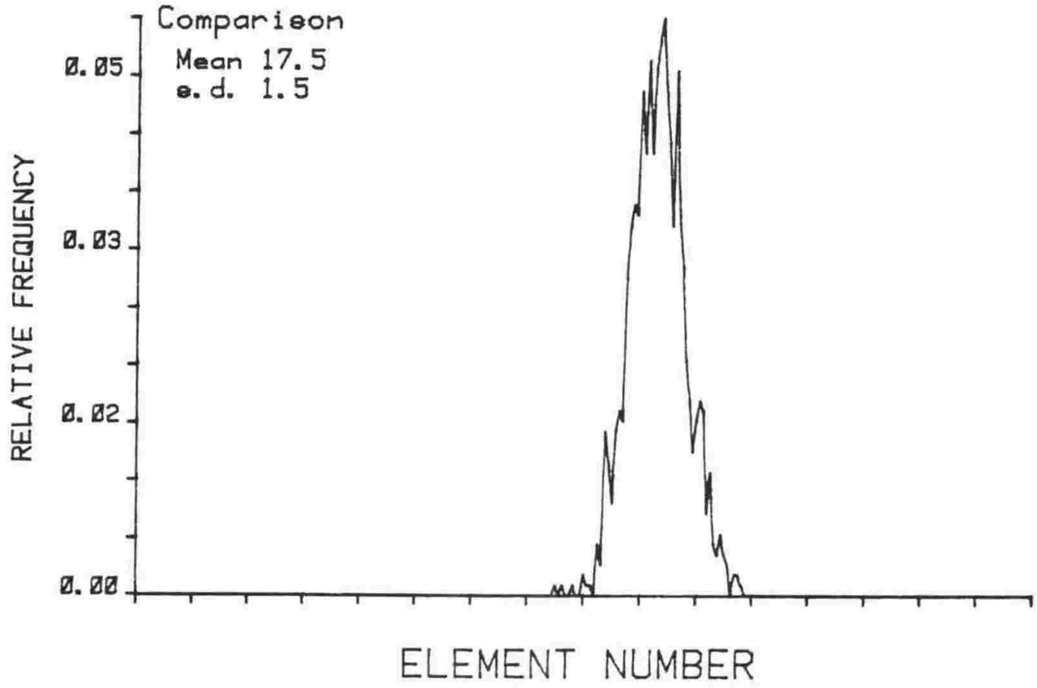
Figure 5 shows a pair of unequal variance distributions with means of 13.5 and 17.5 and standard deviations of 4.8 and 1.5 respectively. Each distribution was made up of 1000 samples (making 2000 trials per replication).[4] The corresponding ROC curve, based on a 25-category rating scale, is shown in Figure 6, plotted on both linear and normal-normal coordinates. The latter are scaled in terms of normal deviates.

The great advantage of the simulation was that distributions of both common and unique noise could be generated. Since the characteristics of all distributions were known (and in addition met the assumptions underlying the theory of Swets et al.) the effect of various

---

[3] The computer used in the simulations, a Hewlett-Packard 9826, stored all numbers internally with 12 significant digits in the mantissa and a two-digit exponent (±99).

[4] For the purposes of plotting the sample distributions arising from simulations (but not during the simulations), the values of the evidence variable were scaled onto a zero-to-256 scale. The range over which the evidence variable was scaled extended from five standard deviations below the mean of the standard distribution to five standard deviations above the mean of the comparison distribution.



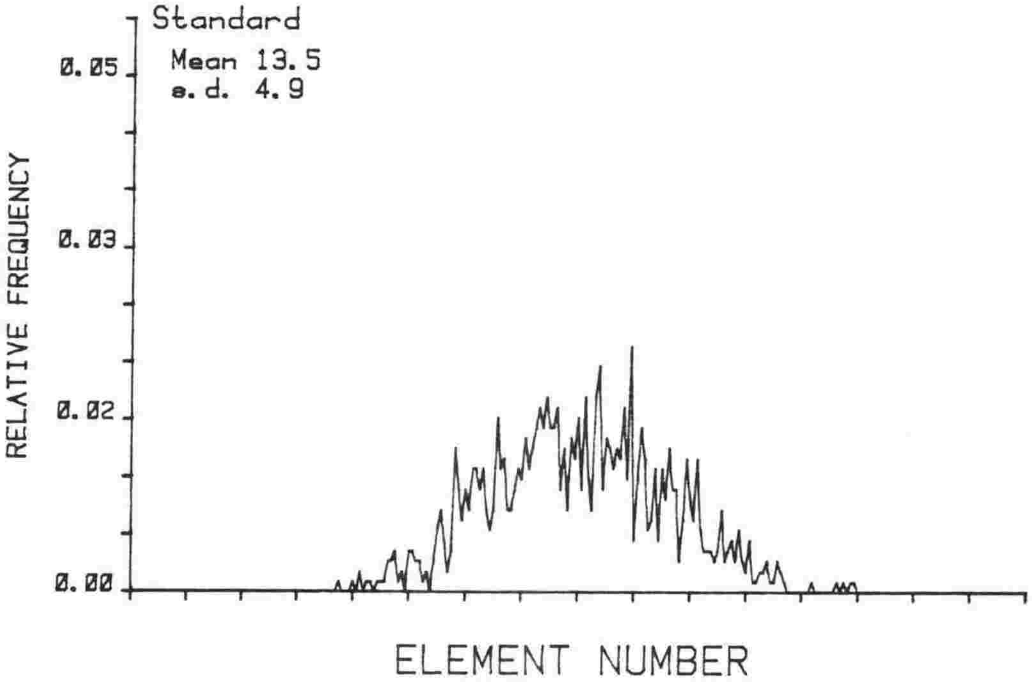


Figure 5. Samples drawn from normal unequal variance distributions produced during a simulation. The simulation contained 2000 trials.

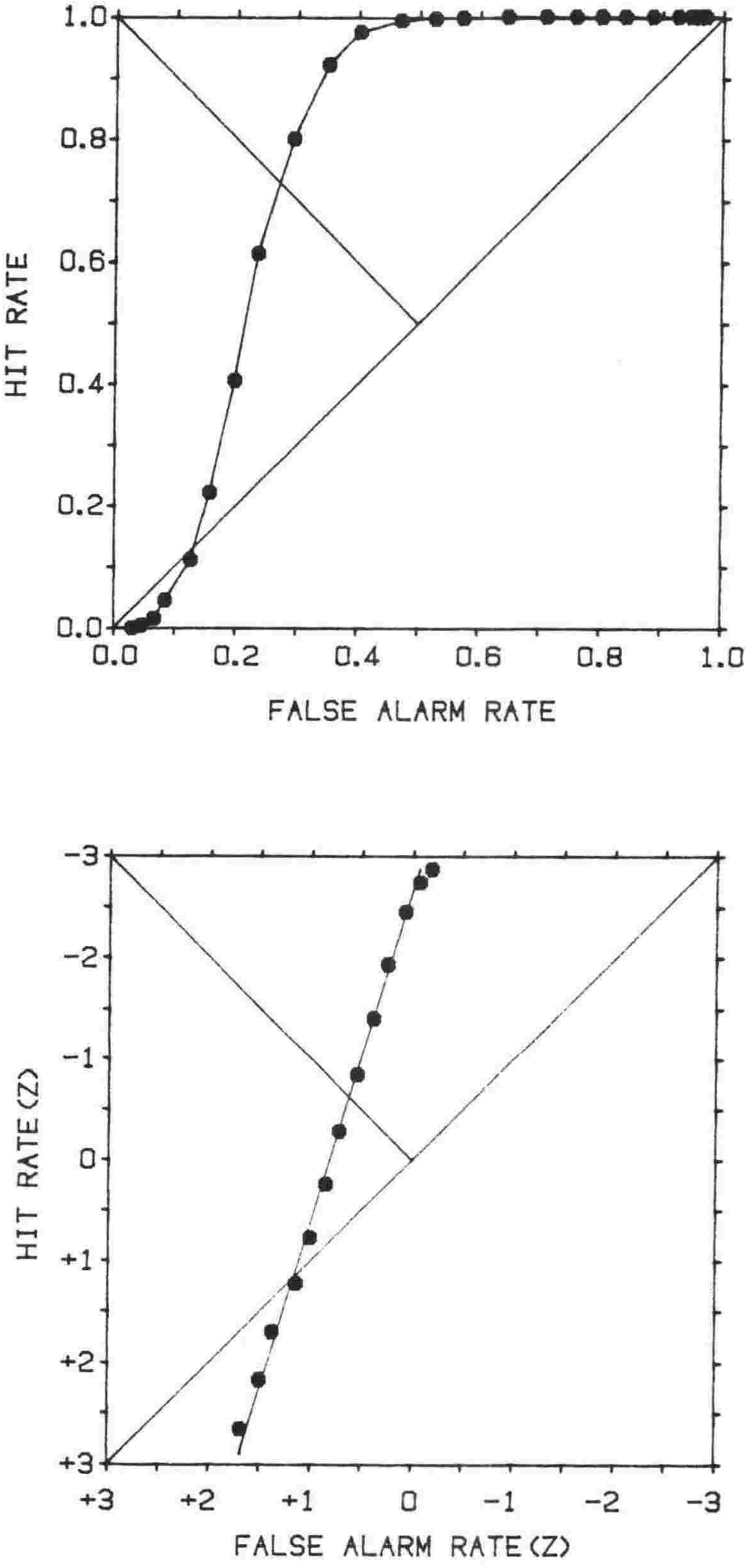


Figure 6. ROC curves for the samples from the normal unequal variance distributions produced during a simulation and shown in Figure 5. In the top graph, the coordinates are linear; in the bottom graph, normal-normal, scaled as z.

manipulations could be assessed unequivocally.

CHAPTER 3

THE RECOVERY OF KNOWN ROC CURVES FROM THE BINARY CHOICE DATA OF PIGEONS

The first part of this chapter describes the apparatus and procedures common to all four pigeon experiments, and the way sound was generated in the first two experiments. The second part of the chapter describes the results of the first experiment.

OBSERVERS, APPARATUS AND PROCEDURES USED IN ALL PIGEON EXPERIMENTS

Observers

Most of the data reported in this thesis were obtained from three racing pigeons, numbered 7, 8 and 18. Birds 7 and 8, a cock and a hen respectively, were three years old at the start of training. Bird 18, a hen, was about fifteen months old.

During initial training, the birds were maintained at 80% of their ad lib. weights. Over the course of experimentation their running weights were increased, so that for most of the duration of the experiments they weighed between 85% and 95% of their ad lib. weights. This had no effect on performance once the birds were well established on a task, and helped to keep the birds healthy over the long periods of experimentation.

In order to increase the number of trials per session without resorting to very short reinforcement times, a mixture of small seeds (mainly millet) was used as the reinforcer. Supplementary food consisted of wheat, maize and partridge peas, occasionally augmented by green vegetables (silver beet and cabbage), cod liver oil and vitamins A, D and C (Vitadol C).

Apparatus

Sound-Attenuated Room

The birds worked in a specially-constructed operant chamber which was housed in a sound-attenuated room. The room was approximately 4.8 m long by 2.4 m high and 3.4 m wide, had multiple walls (consisting of sheet steel, pine panelling, Gibralter Board, fibreglass and air spaces) and was lined with carpet and acoustic tiles. This lining had some absorbent effect at higher frequencies, but for frequencies of 1000 Hz and below, its sound-absorbing characteristics differed markedly from those which would be expected in a free field.[5]

The walls of the room had a good attenuating effect on airborne sound. The results of tests, given in Appendix A, showed that over the range of frequencies used in the experiments, airborne sound originating outside the room was attenuated by at least 45 dB (at around 400 Hz) and at most by almost 60 dB (at around 1000 Hz).

Operant Chamber

The operant chamber, which is pictured in Figures 7 and 8, consisted of a compartment surrounded by a shell of polyurethane foam, acoustic tile and aluminium sheet attached to a steel frame. At one end of the chamber there were three response-keys which could be lit from behind, and a hole giving access to a conventional pigeon feeder. At the other end of the chamber a 200 mm speaker was located outside a hole in the chamber's shell (Figure 7).

The chamber was supplied with a houselight in the ceiling and was ventilated by a whisper fan which produced very low-level noise located in one-third octave bands centred at 100 Hz and 125 Hz.

---

[5] This was measured by comparing the decrease in sound level as a microphone was moved away from a sound source with the decrease which would be expected in a free field (Sabine, 1957).

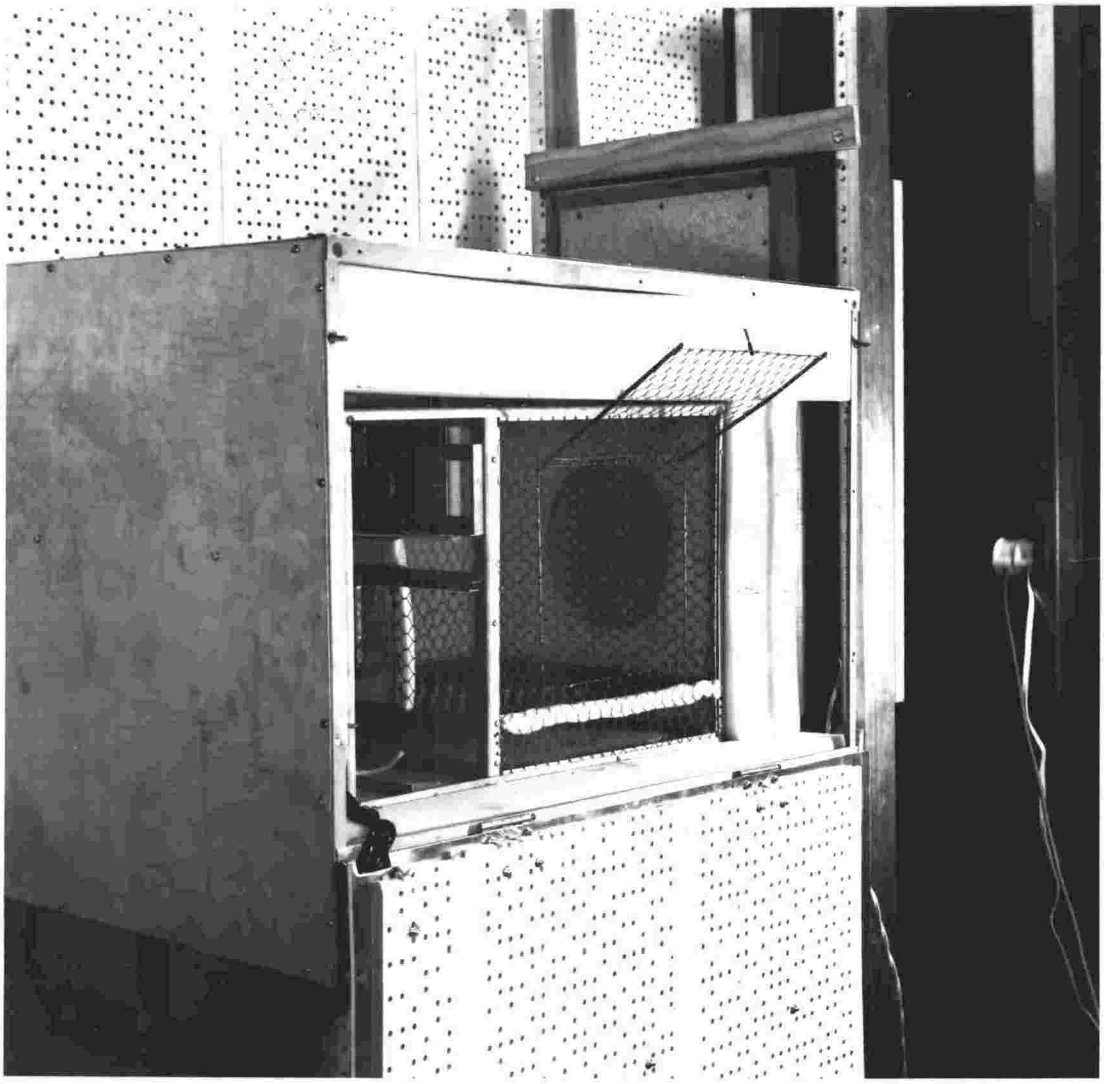


Figure 7. The operant chamber, looking toward the end at which the speaker was mounted.

Figure 8. Bird 8 pecking the centre key in the operant chamber. The houselight, side keys and feeder opening can also be seen.

A detailed description of the chamber is given in Appendix B.

Control of the Experiments

All events in the chamber were controlled by a Hewlett-Packard 9825 desktop computer via a 16-bit interface which controlled a specially-constructed relay box located in the experimental room. Responses were registered using an HP 3437A system voltmeter, which measured voltages associated with different response keys and relayed them to the computer.

Procedures

Trial Procedure

The behavioural procedure eventually used in all the pigeon experiments reported here[6] was based on that described by Krasnegor (1971) and Krasnegor and Hodos (1974).

A trial began with the lighting of the centre key. Ten pecks on this key turned the keylight off, and led to the illumination of the side-keys and the onset of the auditory signal. Pecks on the side-keys while the centre key was illuminated resulted in the cancellation of the trial.

Thirty (not necessarily consecutive) pecks on either side-key switched off the auditory signal and terminated the trial. If a comparison trial (higher frequency)[7] was presented, 30 pecks on the right-hand key ("yes") led to the illumination of the feeder and the raising of the food hopper. Thirty pecks on the left-hand key ("no")

---

[6] A number of training procedures were used in preliminary experiments, which were run over 66 sessions. The main procedure was one in which the birds were required to make one observing response to a centre key and one response to either of two side-keys. The birds showed no sign of learning the frequency discrimination task until the procedure reported here was introduced.

[7] These frequencies varied over experiments.

led to a timeout (TO), during which the houselight, which was on at all other times, was turned off. Thirty left-hand key pecks during a standard trial (lower or lowest frequency) led to food, while thirty right-hand key pecks resulted in a TO. Pecks on the centre key during side-key illumination were recorded but had no scheduled consequences. They were very infrequent.

Trials were separated by inter-trial intervals (ITIs). The length of each ITI was calculated by the computer, utilising its random number generator. Pecks during the ITIs produced a delay in the onset of the next trial if the scheduled delay was greater than the remainder of the ITI.

Each trial lasted 30 seconds. If the bird had not made 30 responses to one or the other key in this time, the trial was repeated after the next ITI. This happened only once or twice in the entire series of experiments.

The durations of reinforcement times, ITIs and TOs varied over birds and experiments, and are given later. Inter-trial intervals also varied from trial to trial between specified maximum and minimum values.

The number of trials varied from experiment to experiment, as did the number of warm-up trials. Trial sequences were generated by the computer, and were subject to restrictions on the number of consecutive standard or comparison trials. These restrictions varied over experiments.

Initial Training

In initial training, the birds were first magazine-trained, then auto-shaped to peck the lighted centre key. After a few reinforcements had been obtained for the pecks at the centre key and both side-keys, the trial sequence was introduced. The number of pecks required on each key, as well as the durations of ITIs and TOs, were gradually increased. Correction trials were used during all the preliminary training sessions, and during the first experiment. In later experiments, correction trials were not used. This will be described in greater detail for each experiment.

Running of Sessions

During both preliminary training and the experiments themselves, sessions were run six days a week, starting at 7 am. The pigeon experiments took almost two years to run (540 actual sessions). The numbers of sessions run for individual experiments are given later in the report.

SIGNAL GENERATION IN THE PIGEON EXPERIMENTS BASED ON KNOWN DISTRIBUTIONS

Apparatus

The equipment used to produce the auditory signals in the pigeon experiments based on constructed distributions is shown in Figure 9.

Sinusoids were provided by an HP 3312A voltage-controlled function generator. The frequency and amplitude of the generator's output was modulated by the computer via two 10-bit HP-IB 59501A D/A Power Supply Programmers. An electronic switch, also operated by the computer, controlled the onset and offset of the sinusoids, applying a linear ramp. A 40 msec rise-fall time was used. From the switch, the signals were fed into two Butterworth low-pass filters (Krohn-Hite 3550, each 24 dB/octave) set at 1000 Hz, in order to minimize harmonic distortion.

A continuous masking noise was produced by a noise generator the output of which had a flat spectrum (±1.5 dB) from 20 Hz to 20 kHz and followed the Gaussian function to a crest factor of at least 4.

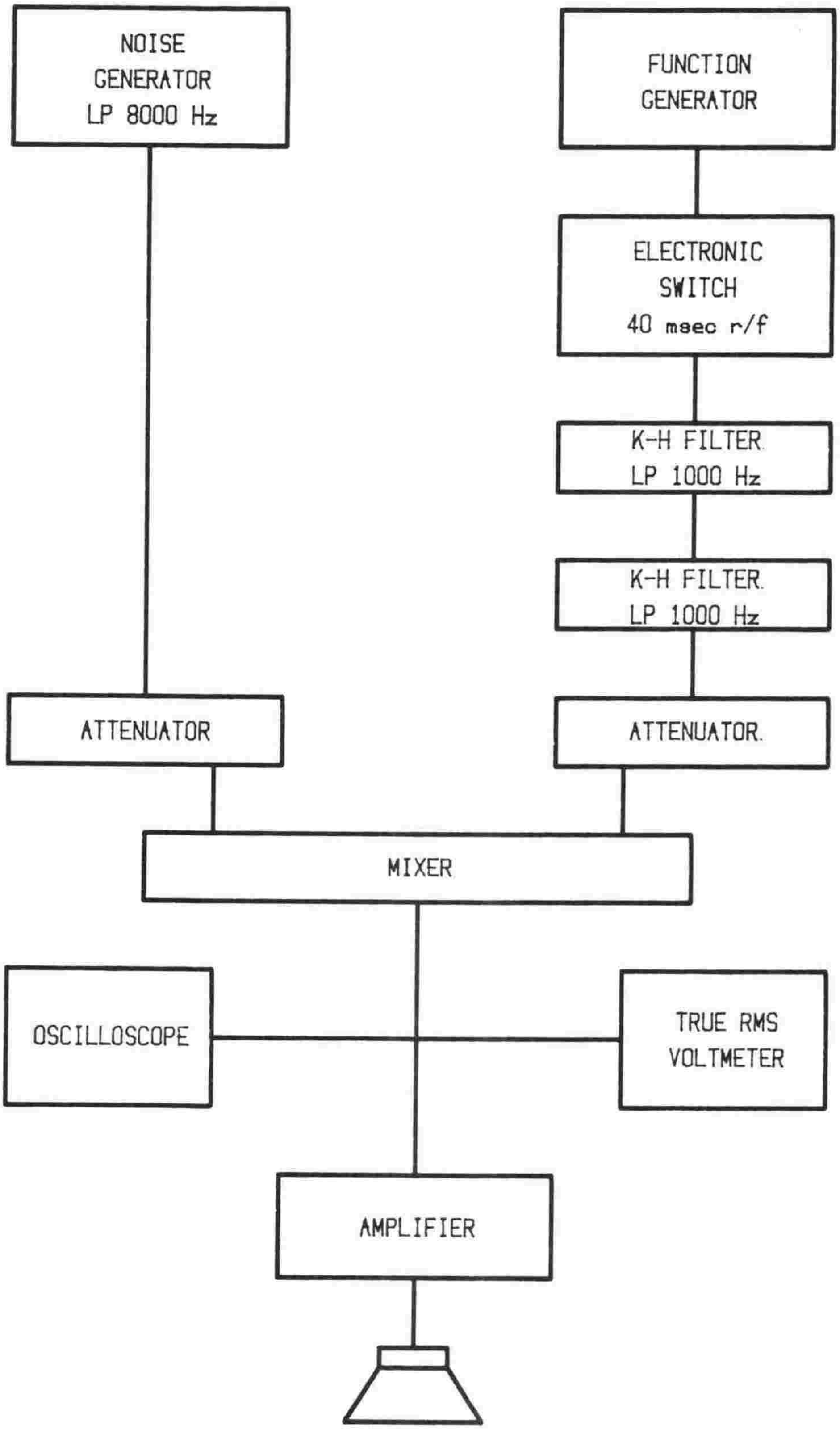


Figure 9. Block diagram of sound-generating equipment used in the first two pigeon experiments.

Both the sinusoids and the noise were fed into a mixer from Hatfield 2050 attenuators, and from there through the wall of the sound-attenuated room into a modified Lafayette model 1421 amplifier. The frequency response of the amplifier was flat from 1 Hz to 20 kHz and the harmonic distortion was 0.6% at 1000 Hz.

Signals

The amplitudes of the sinusoids in the chamber were set at 83 dB SPL.[8] This was measured by a Brüel & Kjaer half-inch (12.7 mm) Type 4134 condenser microphone. For these measurements the microphone was placed in one of two 15 mm diameter holes in the plate housing the response keys. It was set to protrude at the ear level of a freeze-dried pigeon[9] which was placed with its beak touching the centre key in the same attitude and at the same height as the birds used in the experiments. The two microphone holes were located 25 mm above the key slot and 25 mm either side of the centre line.

The codes used by the computer to control the D/A programmers (which in turn controlled the frequency and level of the signals) were determined by an iterative computer program. This program was run with the microphone in six different positions: the head of the microphone was placed at one of three distances from the plate housing the response keys (38, 51 and 64 mm), and measurements were made on both sides of the bird. The frequency and level of the signals were set with an accuracy of less than 1 Hz and 0.5 dB respectively. The mean values of six codes were used for each frequency.

---

[8] All sound levels are expressed in dB SPL relative to $20 \mu N/m^2$.

[9] This pigeon was prepared in the Zoology Department of Victoria University. The freeze drying process hardened the body of the pigeon while preserving the feather characteristics of a live bird.

Subsequent tests showed that the sound levels of the signals were very consistent over different frequencies and at different microphone locations. Systematic measurements at the six locations described above did not vary by more than ±1 dB from 83 dB. The mean level was found to be 82.9 dB, and the standard deviation was 1.3 dB. Less systematic measurements at various points in the chamber generally varied far less than the 15 to 16 dB that Krasnegor and Hodos (1974) suggested was necessary for pigeons to discriminate the intensities of 1000 Hz sinusoids.

A discrete Fourier transform (DFT) of the signals from the speaker showed that at none of the frequencies used was the harmonic distortion greater than 1%. This figure was seen as acceptable since it included distortion introduced by the speaker as well as by the amplifier and the rest of the sound-generating system.

The overall level of the continuous masking noise was 55 dB, and the spectrum level was approximately 16 dB. The signal-to-noise ratio was thus 67 dB.

The calibrations of the tones and the noise were checked periodically during the experiments and were found to be stable over time.

A NOTE ON PIGEONS AS OBSERVERS IN AUDITORY EXPERIMENTS

A number of experimenters have reported difficulties in training pigeons on auditory tasks. These tasks have included both detection (e.g., Heise, 1953; Stebbins, 1970b; Hienz, Sinnott & Sachs, 1977) and discrimination (e.g., Krasnegor, 1971). In experiments involving positive reinforcement, pigeons appear much less ready to attend to aural signals than to visual stimuli; this tendency is reversed when electric shock is used (e.g., Foree & LoLordo, 1973; Delius & Emmerton, 1978). Not surprisingly, measures of sensitivity for the pigeon (Figure

10) have shown considerable variability, although there is no evidence that shock always produces better results.

Heinemann and his colleagues (e.g., Heinemann, Avin, Sullivan & Chase, 1969; Heinemann & Avin, 1973) carried out a series of experiments in which pigeons were trained to discriminate among different intensities of wide-band white noise, using food reinforcement. They suggested that the birds were inattentive to the stimulus on a significant proportion of trials. In the latter study, they also showed how changes in criterion affected the slope of classical psychometric functions. Blough and Blough (1977) have suggested that the inattentiveness demonstrated by Heinemann and his co-workers is one reason why the methods of the theory of signal detectability may be unsuitable for the study of sensory processes in animals:

> the classical view of detection is partly right, for, on some trials, sensory input plays no role in controlling the response. Thus a correction like the classical correction for false reports must be estimated and applied to the data, in effect removing the 'inattentive' trials from the subsequent analysis (p. 536).

GOC analysis may provide a rather more palatable solution to this and other problems <u>within</u> the framework of TSD. The pigeon would seem to be a suitable observer for testing this suggestion, especially in experiments using food as a reinforcer.

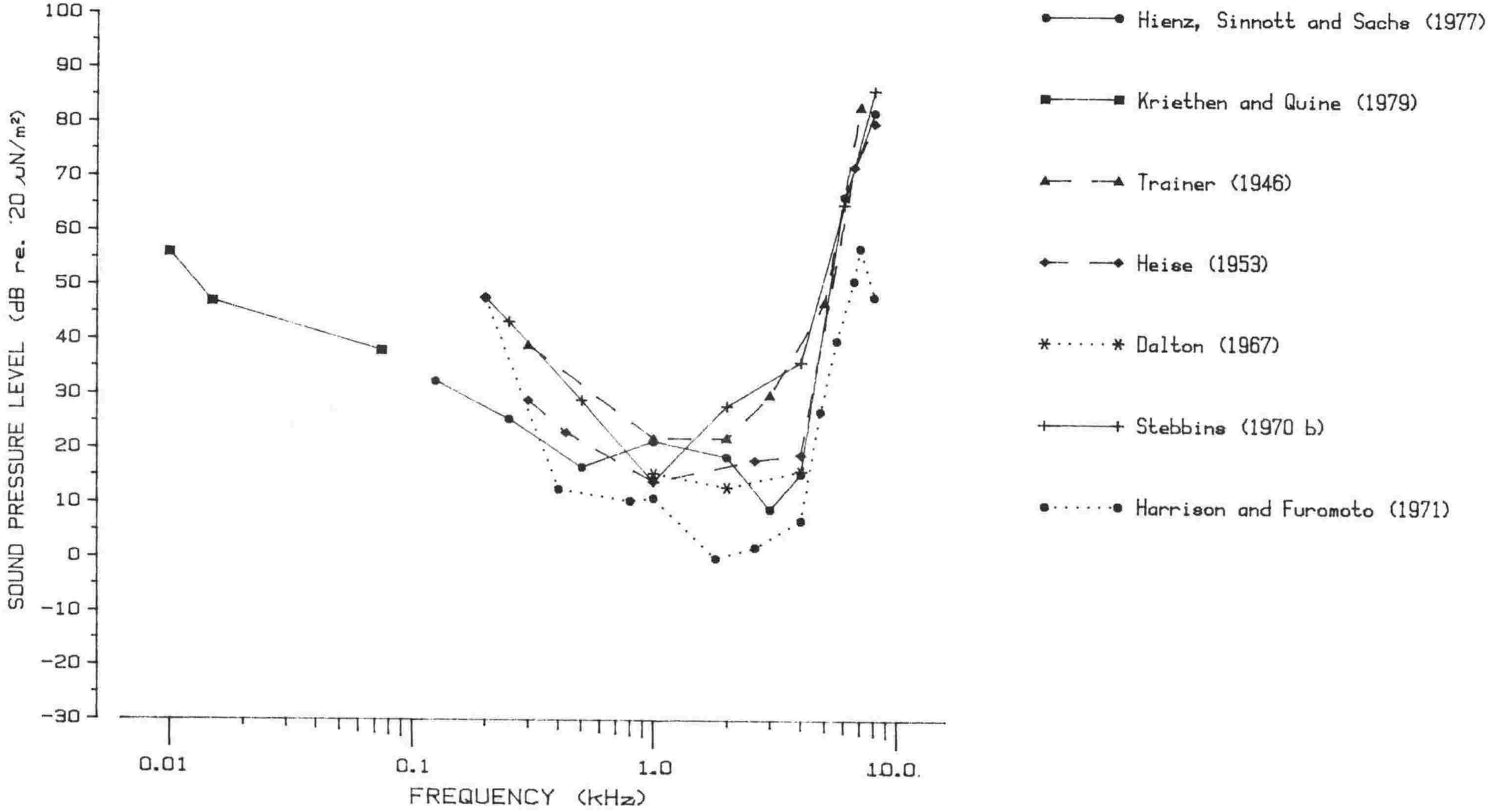


Figure 10. Pigeon audiograms.

THE PSYCHOPHYSICAL PROCEDURE

The procedure used with pigeons in the experiments reported here was equivalent to the yes-no (YN) procedure used with human subjects. In YN tasks, the observer is presented with either a standard or a comparison signal and asked to say whether the standard has occurred or not. This contrasts with the "same-different" (SD) task, in which both the standard and comparison signals are presented, and the observer reports whether there are differences among the signals or not. Research with human subjects (Jesteadt & Sims, 1975; Creelman & Macmillan, 1979) has shown that discrimination performance is better when the SD procedure is used. This also appears to be the case with pigeons: Krasnegor (1971), who used the YN procedure on which the experiments reported here are based, obtained much larger frequency difference limens than the other experimenters who have studied frequency discrimination in pigeons (Price, Dalton & Smith, 1967; Delius & Tarpy, 1974; Sinnott, Sachs & Hienz, 1980), all of whom used techniques analogous to the same-different procedure.

Jesteadt and Sims (1975) concluded that frequency discrimination performance in the YN paradigm is degraded by imperfect memory for the standard frequency and that, within the context of TSD, the effects of such imperfect memory may most appropriately be represented as decision or criterion variability. As discussed earlier, such variability will contribute to unique noise, the effects of which will be removed by GOC analysis.

RECOVERY OF AN ROC CURVE BASED ON TRIANGULAR DISTRIBUTIONS

Method

Subjects

All three birds provided data for this experiment.

The Distributions

The distributions are shown in Figure 3 in Chapter 2. In the experiment, each distribution was realised by 200 stimulus presentations. Thus 400 trials made up one replication of both model distributions. A replication for a given bird was made up of five 80-trial sessions.

The sequence of trials making up each replication was generated independently for each bird and each replication. Trials occurred in a haphazard order determined by the computer,[10] with two restrictions. One was that no more than four '+' or '-' trials[11] could occur consecutively. The other was that no more than 42 '+' or '-' trials could occur in each block of 80 trials. Sequences so generated were stored on magnetic tape and accessed by the computer at the beginning of each session.

---

[10] A starting seed for the computer's random number generator was obtained by using the real-time clock. A candidate seed was generated by combining the date and the second of interrogation. If the resulting number met criteria stipulated by the Hewlett-Packard 9825 manual, it was used; if not, another candidate was generated.

[11] For both the standard and comparison distributions '+' trials were those in which the frequency of the sinusoid was 712 Hz or greater; on '-' trials the frequency was less than 712 Hz. Thus on '+' trials, "yes" responses were followed by food, while on '-' trials "no" responses were reinforced, regardless of whether the sinusoid came from the standard or comparison distribution.

Procedure

As mentioned in Chapter 2, the birds' responses were reinforced according to their relation to a specific cutoff. The cutoff maximised the number of correct classifications of standard and comparison trials.

A large number of training sessions, spread over four months, preceded those which provided data for GOC analysis. At first only a few extreme frequencies (e.g., 400 Hz, 424 Hz, 976 Hz and 1000 Hz) were used. Over sessions, more frequencies were added until the distributions shown in Figure 3 were arrived at.

During the sessions which provided the data reported here, the 80 experimental trials were preceded by 10 warm-up trials, the results of which were discarded. Each of the first and second pairs of warm-up trials involved the highest and lowest frequencies (1000 Hz and 400 Hz) in random order within each pair. The remaining six trials involved a haphazard selection from the three highest frequencies (1000, 976 and 952 Hz) and the three lowest frequencies (400, 424 and 448 Hz).

A correction procedure was used for all warm-up and experimental trials. That is, if a left-hand key response was made on a high-frequency trial (712-1000 Hz) or a right-hand key response was made on a low-frequency trial (400-688 Hz), that trial was repeated until a correct response was made.

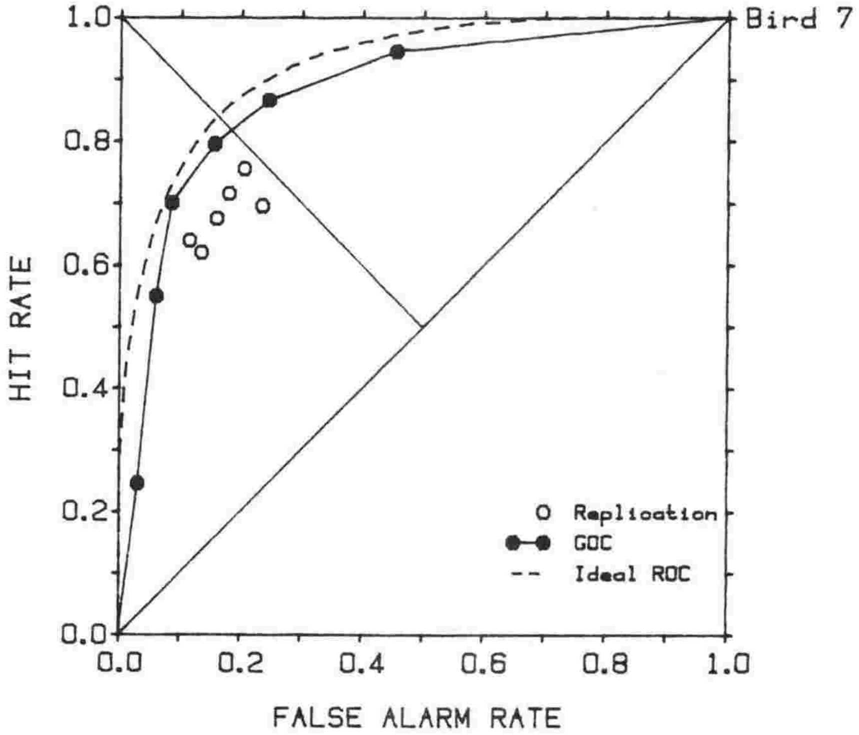
The reinforcement access times for Birds 7, 8 and 18 (9, 5 and 4 seconds respectively) were related to their speeds of eating. The mean ITI was 5 seconds for all birds, and the ITI was uniformly distributed between 4 seconds and 6 seconds. Timeouts and delays for pecks during ITIs all lasted 5 seconds.

Preliminary training took place during approximately 100 sessions run over four months. Experimental sessions began when the birds had been run with the final versions of the distributions for six sessions. A total of six replications (30 sessions) was run for each bird.

Results

GOC curves were produced for each bird by combining the results of equivalent trials over the six replications. Each trial had been labelled as a standard or comparison trial. However, no prior decision was made as to which trials would be treated as equivalent in each replication, and would thus have their outcomes combined over replications. This was done immediately before the GOC analysis by sorting the trials making up each replication into numerical order according to the frequency used on the trials and the distribution to which they had been assigned. The sorting process ensured that results in corresponding positions in different replications came from the same type of trial in terms of sinusoid frequency and standard or comparison distribution.

The main results are shown in Figures 11 and 12. The birds' performances on individual replications are represented by the open circles. For each replication, the hit rate was the proportion of comparison trials on which the bird completed 30 responses on the right-hand key (a "yes" judgement) and the false alarm rate was the proportion of standard trials on which the bird made a "yes" judgement. The hit and false alarm rates for each replication are given for each bird in Table 5 (Appendix C). The results show that the birds (especially Birds 8 and 18) achieved high levels of discrimination and maintained stable criteria over the six replications.



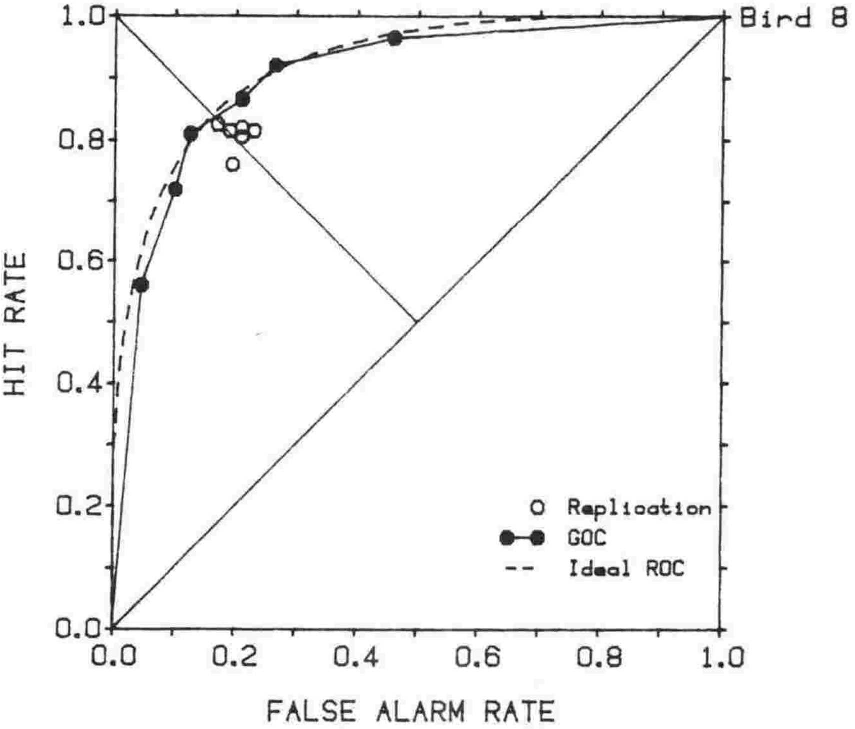


Figure 11. GOCs and individual replication results obtained with the triangular distributions in the first pigeon experiment.

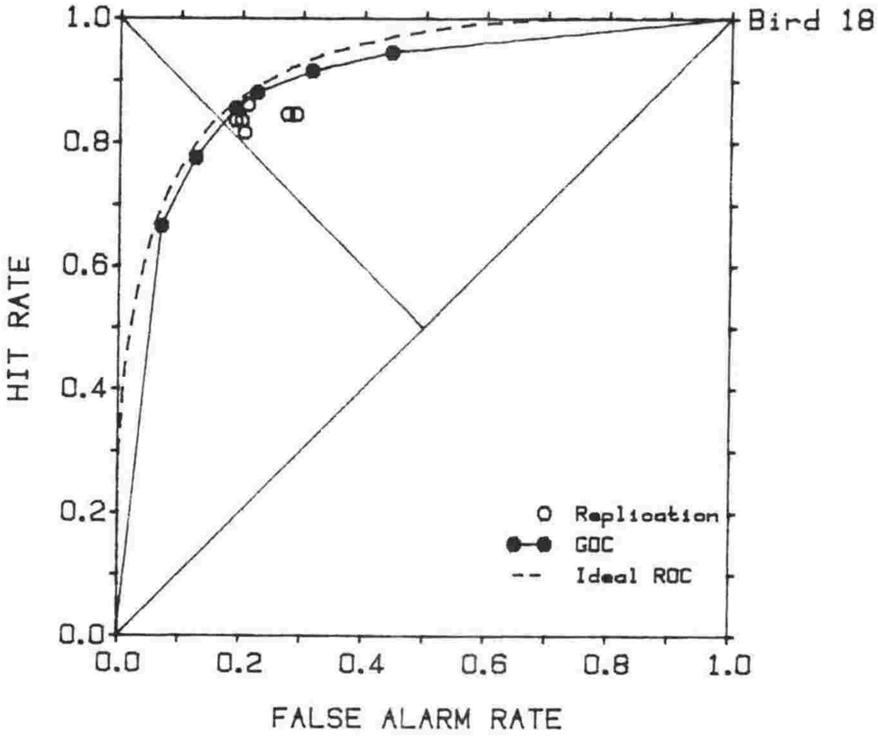


Figure 11 (continued).

In the three graphs in Figure 11, the unbroken lines are six-replication GOC curves. Figure 12 shows the GOC curve for all birds and replications combined, 18 replications in all. These curves can be compared with the ROC curve for the constructed distributions, which is shown by the dashed line.

The GOC curve for each bird is closer to the ROC than are any of the points representing performance on the individual replications. In general, the GOC curves give a reasonable indication of the nature of the ROC; this is especially true for the 18-replication GOC shown in Figure 12.

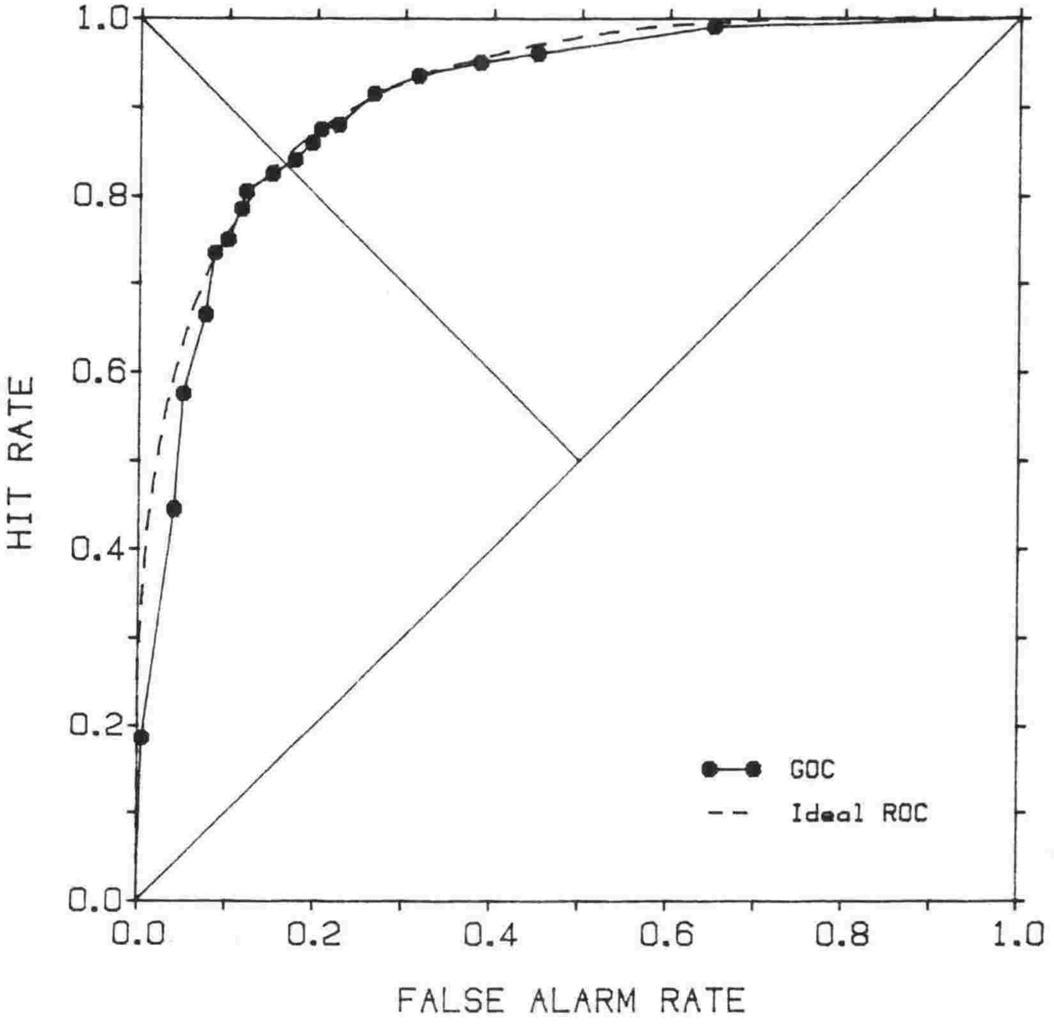


Figure 12. The 18-replication GOC for all birds combined, obtained with the triangular distributions in the first pigeon experiment.

RECOVERY OF AN ROC CURVE BASED ON BIMODAL DISTRIBUTIONS

Although the results obtained with the triangular distributions give some indication of the effects of GOC analysis, the distributions have the disadvantage that the ROC curve is like one based on normal equal variance underlying distributions. This means that normally-distributed unique noise could affect the final GOC for each bird without detracting from the fit of the GOC to the ROC. In fact, the noise could contribute to the fit, giving a false impression of the extent to which the GOC analysis recovered the ROC.

To explore this point without running a further experiment, a new pair of distributions was produced from the original pair. This was possible because the birds' responses were reinforced according to a fixed cutoff.

The new distributions are shown in Figure 13. Their bimodal shapes were designed to produce a relatively complex ROC which would not be compatible with normally-distributed unique noise. They were constrained by the frequency of occurrence of each of the sinusoidal frequencies used in the experiment. As can be seen if Figures 3 and 13 are compared, the sums of the probability of occurrence of each frequency over the standard and comparison distributions are the same in each case.

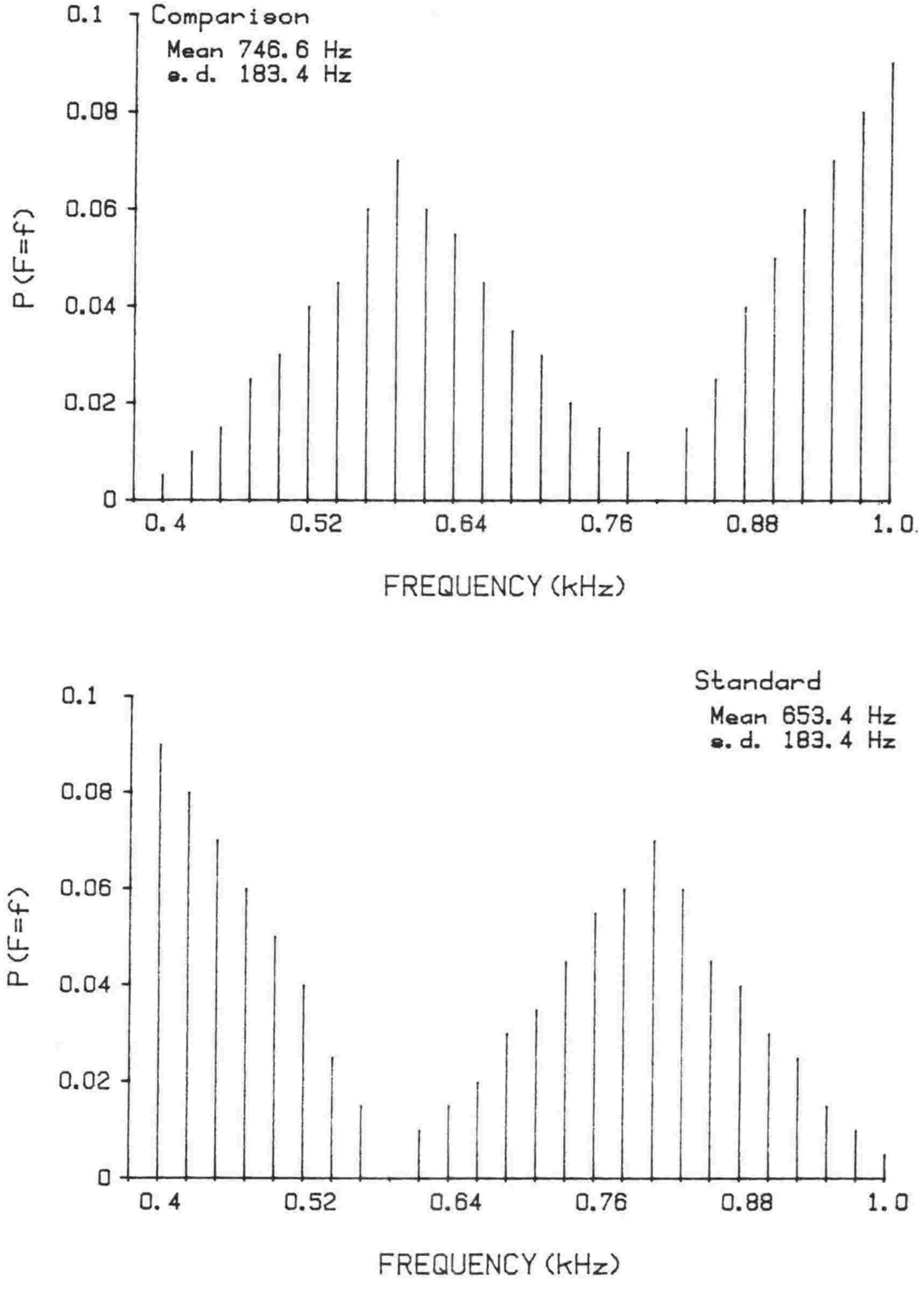


Figure 13. The bimodal distributions used in the second part of the first pigeon experiment.

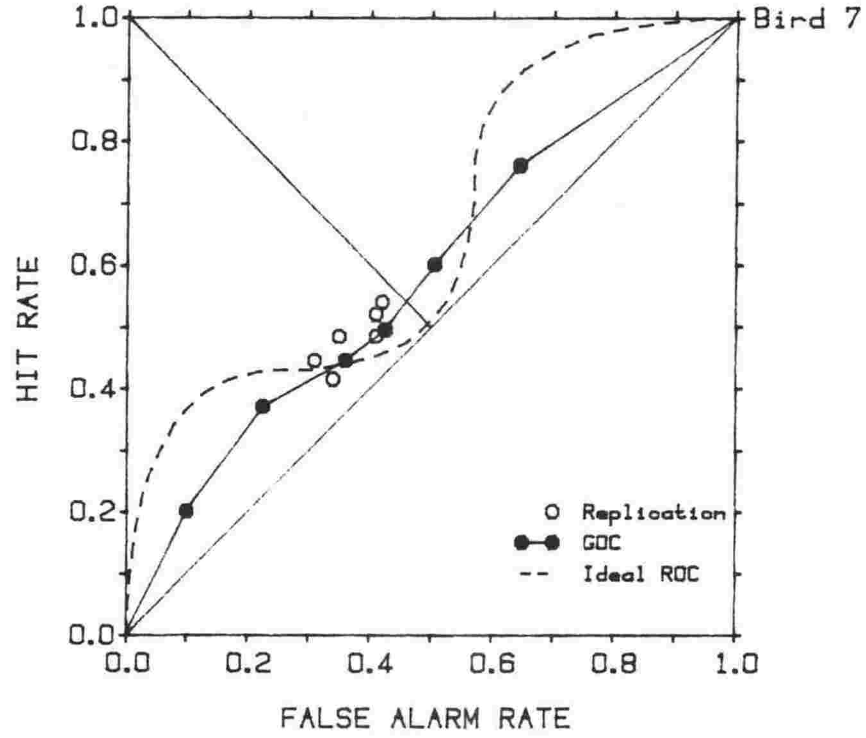
Procedure

The new analysis called for a reassignment of standard and comparison designations to all the trials. The original designations were erased and new comparison tags were, from a haphazardly-chosen starting point, assigned successively for each frequency, according to the requirements of the new model. For example, only one 400-Hz trial was to be designated a comparison trial; the remaining eighteen 400-Hz trials in each replication were to belong to the standard distributions. As can be seen from Figure 13, $P(400\ \text{Hz} \mid \text{comparison}) = 1/200 = 0.005$ and $P(400\ \text{Hz} \mid \text{standard}) = 18/200 = 0.09$. To take another example, 520 Hz was equally likely to occur on standard and comparison trials. Thus $P(520\ \text{Hz} \mid \text{standard})$ and $P(520\ \text{Hz} \mid \text{comparison}) = 8/200 = 0.04$.

Results

The birds' original responses in relation to the newly-defined model were used to calculate hit and false alarm rates as for the triangular distributions. These rates are shown in Table 6 (Appendix C).

The birds' performances on individual replications are shown by the open points in the three graphs in Figure 14. As before, the ROC is shown by a dashed line and the six-replication GOC for each bird is shown by the unbroken line.



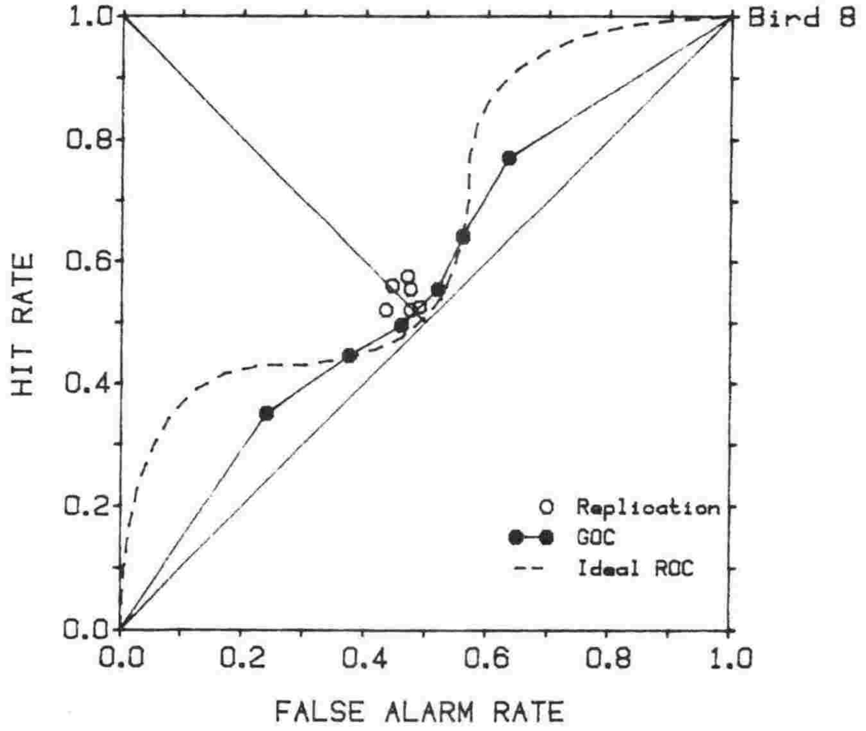


Figure 14. GOCs and individual replication results obtained with the bimodal distributions in the first pigeon experiment.

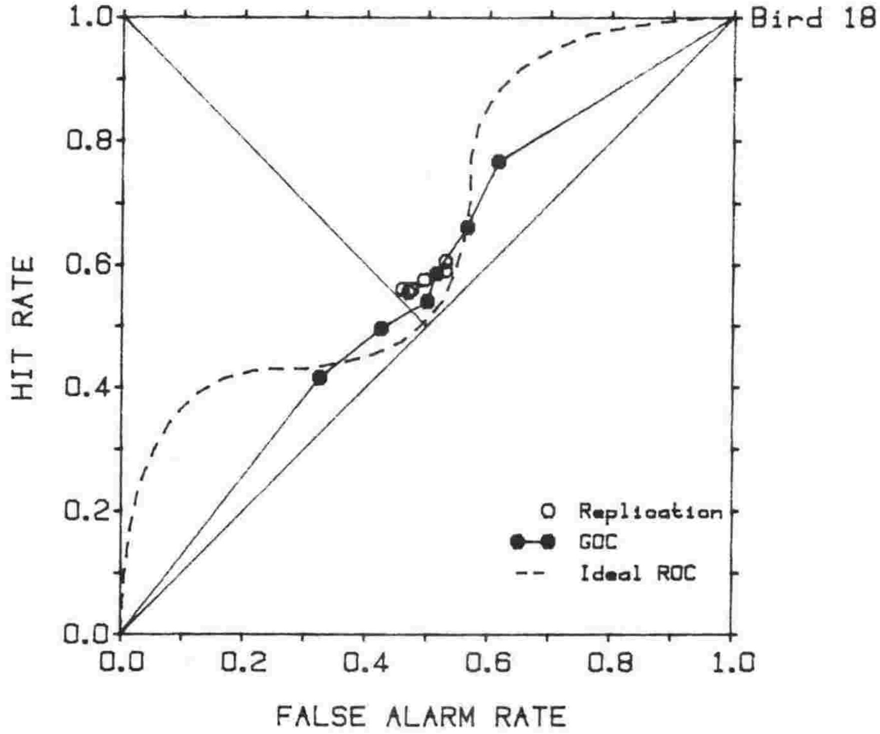


Figure 14 (continued).

These GOC curves approximate the ROC for the bimodal distributions less well than the GOC curves in the original analysis approximated the ROC for the triangular distributions. This suggests that normally-distributed unique noise affects the individual GOC curves. Further replications would be needed if the ROC was to be more nearly approximated.

The data in Figure 15 support this conclusion: the 18-replication GOC, for all birds combined, more clearly reflects the underlying distributions than do any of the individual GOC curves. Even so, the strict and lax criterion regions of the GOC curve are still some distance from those of the ROC curve.

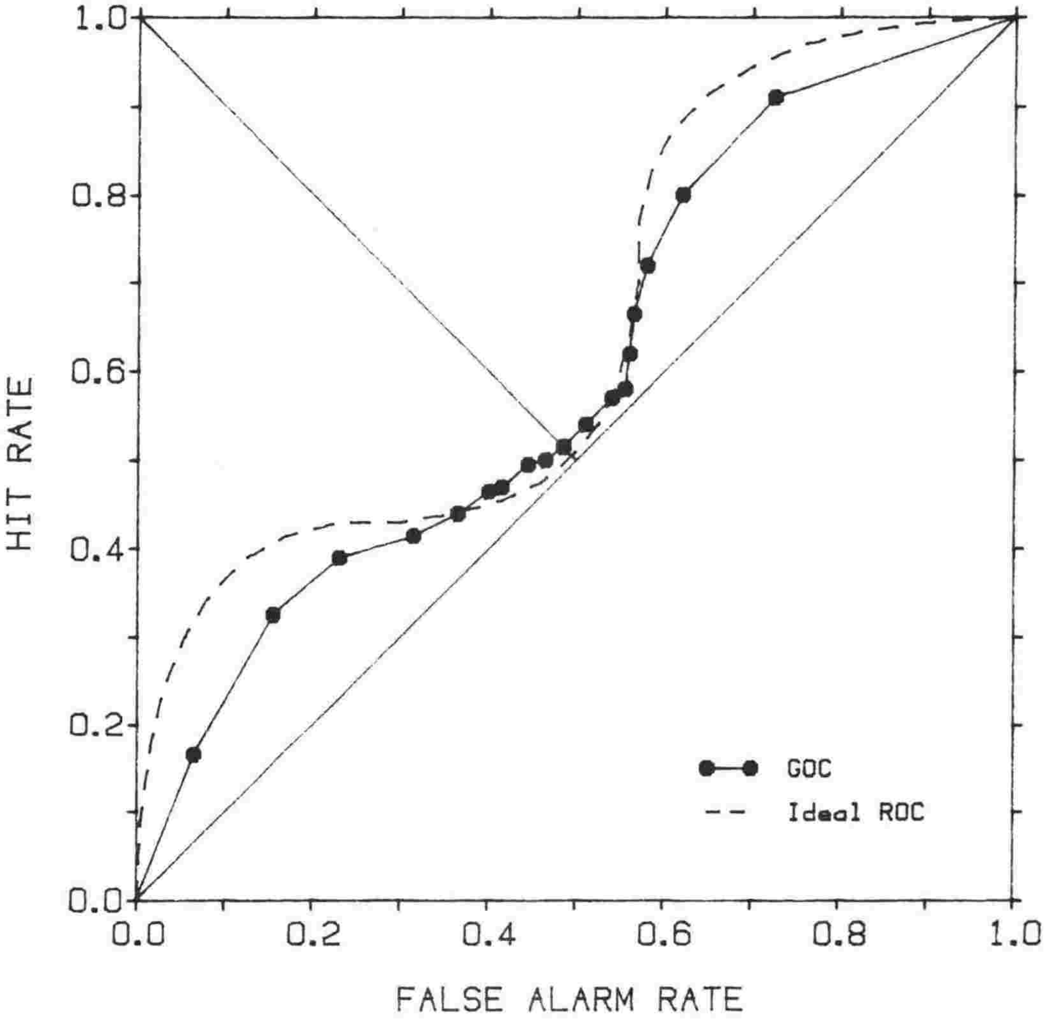


Figure 15. The 18-replication GOC for all birds combined, obtained with the bibimodal distributions in the first pigeon experiment.

CONCLUSION

The results described in this chapter show that the GOC curve provides a better indication of the nature of the ROC for known underlying distributions than do the results for individual replications. This is true even when the shapes of the underlying distributions are complex, and unlike any plausible distributions of unique noise. In the latter case, as Boven noted, GOC analysis is less successful in reducing the effects of the unique noise. Following a description of the computer-simulations which complemented the studies with pigeon observers, the next chapter examines ways of increasing the effectiveness of the GOC analysis of data obtained with complex ROCs.

CHAPTER 4

A COMPUTER MODEL FOR THE STUDY OF GOC ANALYSIS

Auditory experiments with pigeons take a considerable time to set up and run, especially when a number of replications is needed for GOC analysis. Investigating questions arising from one experiment by doing a new experiment is often not practicable. In the present research this limitation was overcome by constructing a computer model for the study of simulated observers.

One of the most important aspects of the model was that the characteristics of both the unique and the common noise could be specified in the knowledge that the observer was not adding common noise from another source. Another advantage was that distributions which were essentially continuous, but whose characteristics were nevertheless known, could be used. A third advantage was that the simulated observer could readily use rating scales with more than two categories.

This chapter describes the computer model, which was implemented on a Hewlett-Packard 9826 desktop computer, and gives the results of simulations based on the bimodal distributions used with the pigeons and on trimodal distributions used by Boven.

Description of the Model

The simulations had three stages. In the first, the common and unique noise distributions were generated. In the second, decisions were made by the simulated observer. Finally a GOC analysis of the observer's responses was carried out.

Generation of the Distributions

In some simulations, discrete distributions of common noise which were similar to those used with pigeons were employed. These were generated by sampling without replacement from a population of elements with values 1 to 26, corresponding to the 26 sinusoids used with the pigeons. The number of elements making up each distribution was around 200, as for the pigeons, but in the simulations the number could be increased to a multiple of the original figure.

The continuous normal distributions of either common or unique noise were produced by the Polar algorithm due to Box, Muller and Marsaglia and described by Knuth (1969). Further details are given in Appendix D. The uniform random variable required by the Polar algorithm was produced either by the computer's pseudo-random number generator or by a software routine based on a method described by Evans, Wallace and Sutherland (1967). The latter is described in Appendix D. The random variable generators were initiated with selections from random number tables which met the specifications for seeds set out in the computer's manual or given by Evans, Wallace and Sutherland (1967).

The values of common noise taken from the discrete distribution or provided by the normal deviate algorithm were placed in an array. Each element in the array corresponded to one trial in a complete replication of a particular experiment. As each trial of a replication was run, a value from the unique noise distribution was added to the appropriate value in the array of samples of common noise. Each addition provided a value of the evidence variable.

During the simulation, the values of the unique and common noise were each cumulated, so that means and standard deviations could be calculated and compared with the nominated values. In some cases the parameters of the normal distributions were chosen to give rise to particular values of k. When the distributions of common noise were of equal variance and the unique noise came from one population, this was simply a matter of nominating standard deviations such that the ratio σ_U^2/σ_C^2 was equal to k, where σ_C is the standard deviation of the common noise distribution and σ_U is the standard deviation of the unique noise distribution. In some simulations the variances of the common noise distributions were not equal, while, as before, the unique noise was sampled from one population. In these cases k was calculated from:

$$\underline{k} = \frac{2\sigma_U^2}{\sigma_{SC}^2 + \sigma_{CC}^2} \qquad (6)$$

where σ_{SC}^2 is the variance of the standard distribution and
σ_{CC}^2 is the variance of the comparison distribution.

One of the most useful types of simulation was one in which there was no common noise (the means and standard distributions of the common noise distributions were set to zero), so that k was very large. The results of this kind of simulation provided a baseline against which the results of the pigeon experiments could be evaluated.

The Decision of the Simulated Observer

When a two-category rating scale (analogous to the two-key responses of the birds) was used, a particular value of the evidence variable was specified as the cutoff. In making the decision, the program simply compared the evidence variable with the cutoff. If the value of the evidence variable was greater, a "yes" response was stored in the appropriate element of the array for that replication, otherwise a "no" response was stored.

When the simulated observer used more than two rating-scale categories, the cutoffs were equally spaced over a range from two standard deviations below the mean of the standard common noise distribution to two standard deviations above the mean of the comparison common noise distribution. The standard deviation used was that of the combined common and unique noise distributions.

The simulation program was designed so that results could be obtained simultaneously for a rating scale with two categories and another with more than two categories. Scales with different numbers of categories could be applied to identical noise data by running the simulation twice and using the same parameters for the random number generator.

GOC Analysis

The simulation program incorporated a section which carried out GOC analysis as described in Chapter 1. Up to 10 replications could be analysed. Details of an extended analysis which was carried out with both pigeon and simulated data in order to estimate *k* are given in Chapter 5.

GOC Analysis with Multi-Modal Distributions

Bimodal Distributions

With the bimodal distributions used in the experiment described in the previous chapter, the GOC results for individual birds were not very impressive: the GOC curves gave only a hint of the nature of the ideal ROC. When the results for all three birds were combined, the GOC curve approximated the ROC more closely, illustrating the point that when underlying distributions are complex, a larger number of replications may be needed.

One way of improving the results of GOC analysis without increasing the number of replications is to decrease the loss of information by having observers use a rating scale with a relatively large number of categories. The pigeons used only two, but with the computer model greater numbers were possible. Figure 16 gives the results of a simulation in which unique noise sampled from a normal distribution with mean zero and a standard deviation of five was added to the bimodal distributions shown in Figure 13 in the previous chapter. Two simulation experiments, each consisting of 10 replications, were run. In one, the observer used a two-category rating scale with a cutoff set mid-way between the means of the standard and comparison distributions. In the second experiment, the observer used a scale with 10 categories. The 10-replication GOC results for the second experiment (unbroken line) are superior to those based on the two-category scale (dotted line). However, the simulated observer maintained its multiple cutoffs with perfect consistency; a flesh-and-blood observer would be unlikely to do so, thus introducing unique noise which would go some way towards offsetting the benefits of the extra rating-scale categories.

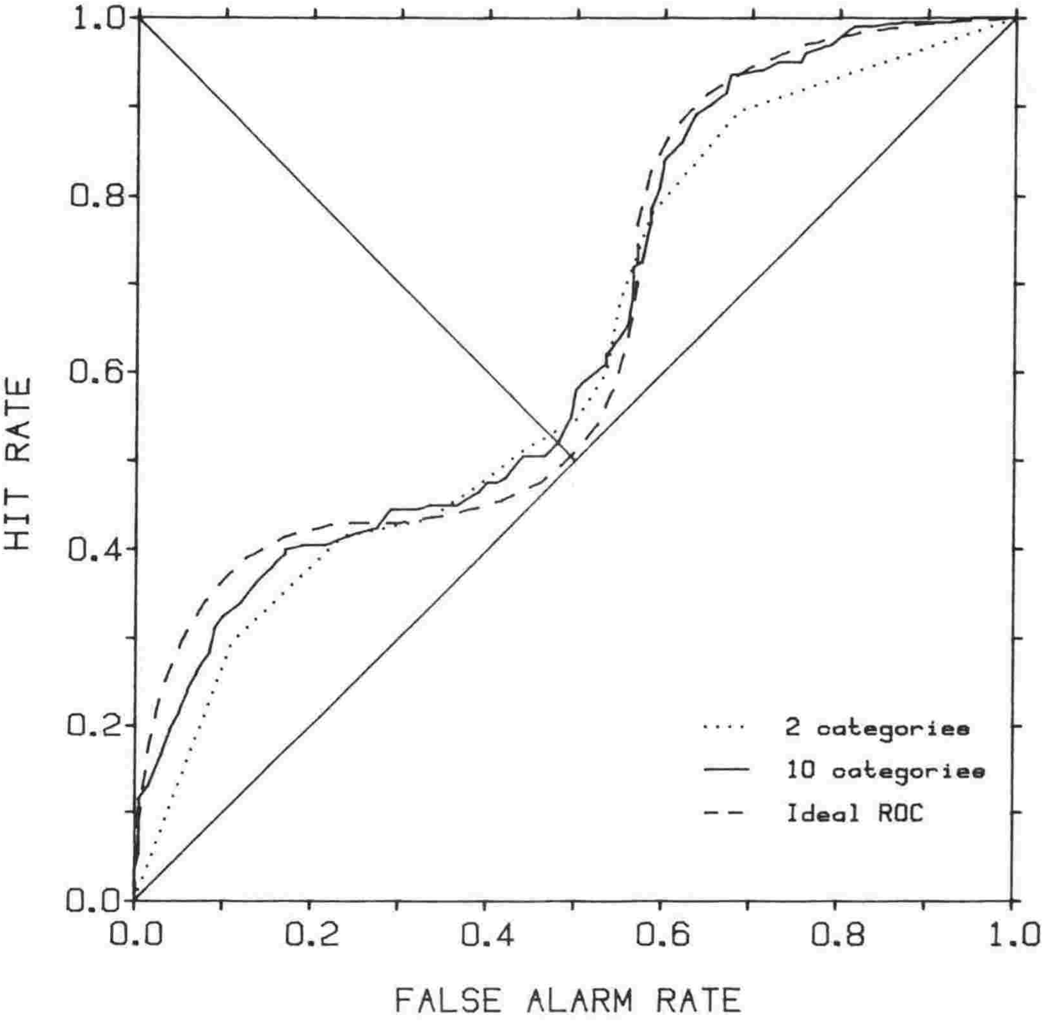


Figure 16. GOC results obtained from simulated observers using two- and 10-category rating scales respectively. The simulations were based on the bimodal distributions shown in Figure 13. There were 10 replications, and the standard deviation of the normally-distributed unique noise was five.

Trimodal Distributions

Boven (1976) found GOC analysis to be relatively unsuccessful in capturing the ROC for the trimodal distributions in Figure 17. He attributed this lack of success to insufficient data, suggesting that the complex nature of the distributions made them more sensitive than were the simpler (normal and rectangular) ones to chance effects in the placement of the points on the GOC curve. He concluded that, when the shape of the ROC curve for the underlying distributions is thought to be complex, relatively large numbers of experimental trials should be used, the exact number depending on the complexity of the ideal ROC curve.

Boven is here talking about insufficient trials, not insufficient independent stimuli, since with the constructed distributions used by Boven and in the present work the samples of stimuli are exhaustive. Unrepresentative sampling of the independent stimuli would have led to ROC curves which gave a misleading impression of the nature of the population, even when the effects of the unique noise had been removed. The effects of insufficient trials, on the other hand, should be reduced by either running more trials per replication, as suggested by Boven, or simply adding more replications. Since which of these two strategies would be the most effective was not clear, several simulations based on the trimodal distributions shown in Figure 17 were run. The results were quite clear-cut and can be represented by one example. Figure 18 shows the GOC curves for two simulation experiments, one of which (dotted line) consisted of five 864-trial replications and the other (unbroken line) of ten 432-trial replications. In order to remove any confounding effects, the unique noise added to the standard and comparison distributions was the same in each experiment. The simulated observer used a rating scale with four categories, a number which could readily be handled by a human observer.

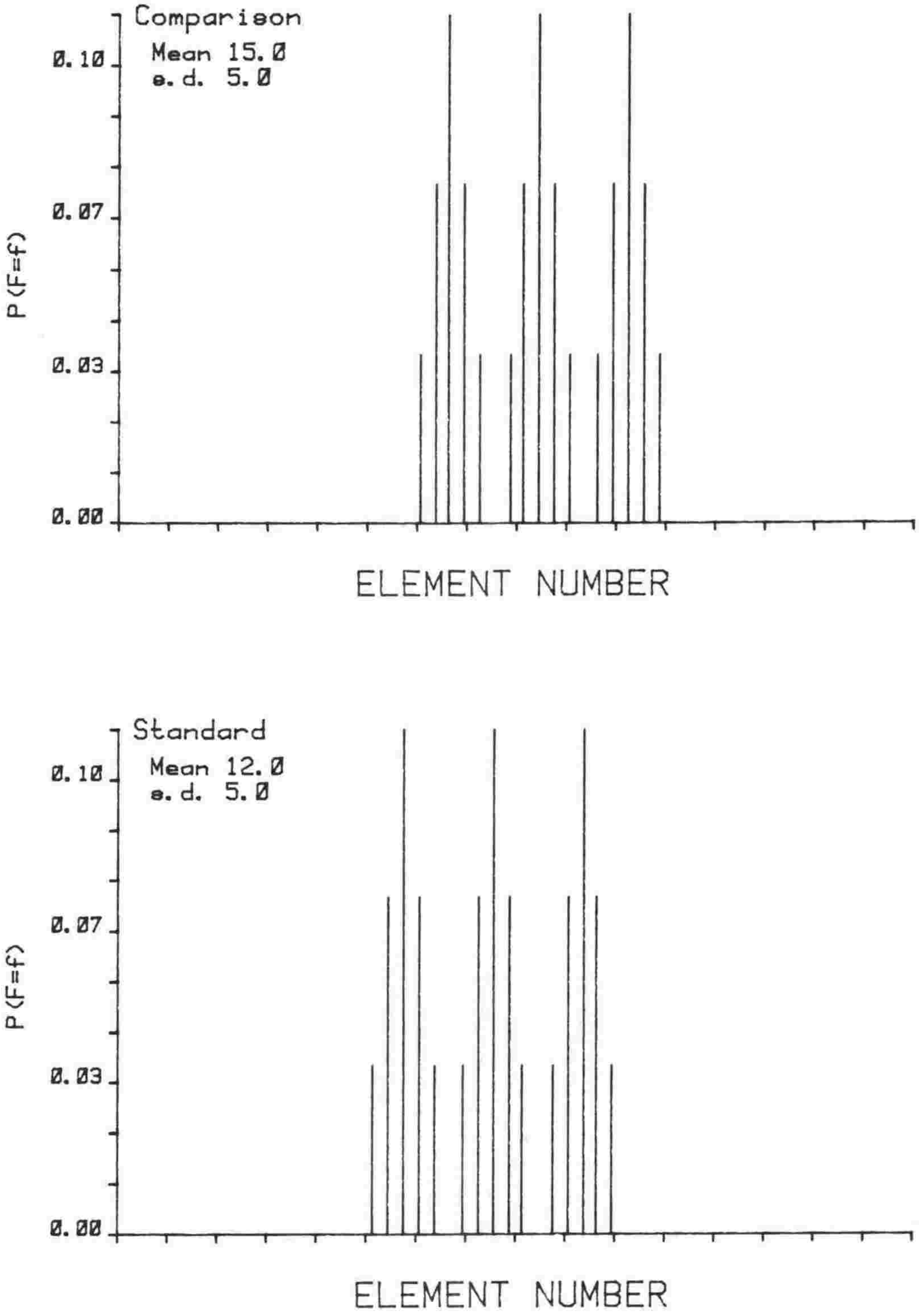


Figure 17. Trimodal distributions similar to those used by Boven (1976).

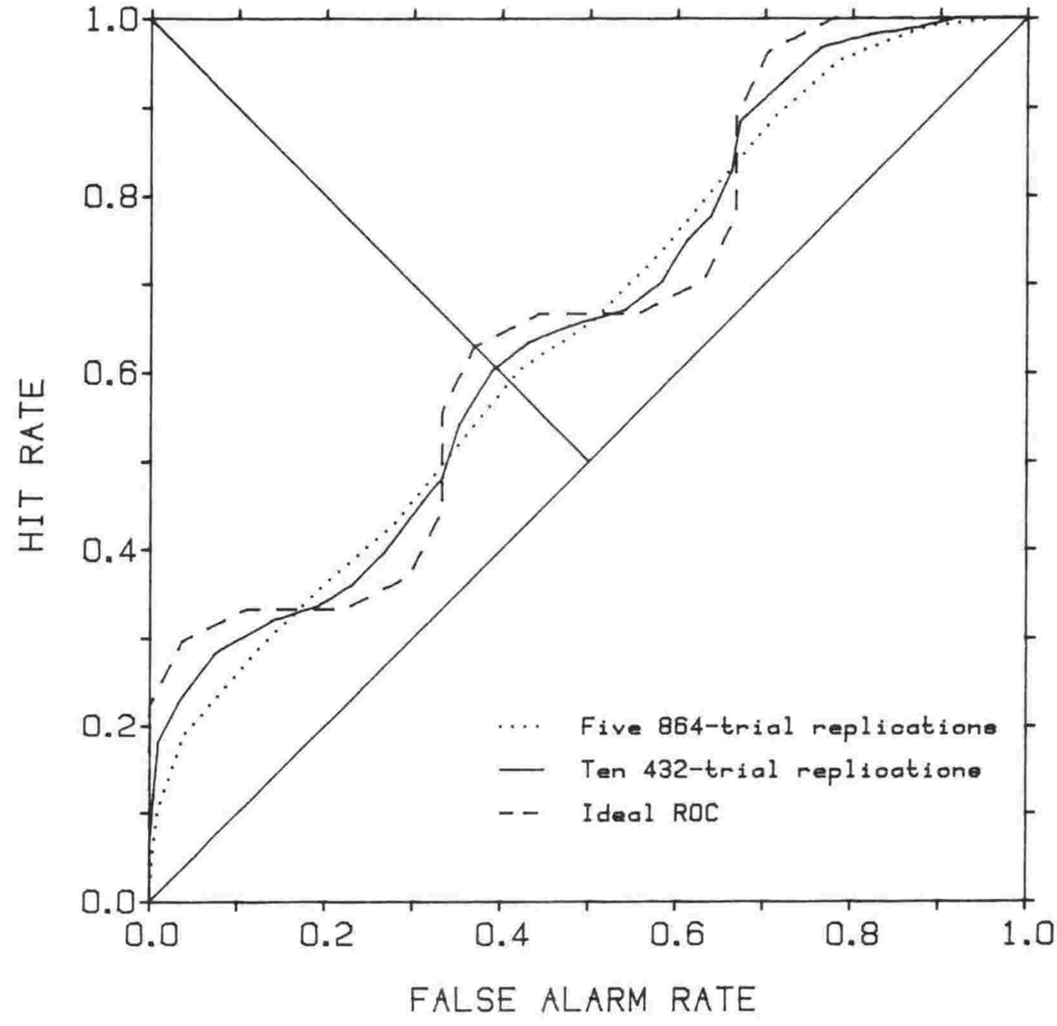


Figure 18. GOC results showing the effects of adding more replications as against increasing the number of trials per replication. The results are from simulations based on the trimodal distributions shown in Figure 17.
The standard deviation of the normally-distributed unique noise was 3.5.

The GOC for the greater number of replications gives the best indication of the shape of the ideal ROC. When the observer is entirely consistent in its placement of cutoffs, and the common noise is adequately sampled, adding trials by way of further replications is evidently better than running more trials in each replication.

Summary

With the computer simulation, experiments consisting of up to 10 replications and up to 4000 trials per replication were run. The main advantages of the simulation were that experiments could be run relatively quickly, the characteristics of both the common and unique noise could be specified, and the observer could use any number of rating scale categories.

Simulations involving complex distributions were used to demonstrate the effects of using more than two rating scale categories and to find out whether, in order to recover a complex ROC, using more trials per replication was better than adding more replications.

CHAPTER 5

PROBLEMS IN MEASURING THE AMOUNT OF UNIQUE NOISE BY GOC ANALYSIS

If the assumptions outlined in Chapter 1 are met, k, the ratio of the variances of the unique and common noise, can readily be obtained from equation 4, first given in Chapter 1. With the index d_z, which is applicable when the underlying normal distributions are not of equal variance:

$$\underline{k} = \frac{n[\{d_{z(n)}\}^2 - \{d_{z(1)}\}^2]}{n\{d_{z(1)}\}^2 - \{d_{z(n)}\}^2} \qquad (4)$$

where n is the number of replications added together in the GOC analysis,

$d_{z(1)}$ represents the observer's performance on one replication [either $d_{z(1)}$ averaged over replications, or $d_{z(1)}$ for the HR and FAR based on response frequencies pooled over replications], and

$d_{z(n)}$ is the performance represented by the results of a GOC analysis based on n replications.

This equation can be expressed in a more general form which allows the estimation of k from any pair of the $d_{z(1)}$ to $d_{z(N)}$ values (where N is the total number of replications). In this expression (derived in Appendix E):

$$\underline{k} = \frac{in[\{d_{z(n)}\}^2 - \{d_{z(i)}\}^2]}{n\{d_{z(i)}\}^2 - i\{d_{z(n)}\}^2} \qquad (4a)$$

where i is the number of replications on which the first d_z is based (0<i<N), and

n is the number of replications on which the second d_z value is based (n>i).

Two of the assumptions given in Chapter 1 have been mentioned already. One is that the underlying distributions (both common and unique noise) are normal. The other is suggested by the definition of $d_{z(1)}$, which implies that the observer's overall performance before GOC analysis can be measured by the performance on any one replication; performance is assumed to be constant across replications. This assumption is probably easier to accept if the replications are all for one observer, as in the Swets et al. (1959) experiments, rather than for a number of observers.

The third assumption, that the observations are combined with no loss of information, is the one which is hardest to meet. This difficulty was mentioned in Chapter 1, and the difference between GOC results obtained with two- and 10-category rating scales was shown in the previous chapter. As a preliminary to considering the problem and possible solutions further, the following sections describe (a) how d_z was calculated from the GOC results in the experiments reported here, and (b) a technique for removing the effects of variations in the observer's performance over replications on d_z and therefore on the estimates of $\underline{k}$.

The Calculation of dz

The formula for d_z given in Chapter 1 (equation 5) is:

$$d_z = \frac{(2)^{1/2} m}{(\sigma_S^2 + \sigma_N^2)^{1/2}} \quad . \qquad (5)$$

This includes terms for the difference between the means of the distributions (m) and for the variances of the standard and comparison distributions. However d_z may be calculated without knowing the values of these parameters directly. If the underlying distributions are normal, ROC (or GOC) points plotted on normal-normal coordinates can be fitted with a straight line. The slope and intercept of this line can then be used to calculate d_z. As shown in Appendix F,

$$d_z = \frac{-c(2)^{1/2}}{(s^2 + 1)^{1/2}} \qquad (7)$$

where c is the intercept of the line fitted to the ROC (or GOC),

and

s is the slope of the line.

In most of the work reported here, d_z was not calculated directly from this formula, but from d_s (Egan, Greenberg & Schulman, 1961). This index, which is perhaps the most frequently used when the underlying distributions are assumed to be normal but of unequal variance, was found to be inappropriate for use in equation 4 for calculating <u>k</u>. The formula used to calculate d_z from d_s was:

$$d_z = \frac{2^{1/2}[d_s(s + 1)]}{2(s^2 + 1)^{1/2}} \quad . \qquad (8)$$

The derivation of this equation is also given in Appendix F.

Line Fitting

In order to calculate d_s or d_z a straight line was fitted to a number of points in an ROC space with normal-normal coordinates. In the past (e.g., Egan, Schulman & Greenberg, 1959), the method of least squares has been used for this purpose, but as various writers (e.g., Isaac, 1970) have pointed out, this method is inappropriate when there is error in the measurement of both the X-variable (the false alarm rate, in the present case) and the Y-variable (the hit rate). In the analyses reported here, the slope and intercept parameters were estimated using the formulae for structural relations provided by Isaac (1970) and Madansky (1959). The slope was calculated as:

$$\text{Slope} = \frac{\text{var}(y) - \underline{e}.\text{var}(x) + \{[\text{var}(y) - \underline{e}.\text{var}(x)]^2 + 4\underline{e}.\text{cov}^2(x,y)\}^{1/2}}{2\text{cov}(x,y)}$$

where X is the false alarm rate,

Y is the hit rate, and

$\underline{e}$ is the ratio of the error variances of the two rates, assumed to be unity.

The intercept was: intercept = mean(HR) - slope.mean(FAR).

In order to reduce the effects of error in extreme values on the estimates of slope and intercept, points with z(HR) and z(FAR) greater than 2.5 were excluded from the line-fitting procedure. Results very similar to those reported here were obtained when the least-squares method was applied to the same data and extreme points were not excluded.

The Calculation of dz(1)

When a simulated observer used a rating scale with more than two categories, calculation of the $d_{z(1)}$ value for individual replications was straightforward: since each ROC was defined by two or more points, a line could be fitted as described above. However, when a two-category scale was used, the ROC was defined by a single point, and its slope was unknown.

If the underlying distributions of common and unique noise were known to be normal and of equal variance, there was little doubt that the ROC slope (like those of the GOCs) was close to unity. In the second pigeon experiment, which was based on normal unequal variance distributions, the ROC slope was not unity, and without knowing the variances of the unique noise added to the standard and comparison distributions, an estimate of the ROC slope was difficult to arrive at. In this case, the resulting uncertainty about the value of $d_{z(1)}$ could be sidestepped by using only $d_{z(2)}$ to $d_{z(N)}$ in equation 4a in order to estimate _k_.

For the less critical purpose of estimating the basic performance of the observers, the slope of the ROC was estimated by extrapolation of the mean slopes of the GOCs obtained during GOC analysis. This was made possible by a method of combining the results of a series of GOC analyses which gave rise to very orderly changes in GOC slopes as replications were added. This method is described in the next section; the estimation of $d_{z(1)}$ is described at the end of the section.

All Combinations Analysis

Consider a GOC experiment in which five replications with one observer are run. The simplest way of obtaining an estimate of _k_ would be to enter the observer's performance in the absence of GOC analysis $[d_{z(1)}]$ and the performance corresponding to the GOC for the five replications combined $[d_{z(5)}]$ into equation 4. Fuller use would be made of the data if the d_z values based on the combination of two, three and four replications were also used. Estimates of _k_ could then be obtained by using $d_{z(1)}$ and $d_{z(2)}$, $d_{z(1)}$ and $d_{z(3)}$, and $d_{z(1)}$ and $d_{z(4)}$, as well as $d_{z(1)}$ and $d_{z(5)}$, in equation 4, not to mention the other pairs made possible by the more general equation 4a. If all the assumptions on which the Swets et al. equations are based were met, and there was no experimental error, all estimates of <u>k</u> would be the same. However, the ways in which estimates of <u>k</u> obtained from different analyses of the data differ may tell us something about the way in which the assumptions are not being met.

If the results of the intermediate GOC analyses are to be used, another problem arises: when the observer's performance varies from replication to replication, the changes in d_z, and therefore in the estimates of <u>k</u>, will depend on the order in which the replications are combined. For example, if the observer's performance had improved as replications were run, and $d_{z(1)}$ was based on the first replication, <u>k</u> would tend to be overestimated when $d_{z(1)}$ was compared with $d_{z(2)}$, $d_{z(3)}$, etc.

One way to eliminate order effects would be to combine the replications in all possible orders. This is not very practicable. For example, there are 10!, or 3,628,800, permutations of ten replications. Another solution along these lines would be to carry out analyses for a sample of all possible orders. This approach was tried, but the results showed that a prohibitive number of samples would be necessary in order to obtain acceptably orderly results.

The solution finally adopted was to use all possible <u>combinations</u> of replications to obtain an estimate of d_z for each possible number of replications. To give an example for the case of four replications: There are four permutations of four replications in which replications 1 and 2 come first. These permutations are [1,2,3,4], [1,2,4,3], [2,1,4,3] and [2,1,3,4]. Since the order in which replications 1 and 2 are combined does not affect the value of $d_{z(2)}$ derived from the GOC analysis, the combination of replications 1 and 2 need only be considered once in calculating $d_{z(2)}$. This is also true for all other combinations of two replications, namely [1,3], [1,4], [2,3], [2,4] and [3,4]. Thus in order to obtain an estimate of $d_{z(2)}$ based on all possible combinations of replications, only 4!/(4-2)!2!, or six, GOC analyses need to be carried out.

This method was used to obtain a mean value of d_z for each number of replications from two to N. A mean GOC slope for each number of replications was also obtained.

<u>The Estimation of dz(1) for Two-Category Rating Scale Data</u>

When the ROC from which $d_{z(1)}$ was to be derived consisted of one point, the set of mean GOC slopes obtained during the all-combinations analysis was used to estimate the slope necessary for the calculation. This process can be illustrated by reference to Figure 19, which shows results obtained from simulations based on normal unequal variance distributions. The circles and squares in each graph represent the mean GOC slopes obtained with two- and 20-category rating scales respectively, while the lines are quadratic functions[12] fitted to the mean slopes for two to eight replications. With the continuous model (lower graph), initial slopes based directly on the 20-category scale, and those estimated from extrapolation of the functions, match the actual slope calculated directly from the parameters of the simulation (not shown) fairly well; the correspondence for the discrete model is not so good. However, in most cases such discrepancies were small (varying from 2% to 9%) and, more importantly, led to even smaller error

---

[12] A number of functions were fitted in preliminary work. For the distributions used here, quadratic functions usually gave the best results.

in the estimates of $d_{z(1)}$.

When a two-category rating scale was used, $d_{z(1)}$ was calculated by expressing the intercept c in equation 7 in terms of the estimated slope and a pair of hit and false alarm rates, so that

$$d_z = \frac{2^{1/2}[s\{z(FAR)\} - z(HR)]}{(s^2 + 1)^{1/2}} \quad . \qquad (9)$$

The initial estimates of $d_{z(1)}$ were based on hit and false alarm rates pooled over all replications. However, an overall value of $d_{z(1)}$ was also obtained by averaging the $d_{z(1)}$ values for individual replications. The individual values of $d_{z(1)}$ were based on the hit and false alarm rates for each replication, although the same slope estimate was used in each case. The estimate based on averaging was obtained to check on the possible deflating effects of pooling, which may occur when there is some criterion variability over replications (Swets & Pickett, 1982). In the present example, the estimates of $d_{z(1)}$ based on pooled and mean data were very similar, the latter being only 0.1% and 0.6% higher for the discrete and continuous distributions respectively. Together they differed from the simulated distribution parameters by around 3% and less than 1% respectively. In other work these errors usually fell in the 1% to 5% range.

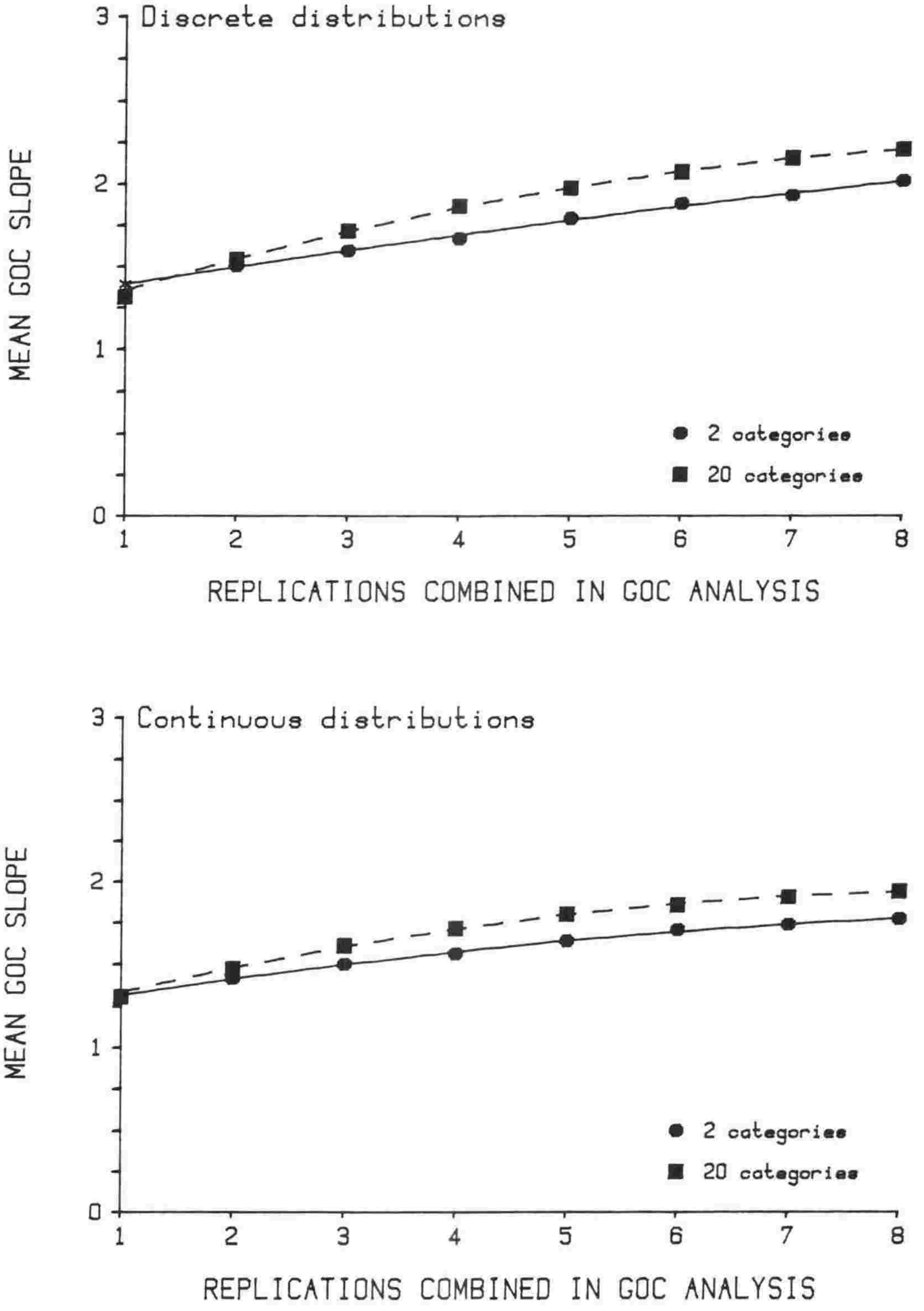


Figure 19. The mean slopes of GOCs based on different numbers of replications, obtained during a simulation in which observers used two- and 20-category rating scales. The nominated parameters of the underlying normal unequal variance distributions were the same as those for the distributions shown in Figure 5.
The points are fitted with quadratic functions.

The Estimation of k

The preceding sections showed how d_z was calculated for GOCs, introduced a method for eliminating the effects of changes in observers' performances over replications and finally examined how, when a two-category rating scale was used, the slope of the ROC could be estimated and combined with the hit rate and false alarm rate in order to estimate $d_{z(1)}$. We now return to the main topic introduced at the beginning of the chapter, the estimation of k.

Two main methods were investigated during the course of this work. Both used all, or most, of the set of $d_{z(1)}$ to $d_{z(N)}$ values obtained from the all-combinations analysis.

For the first method a series of lines were generated by systematically inserting various values of k in a variant of equation 3, namely:

$$d_{z(n)} = \frac{d_{z(1)}[1 + \underline{k}]^{1/2}}{[1 + \underline{k}/n]^{1/2}} \quad ,$$

and taking as k the value which gave rise to the best-fitting line. This method was found to work reasonably well when rating scales with large numbers of categories (10 or more) were used. With two-category scales, the estimates of k depended very much on the number of replications run (or on the number of $d_{z(i)}$ values to which the lines were fitted). This method is not discussed further.

The method of estimating k reported here involved the generation of a series of values of k using $d_{z(1)}$ [or $d_{z(2)}$] as $d_{z(i)}$ and $d_{z(2)}$ [or $d_{z(3)}$] to $d_{z(N)}$ as $d_{z(n)}$ in equation 4a, in which

$$\underline{k} = \frac{in[\{d_{z(n)}\}^2 - \{d_{z(i)}\}^2]}{n\{d_{z(i)}\}^2 - i\{d_{z(n)}\}^2} \quad . \qquad (4a)$$

This method will be illustrated by applying it to the results of the simulations mentioned in the previous sections (Figure 19).

Figure 20 shows the values of d_z obtained for the discrete and continuous normal unequal-variance distributions. In both cases the values obtained with the larger number of rating scale categories are higher, although in neither case are the initial values of $d_{z(1)}$ very different. This illustrates the effects of the loss of information.

When the observer used a small number of rating-scale categories, the addition of the first few replications resulted in smaller increases in d_z than were obtained when large numbers of categories were used. However, the addition of the last few replications led to relatively <u>greater</u> increases. This can be seen to some extent in the results for the continuous distribution in Figure 20. The effect was most obvious when theoretical functions were fitted to a series of d_z values obtained from an observer using a two-category rating scale. These functions, generated on the assumption that no information was lost during GOC analysis, tended to fall above the first few d_z values and below the later d_z values. Because of this effect, the use of a small number of rating scale categories could lead to the under- or overestimation of $\underline{k}$, depending on the values of d_z used in equations 4 and 4a.

Figure 21 shows a series of estimates of $\underline{k}$ obtained by entering the $d_{z(1)}$ to $d_{z(N)}$ values into equation 4. The two-category results (circles) vary according to which of the $d_{z(2)}$ to $d_{z(8)}$ values was entered into the equation with $d_{z(1)}$. This was not the case for the 20-category data which, except for one point, were quite consistent and, compared with the two-category data, more accurate.

The most satisfying solution to the problems of estimating $\underline{k}$ when only two rating scale categories are used would probably be to develop a theoretical model which deals specifically with this case. In the

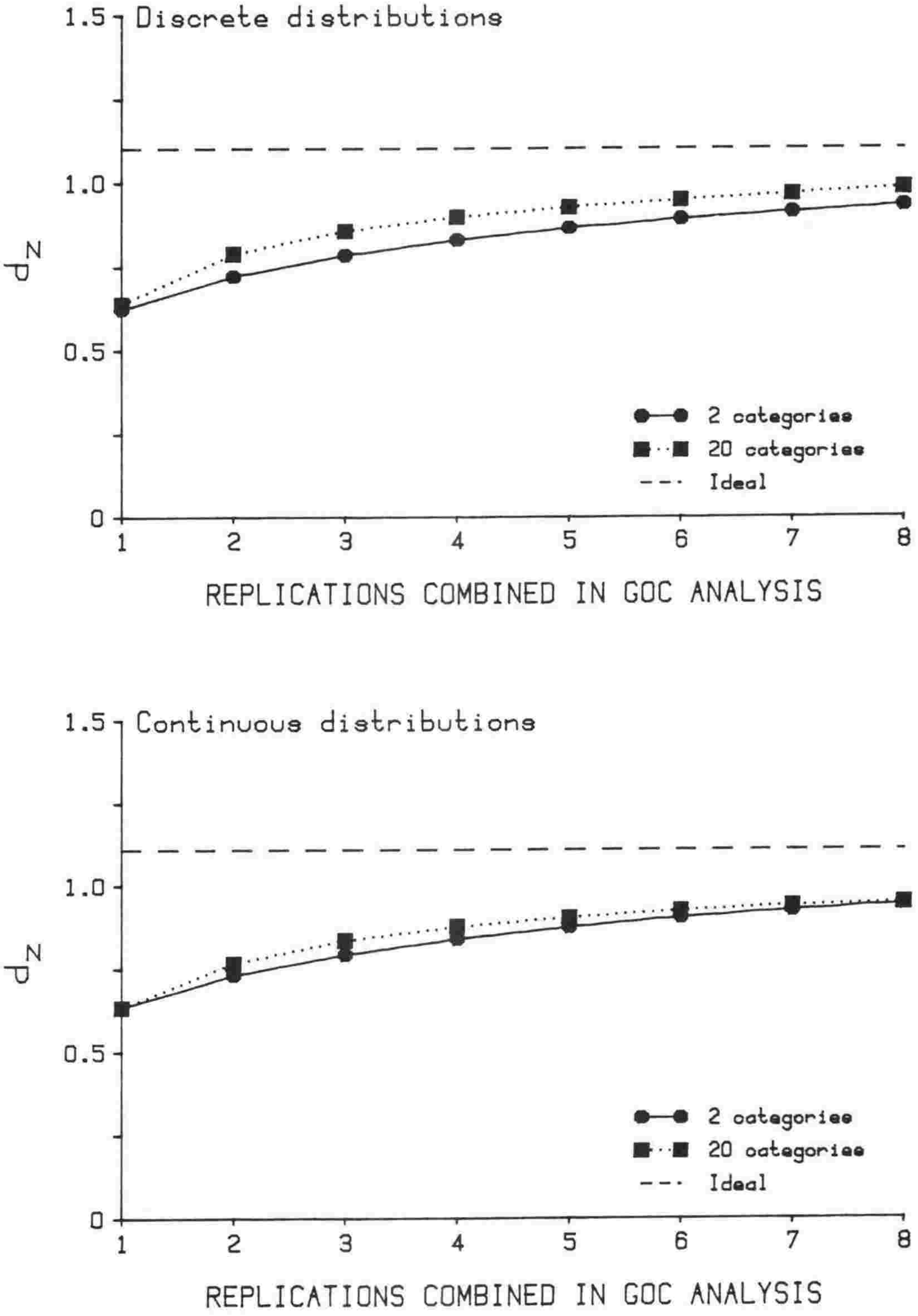


Figure 20. The mean dz values for GOCs based on different numbers of replications, obtained during a simulation in which observers used two- or 20-category rating scales. The nominated parameters were the same as those for the distributions shown in Figure 5.

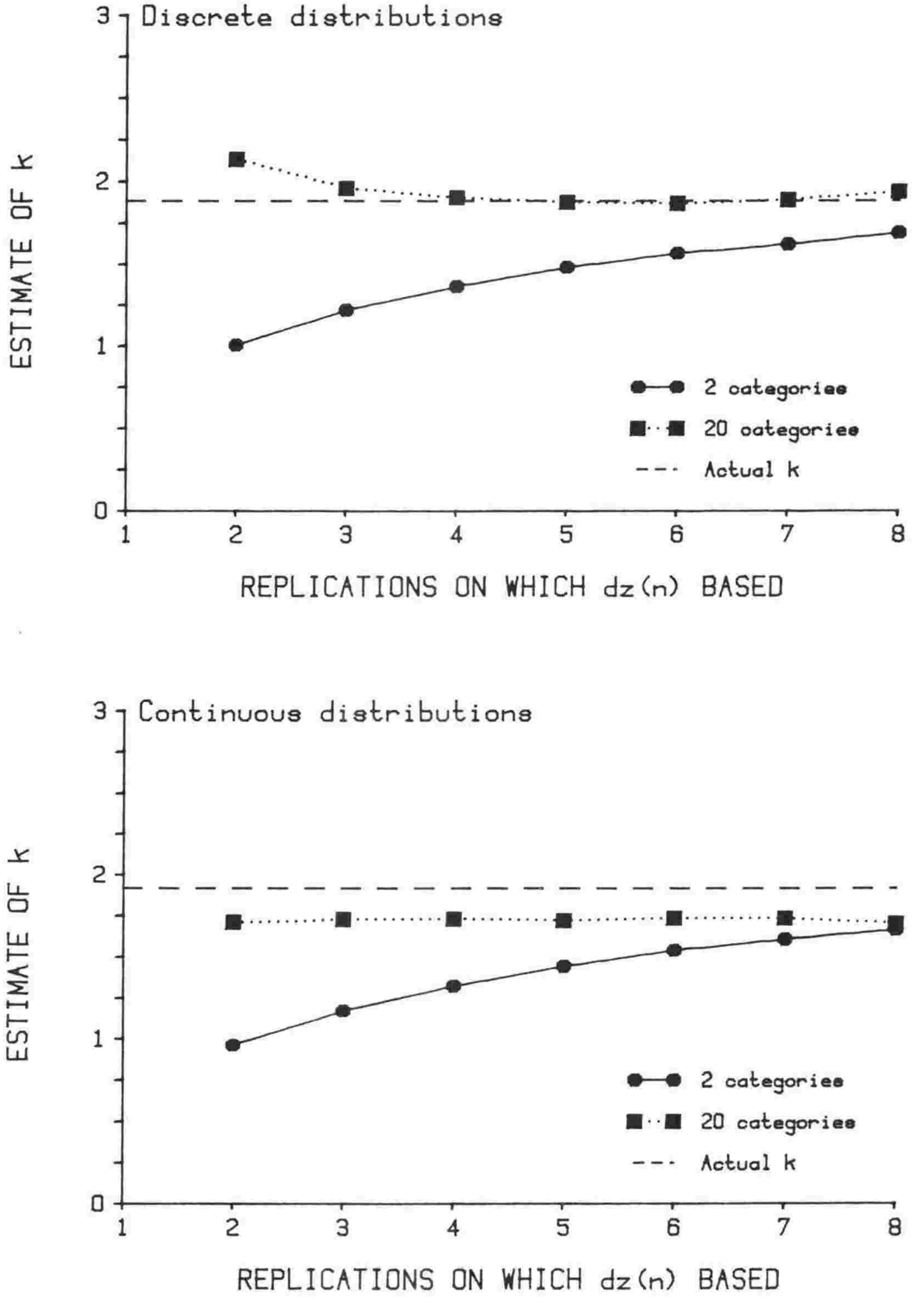


Figure 21. Estimates of k obtained from simulations in which observers used two- or 20-category rating scales. The estimates were based on comparisons of dz(1) with dz(2) to dz(8). The nominated parameters of the underlying normal unequal variance distributions were the same as those for the distributions shown in Figure 5.

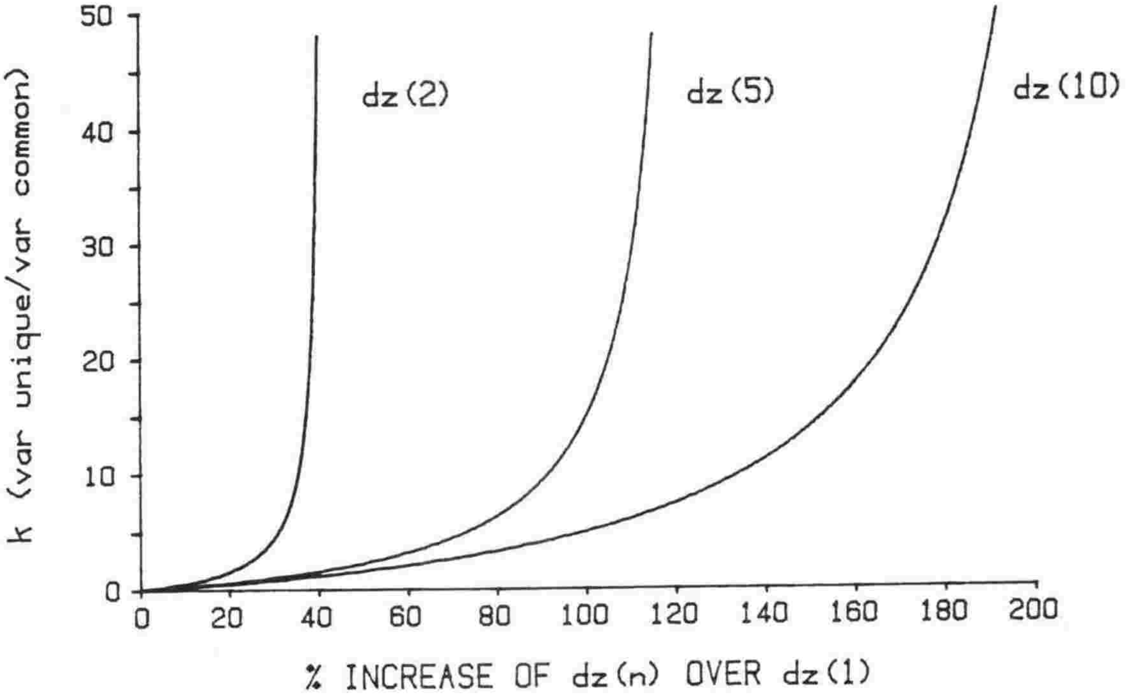
absence of this solution, the less elegant approach of comparing the results obtained from the pigeons with a series of simulation results, for which k was known, was adopted. As with the building of a theoretical model, this approach is restricted by the assumptions which must be made about the underlying distributions. When the method was applied to the pigeon data, these restrictions were not critical. To anticipate, the results of simulations in which there was no common noise (k very large) proved to be the most useful baselines in the later work with pigeons.

Errors in Estimating k

A final point regarding the estimation of k by means of equations 4 and 4a concerns the effects of error in the d_z values on the estimate of k. Investigation showed that at higher levels of k quite small variations in $d_{z(i)}$ and $d_{z(n)}$ could produce marked changes in the value of k.

The reason for this can be seen in Figure 22. This shows the relation between percentage increases in initial d_z values and the estimates of k derived from equation 4a. The upper graph gives the results for pairings of $d_{z(1)}$ with $d_{z(2)}$, $d_{z(5)}$, and $d_{z(10)}$, and the lower graph the results for pairings of $d_{z(2)}$ with $d_{z(3)}$, $d_{z(5)}$, and $d_{z(10)}$. To take an example from the upper graph: if the d_z value for a GOC curve based on 10 replications $[d_{z(10)}]$ was twice that for a single replication $[d_{z(1)}]$, an increase of 100%, equation 4 would give a k value of 5.

Although the functions increase gradually for small percentage increases in d_z, beyond certain levels very small increases in d_z values are associated with very large increases in k. Around these levels [e.g., a 41% increase in $d_{z(2)}$ relative to $d_{z(1)}$, or a 21% increase in $d_{z(3)}$ relative to $d_{z(2)}$] estimates of k are particularly vulnerable to



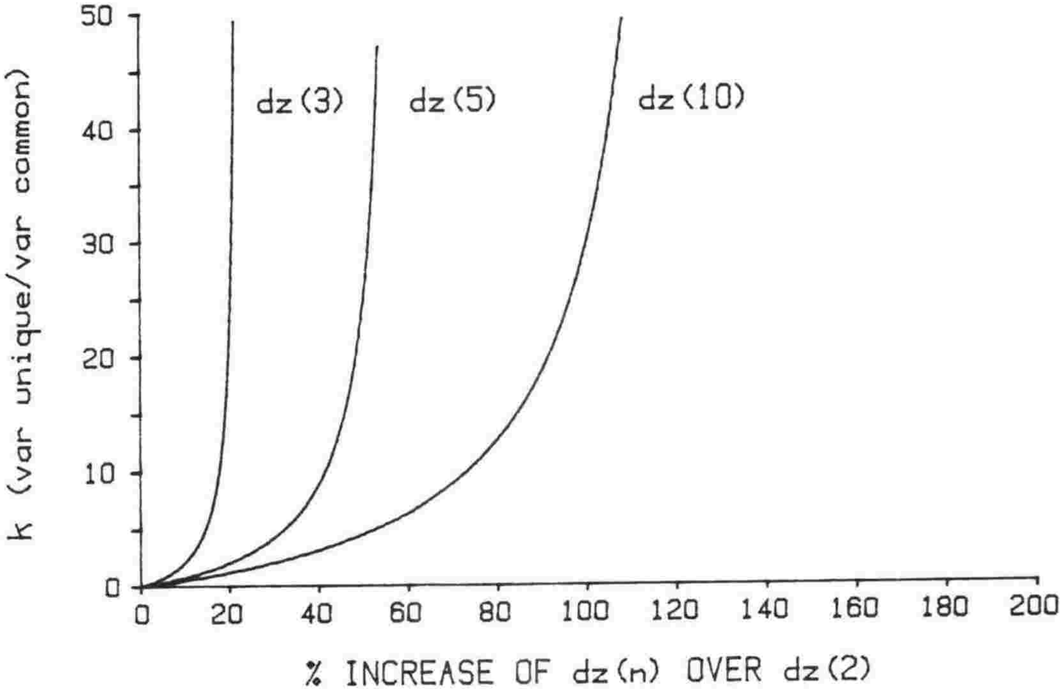


Figure 22. The theoretical relationship between k and the percentage increase in dz(n) relative to dz(1) [upper graph] and dz(2) [lower graph].

error in measures of d_z. Since large increases in d_z will tend to occur when k is large, large values of k will be difficult to estimate with any accuracy.

Because the graphs suggested that there was less room for error when $d_{z(2)}$ rather than $d_{z(1)}$ was used as the base value in equation 4a, $d_{z(1)}$ was used in analysing the pigeon results except when the estimates of $d_{z(1)}$ were subject to systematic error.

The Estimation of k when the Underlying Distributions are Known

In simulations, the characteristics of the distributions of both the common and unique noise were known. In the first and second experiments with pigeons the characteristics of the common noise distributions were known, but those of the unique noise distributions were not. However, if the unique noise added on both standard and comparison trials is assumed to be sampled from one normal distribution of unique noise with mean zero, the following expression holds (Appendix G):

$$\underline{k} = [d_{z(n)}/d_{z(1)}]^2 - 1 \quad , \qquad (10)$$

where $d_{z(n)}$ is the value for the unknown underlying distributions, and

$d_{z(1)}$ gives observer's sensitivity in the absence of GOC analysis.

This k can be checked against estimates obtained with other methods.

Summary

Two main problems were encountered in the estimation of k, the ratio of the variances of the unique and common noise. One was that, when two-category rating scales were used, only one hit and false alarm rate was obtained for each replication. If the ROC and GOC slope departed from unity, the value of $d_{z(1)}$ became uncertain. When estimating <u>k</u>, this problem could be sidestepped by using only $d_{z(2)}$ to $d_{z(N)}$. For other purposes reasonably accurate estimates of $d_{z(1)}$ could be obtained by extrapolating the mean slopes of the GOCs obtained during all-combinations GOC analysis. This form of analysis had the effect of reducing variations in d_z caused by fluctuations in the observers' performance over replications.

A second problem in estimating <u>k</u> was also associated with the use of rating scales with relatively few categories. Because of information loss, increases in d_z as replications were added in GOC analysis tended to be either attenuated or amplified, so that the <u>k</u> values could not be taken at face value. The solution to the problem was to compare the results obtained in the experiments with the results of computer simulations for which <u>k</u> was known. This method was applied in the analysis of the experiment described in the next chapter. Several limitations of the method are discussed there.

CHAPTER 6

THE ESTIMATION OF $\underline{K}$ FROM PIGEON DATA WITH KNOWN DISTRIBUTIONS

The known distributions used in this experiment were constructed in the same way as those used in the first experiment with pigeons. However, in this case discrete analogues of normal distributions were used, so that they had some of the properties of normal distributions.

A difference from the earlier work was that the birds' decisions were reinforced according to the distribution (standard or comparison) from which each element was drawn. In the first experiment, decisions were reinforced in relation to an experimenter-defined cutoff so that, at a given frequency, reinforcement occurred for only one type of decision. In the present experiment, identical values of the evidence variable could be associated with the reinforcement of different decisions.

Although observers in experiments in which the underlying distributions are not known may be encouraged to adopt stricter or more lax criteria, these can only be specified in terms of the frequency of "yes" or "no" decisions (e.g., a fixed false alarm rate), since the experimenter does not usually know what value of the evidence variable has occurred on a particular trial.

Method

The apparatus and methods of signal generation were the same as those in the first experiment (Chapter 3). Changes in the procedure are described below, together with the details of the distributions.

Subjects

Birds 7, 8 and 18 started on the experiment, but Bird 7 was withdrawn because of illness.

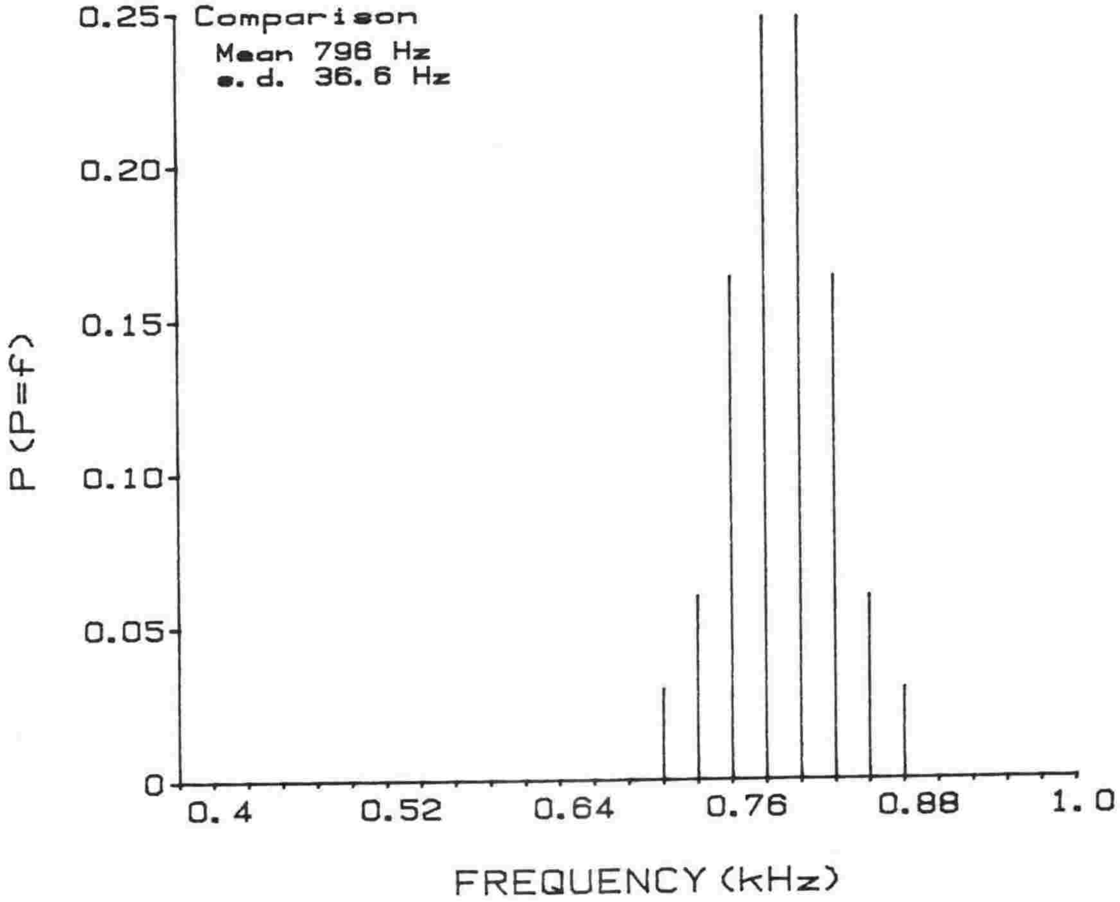
Probability Distributions

The probability distributions, which were constructed from the same sinusoids used previously, are shown in Figure 23. They are rough discrete analogues of normal unequal variance distributions. The probabilities of the sinusoids were approximately the same as the heights of ordinates of the standard normal distribution at similar values of z.

The comparison distribution was made up of eight elements ranging in frequency from 712 Hz to 880 Hz. It had a mean of 796 Hz and a standard deviation of 36.6 Hz. The standard distribution was made up of all 26 elements from 400 Hz to 1000 Hz and had a mean of 700 Hz and a standard deviation of 117.6 Hz.

In the experiment, each probability distribution was defined by 202 stimulus presentations [i.e., P(comparison)=P(standard)], so that 404 trials made up a complete presentation of both distributions. One replication consisted of four 101-trial sessions.

The trial sequences were made up in the same manner as those in the first experiment, except that the restriction on the number of consecutive standard and comparison trials was raised to five.



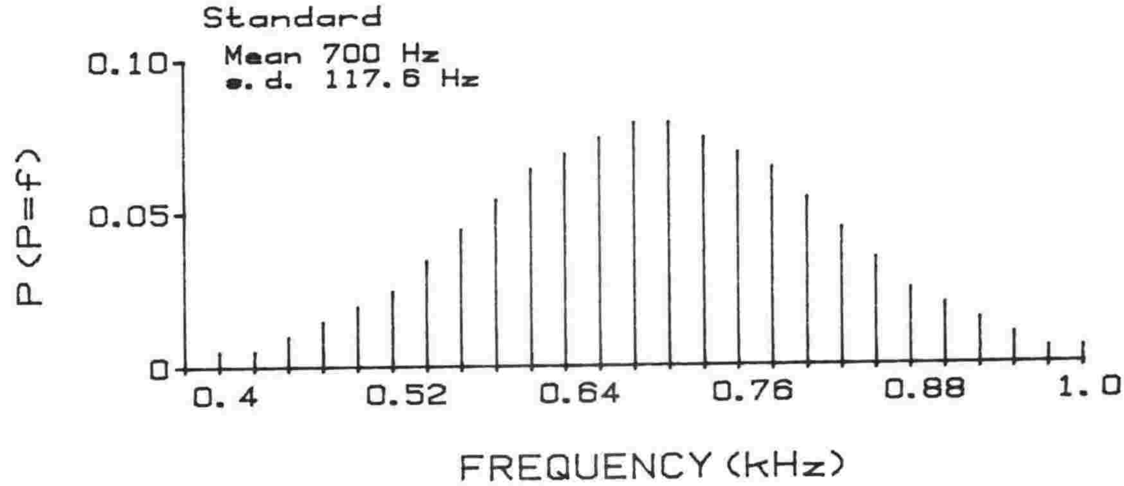


Figure 23. The analogues of normal unequal variance distributions used in the second pigeon experiment.

Procedure and Training

Correction trials were not used in this experiment, since they would have increased the number of trials which involved apparently inconsistent reinforcement. An advantage of not using correction trials was that more trials could be run in each session without satiation.

Following the first experiment, training with the triangular distributions continued, but correction trials ceased, the number of trials was increased to 100, the run restriction was made more liberal, and non-reinforcement of correct responses was gradually introduced. Over the course of 12 sessions, the probability of reinforcement for hits and correct rejections was reduced from 1 to 0.75. This was done to lessen any disturbing effects of apparently inconsistent reinforcement during the experiment.

The new distributions were then introduced and nine replications (36 sessions) were run for each bird. Twenty warm-up trials preceded the 101 experimental trials in each session. For the first three replications, the warm-ups consisted of a haphazard selection from the trials which were to be run that day, and there was no correction. However, performance with this procedure tended to be inconsistent, so correction trials were reintroduced following the third replication. The first and second pair of warm-up trials always involved one frequency (832 Hz) for which right-hand key ("yes") responses were reinforced and one frequency (640 Hz) for which left-hand key ("no") responses were reinforced. The remaining 16 warm-up trials consisted of a haphazard selection from the following frequencies: 832, 808, 784, 688, 664 and 640 Hz. During the first three trials, "yes" responses were followed by food, while during the second three, only "no" responses were reinforced.

Results

GOC Analysis

The results of trials were combined over replications as described in Chapter 1.

The performances of the two birds on each of the nine replications are shown by the circles in Figure 24 and Figure 25. Table 7 (Appendix C) shows the corresponding hit and false alarm rates.

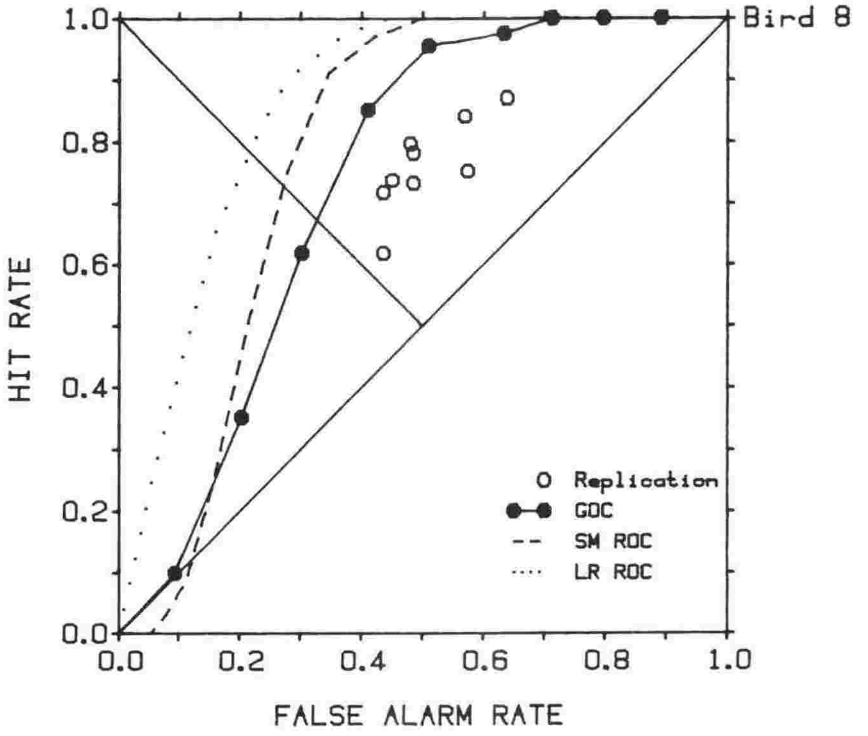
Nine-replication GOC curves for the two birds are shown by the unbroken lines. These may be compared with the ROC curves for the model, which are shown by the broken lines. Two ROC curves are shown. The dashed one is what would be expected for an observer whose decision axis was stimulus magnitude (sinusoid frequency). The dotted ROC curve represents the performance of an observer who used likelihood ratio to order the values of the evidence variable.

The birds made choices according to sinusoid frequency, as would be expected from their training. This was reflected in the GOC curves--especially that for both birds combined (18 replications in all)--which are shown in Figure 24 and Figure 25.

The GOC curves are also shown plotted on normal-normal coordinates in Figure 26 and Figure 27 respectively. In each case the points are well fitted by a straight line, as is the stimulus magnitude ROC for the model, which is also shown in the graphs.

The data in Table 1 show that the slope of the ideal ROC was approximately what would be expected from the standard deviations of the distributions (σ_S/σ_C = 117.6/36.6 = 3.2), and that the d_z value for the ideal ROC was 1.08.

For each bird, the d_z value for the GOC was clearly higher than that based on pooled or mean data for individual replications; in each case $d_{z(9)}$ exceeded all of the individual $d_{z(1)}$ values. The 18-replication



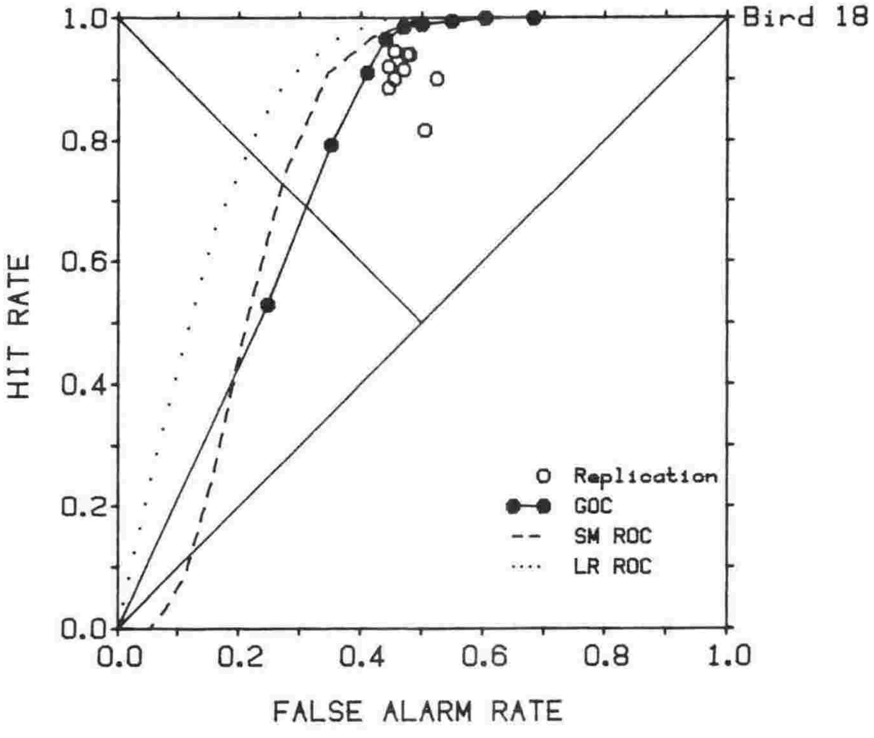


Figure 24. GOCs and individual replication results obtained in the second pigeon experiment.
The dashed and dotted lines are, respectively, the stimulus magnitude (SM) and likelihood ratio (LR) ROCs for the analogues of normal unequal variance distributions.

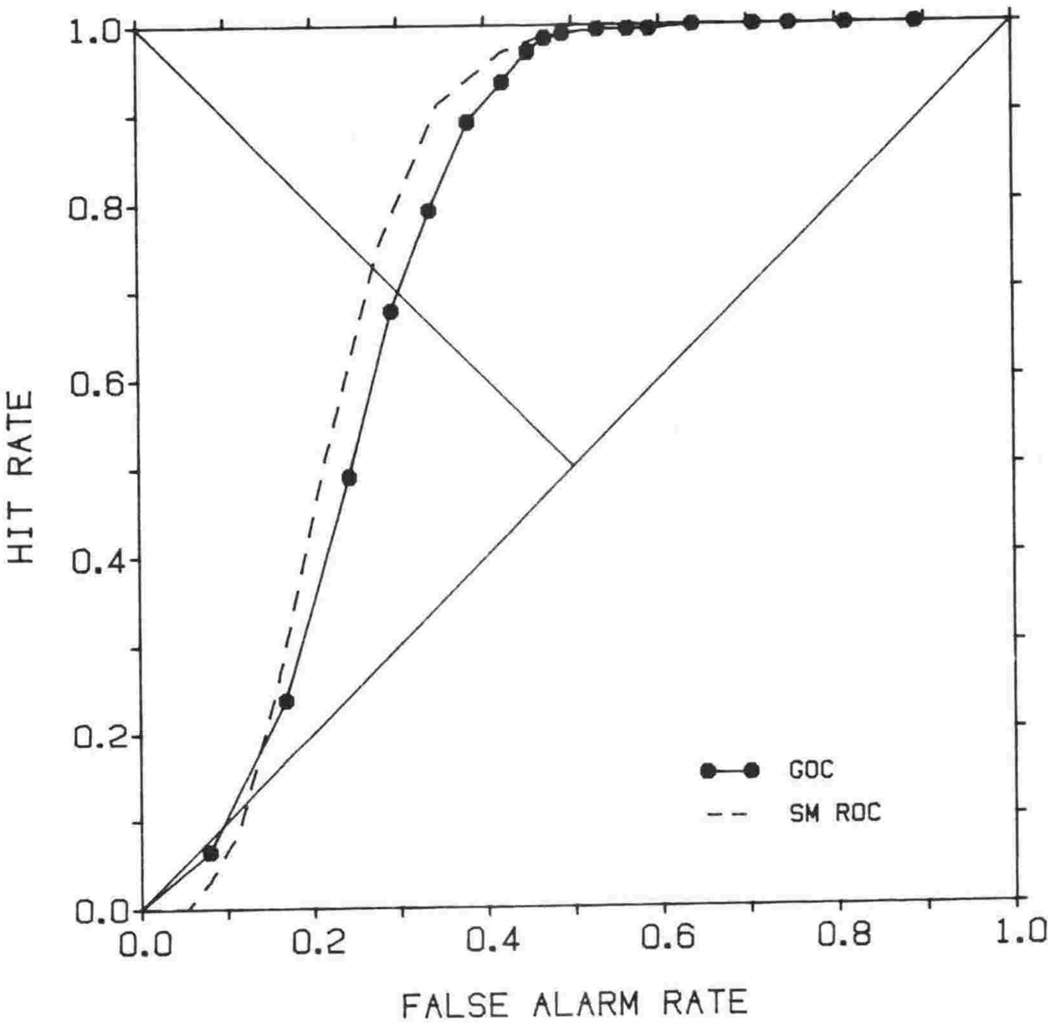
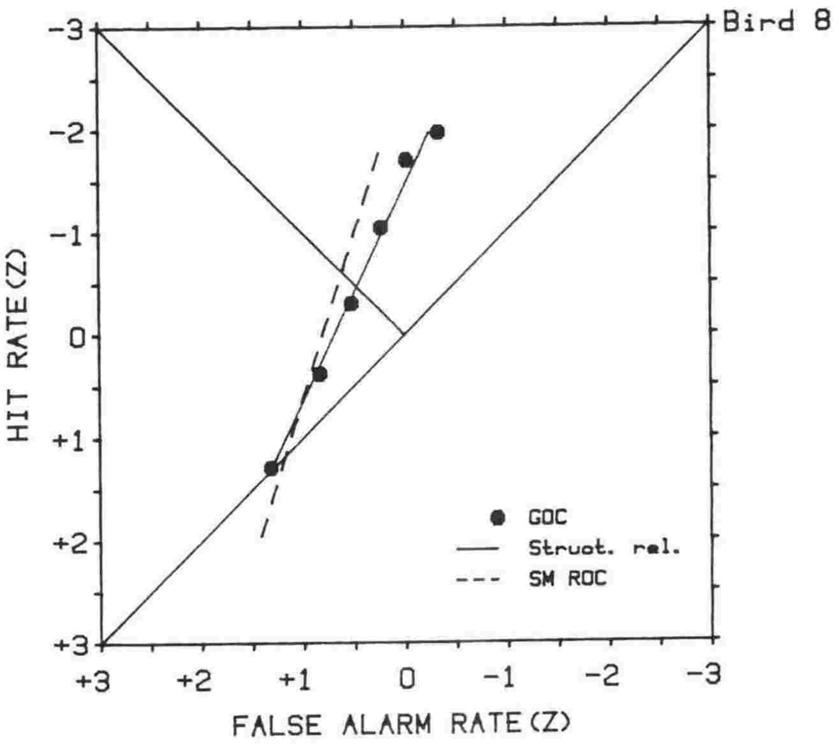


Figure 25. The 18-replication GOC for both birds combined, obtained in the second pigeon experiment.



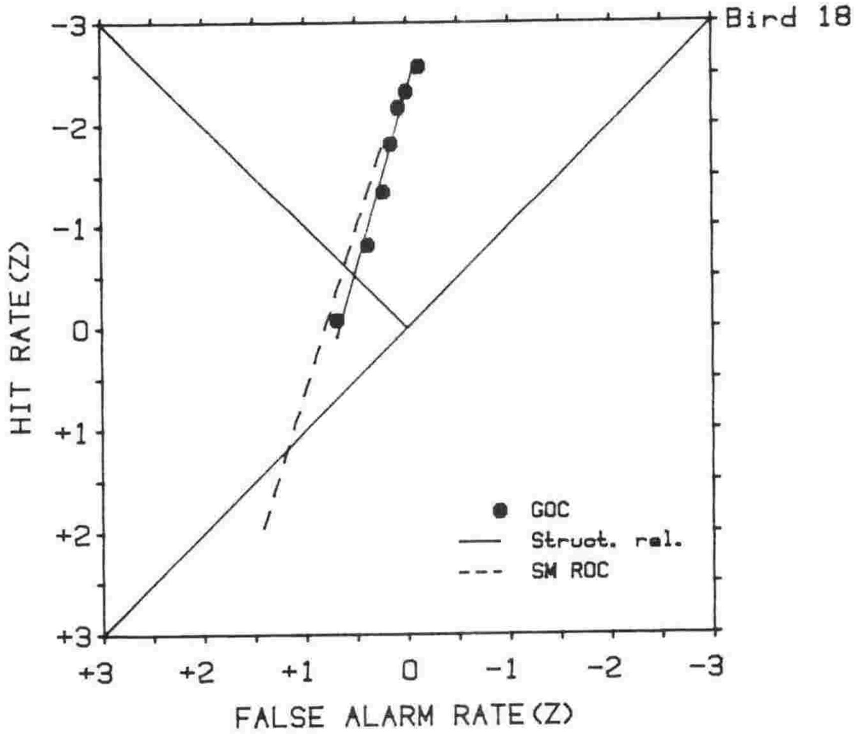


Figure 26. Individual GOCs obtained in the second pigeon experiment plotted on normal-normal coordinates.

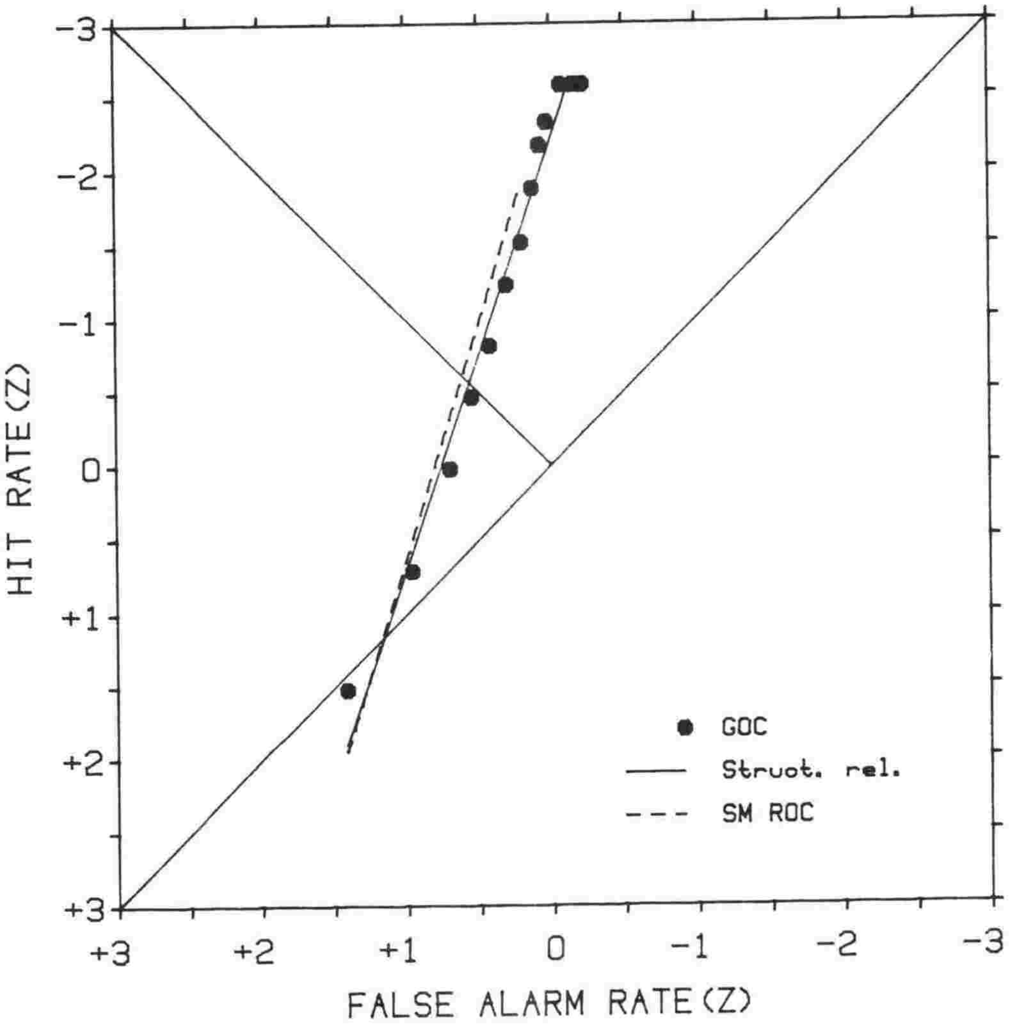


Figure 27. Eighteen-replication GOC for both birds combined, obtained in the second pigeon experiment, plotted on normal-normal coordinates.
Points representing values greater than 2.5, though plotted, were not included in the fitting.

TABLE 1

The slope, d_z values and fits for ROCs and GOCs in the second pigeon experiment (normal unequal variance)

	Slope	d_z	r^2
Ideal ROC	3.180	1.079a	0.999
<u>Bird 8</u>			
ROC -pooled data	1.437	0.556	-
-mean data	"	0.571	-
GOC	2.100	0.876	0.988
<u>Bird 18</u>			
ROC -pooled data	2.748	0.731	-
-mean data	"	0.746	-
GOC	3.521	0.894	0.974
<u>18-replication</u>			
<u>GOC</u>	2.904	1.013	0.977

a The dz value for the ideal ROC obtained by direct calculation was 1.1012. The difference probably arose because the distributions were not normal.

GOC was the best approximation to the ideal ROC; the d_z value for this curve was higher than that for either of the individual GOCs.

The slope of Bird 18's GOC was greater than that of the ideal ROC (although the d_z value for the ROC was smaller than that for the GOC). This can be seen quite clearly in Figure 26. The reason for the sharper slope is considered at the end of the next section.

<u>The Estimation of k</u>

Figure 28 shows the d_z values obtained from the all-combinations GOC analysis. For each bird, $d_{z(1)}$ is the mean of the values for the nine replications. As can be seen in Table 1, the mean values were slightly higher than those based on pooled data. The ROC slope was estimated from the data shown in Figure 29.

The estimates of <u>k</u> obtained from equation 4a are shown by the points in Figure 30. The lines in this figure are from simulations based on

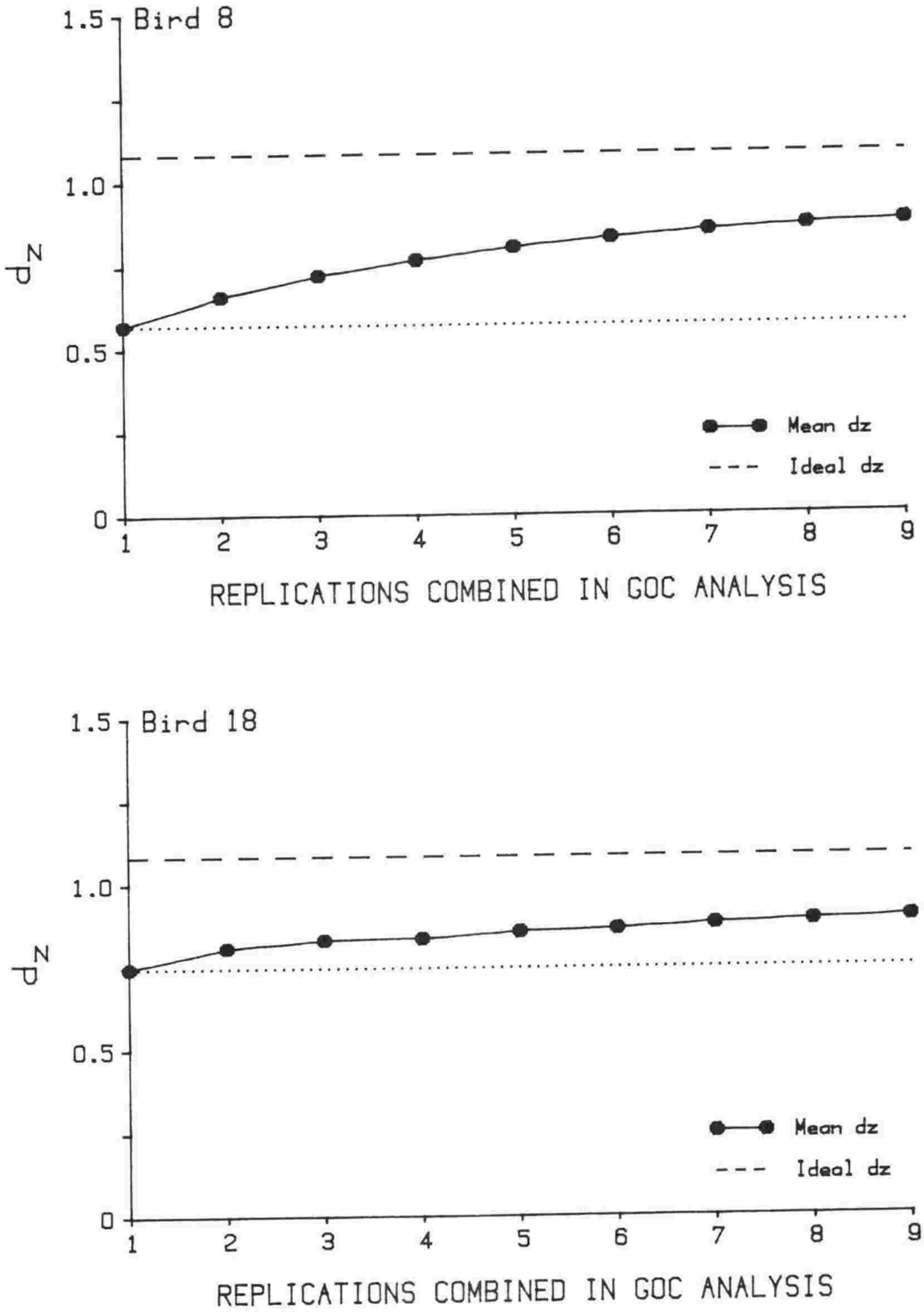
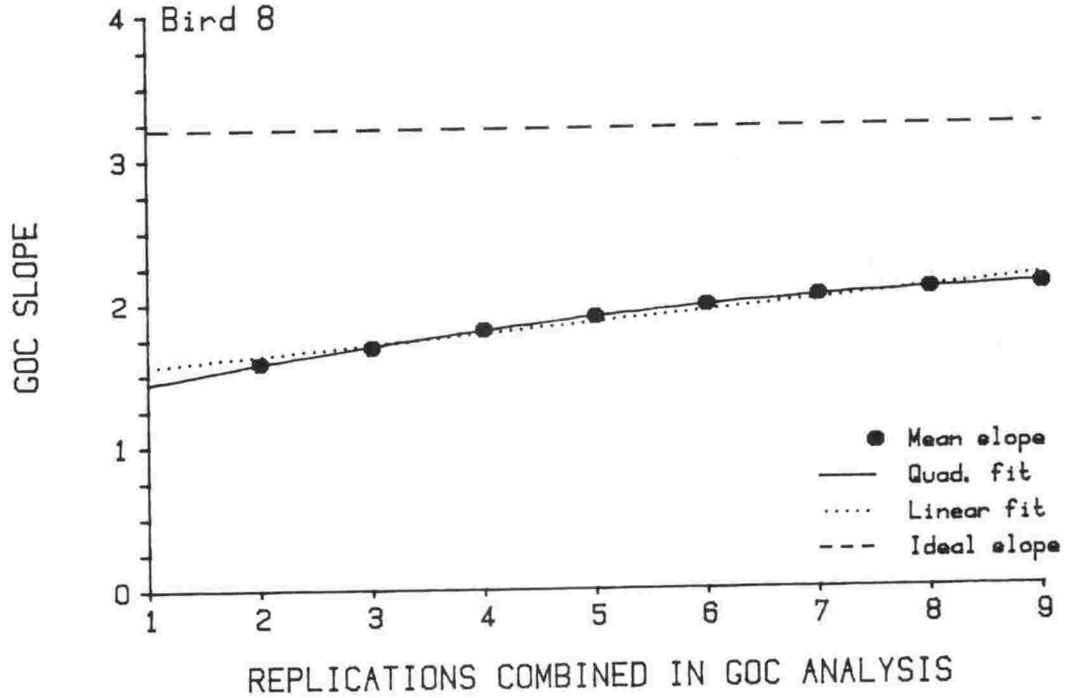


Figure 28. The mean dz values for GOCs based on different numbers of replications, obtained in the second pigeon experiment. The dashed line shows the dz for the stimulus magnitude ROC.



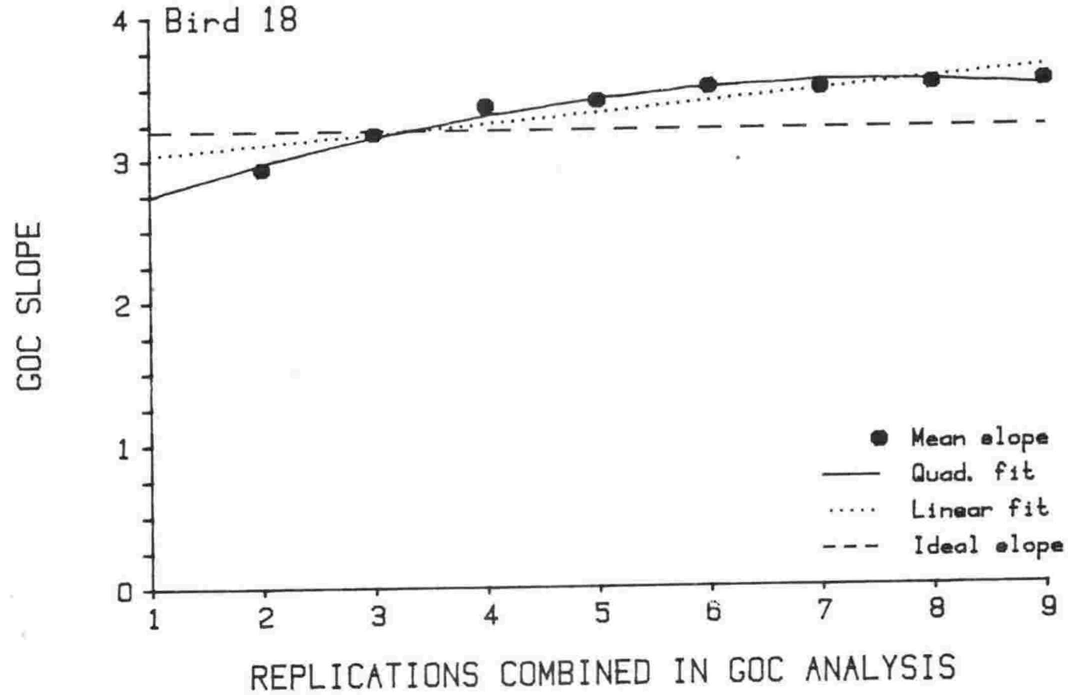


Figure 29. The mean slopes of GOCs based on different numbers of replications, obtained in the second pigeon experiment.
The dashed line is the slope of the stimulus magnitude ROC.

standard and comparison distributions identical to those used in the experiment.

These simulations consisted of nine 1212-trial (3 x 404) replications. The more than 10,000 samples of unique noise in each simulation were drawn from a normal distribution with a mean of zero, and the cutoff was set at the mean of the standard distribution (700 Hz).

Because of uncertainty about the slope of the ROCs for individual cycles, $d_{z(2)}$ was used as the base figure in equation 4a.

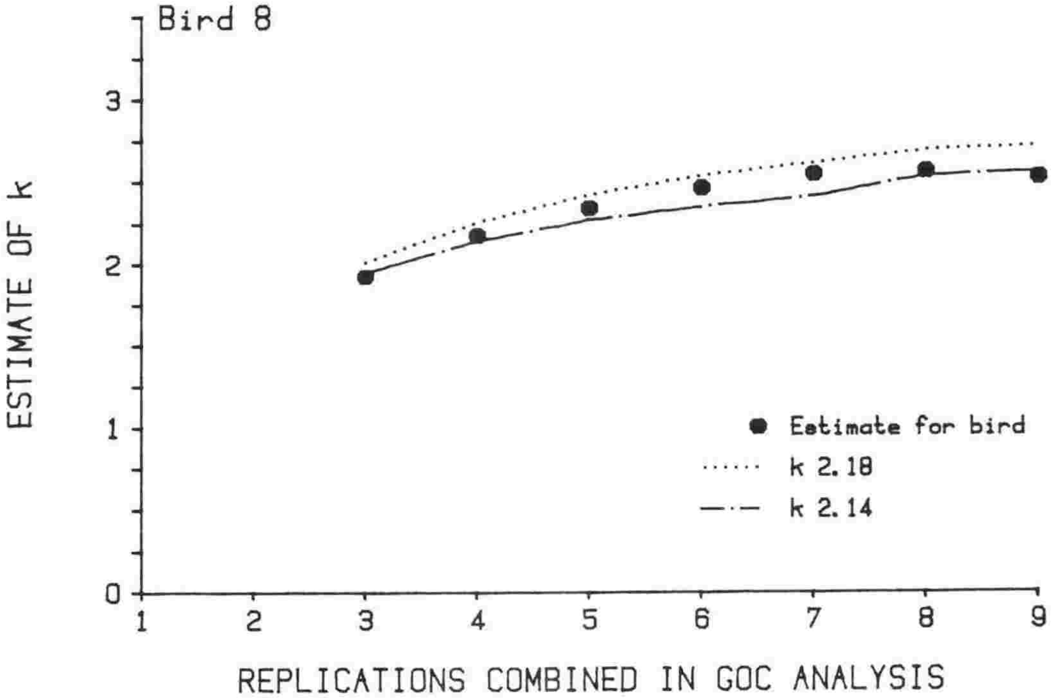
The results suggest that for Bird 8, $\underline{k}$ was around 2.1, while for Bird 18 it was around 0.5. These estimates may be compared with the values obtained from equation 10, namely:

$$\underline{k} = [d_{z(n)}/d_{z(1)}]^2 - 1 \quad . \qquad (10)$$

The estimates from equation 10 were 2.6 for Bird 8 and 1.1 for Bird 18. The first value agrees tolerably well with the previous result, but that for Bird 18 suggests that the first method led to a significant underestimate of the $\underline{k}$ value.

Further investigation showed that the results for Bird 18 were what would be expected if the effective common noise variance of the standard distribution was around twice the value for the constructed distribution (which was $\{117.6\}^2$). This is of a piece with the finding, noted earlier, that the GOC for Bird 18 was steeper than the ideal ROC. The discrepancy between the two estimates of $\underline{k}$ arose because equation 10 is based on the assumption that the difference between $d_{z(1)}$ and the d_z value for the model distributions is entirely due to unique noise.

Common noise over and above that included in the signals can arise in various ways. Several possible sources were eliminated in this and the other pigeon experiments. The trials were run in different haphazard



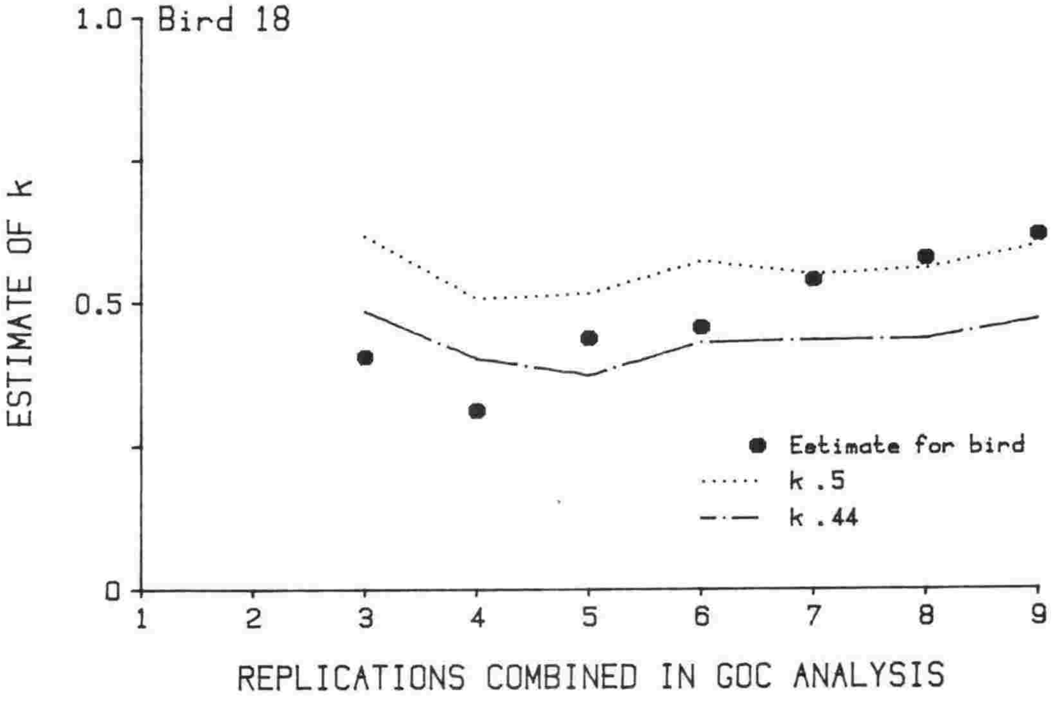


Figure 30. Estimates of k obtained in the second pigeon experiment and in simulations in which the true value of k was known. The estimates were based on comparisons of dz(2) with dz(3) to dz(9).

orders within each replication so that, like the low-level masker (and any other extraneous noise), sequential dependencies contributed to the unique rather than common noise. Examination of the data did not reveal any other sources of common noise.

However, comparison of the data for the two birds led to an explanation in terms of the high initial performance and cutoff placement of Bird 18. This is best explained by reference to Figure 31. The figure shows the distributions of responses underlying the nine-replication GOC curves shown in Figure 24 and Figure 26. The abscissa shows the random variable created by adding together the responses over replications in GOC analysis. For convenience, the variable, whose values in fact ranged from nine to 18 over nine replications ("yes" was stored as 1 and "no" as 2) has been translated into the number of "yes" responses. Thus, if a bird made a left-hand key response--"no"--every time a nominally identical stimulus element was presented over the nine cycles, the number of "yes" responses would be zero.

The difference between the distributions for the two birds is clearly apparent. The distribution for Bird 18's responses on standard trials is bimodal, whereas that for Bird 8 is more or less flat. Similarly, the distribution for Bird 18's responses on comparison trials is much more skewed than that for Bird 8. These differences are partly due to the relatively high initial performance of Bird 18 and partly due to the placement of Bird 18's cutoff. If the cutoff had fallen somewhat more towards the high frequency end of the distribution, the response distribution for comparison trials would not be so dramatically skewed; rather it would tend to be bimodal like the distributions for standard trials, and the ratio of the standard deviations (and hence the slope of the GOC) would be lessened.

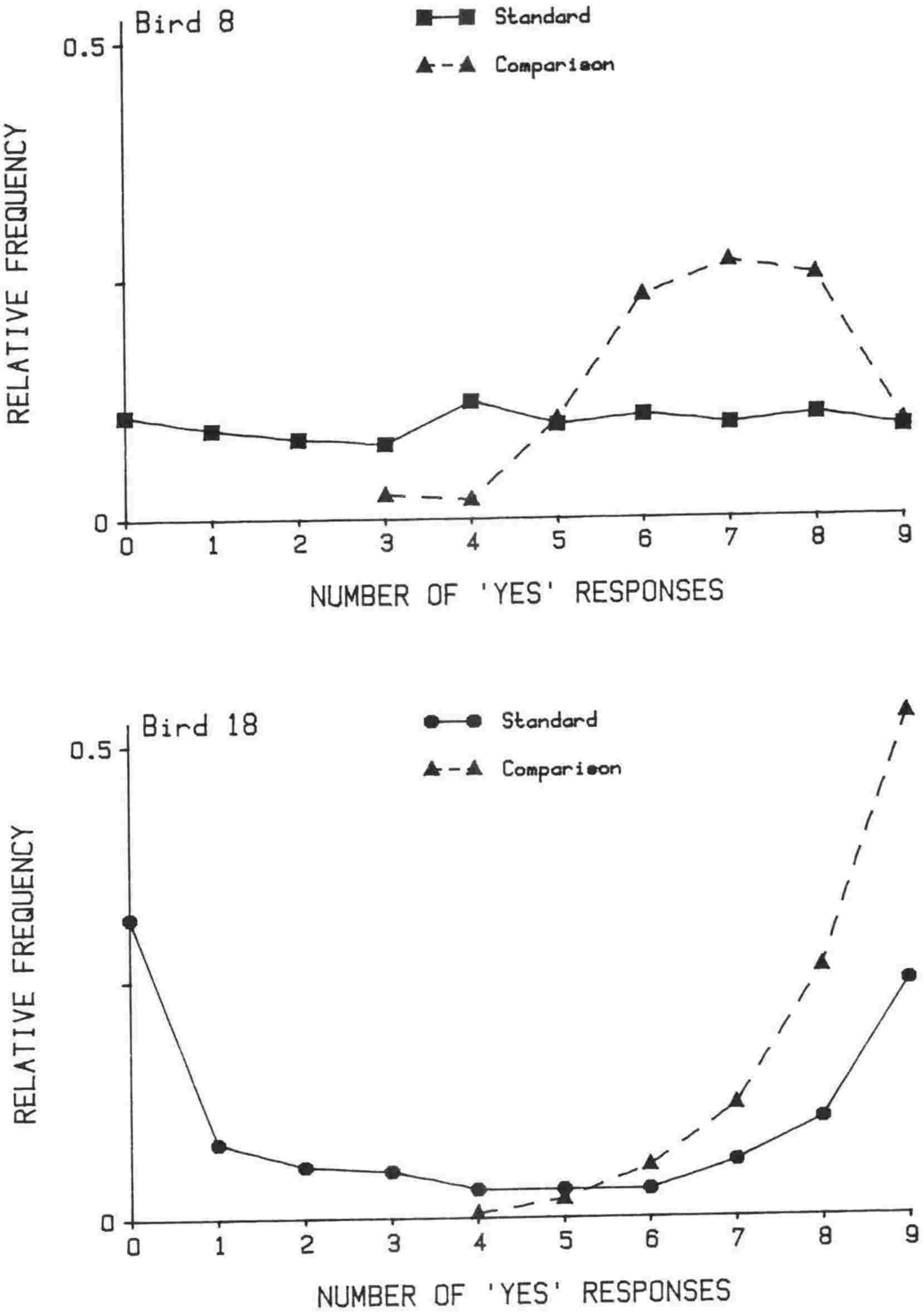


Figure 31. The relative frequency of different numbers of "yes" responses on standard and comparison trials in the second pigeon experiment.

While the effect described above could occur in any GOC experiment, it is probably exaggerated in the present case by the rather truncated comparison distribution.

Conclusion

The GOC analysis was again successful in removing the effects of unique noise, leading to estimates of d_z which more and more closely approximated that for the ideal ROC.

Comparison of the birds' results with those of a series of computer simulations led to a reasonably accurate estimate of <u>k</u> for Bird 8, but not for Bird 18. The result for the latter bird suggested that common noise had been added to the standard distribution. Investigation showed that the joint effect of Bird 18's high level of performance and cutoff placement led to a higher GOC slope.

CHAPTER 7

INTRODUCTION TO AN APPLICATION OF GOC ANALYSIS TO FREQUENCY DISCRIMINATION IN PIGEONS

The main aims of the first two experiments with pigeons were to establish the applicability of GOC analysis to the study of hearing in pigeons and to explore some methodological and technical problems. To this end, known underlying distributions were used. The experiments described in this and the next chapter investigated the effects of GOC analysis on measures of performance obtained under more usual conditions: the birds were trained to discriminate between one standard frequency and two or more comparison frequencies, and the nature of the underlying distributions was, of course, unknown.

In the first of these experiments the standard frequency was 450 Hz and there were five comparison frequencies (500, 600, 700, 800 and 900 Hz). In the second experiment, the same standard frequency was used, but there were only two comparison frequencies, 650 and 900 Hz, with the latter occurring only occasionally. The aim of this second experiment was to obtain a GOC curve sufficiently well-defined to give some information about the shapes of the distributions underlying the birds' discrimination performances. Inferences about the specific probability functions can never be conclusive, but the elimination of some families of functions might be possible.

Although the general procedure followed in these experiments was similar to that used previously, the stimuli were quite different in that they consisted of repeated presentations of relatively brief pulses of masked sinusoids, separated by longer intervals.

The first part of this chapter discusses the reasons for the use of these stimuli. The second part describes the way in which the digital stimuli used in the experiments were generated, gives details of the subjects, apparatus and general procedures, and provides some information about the ability of pigeons to discriminate auditory frequencies. The experiments themselves are described in Chapter 8.

Some Stimulus Considerations

In the first two pigeon experiments (Chapters 3 and 6) the common noise was provided by the constructed distributions, and an upper limit was placed on the birds' performances. If there is no common noise, use of the GOC technique will lead to perfect performance, since all the noise will be unique, and will be removed. In the third and fourth pigeon experiments, a basis for common noise was provided in the form of bandpass filtered noise. This noise, together with digitally-produced sinusoids of the appropriate frequencies, was digitally recorded, and was thus reproducible both within and over trials.

The need for common noise led to the use of relatively brief presentations of the noise and sinusoid. When observers can sample from different points within signals, all noise becomes unique. To avoid this, signals should be shorter than the integration time of the observers.

Only one study of temporal integration in birds has been published. This was carried out by Dooling (1979) with budgerigars and field sparrows. The results suggested that the integration time for these birds was about 230 msecs, roughly similar to that of human subjects. To be on the safe side for these birds (and presumably for pigeons), a signal pulse would need to be clearly less than 230 msecs. If a series of transients were to be presented, the gaps between the transients would need to be somewhat greater than this.

Since the most desirable procedure would have been to present only one transient per trial, considerable time was devoted to preliminary work with this method. The results of this experimentation suggested that the birds were unlikely to maintain stable discriminations with single transients of the duration, signal-to-noise ratio and frequency differences required, so the final experiments used repeated stimuli.

Work prior to the first experiment[13] led to the use of six frequencies--450 Hz (the standard) and 500 Hz to 900 Hz in 100 Hz steps (the comparisons). The reproducible masking noise had a 3 dB bandwidth of 800 Hz (200 Hz to 1000 Hz), and the signal-power-to-noise spectral density ratio was 40 dB for all frequencies. The absolute levels of the sinusoids ranged from 77.5 dB to 80 dB.

A signal-to-noise ratio of 40 dB is high by human standards. Two factors played a part here. One was the need to fix on a value which would result in stable performances during the experiments. These were more difficult, in terms of the frequencies to be discriminated, than the preliminary training task. Also, for the reasons given earlier, there was a constraint on the durations of the signals and of the gaps between them. Another factor was the likelihood that in the frequency range used, the critical bands of pigeons are considerably wider than those of human subjects. Work with budgerigars (e.g., Saunders, Denny & Bock, 1978) suggested that at around 500 Hz the critical bands of these birds are between two and three times wider than human critical bands in the same region. If pigeons' critical bands are similar to those of budgerigars, the masking noise used in the present experiments would be much more effective in masking tones with the pigeons than it would be with humans.

---

[13] The preliminary work with both single and repeated stimuli is described in Appendix H.

The Generation of the Signals

In these experiments the gated sinusoids and masking noise originated from digital codes. The generation of these codes, which was done before each experiment began, will be described first. The production of signals from the digital codes during each session will then be outlined.

Generation of the Digital Codes

Codes for sinusoids of appropriate amplitude and frequency were calculated directly by the computer. However, the codes for the gated masking noise were derived from analogue generators.[14] The two sets of codes were then combined to provide signal-plus-noise samples. The apparatus used to generate the stimuli is shown in Figure 32.

The masking noise originated from two different Gaussian noise generators, both passive low-pass filtered at 2000 Hz (Butterworth 24 dB/octave). The output of one of these generators was passed through a Krohn-Hite (Model 3550) filter in band-pass mode (Butterworth 24 dB/octave) set at 625 Hz (high pass) and approximately 900 Hz (low pass). The noise sources were then mixed. A second Krohn-Hite filter, set at 390 Hz and 790 Hz (Bessel), filtered the output of the mixer. This arrangement was arrived at after considerable experimentation with different noise sources and filter settings. The characteristics of the noise recorded in the experimental chamber are described in the next section.

During the recording of the digital signals the output of the second Krohn-Hite (Figure 32), instead of being passed to the experimental chamber, was fed via two amplifiers straight to a Hewlett-

---

[14] In order to provide noise which would have a reasonably flat spectrum in the experimental chamber, the noise produced by the equipment had a complex spectrum. This would have been difficult, if not impossible, to produce digitally.

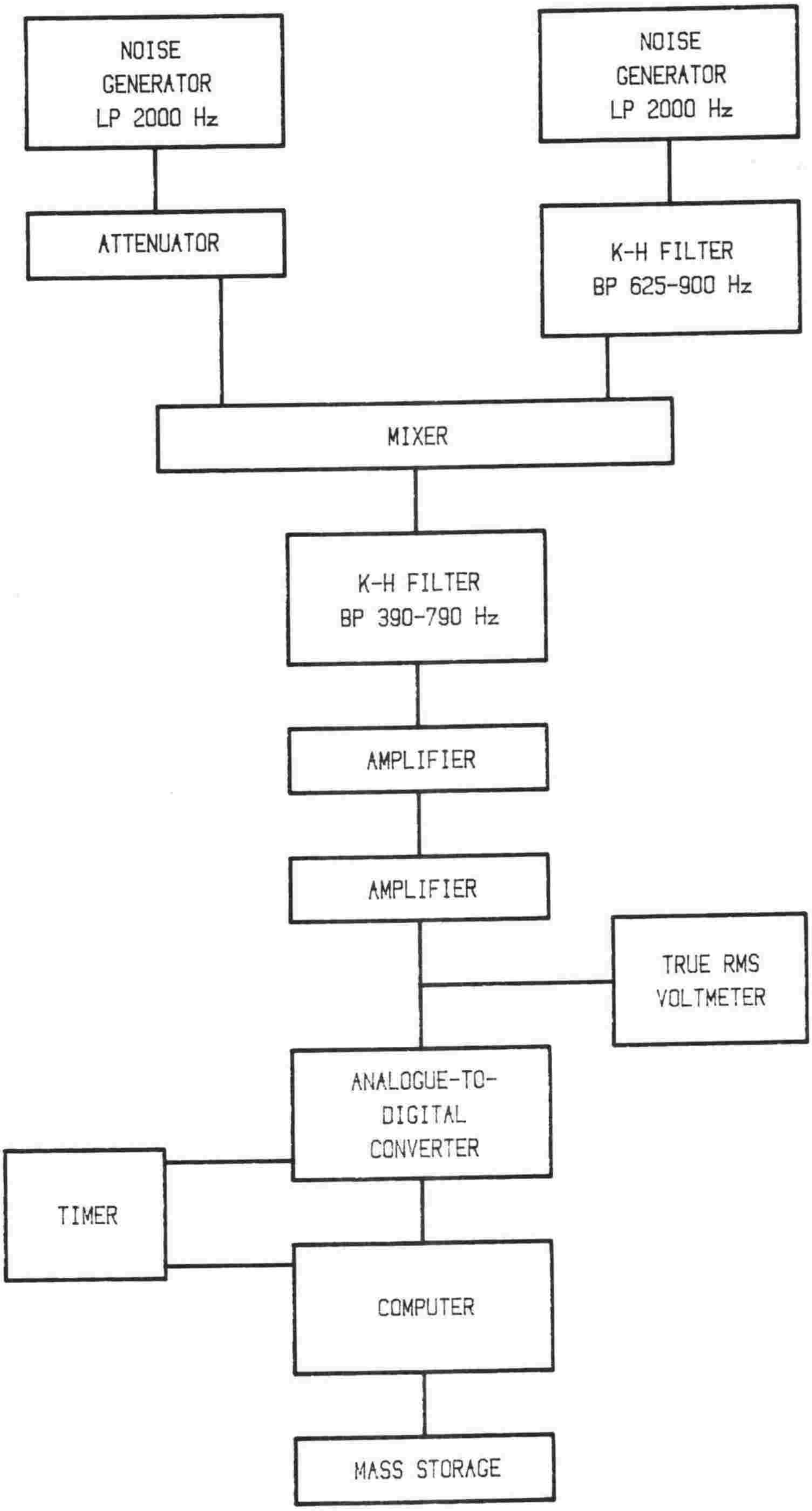


Figure 32. Block diagram of apparatus used to generate and record digital signals for the third and fourth pigeon experiments.

Packard 6940B multiprogrammer containing the 12-bit ADC card (HP 69422A). The multiprogrammer was clocked by a Hewlett-Packard 59308A timing generator controlled by the computer. The computer's program adjusted the level of the noise, added a sinusoid of the appropriate frequency and amplitude, and applied a 40 msec Hanning window (of 101 msec total duration) to the waveform. The resulting digital code was recorded on magnetic tape for subsequent use.

The analogue noise waveform was sampled at a rate of 7246 Hz (138 µsec between samples), and 1024 samples were taken from each signal; consequently the sinusoid also was defined by 1024 points. In conjunction with the filters used during the experimental sessions, the sampling rate was sufficiently high to avoid any problems due to aliassing. Also, given the number of points, the sampling rate produced a signal of approximately the correct length (0.000138 secs x 1024 = 141 msecs). The equivalent rectangular duration was approximately 100 msecs.

A total of 792 independent 141-msec samples of noise were used in each experiment. The 396 samples combined with 450-Hz sinusoids were used in both experiments, while the balance of samples was in each case combined with sinusoids of frequencies appropriate for the experiment.[15] Since the sessions contained 132 trials, the entire set of samples could be run over six sessions. All the codes for one session were stored on a single tape cartridge.

Figure 33 shows an example of the signals produced during the recording of the digital codes, and its spectrum, based on one 1024-point sample. In this case the frequency of the sinusoid was 450

---

[15] In the first experiment, 54 samples were combined with 500-Hz sinusoids, 60 with 600-Hz sinusoids, 72 with 700-Hz sinusoids, 90 with 800-Hz sinusoids, and 120 with 900-Hz sinusoids. In the second experiment, 336 samples were combined with 650-Hz sinusoids and 60 with 900-Hz sinusoids.

Hz. During the experiments the signal and noise were passed via the amplifier and loudspeaker into the experimental chamber. The acoustic characteristics of the room and the chamber altered the shape of the spectrum, as demonstrated in the next section.

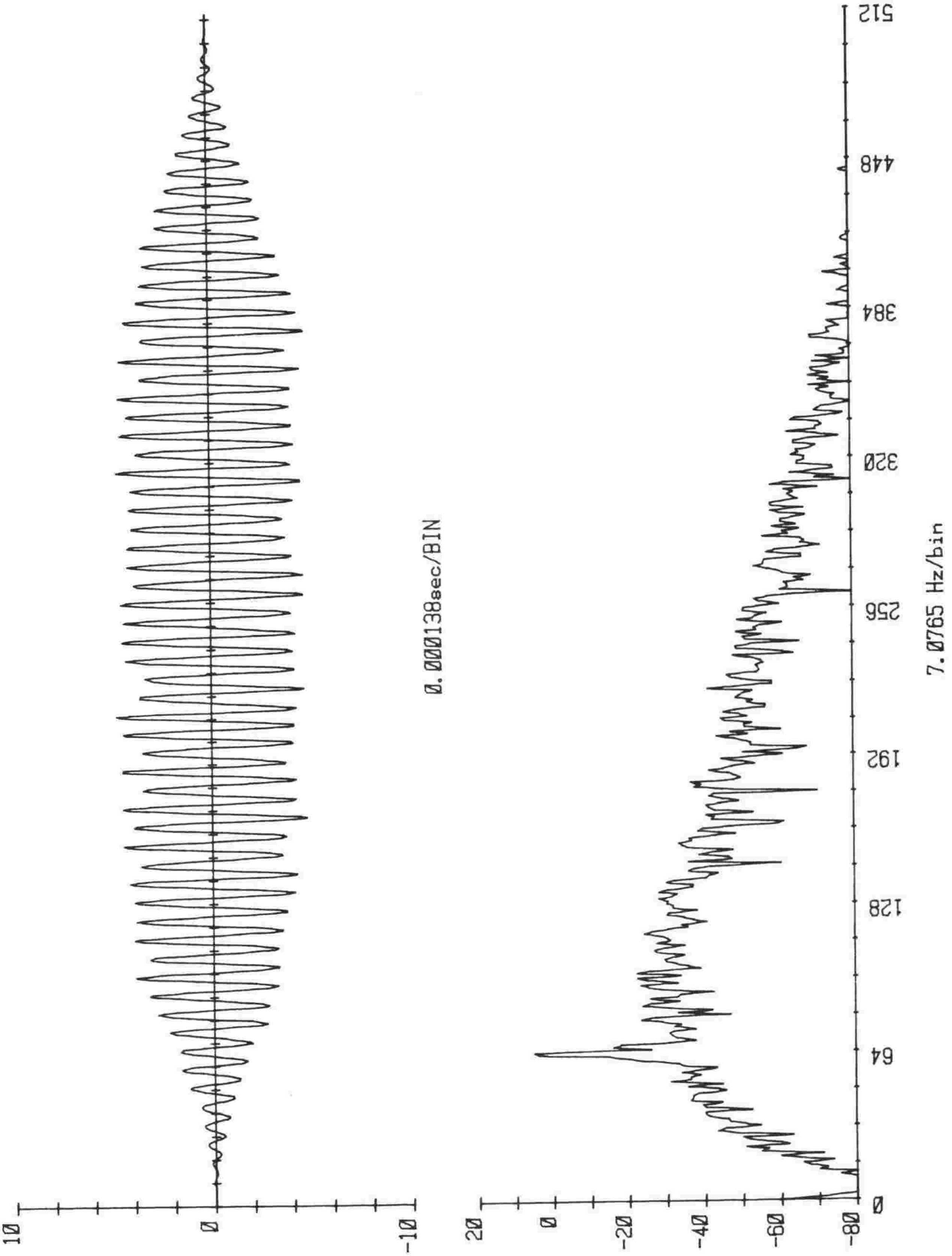


Figure 33. The waveform and spectrum of a 450-Hz sinusoid and masking noise produced during the recording of the digital signals used in the third and fourth pigeon experiments.

Characteristics of the Masking Noise in the Chamber

Figure 34 shows the averaged (vector addition) results of fast Fourier transforms (FFTs) carried out on thirty-two 512-point samples of the noise in the experimental chamber. During this recording, the chamber's whisper fan was operating (hence the peak in the spectrum at around 100 Hz), and the low-level 8000-Hz low-pass noise was also present. For the purposes of the FFT analysis, the high-level masking noise was on continuously, rather than being gated as it was during experimental sessions.

The FFTs were performed on signals from the 12.7 mm B & K microphone (placed in a standard position in the operant chamber) after they had been passed through a 50-Hz high-pass filter (to reduce the effects of ambient noise on the lower part of the spectrum) and a 1250-Hz low-pass filter with a roll-off of approximately 200 dB/octave. An analogue-to-digital converter, which was clocked at 3846 Hz (260 µsec between points), then passed the signals to the computer.

The relatively flat nature of the spectrum from about 240 Hz (bin 32) to 1000 Hz (bin 133), and the fall-off above 1000 Hz, can be seen in the figure.

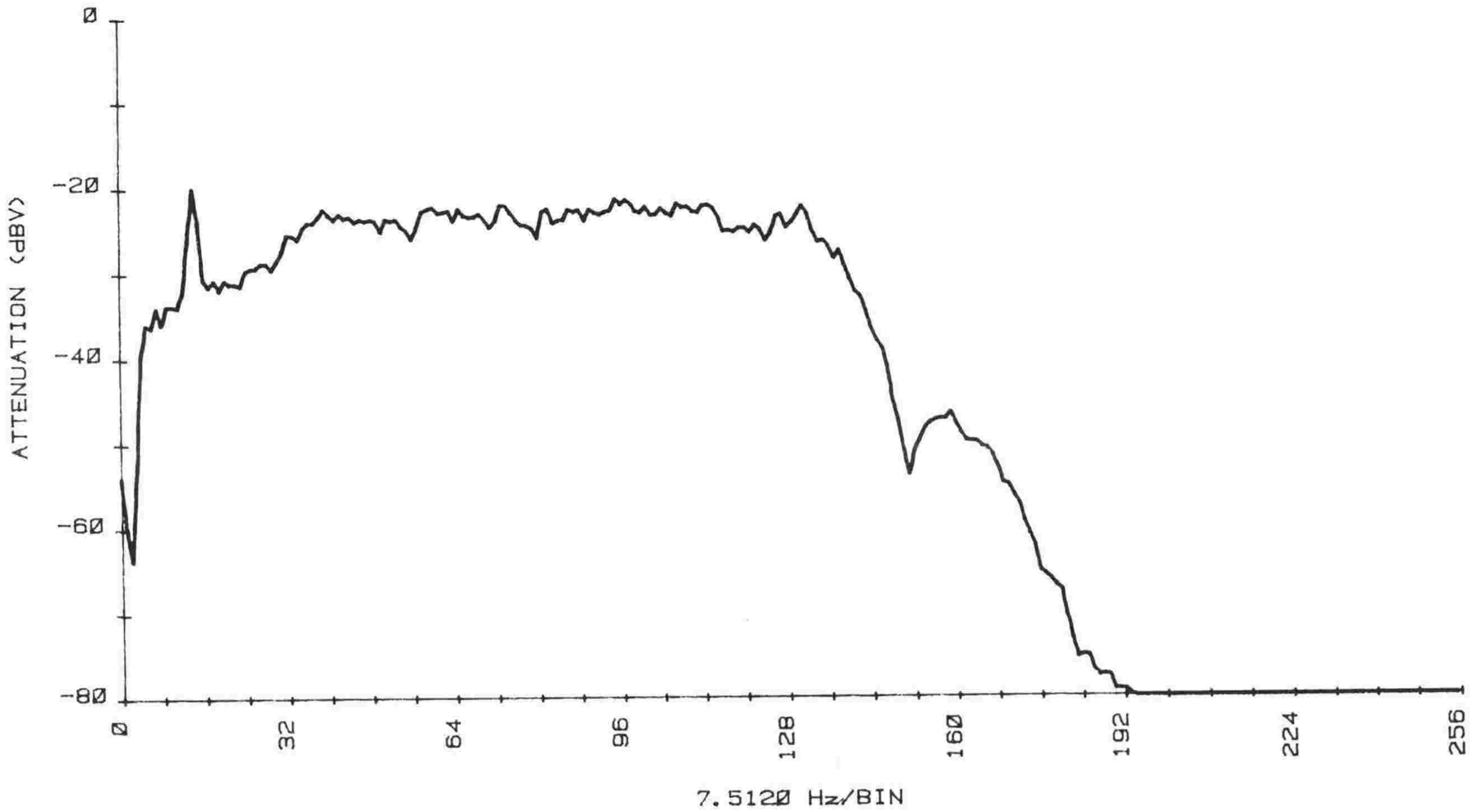


Figure 34. The waveform and spectrum of the masking noise which was sampled and digitally recorded for the third and fourth pigeon experiments, recorded in the experimental chamber. The peak at around 100 Hz was produced by the chamber's ventilation fan.

The Generation of Signals During Sessions

The apparatus used to generate the signals during experimental sessions is shown in Figure 35. Before each trial, the appropriate file on the magnetic tape was loaded into the computer. During the presentation of the signal, the digital codes were passed to the multiprogrammer's digital-to-analogue converter (DAC). A Hewlett-Packard 59308A timing generator controlled the rate at which the signal was generated. The effective clocking rate was 8333 Hz (120 µsec between points), faster than the recording rate of 7246 Hz. The higher rate compensated for slight lags in the output of the multiprogrammer. The final signals were 140 msecs in duration. Gaps between repetitions of the signals during each trial (360 msec) were timed by a Hewlett-Packard 98035A real-time clock controlled from the computer.

The output of the multiprogrammer was passed successively through a passive low-pass filter set at 1250 Hz (with a roll-off of approximately 200 dB/octave), a Krohn-Hite 3550 24 dB/octave filter set at low-pass 2000 Hz, and an attenuator. It was then mixed with the low-level continuous noise and passed to the audio amplifier and speaker in the sound-attenuated room.

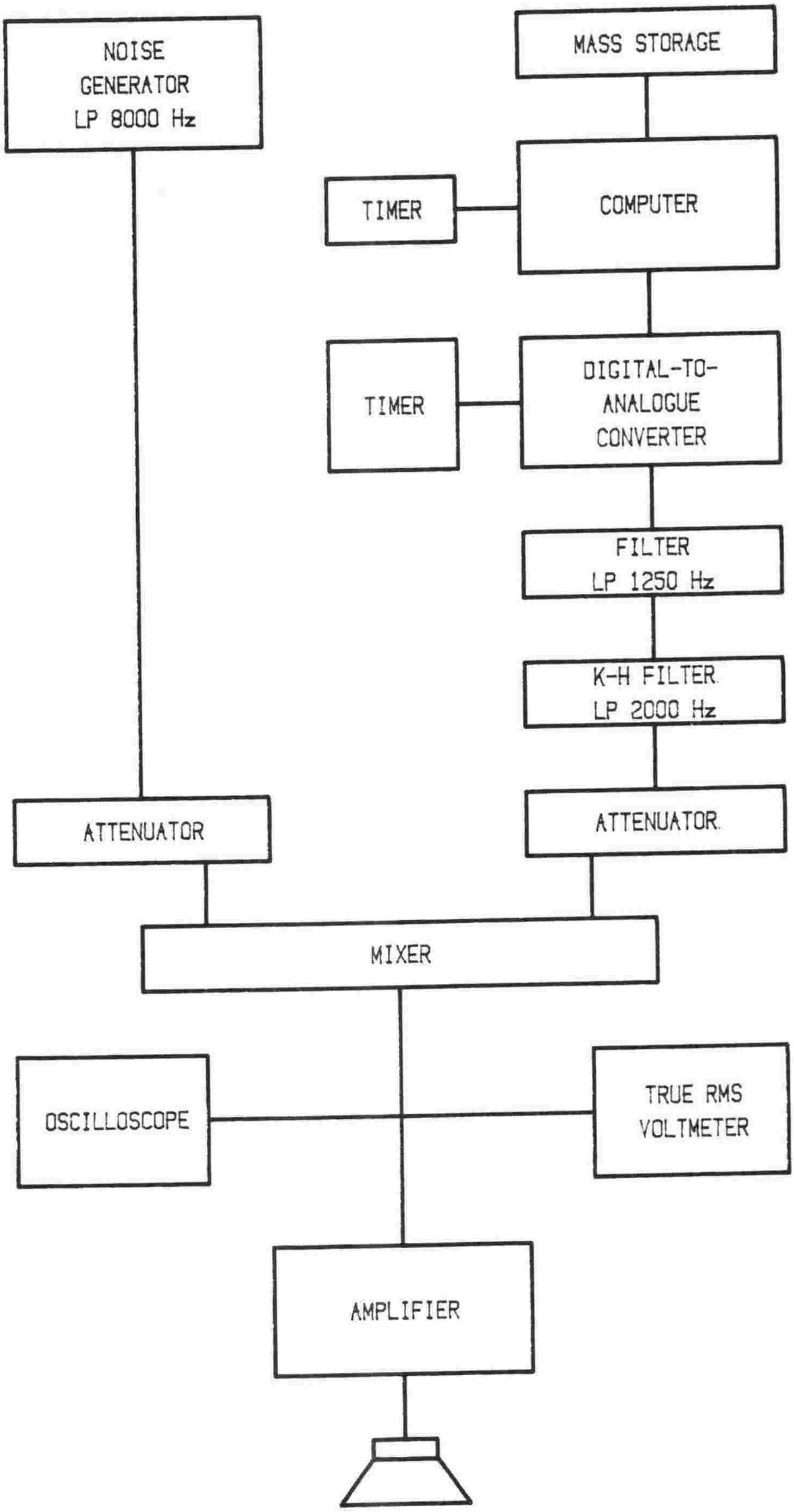


Figure 35. Block diagram of the sound-generating equipment used during the third and fourth pigeon experiments.

Calibration

The frequencies of the signals produced by the multiprogrammer were checked by generating sinusoids of appropriate frequencies and measuring their periods with a Philips PM 6620 counter-timer. These measurements were made on the output of the first (1250 Hz low-pass) filter (Figure 35).

Levels were checked by measuring peak voltages of the brief signals via the microphone placed in a standard position in the experimental chamber. The voltages from the microphone were measured with a B & K RMS voltmeter (Type 2425) in peak hold mode, and could be compared with the peak voltages of analogue signals of similar frequency and duration.

The strategy of measuring peak levels was necessary because, when the multiprogrammer was used to produce the digital signals, no ADC was available to make RMS measurements on the brief signals. No comparable problem had arisen with the measurement of the analogue signals, since, for the purposes of calibration, continuous signals could easily be produced.

Finally, the signal-plus-noise as measured by the microphone in the experimental chamber was analysed with a Hewlett-Packard 3580A spectrum analyser. A typical output from the analyser can be seen in Figure 36, which shows a 450 Hz signal. During this analysis, six different digital signals were repeatedly presented in a haphazard order. In the present case, the chamber's whisper fan was operating and the continuous noise was on. Also, the output from the microphone was unfiltered, leading to the relatively high level of noise below 50 Hz.

The spectrum in Figure 36 can be compared with that in Figure 34. As in that case, the flat portion of the gated masking noise went from approximately 200 Hz to around 1000 Hz, a bandwidth of 800 Hz. The slight upward slope of the noise spectrum from lower to higher

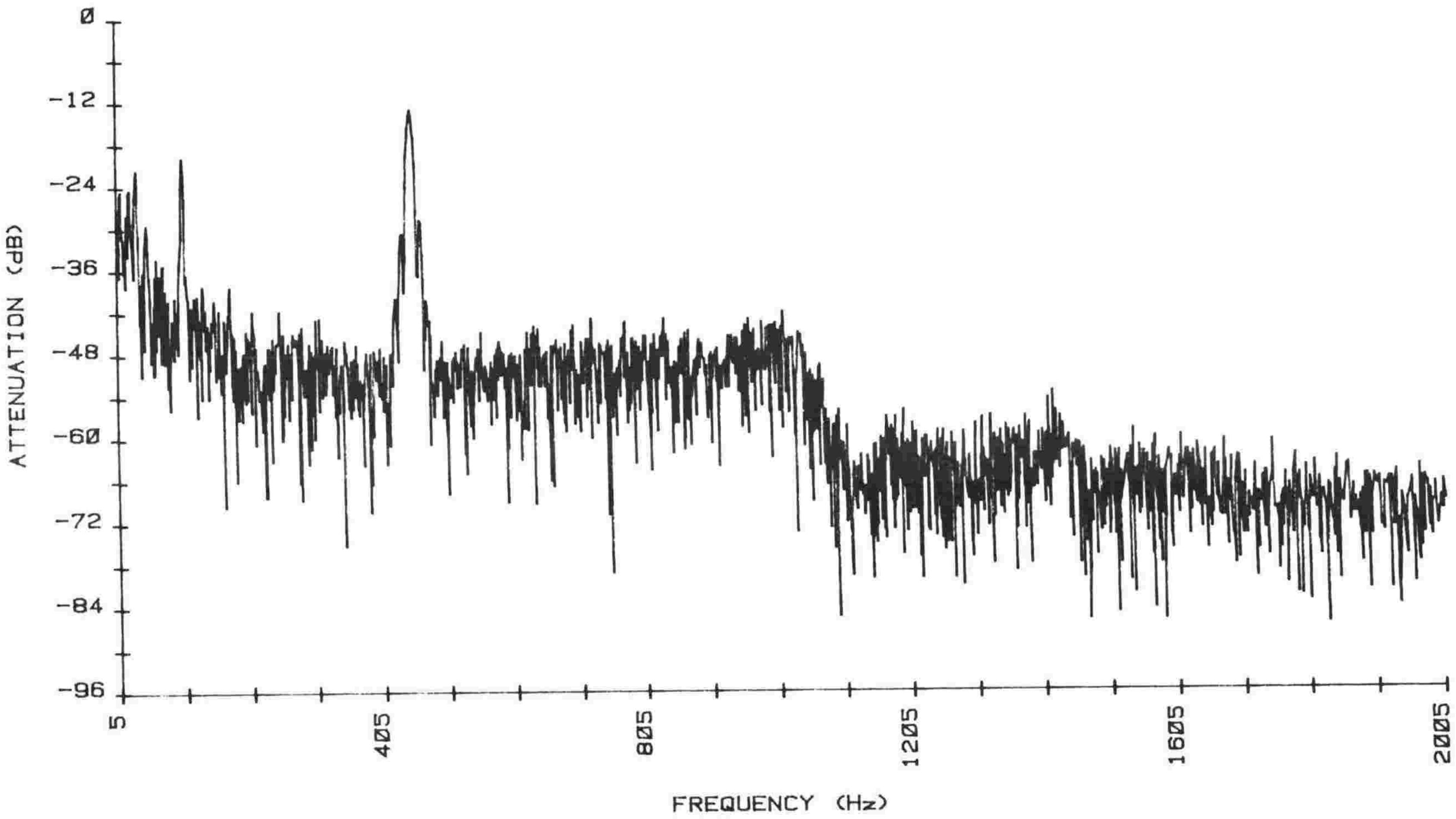


Figure 36. The spectrum of digitally-recorded 450-Hz sinusoids and masking noise (plus continuous masking noise and fan) used in the third and fourth pigeon experiments, recorded in the experimental chamber. Signals based on six different codes were repeatedly presented in haphazard order during the construction of the spectrum.

frequencies, which led to the adjustments of the signal levels described earlier, can also be seen in both figures. Between 1000 Hz and 1100 Hz the level of the gated masking noise fell off sharply to a level below that of the continuous masking noise.

Observers, Apparatus and Procedure

Birds 8 and 18 were used in both experiments. The chamber and controlling equipment were as for the first two pigeon experiments. The only difference was that after recalibration the minimum force needed to operate the side keys rose from 0.15 N (in the first two experiments) to 0.20 N.

The behavioural procedures were generally the same as those used previously. Trials lasted 30 seconds, or until the bird had made 30 responses to one key or the other, whichever was the shorter. Haphazard inter-trial interval durations were determined by the computer, in this case between the limits of one and three seconds, with a mean of two seconds. Pecks during ITIs resulted in a three-second delay of trial onset. For Bird 8, timeouts for errors lasted 15 seconds, while those for Bird 18 lasted five seconds.

Daily sessions were made up of warm-up and experimental trials. The number and nature of the warm-up trials differed slightly over the two experiments, so they will be described in the introductions to the respective experiments. However, correction was always used during the warm-up trials, and there were never fewer than 20 such trials. In each experiment the 132 experimental trials consisted of 66 standard (450 Hz) trials, and 66 comparison trials. No correction was used during the experimental parts of the sessions.

The order of experimental trials was determined independently for each bird and session, using the random number generator on the computer, with the restriction that no more than four of either kind of trial (standard or comparison) could occur in succession.

The Ability of Pigeons to Discriminate Frequency

Probably because of the difficulties involved, relatively few studies of pigeon auditory frequency discrimination have been published. Figure 37 shows the results of most of those which have been carried out with modern equipment.[16] The most comprehensive study, and that which gave the smallest Weber fractions ({delta-f}/f), was carried out by Sinnott, Sachs and Hienz (1980). From 500 Hz to 4000 Hz, the fraction was 0.02 or less, giving a delta-f value of about 10 Hz at 500 Hz, the closest frequency to the 450 Hz standard used in the experiments reported here. The two other studies gave quite different results, especially at 500 Hz: Price, Dalton and Smith (1967) reported a delta-f value of around 20 Hz, and Delius and Tarpy (1974) a value of approximately 40 Hz.

The fraction of 0.15 (150 Hz) at 1000 Hz obtained by Krasnegor (1971) using the methodology adopted in the present work is particularly discrepant. As discussed in Chapter 1, the level of Krasnegor's results appears to be largely due to the use of a yes-no (YN) procedure in which only the standard or a comparison frequency was presented on each trial. All the other research used variants of the same-different procedure, in which standard and comparison frequencies are both presented on each trial. Other variations among results may be due to the nature of the stimuli and to differences in the acoustic environments, in the way difference limens were calculated, and the way "yes" and "no" responses were defined.

---

[16] Not shown in the figure are some recent data for frequencies of 1 to 20 Hz, which were obtained using classical conditioning of heart-rate (Quine & Kriethen, 1981).

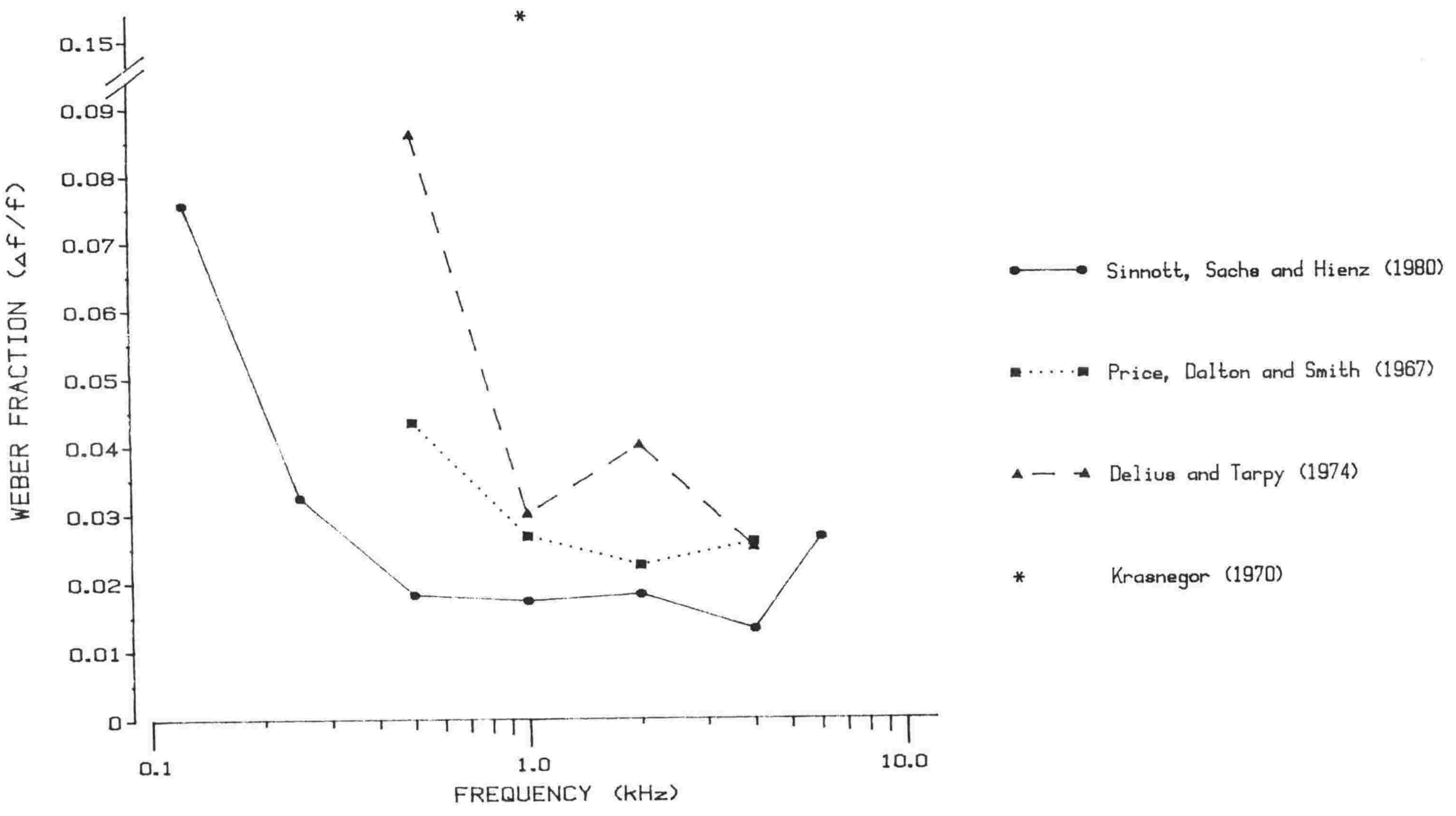


Figure 37. Frequency difference limens for pigeons.

CHAPTER 8

AN APPLICATION OF GOC ANALYSIS TO FREQUENCY DISCRIMINATION IN PIGEONS

The first experiment described in this chapter was based on a single standard frequency and a number of comparison frequencies, and was thus typical of many experiments which have sought to measure difference limens with the method of constant stimuli. In the second experiment, only one comparison frequency was used, with the aim of obtaining a well-defined GOC curve for that frequency.

MULTIPLE COMPARISON FREQUENCIES

Method

Details common to both the multiple and single comparison experiments were given at the end of the previous chapter.

Signals

The characteristics of the auditory stimuli can be summarised as follows: The standard frequency was 450 Hz (with a level of 76 dB SPL) and the comparison frequencies were 500 Hz (77.5 dB), 600 Hz (78 dB), 700 Hz, 800 Hz and 900 Hz (all 80 dB). The gated masking noise was 800 Hz in bandwidth and its overall level was approximately 60 dB, giving a spectrum level of 30 dB and a signal-to-noise ratio of 40 dB. Each acoustic transient consisted of a sinusoid in gated noise, and was 140 msec in total duration, with rise-fall times provided by 40 msec Hanning data windows. The gaps between transients lasted 360 msec. The continuous masking noise, as in previous experiments, was about 8000 Hz in bandwidth and had a spectrum level of 16 dB SPL.

Procedure

Daily sessions were made up of 152 trials. The first 20 of these were warm-ups, during which each successive pair of trials consisted of one 450-Hz and one 900-Hz trial. This meant that any one kind of trial could not occur more than twice in succession. Incorrect responses (right-hand key or "yes" responses on 450 Hz trials, and left-hand key or "no" responses on 900 Hz trials) were followed by repetitions of the trial. This continued until a correct response occurred.

The 132 experimental trials consisted of 66 standard trials, all of 450 Hz, and 66 comparison trials. The latter were made up of nine 500-Hz trials, ten 600-Hz trials, twelve 700-Hz trials, fifteen 800-Hz trials and twenty 900-Hz trials. No correction trials were used during this part of the sessions.

Each replication consisted of 792 trials, with a different independent sample of gated noise on each trial. Five replications were run during the experiment, making a total of 30 sessions in all.

Prior to the experiment, preliminary sessions were run on 90 days over a period of four months.

Results

GOC Analysis

GOC curves were produced by combining the results for equivalent trials over replications. An important difference between the present analysis and the ones based on the known distributions was that each trial in a replication could be matched with only one trial in any other replication. Thus any arbitrariness present in the earlier analyses did not exist here, and there was only one possible outcome when all replications were combined.

The results for individual replications[17] and the five-replication GOCs for each bird are shown in Figure 38, while Figure 39 shows the GOCs for both birds combined (10 replications in all).

Figures 40 and 41 show the data from the two preceding figures plotted on normal-normal coordinates and fitted with straight lines. The slopes of these lines, their fit, and the d_z values for individual replications and for the GOCs are given in Table 2.

As in previous experiments, the GOC curve for each bird represented a higher level of performance than that obtained on individual replications. The result for Bird 8 at 500 Hz is ambiguous because on some replications the ROC point fell on the chance line and on others somewhat below it. Overall, however, the bird tended to treat the 500 Hz signal as a lower frequency than the 450 Hz standard, and the d_z value for the GOC reflects this.

The differences between the performances on individual replications and those represented by the GOC curves were much greater than those in previous experiments: for both birds most of the GOC values were around twice the initial values.

The plots in Figures 40 and 41 and the squared correlation coefficients in Table 2 show that the GOC points were reasonably well fitted by straight lines. However, for Bird 18 the points excluded from the fits for the 700 Hz, 800 Hz and 900 Hz stimuli (because z[FAR] was greater than 2.5) were quite out of line with the other points. Their inclusion would have had the effect of increasing the slopes of the GOCs by about 20%, 30% and 40% respectively. There is thus room for some doubt about the true slopes, although the inclusion of the extra points turned out to have very little effect on the values of d_z.

---

[17] The figures on which the graphs are based are given in Table 8 (Appendix C).

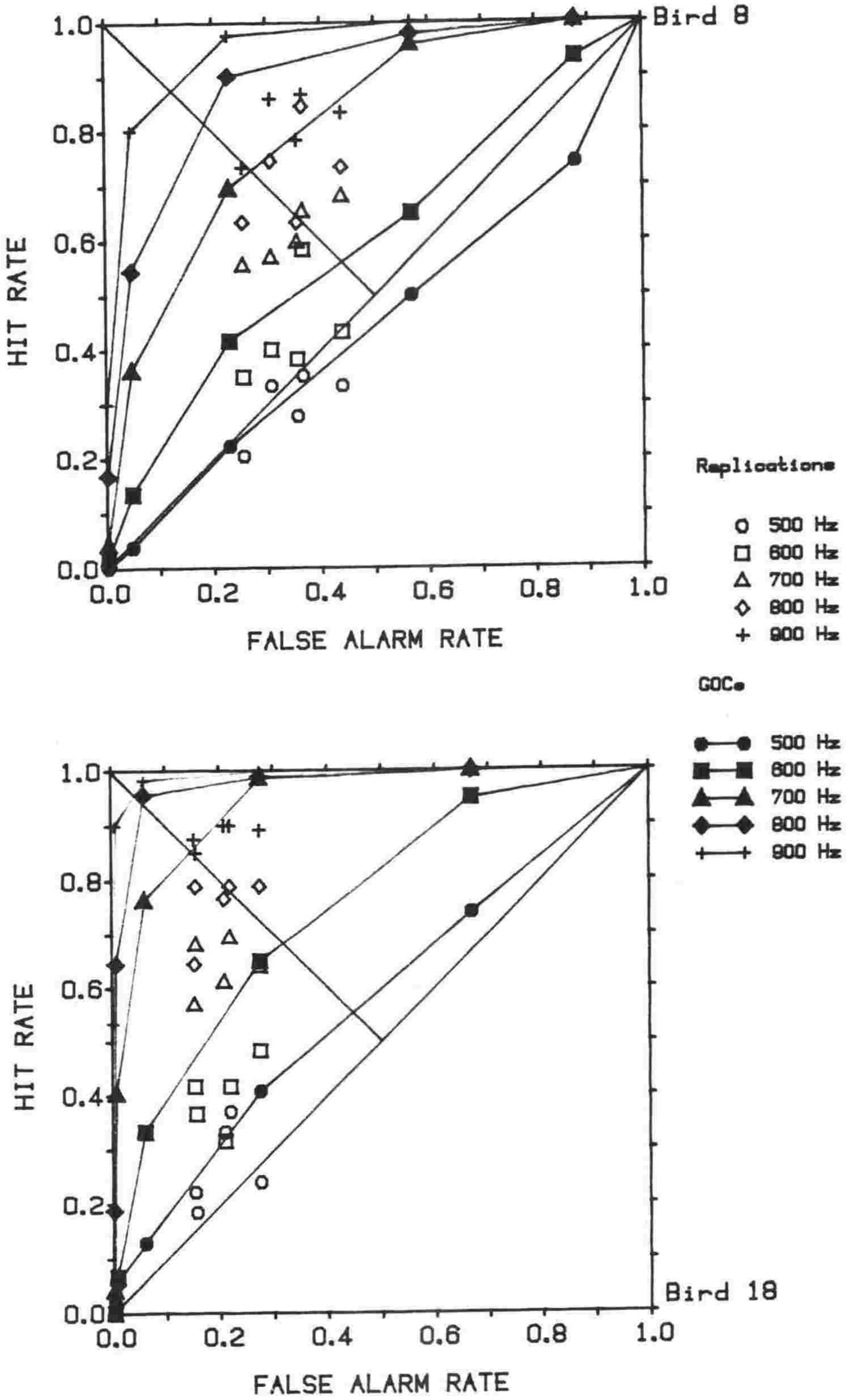


Figure 38. GOCs and individual replication results obtained in the third pigeon experiment.

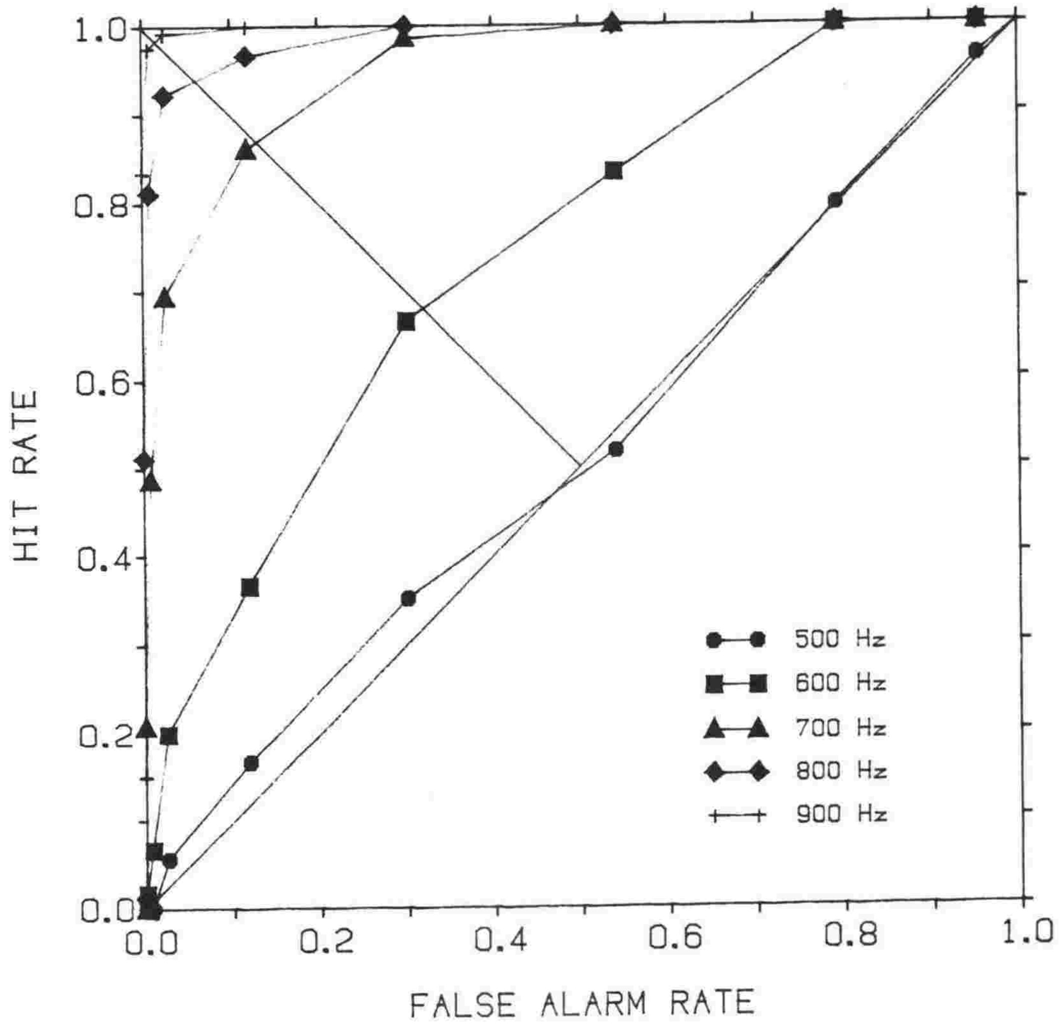


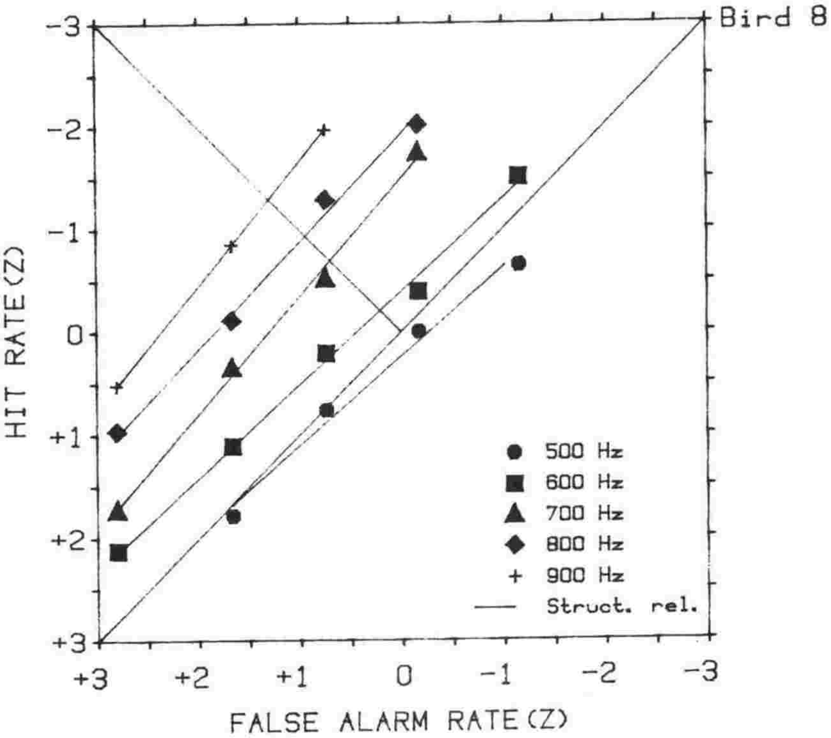
Figure 39. Ten-replication GOCs for both birds combined, obtained in the third pigeon experiment.

TABLE 2

The slope, d_z values and fits for ROCs and GOCs in the third pigeon experiment (multiple comparisons)

		Slope	d_z	r^2
Bird 8				
500 Hz	ROC -pooled data	0.948	-0.150	-
	-mean data	"	-0.152	-
	GOC	0.864	-0.261	0.986
600 Hz	ROC -pooled data	0.972	0.214	-
	-mean data	"	0.218	-
	GOC	0.905	0.403	0.990
700 Hz	ROC -pooled data	1.062	0.684	-
	-mean data	"	0.693	-
	GOC	1.143	1.377	0.989
800 Hz	ROC -pooled data	1.087	0.967	-
	-mean data	"	0.990	-
	GOC	1.043	1.870	0.983
900 Hz	ROC -pooled data	1.166	1.254	-
	-mean data	"	1.273	-
	GOC	1.209	2.573	a
Bird 18				
500 Hz	ROC -pooled data	0.913	0.154	-
	-mean data	"	0.148	-
	GOC	0.803	0.279	0.989
600 Hz	ROC -pooled data	1.039	0.598	-
	-mean data	"	0.606	-
	GOC	1.076	1.095	0.995
700 Hz	ROC -pooled data	1.052	1.198	-
	-mean data	"	1.210	-
	GOC	1.347	2.483	0.990
800 Hz	ROC -pooled data	1.134	1.529	-
	-mean data	"	1.548	-
	GOC	1.074	2.979	0.942
900 Hz	ROC -pooled data	1.047	2.014	-
	-mean data	"	2.029	-
	GOC	0.963	3.688	a
10-replication GOC				
500 Hz		0.920	0.119	0.993
600 Hz		0.956	0.907	0.991
700 Hz		1.139	2.495	0.971
800 Hz		0.753	3.136	0.955
900 Hz		0.917	4.364	a

a Only two points available.



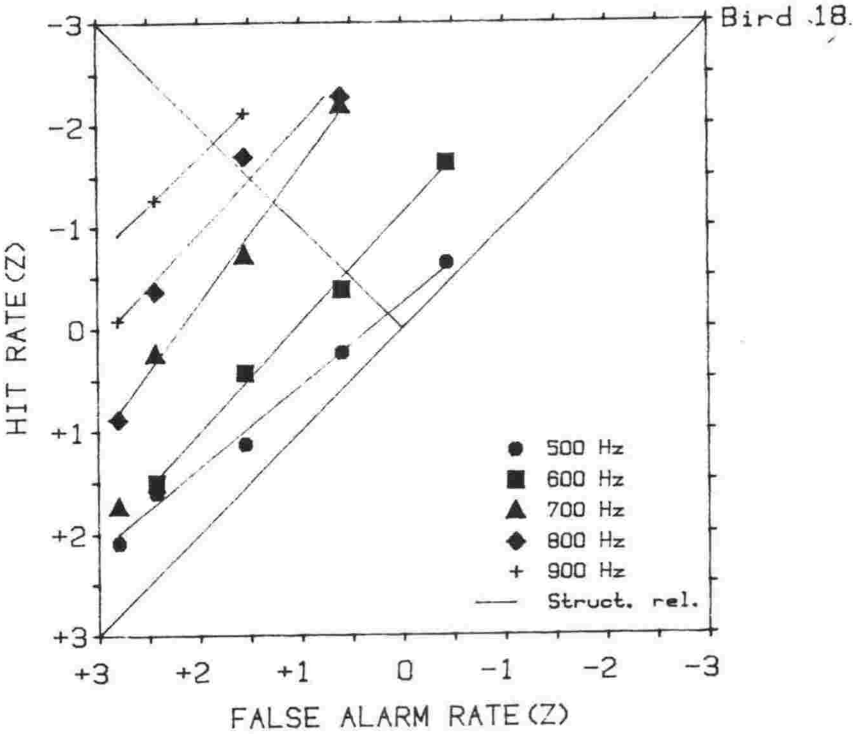


Figure 40. Individual GOCs obtained in the third pigeon experiment plotted on normal-normal coordinates.
Points representing values greater than 2.5, though plotted, were not included in the fitting.

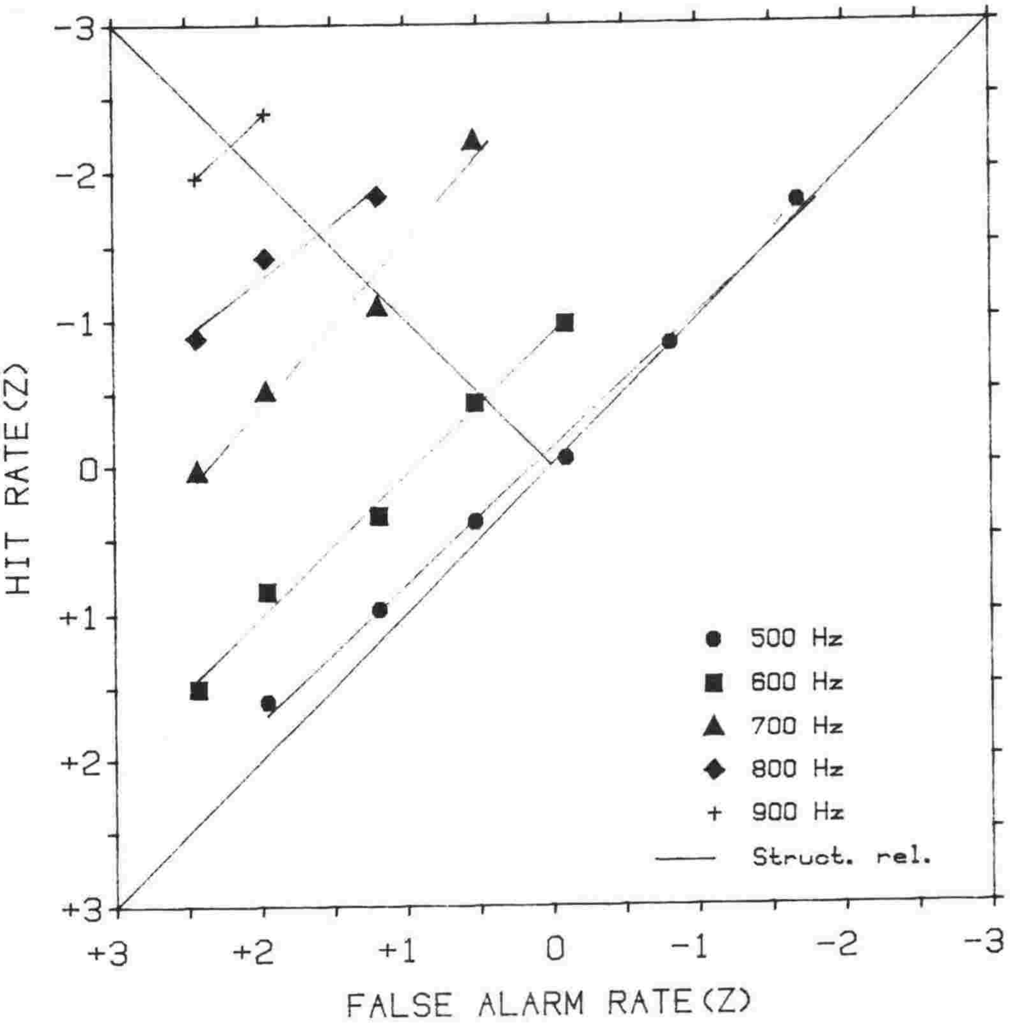


Figure 41. Ten-replication GOCs for both birds combined, obtained in the third pigeon experiment, plotted on normal-normal coordinates.

The Estimation of k

Figure 42 shows how d_z changed as replications were added in the experiment. All possible combinations of replications were used to estimate each point.

Because of the small number of points and their variability, linear functions fitted by the method of weighted least-squares, rather than quadratic functions,[18] were used to estimate the slopes of the ROCs and thus the values of $d_{z(1)}$. Linear functions were not as drastically affected by the deviations of individual points. The weights were the number of combinations of replications on which each mean slope was based. The main effect of the weights was to reduce the contribution of the mean slope of the single five-replication GOC, which in some cases was based on only two or three points.

Estimates of k were obtained using both $d_{z(1)}$ and $d_{z(2)}$ as base values in equation 4a, namely:

$$\underline{k} = \frac{in[\{d_{z(n)}\}^2 - \{d_{z(i)}\}^2]}{n\{d_{z(i)}\}^2 - i\{d_{z(n)}\}^2} \quad . \qquad (4a)$$

The results obtained with $d_{z(2)}$ were equivalent to those obtained with $d_{z(1)}$. Only one set of results, those based on $d_{z(1)}$, are presented.

The estimates obtained with $d_{z(1)}$ are shown by the points in Figure 43. The lines in the graphs show the highest and lowest estimates of k obtained from a series of 10 simulations in which there was no common noise. Each simulation consisted of five 4000-trial replications and was based on continuous normal distributions of unit variance. Differences between the means were set equal to each of the initial d_z values obtained in the experiment. There was no systematic relation

---

[18] The functions are shown in Figure 60 (Appendix C).

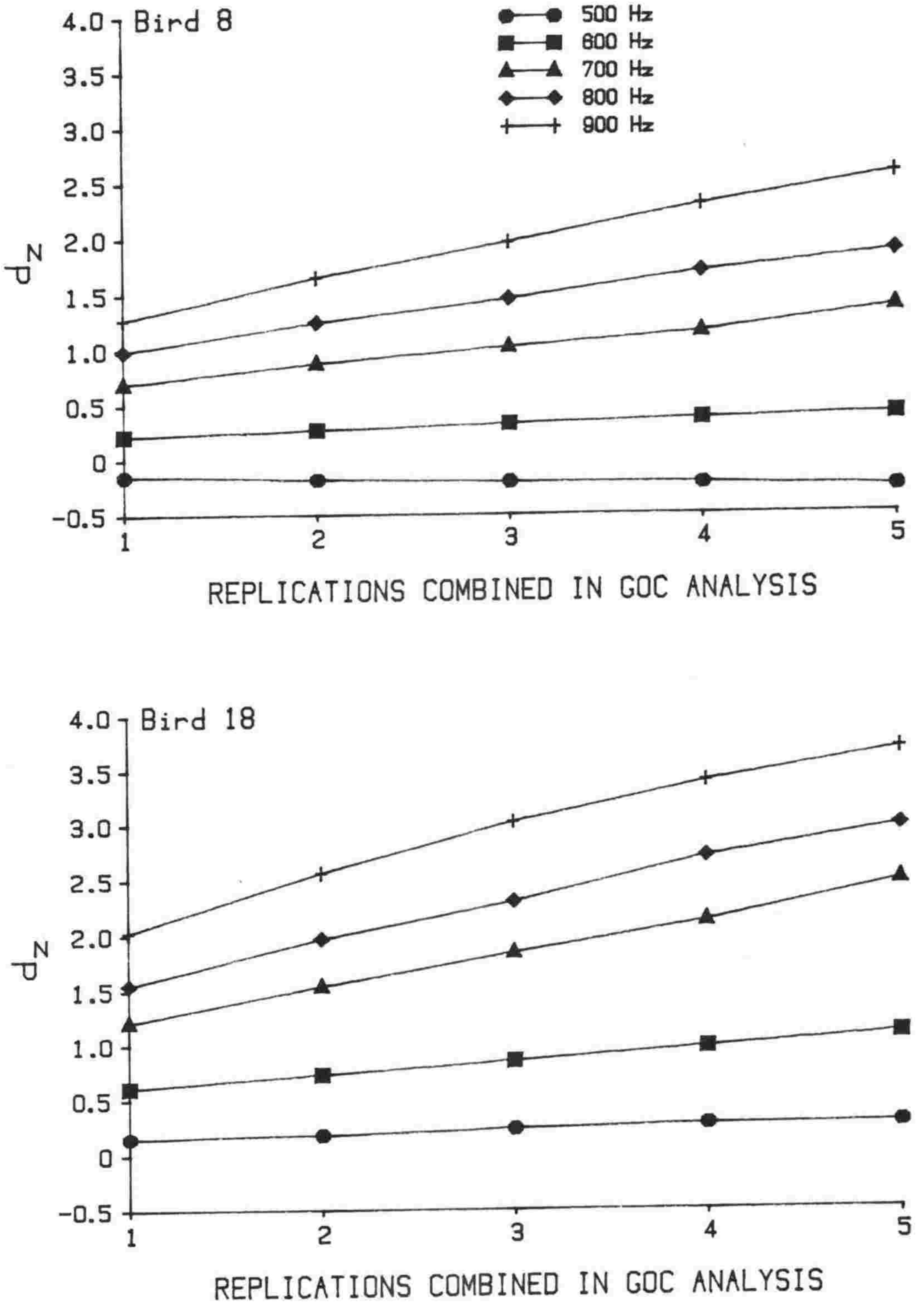


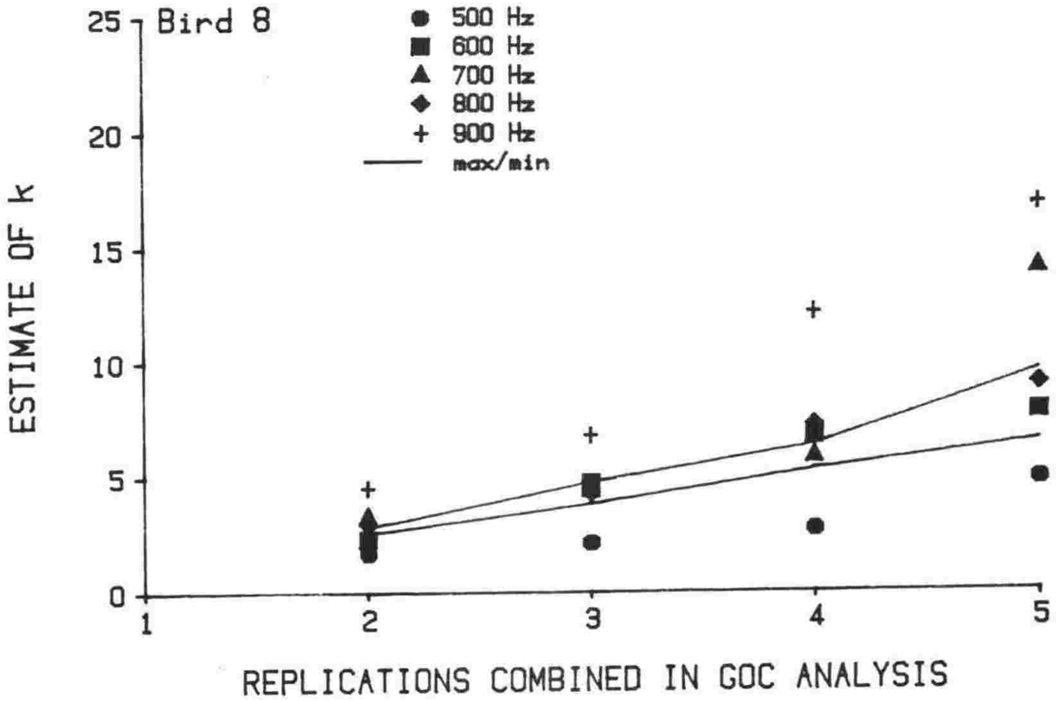
Figure 42. The mean dz values for GOCs based on different numbers of replications, obtained in the third pigeon experiment.

between these values and the k estimates in these or any of the other simulations carried out.

The graphs show that the results for the birds varied over frequency and that the estimates of k would be difficult to discriminate from those which would be expected if there was no common noise present in the experiment.

The variability of the pigeon results was due to the large sampling error associated with both the small number of replications (and the variability of performance over replications) and the small number of trials for each frequency within the replications. There is no obvious explanation for the apparent relation between the estimates of k and frequency which is suggested by the results for the 500 Hz and 900 Hz signals for Bird 8.

Given the amount of variation in the estimates of k, establishing that there was any common noise in the experiment would be impossible. If anything, the k estimates from the experiment were higher than those arising from the simulations in which there was no common noise. In the second experiment, more replications were run, and there were many more samples of common noise for the main comparison frequency, thus reducing the variation in the estimates of k.



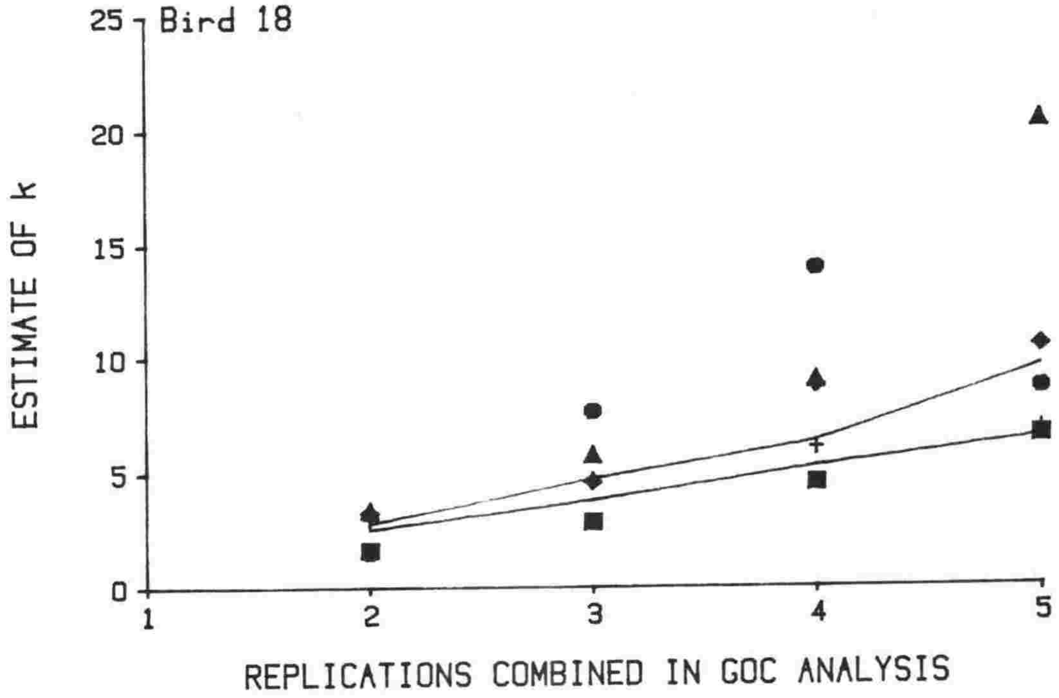


Figure 43. Estimates of k obtained in the third pigeon experiment and the highest and lowest estimates of k from simulations in which there was no common noise.
The estimates were based on comparisons of dz(1) with dz(2) to dz(5).

A SINGLE COMPARISON FREQUENCY

The main aim of this experiment was to obtain enough data to define two curves as well as possible: the GOC curve itself, and the curve showing the effect of adding replications. Only one comparison frequency was used, so that frequency was represented by as many independent samples as were practicable within the limits of the experiment, reducing the effects of sampling error on the shape of the GOC curve. The object of running many replications (78 sessions) was to reduce the effects of unique noise as much as possible, and to provide better estimates of *k* and of the asymptotic value of d_z than could be obtained from the results of the previous experiment. A particular interest lay in the presence or absence of common noise.

The experiment followed the previous one without a break and without any immediate changes in procedure. The only important difference was the change in the comparison stimuli. This will be described below.

Method

Subjects

Birds 8 and 18 were used. Bird 18 completed 13 six-session replications but Bird 8 completed only eight usable replications. Following poor performance in the third replication, and other signs of illness, the latter bird was removed from the experimental conditions and allowed ad lib. access to food. She was diagnosed by a veterinarian as having coccidiosis and was treated with a proprietary medicine (Scourban). Eventually she was put on a light food deprivation schedule (greater than 90% of her ad lib. weight) and after six sessions consisting only of correction trials she was returned to the experimental procedure. She remained healthy and completed six more replications.

Stimuli

A frequency of 650 Hz was chosen for the main comparison signal. The results of the previous experiment suggested that this frequency would be high enough to ensure stable performance over a long period but low enough to leave room for substantial improvement. Signals of 650 Hz occurred on 56 out of the 66 comparison trials in each session; 900-Hz signals occurred on the remaining 10 comparison trials. This meant that each replication contained 336 650-Hz trials and 60 900-Hz trials. The 900 Hz signals were used in the present experiment to contribute to stable performance.

The 650 and 900 Hz signals were generated and recorded using the equipment and procedures described in Chapter 7. In order to maintain the 40 dB S/N ratio, the 650 Hz signals were recorded so as to have an absolute level of 79 dB SPL. The 450 Hz signals were the ones used in the previous experiment.

Procedure

The only procedural modification concerned the warm-up trials. Initially the warm-ups consisted of 20 pairwise 450 Hz and 900 Hz trials with correction, exactly as in the previous experiment. During Bird 8's third replication (which was later discarded) 30 correction trials were introduced for this bird. The first 10 trials consisted of pairwise 450 Hz and 900 Hz trials as before, but during the remaining 20 trials a random half of the comparison trials involved 650 Hz signals instead of 900 Hz signals. This modified warm-up procedure, which was used for the remainder of Bird 8's sessions, was also brought into use for Bird 18. It was introduced during her sixth replication in order to prevent a key preference from leading to exclusive responding on one key, and was used for that bird from then until the end of the experiment.

Results

GOC Analysis

Figure 44 shows the results for individual replications[19] and the eight- and thirteen-replication GOC curves for Birds 8 and 18 respectively. Figure 45 shows the GOC curves for both birds combined (21 replications in all).

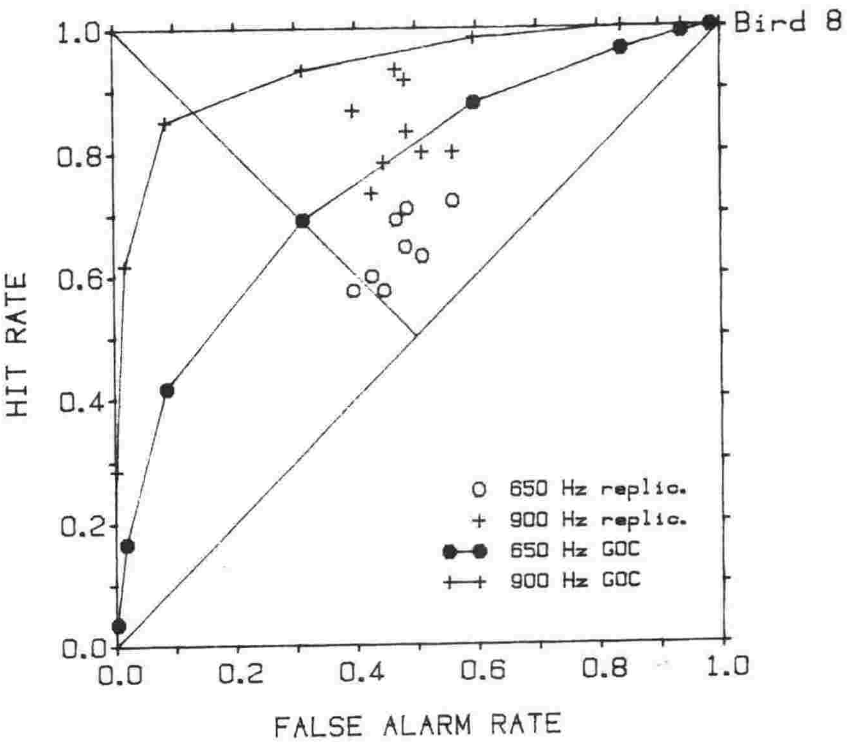
The rest of the analysis is based on the birds' performance at 650 Hz, but for the sake of completeness the 900-Hz GOC curves are included in these graphs. For Bird 18, and for both birds combined, these curves reached the highest possible level.

Figures 46 and 47 show the 650 Hz data from Figures 44 and 45 plotted on normal-normal coordinates and fitted with straight lines. The slopes of these lines, their fit and the d_z values for individual replications and for the GOCs are shown in Table 3. The ROC curve slopes given here were derived from extrapolations of quadratic functions fitted to the GOC slopes over replications; very similar ROC slopes, and hence d_z values, were obtained with extrapolations of linear functions (Figure 61, Appendix C).

One of the aims of this experiment was to obtain well-defined GOC curves which might give some indication of the general class of the underlying distributions. As Figures 44 to 47 show, the curves were about what would be expected if the distributions were normal with approximately equal variance. This raises the question of whether the shapes of the GOC curves reflected normally distributed unique noise more than common noise. This question will be taken up in the next section.

---

[19] Individual replication results are given in Table 9 (Appendix C).



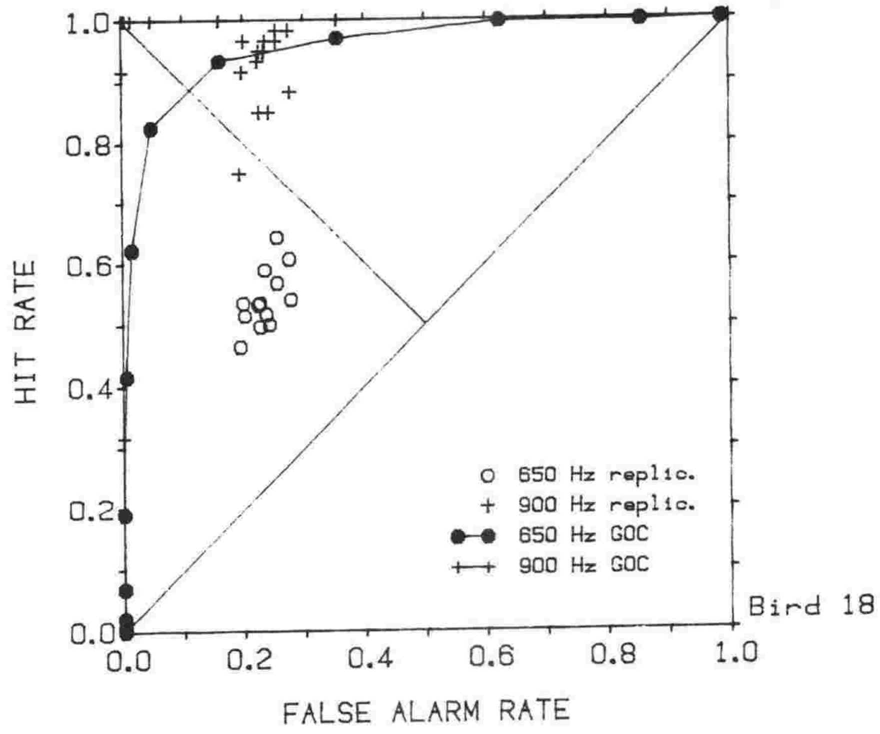


Figure 44. GOCs and individual replication results obtained in the fourth pigeon experiment.

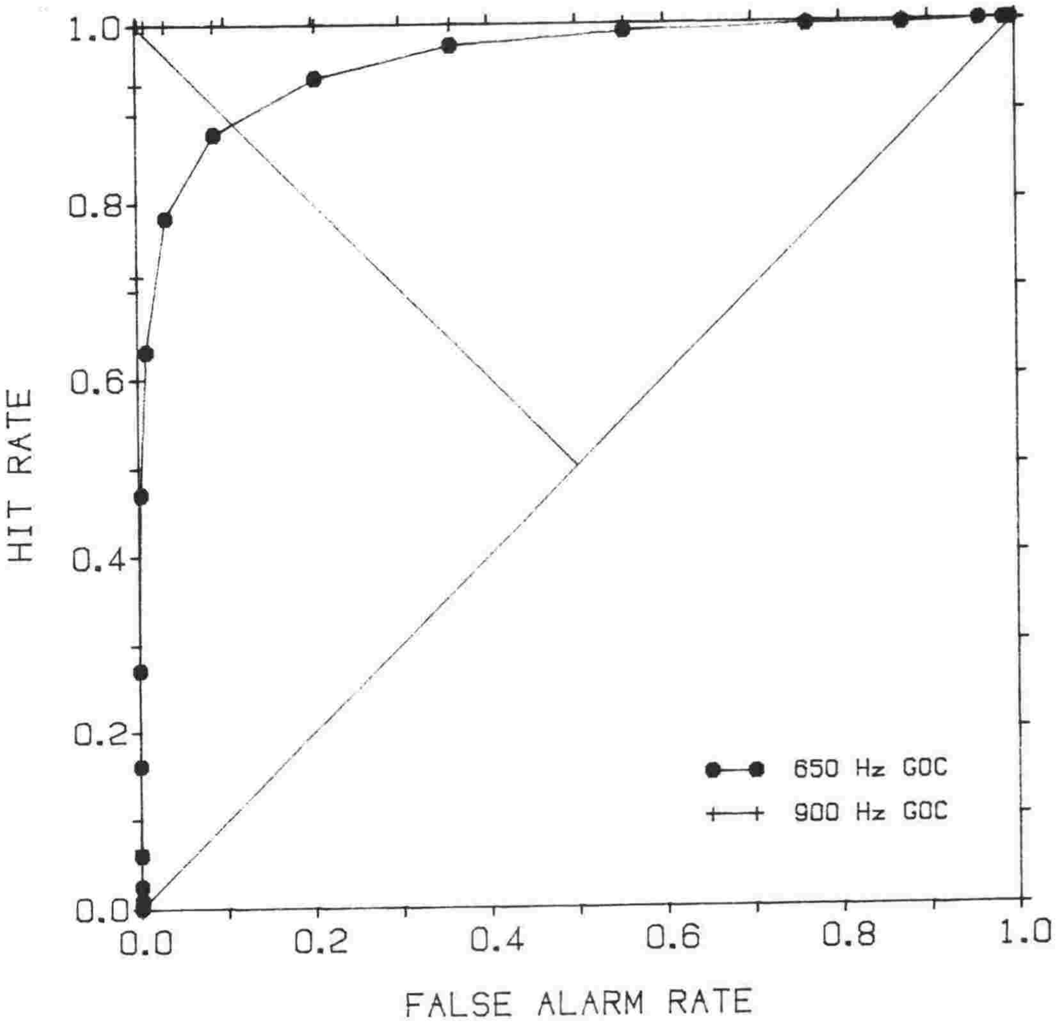
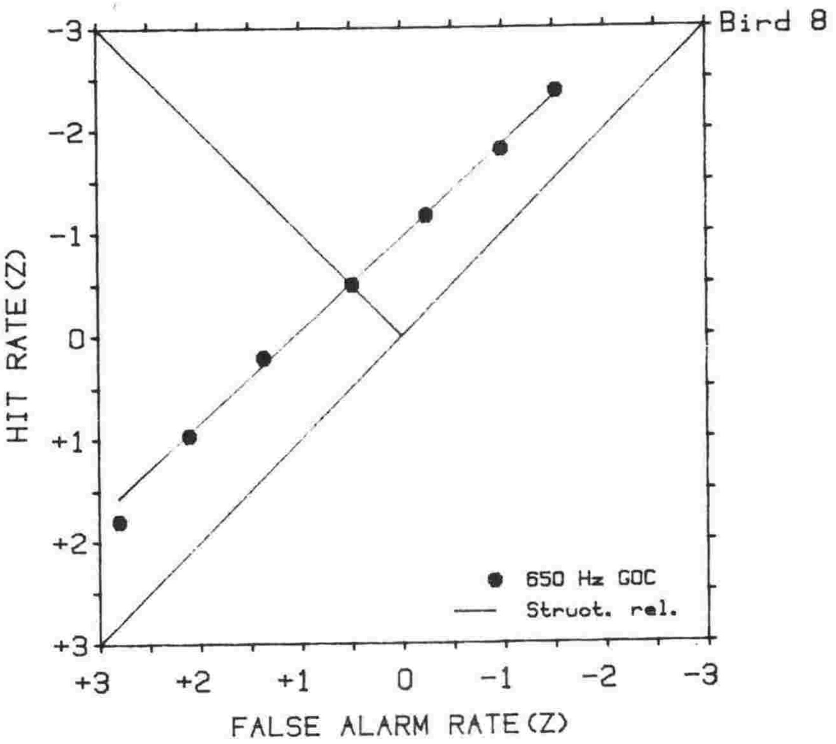


Figure 45. Twenty-one replication GOCs for both birds combined, obtained in the fourth pigeon experiment.

TABLE 3

The slope, d_z values and fits for ROCs and GOCs in the fourth pigeon experiment (single comparison)

	Slope	d_z	r^2
Bird 8			
ROC -pooled data	1.002	0.436	-
-mean data	"	0.440	-
GOC	0.900	1.005	0.999
Bird 18			
ROC -pooled data	0.979	0.822	-
-mean data	"	0.826	-
GOC	0.876	2.442	0.978
21-replication			
GOC	0.820	2.473	0.999



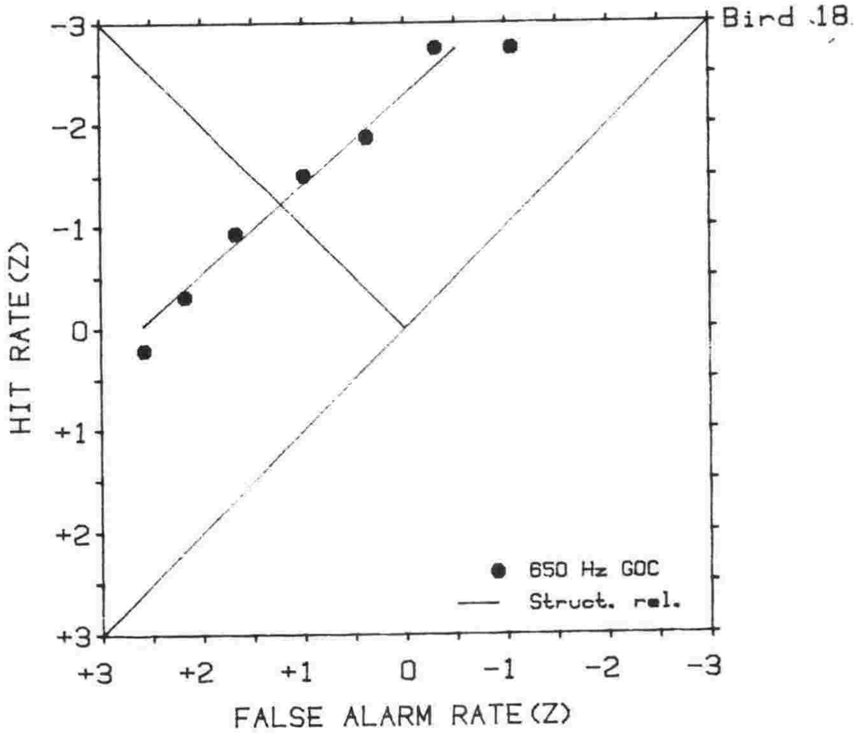


Figure 46. Individual GOCs obtained in the fourth pigeon experiment plotted on normal-normal coordinates.
Points representing values greater than 2.5, though plotted, were not included in the fitting.

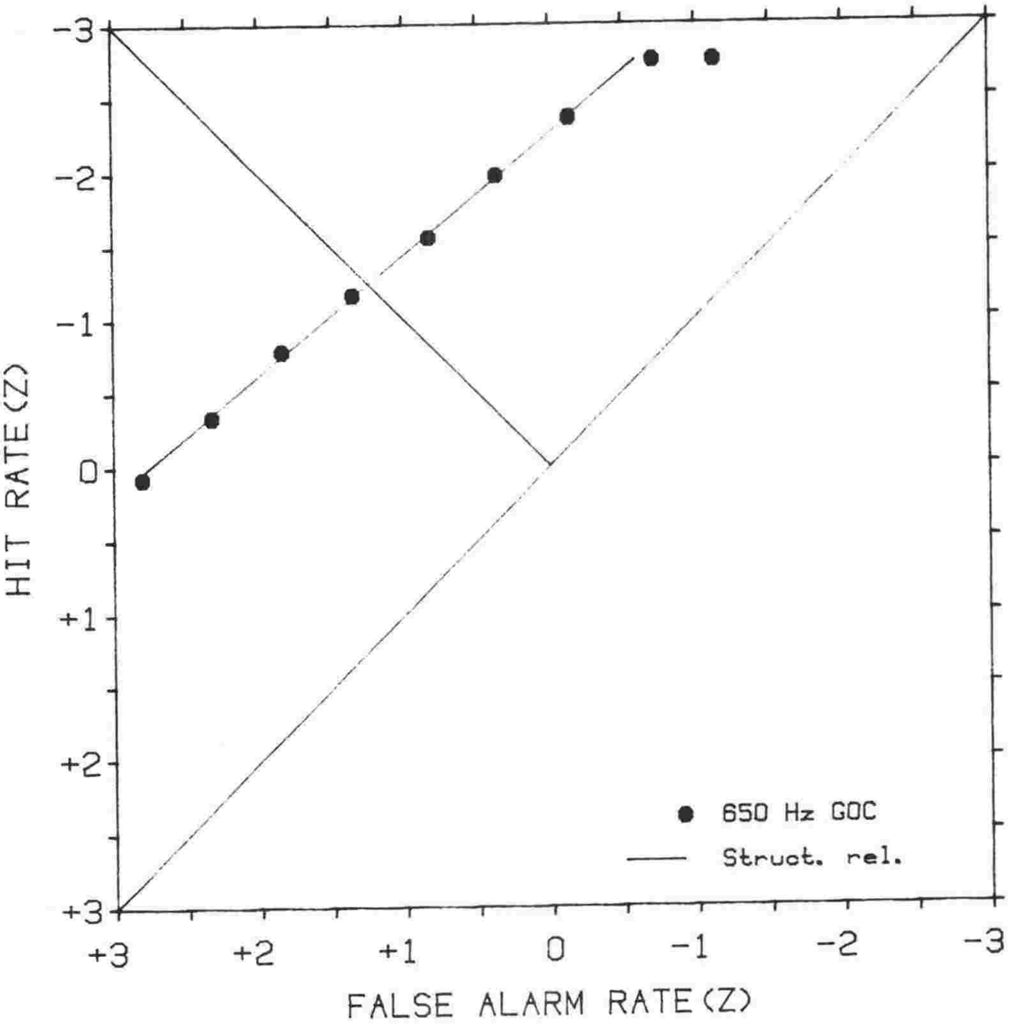


Figure 47. Twenty-one replication GOC for both birds combined, obtained in the fourth pigeon experiment, plotted on normal-normal coordinates.
Points representing values greater than 2.5, though plotted, were not included in the fitting.

The Estimation of k

Figure 48 shows how d_z for the 650 Hz stimulus changed as the replications were added, while the points in Figure 49 show the estimates of k for the different numbers of replications. The lines in Figure 49 show the mean results of seven 10-replication simulations in which each replication consisted of 2000 trials. The distributions of unique noise were normal with unit variance. There was no common noise. The differences between the means of the distributions were similar to the $d_{z(1)}$ values obtained in the experiment (0.4 and 0.8, as can be seen in Table 3). The mean results of these simulations, and the 95% confidence limits, were very well represented by functions of the form

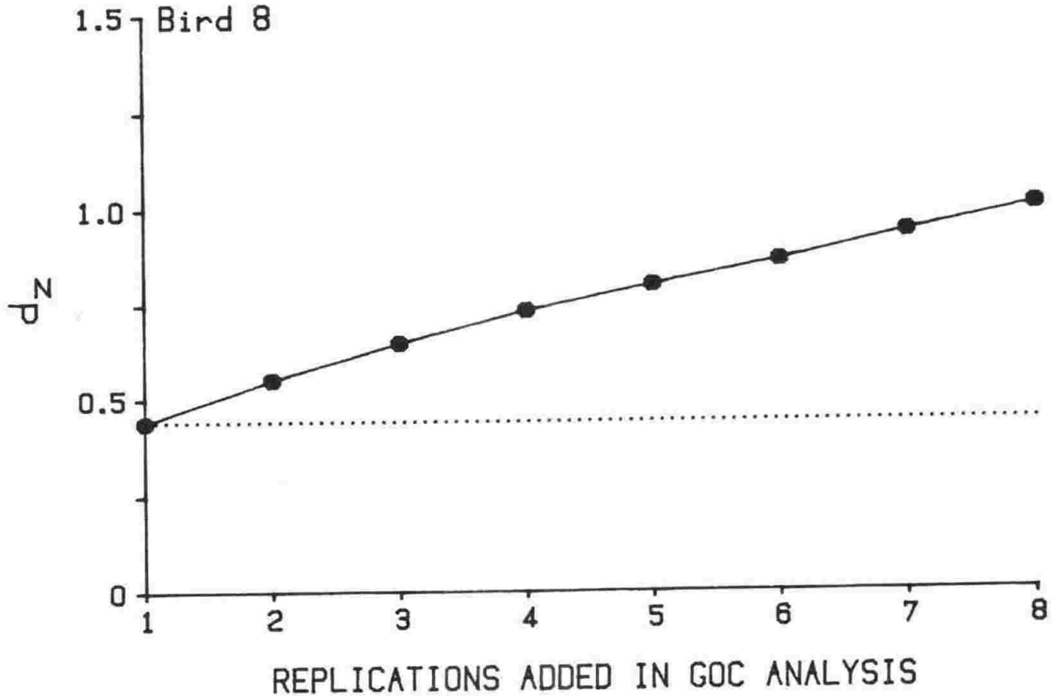
$$\underline{k} = 1/(a + b/N),$$

where a and b are constants and N is the number of replications. These functions, which were selected empirically, are shown in the graph. The results beyond the tenth replication are extrapolations.

The estimates of k shown in the figure were obtained using $d_{z(1)}$ as the base in equation 4a. Similar, but more variable, results were obtained when $d_{z(2)}$ was used as the base.

In the simulations there was no common noise. The results for Bird 18 may be seen to depart a little from the line representing the mean of the simulation results, but only one point for each bird falls outside the 95% confidence limits. Together with the error associated with the estimates of k for the pigeons, this suggests that there was no common noise in the experiment. However, the results are also compatible with the possibility that there is common noise which is swamped by a very high level of unique noise.

The frequency difference limens suggested by the results of the two experiments were very much larger than those reported by Sinnott, Sachs



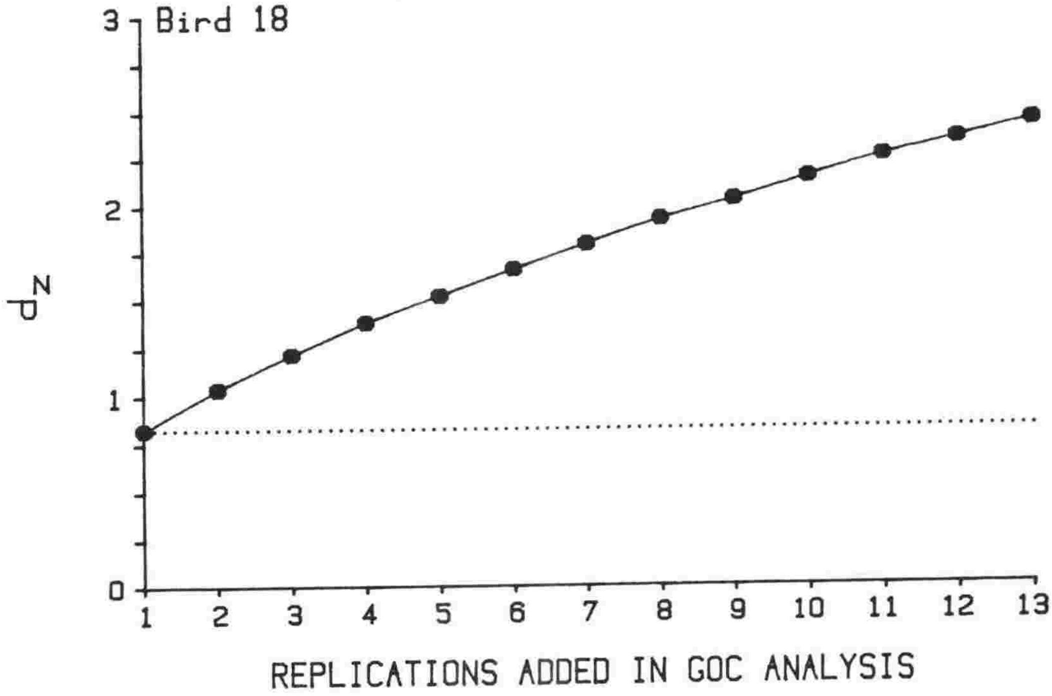


Figure 48. The mean dz values for GOCs based on different numbers of replications, obtained in the fourth pigeon experiment with the 650-Hz stimulus.

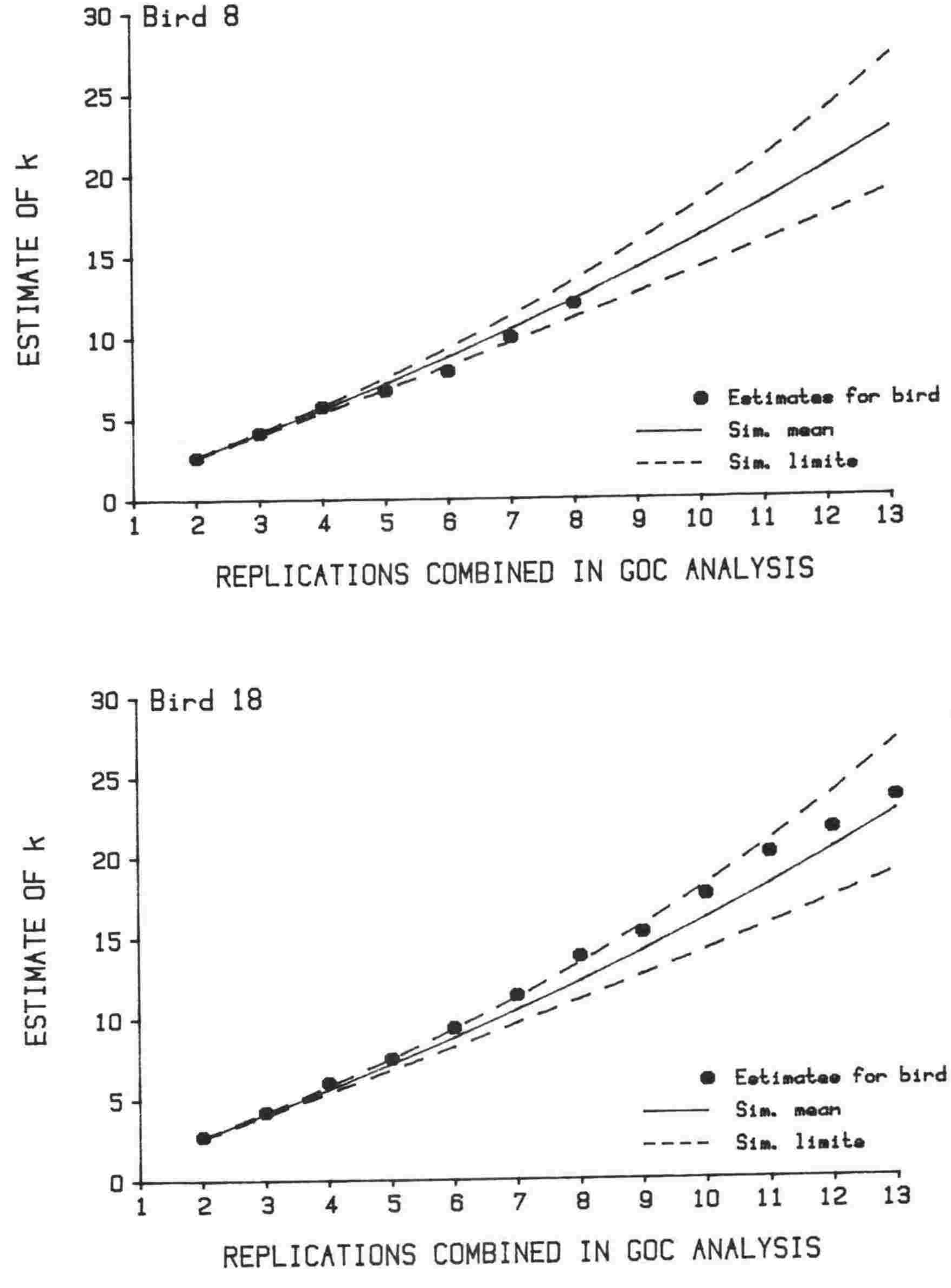


Figure 49. Estimates of k obtained with the 650-Hz stimulus in the fourth pigeon experiment and mean estimates of k from simulations in which there was no common noise.
The dashed lines are 95% confidence limits for the simulation means.

and Hienz (1980). Some of the difference is probably due to the masking noise used in the present experiments (none was used by Sinnott, Sachs and Hienz), and the brevity of the signals. Most of the difference, however, is attributable to the YN methodology, which leads to high levels of unique noise. If the levels of this noise are very high evidence of common noise may well be masked by unique noise.

DISCUSSION

Two interpretations of the results of these experiments are possible. One is that there was no common noise in the experiments, and perhaps in pigeon frequency discrimination in general: that noise reproduced on different occasions was not treated in the same way on each occasion by the birds' ears. The other possibility is that there is common noise, but that the level of the unique noise in these experiments was so high that the presence of common noise was difficult to detect. Without further experimentation, choice between these alternatives is not possible. This question is taken up again in the next chapter.

An important point, mentioned in Chapter 5 (Figure 22), is illustrated by the results of the first experiment. In the absence of large differences between $\underline{k}$ values, large numbers of trials may be necessary to obtain stable performance over replications, if high $\underline{k}$ values are to be differentiated. No systematic investigation was carried out, and the actual figures would depend on the $\underline{k}$ values involved. However, in simulations 1000 to 2000 trials over eight to 10 replications produced consistent results.

CHAPTER 9

THE APPLICATION OF GOC ANALYSIS TO A SIMULATION OF PIGEON FREQUENCY DISCRIMINATION

The main aim of the simulation was to examine how the performance of a temporal model of discrimination could be affected by unique noise from various sources. Another aim was to evaluate the results of the pigeon experiments described in the previous chapter, especially the evidence for the presence or absence of common noise. A more general aim was to see how hardware models might be used in conjunction with GOC analysis to study hearing.

An Overview of the Simulation

The overall approach of the simulation was similar to that adopted by Jeffress (1964, 1967, 1968) in his investigations of a stimulus-oriented approach to detection. Signals were generated in the same way as they had been during the pigeon experiments. They were recorded by a microphone in the experimental chamber and processed by electronic devices which produced an output which a computer program used to make a "yes" or "no" decision.

The model of frequency discrimination was similar to one investigated by McAuley (1978). It consisted of a band-pass filter, analogous to a critical-band mechanism, and a device for measuring temporal intervals between axis-crossings of the stimulus waveform. The model was not evaluated except in a very general way, but, as McAuley points out, temporal models of this sort can be justified in terms of both psychophysical (e.g., Moore, 1973a, 1973b) and physiological evidence (e.g., Siebert, 1970), especially at lower frequencies. The main difficulties arise from uncertainty about the shape and bandwidth of the filter and the way in which the temporal information is processed (for example, the amount of averaging). McAuley varied both of these parameters, but systematic investigations were not possible in the present work.

While the nature of human critical bands is a matter of continuing debate, the topic has scarcely been touched in the field of bird hearing except with the budgerigar, which has been the subject of a number of investigations (e.g., Dooling & Saunders, 1975; Saunders, Denny & Bock, 1978). The results suggest that at frequencies around 300, 600 and 1000 Hz the critical bandwidths of budgerigars are approximately 750, 300 and 360 Hz respectively.

Figure 50 shows the main features of the simulation. Reproducible 450 Hz and 650 Hz signals were fed into the experimental chamber as they had been for the pigeons and were monitored by a half-inch microphone. The output of the microphone was passed through the filter which simulated the auditory filter, then analyzed to give rise to distributions of the intervals between the near-zero-axis crossings of the waveform. A parameter of the distributions was used to decide whether a standard (450 Hz) or comparison (650 Hz) trial had occurred.

Four types of degrading noise were investigated. These are shown in the dashed boxes in the figure. One source was wide-band Gaussian noise which was fed into the chamber along with the reproducible signals. This noise source represented the biological noise mentioned in Chapter 1, and also extraneous external noise. Another type of noise arose from variations in the filter settings, sometimes referred to as filter jitter (e.g., Henning, 1967b). A third type of noise resulted from inattention on the part of the simulated observers: on some trials decisions were made without reference to the parameter derived from the

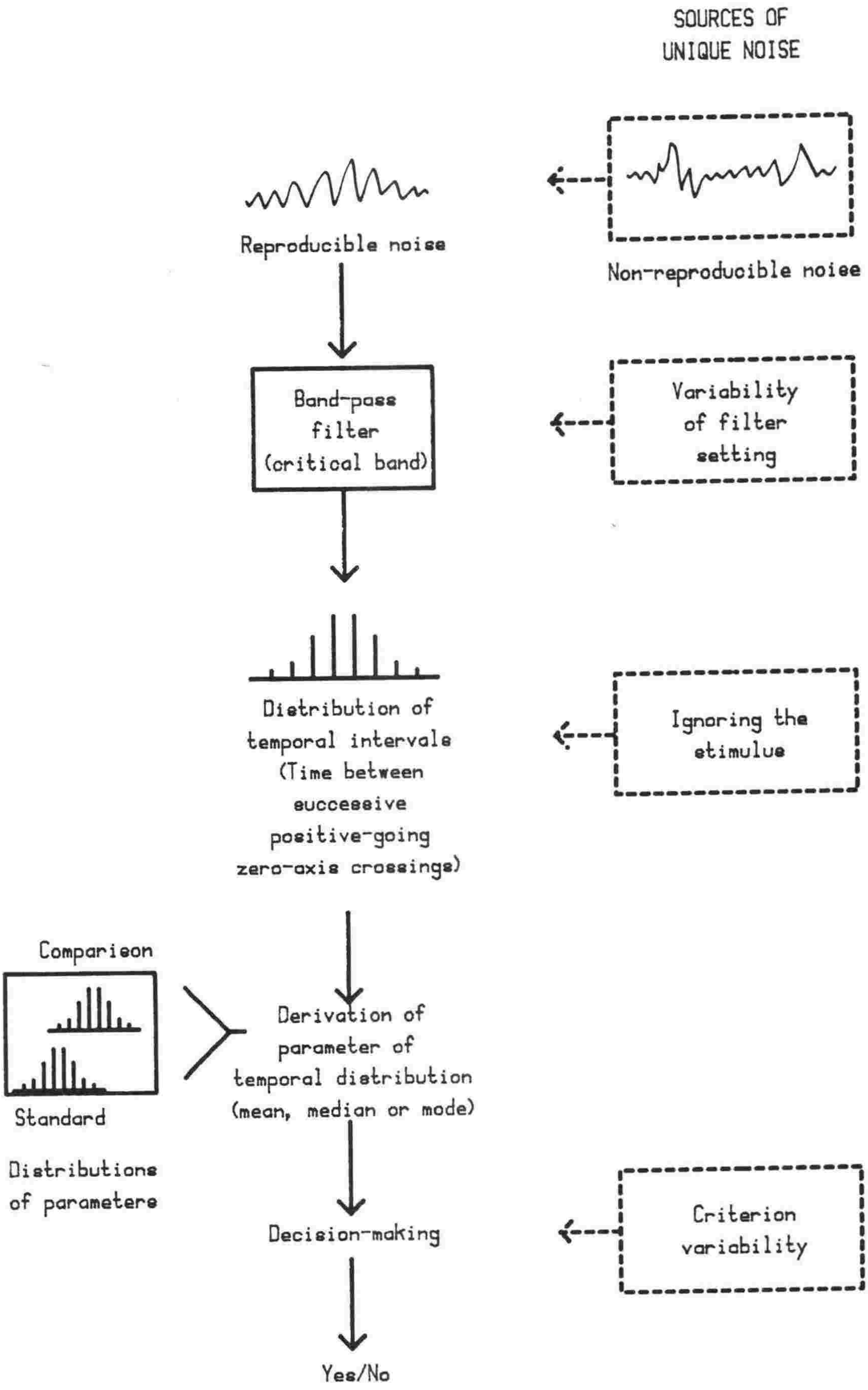


Figure 50. Block diagram giving an overview of the hardware simulation, showing the various sources of unique noise (dashed boxes).

distribution of temporal intervals. Finally, performance was degraded by trial-by-trial variations in the criterion held by the observer.

The sources of noise included in the simulation were not intended to exhaust the possibilities or to be detailed representations of the types of noise actually found in psychophysical experiments. The wide-band noise, in particular, stood in for a variety of sources (Soderquist & Lindsey, 1972). Inattention and criterion variability were included because observation and the results of earlier research (e.g., Heinemann & Avin, 1973; Blough & Blough, 1977) suggested that these have an important effect on pigeons' performances in auditory experiments.

The simulated observers' decisions were stored in the same manner as those made by the pigeons. The GOC and subsequent analyses took the same form as those carried out earlier.

Apparatus

Figure 51 shows the equipment used in the simulation. As noted at the top of the figure, the sound input to the system was provided by the apparatus pictured in Figure 35, supplemented for the simulation by a further Gaussian noise generator, low-pass filtered at 2000 Hz and again at 1000 Hz, which provided a source of unique noise.

The signals were monitored by a B & K 12.7 mm microphone, the output of which passed, via a B & K 2121 frequency analyzer located in the sound-attenuated room, and an attenuator, to two voltage-programmable band-pass filters (Frequency Devices models 301VT 1K25 [1250 Hz maximum frequency] and 2K5 [2500 Hz maximum]). The centre frequencies of these filters were controlled by the computer through two DACs. The resolution of the DACs gave a maximum setting error of 1 Hz, while the lower frequency of the pass band, set by the second filter, was 125 Hz.

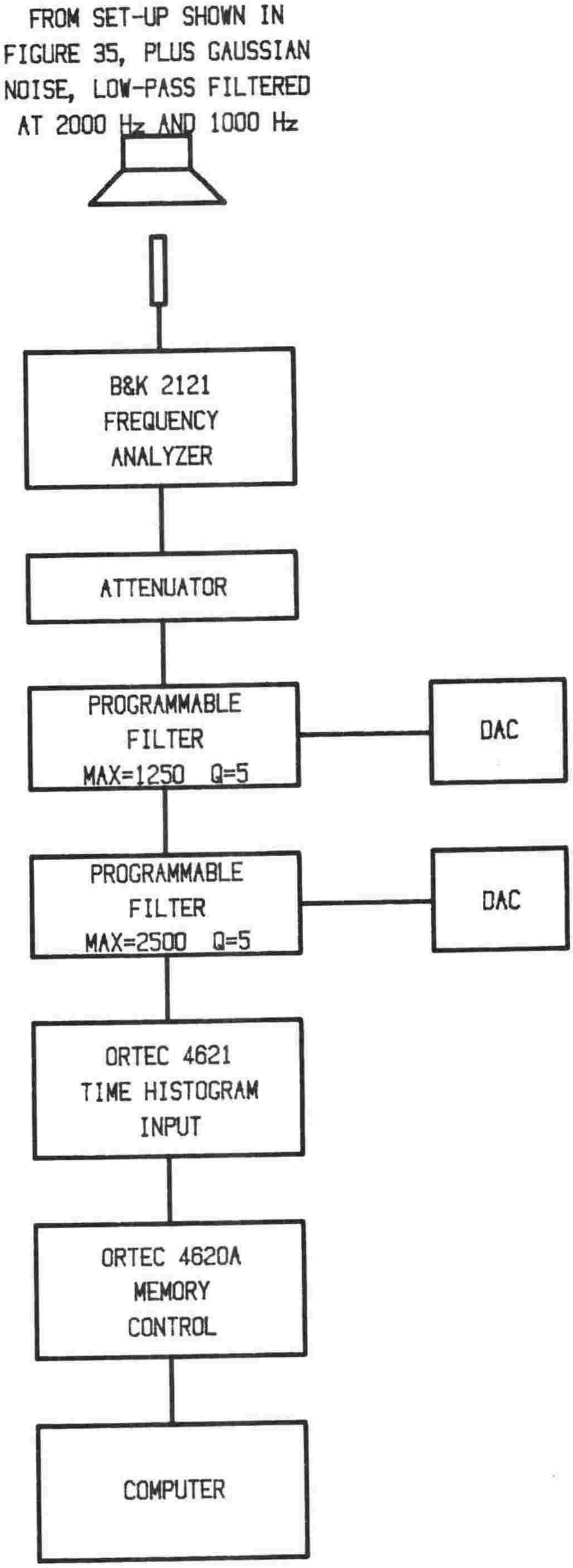


Figure 51. Block diagram of apparatus used in hardware simulation.

After some preliminary experimentation, the filters were set to Q=5. The shape of the output of the resulting filter system at a centre frequency of 450 Hz can be seen in Figure 52. This graph shows the results of a test based on sinusoids ranging in frequency from 14 Hz to 7192 Hz in 14-Hz steps. The 3-dB bandwidth of the filter system was approximately 13% of the centre frequency--59 Hz at 450 Hz and 89 Hz at 650 Hz. The equivalent rectangular bandwidth was 16%. The bandwidth was probably much narrower than the critical band for pigeons.

The output of the filter system was passed into an ORTEC 4621 Time Histogram Input analyzer. This device was set to respond to each positive-going near-zero-axis crossing of the input waveform.

An ORTEC 4620A Memory Control built up a distribution of temporal intervals in 255 20-μsec bins. Following each 140 msec signal, the counts in each bin of the memory were transferred to the controlling computer. Subsequent analyses were based on bin numbers. For example, distributions built up during 450 Hz signals would be expected to centre around bin number $1/(450 \times 20\text{x}10^{-6})$ or 111. The corresponding figure for 650 Hz signals was 77.

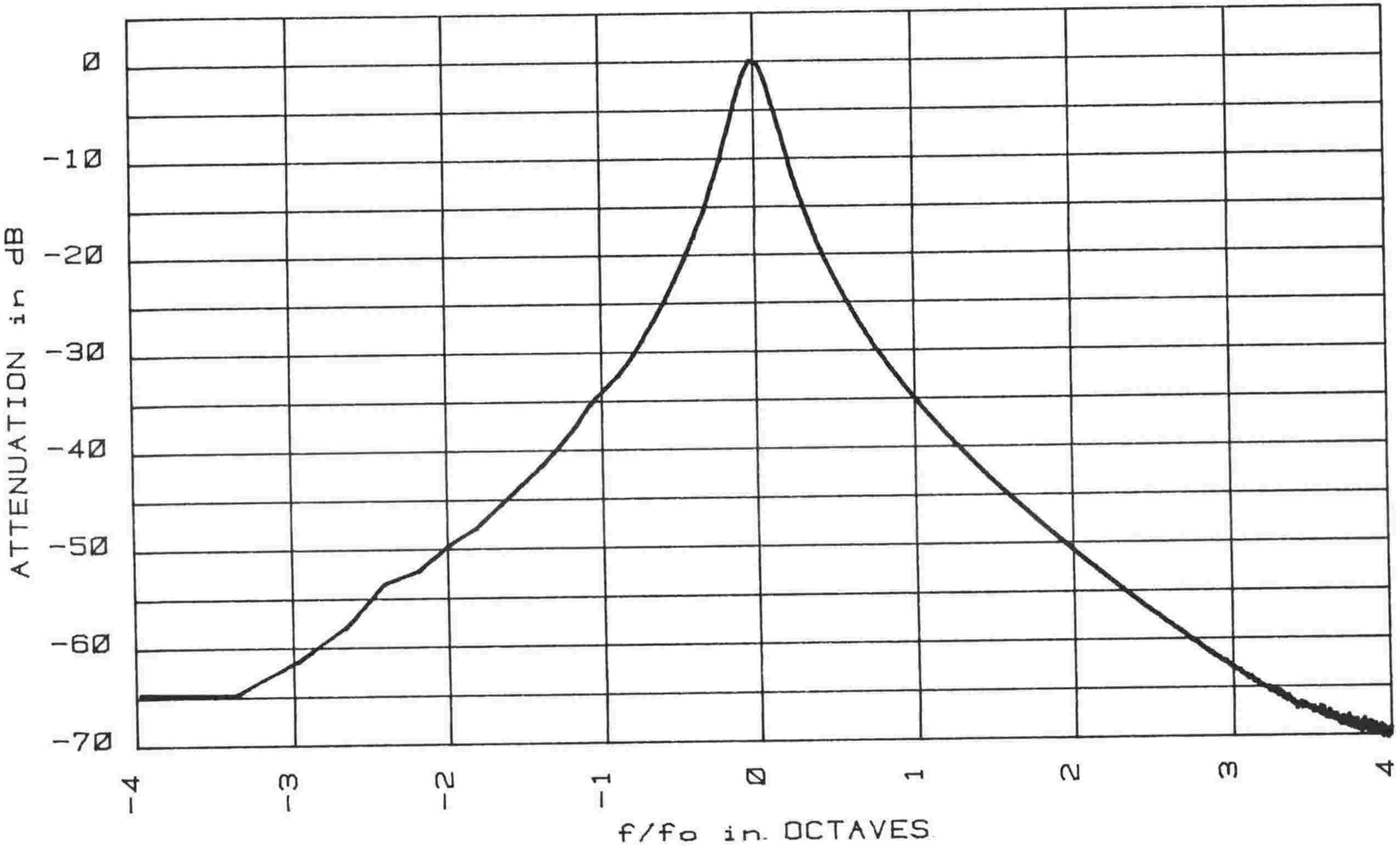


Figure 52. The output of the filter system used in the. hardware simulation at a centre frequency of 450 Hz..

The Simulation Program

Each replication in the simulation consisted of 396 450-Hz trials and 336 650-Hz trials (the 900 Hz signals used in the pigeon experiment were omitted). At the beginning of each replication a number of parameters were specified. These were:

1. the criterion, in bins (time);
2. the standard deviation of criterion variability, in bins (time);
3. the probability of ignoring the stimulus;
4. the probability of a "no" response given that the stimulus was ignored;
5. the filter setting at 450 Hz;
6. the s.d. of the filter setting around 450 Hz;
7. the filter setting at 650 Hz, and
8. the s.d. of the filter setting around 650 Hz.

In the final simulation, each trial in a replication consisted of a number of steps. The steps are described, then a more detailed account of some points is given.

a) The programmable filters were set near 450 Hz or 650 Hz, depending on the type of trial. This first setting introduced jitter in accordance with (6) and (8) above.
b) The signal was output to the operant chamber, recorded by the microphone, and the results processed by the equipment shown in Figure 51. The distribution of the temporal intervals was sent to the computer, and the mean value calculated.
c) A "yes" or "no" decision was made by the simulated observer after comparison with the cutoff nominated under (1) above.
d) The programmable filter was set exactly to the appropriate frequency (no filter jitter).

e) The same signal was again output to the operant chamber and processed as in b).

f) The distribution of temporal intervals was sent to the computer and the mean calculated.

g) With the probability nominated under (3) above, the observer ignored the mean and, with the probability nominated under (4), made a "yes" or "no" decision. When the observer did not ignore the information about the stimulus, it made a decision using the cutoff nominated under (1) with no criterion variability.

h) Criterion variability was studied independently of the effects of inattention. On each trial the mean of the temporal intervals used by the inattentive observer was also used by an independent observer with a variable cutoff. The cutoff was placed near the value set under (1) with an accuracy determined by the standard deviation nominated under (2). The mean value of the temporal intervals was then compared with the cutoff, and a decision made.

At the end of each trial, five items of data were stored. Means of temporal intervals were added to two distributions of the evidence variable, one affected by filter jitter (a. to c.) and the other unaffected by filter jitter (d. to h.). The "yes" and "no" decisions for observers affected by filter jitter, inattention and criterion variability were each stored independently.

During the simulation the variable filter and cutoff settings were chosen from an approximately normal distribution with mean zero and standard deviation equal to the nominated value. The distribution was provided by the Polar algorithm described in Chapter 4 and Appendix D.

In early versions of the program, the mean, median and mode of the distributions of temporal intervals were calculated. As discussed below, the results based on each of these parameters were generally very similar, so in the final version only the mean was used.

Preliminary Experimentation

The initial work with the simulation was exploratory and rarely involved the running of full replications. It served to establish several points. These will be summarised briefly.

Discrimination Performance Without Added Noise

Without any noise other than that presented in the pigeon experiments (the reproducible noise and the continuous low level masking noise) the simulated observer could discriminate perfectly. The parameters of the distributions of temporal intervals varied little, and there was no overlap between their distributions on standard and comparison trials. For example, on one run the means for 450-Hz trials were spread over five bins (s.d. 0.86) and those for 650-Hz trials over four bins (s.d. 0.64). The gap between the distributions was 30 bins. Under these circumstances a wide range of cutoffs could be used without affecting the perfect performance.

Noise in the Chamber

In an early attempt to degrade peformance, within-chamber noise was set at an overall level of 93 dB SPL, giving a spectrum level of 63 dB SPL with the 1000 Hz bandwidth. Although this noise led to an increase in the variability of the mean and the median, the performances based on them were not affected, since their distributions on standard and comparison trials still did not overlap.

The variability of the mode, however, was increased to such an extent that the 450 Hz and 650 Hz distributions did overlap. On one run an ROC curve not unlike those obtained with normal unequal variance distributions was obtained.

The use of high level noise in the chamber as the only source of degrading noise was not pursued, since it was clearly much more intense than any normal biological noise, and did not provide a very convincing analogue of other sources of unique noise. However, during all subsequent tests some additional noise was present in the chamber. Its overall level was around 70 dB SPL, giving a spectrum level of 40 dB SPL. It had the effect of reducing the signal-to-noise ratio from its original level of 38 dB to about 27 dB. This level of noise by itself had little effect on the performance of the discrimination.

Filter Jitter

Early experimentation showed that extreme filter jitter was needed to degrade performance significantly. For example, standard deviations of 300 Hz and 452 Hz at the 450 Hz and 650 Hz filter settings respectively were needed to obtain discrimination performances similar to those of the actual birds. Besides being rather unrealistic (presumably), this sort of jitter produced a significant number of filter settings which were out of the ORTEC's range, mainly at the lower frequencies.

The Final Simulations

The final version of the simulation program, described above, was used to study the effects of GOC analysis on discrimination performances degraded by two different types of noise--criterion variability and inattention--each added independently. With criterion variability, levels of initial performance similar to those of Birds 8 and 18 with the 650 Hz signal were produced. SO 8 and SO 18,[20] like their flesh-

---

[20] The two simulated observers were given the same numbers as the pigeons because an attempt was made to reproduce the performances of

and-blood counterparts, were run for eight and 13 replications respectively. Filter jitter was also used but the levels of performance were so high that the effects of GOC analysis could not be studied.

Gaussian noise with a spectrum level of 40 dB was fed into the chamber throughout the simulations. Other sources of noise were as shown in Table 4. For both simulated observers the setting of the cutoff varied between replications and, in the investigation of criterion variability, within replications. For SO 18 the setting of the criterion variability and of the filter variability also varied over replications.[21] Replication-to-replication variability was introduced in an attempt to copy the variability shown by the real birds. The cutoff chosen for a given replication was used for all decision-making within that replication; it remained constant within the replication for the observers affected by inattention and filter jitter, but varied for the observer whose criterion varied.

TABLE 4

Nominated values of the parameters of the final simulation

	SO 8		SO 18	
	Mean	s.d. over replications	Mean	s.d over replications
Cutoff (bins)	110.0	2.2	81.8	2.7
Criterion var. (s.d.)	80.0	0.0	47.1	4.5
P(ignore stimulus)	0.5	0.0	0.3	0.0
P("no"¦ignore)	0.5	0.0	0.5	0.0
s.d. of 450 Hz filter	100.0	0.0	98.9	24.8
s.d. of 650 Hz filter	85.0	0.0	82.1	7.2
Number of replications		7		12

---

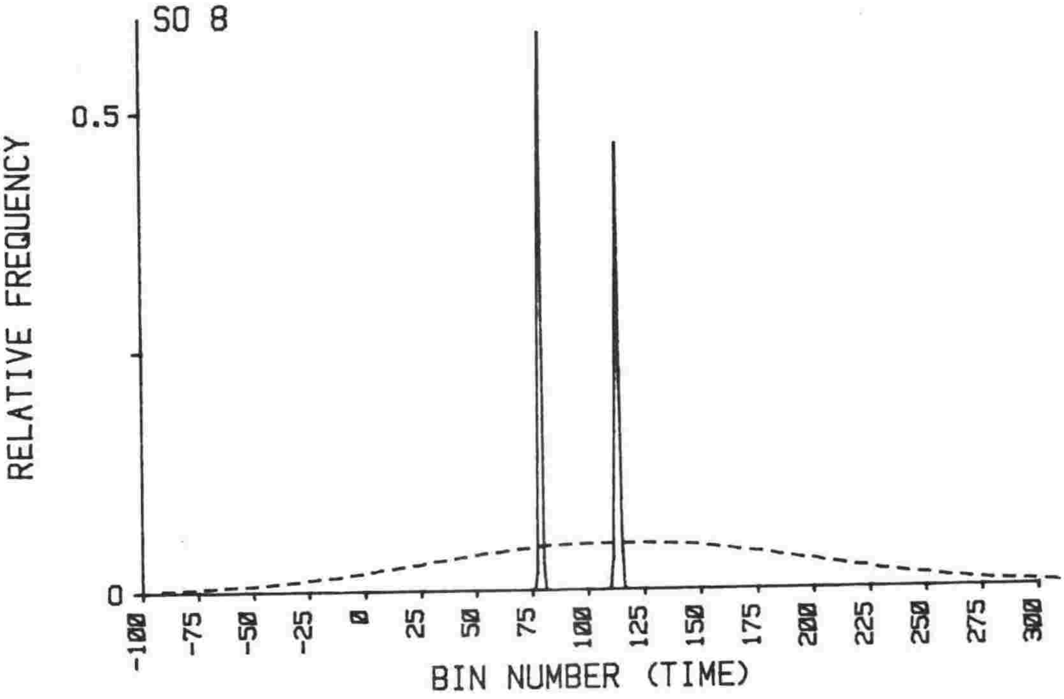
the real birds; the prefix "SO" is used for the simulated observers.

[21] The parameters and performance on each replication are given in Tables 10 and 11, Appendix C.

Results

Discarding of Replications. Initial GOC analyses for both SO 8 and SO 18 led to estimates of k which were much lower than would have been expected given the actual underlying distributions. Further investigations revealed very high correlations (0.95 to 1.00) between the responses on two pairs of replications for each observer (replications 5 and 7 for SO 8, and 6 and 8 for SO 18). The average correlations between responses on different pairs of replications were 0.02 and 0.09 for criterion variability and 0.23 and 0.25 for inattention for SO 8 and SO 18 respectively. Therefore the high correlations were unlikely to be due to chance, and must have arisen from accidental resetting of the random number generator or a re-cycling of the sequence. The true explanation was not established, but for all the analyses reported here the second replication in each of the highly correlated pairs was omitted. This had a dramatic effect on the estimates of k.

Criterion Variability. Figure 53 shows how the means of the distributions of the temporal intervals were distributed. The graphs are based on six replications for each simulated observer. Even with the noise in the chamber the distributions are very narrow (the standard deviation of the 450 Hz distribution was around 0.88 bins and that for the 650 Hz distribution was around 0.56) and there was no overlap. In order to produce levels of performance similar to those of the actual birds, considerable criterion variability--represented in the figure by the idealized normal distributions--was introduced. Also, as can be seen both from the graph and Table 4, the mean cutoff differed for the two simulated observers.



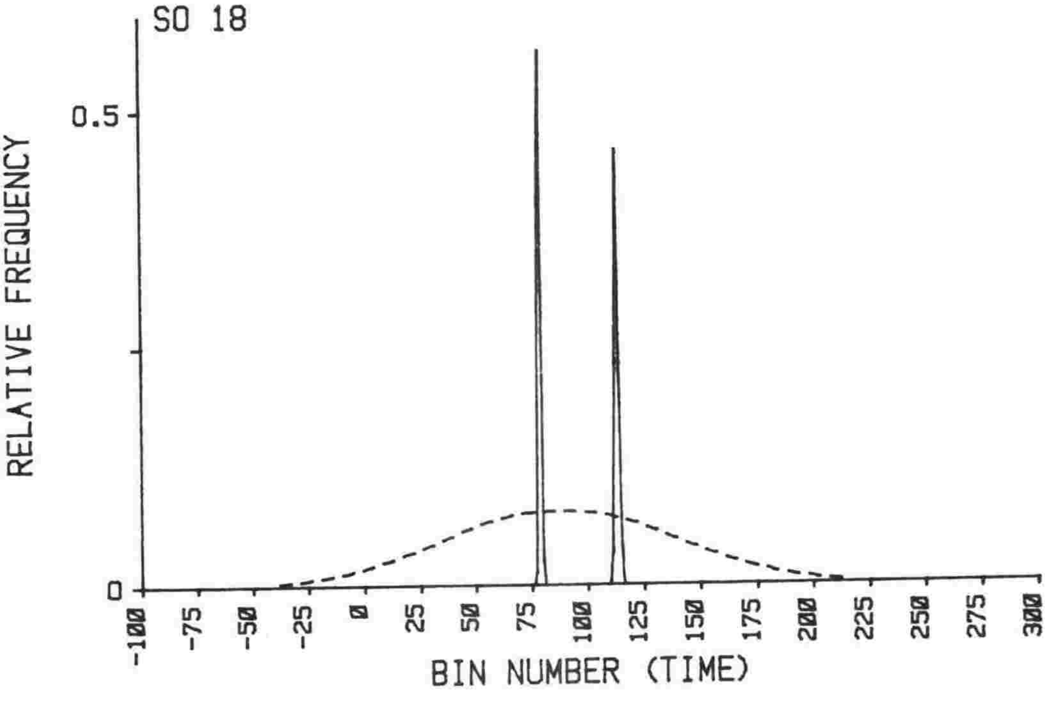


Figure 53. The underlying distributions and criterion variability in the hardware simulations of the fourth pigeon experiment.

The top left graphs in Figures 54 and 55 show the simulated observers' performances on each replication. As can be seen by comparing these results with those shown in Figure 44 (Chapter 8), the hit and false alarm rates generally fell in the same region as those for the real birds. The mean $d_{z(1)}$ values for the simulated and real observers were 0.378 and 0.440 respectively for SO 8 and Bird 8, and 0.788 and .826 for SO 18 and Bird 18.

Figures 54 and 55 also show the final GOC curves on linear coordinates (top left) and normal-normal coordinates (top right).

The estimates of $\underline{k}$, obtained with $d_{z(1)}$ as the base value in equation 4, are shown by the points in the bottom graph of each figure. The lines in these graphs represent the means and confidence limits of estimates derived from the seven 10-replication simulations with no common noise, described in the previous chapter.

For SO 8 the estimates of $\underline{k}$ are difficult to distinguish from the results expected in the absence of common noise. The results for SO 18 indicated a somewhat lower $\underline{k}$. Direct calculations, however, suggested a true $\underline{k}$ of around 4000. This figure was derived using equation 6 (Chapter 4), which gives an expression for $\underline{k}$ in terms of the variances of the unique and common noise distributions:

$$\underline{k} = 2\sigma_U^2/(\sigma_{SC}^2 + \sigma_{CC}^2)$$

where the subscript U stands for the unique noise, and SC and CC for the common noise of the standard and comparison distributions respectively. Values of 0.88 and 0.56 for the standard deviations of the unique noise (Table 4) give a $\underline{k}$ value of 4061. This value cannot be taken very seriously, because the distributions of common noise are not normal. However, it is clearly incompatible with the results in Figure 55, which suggest a $\underline{k}$ somewhere between 50 and 100.

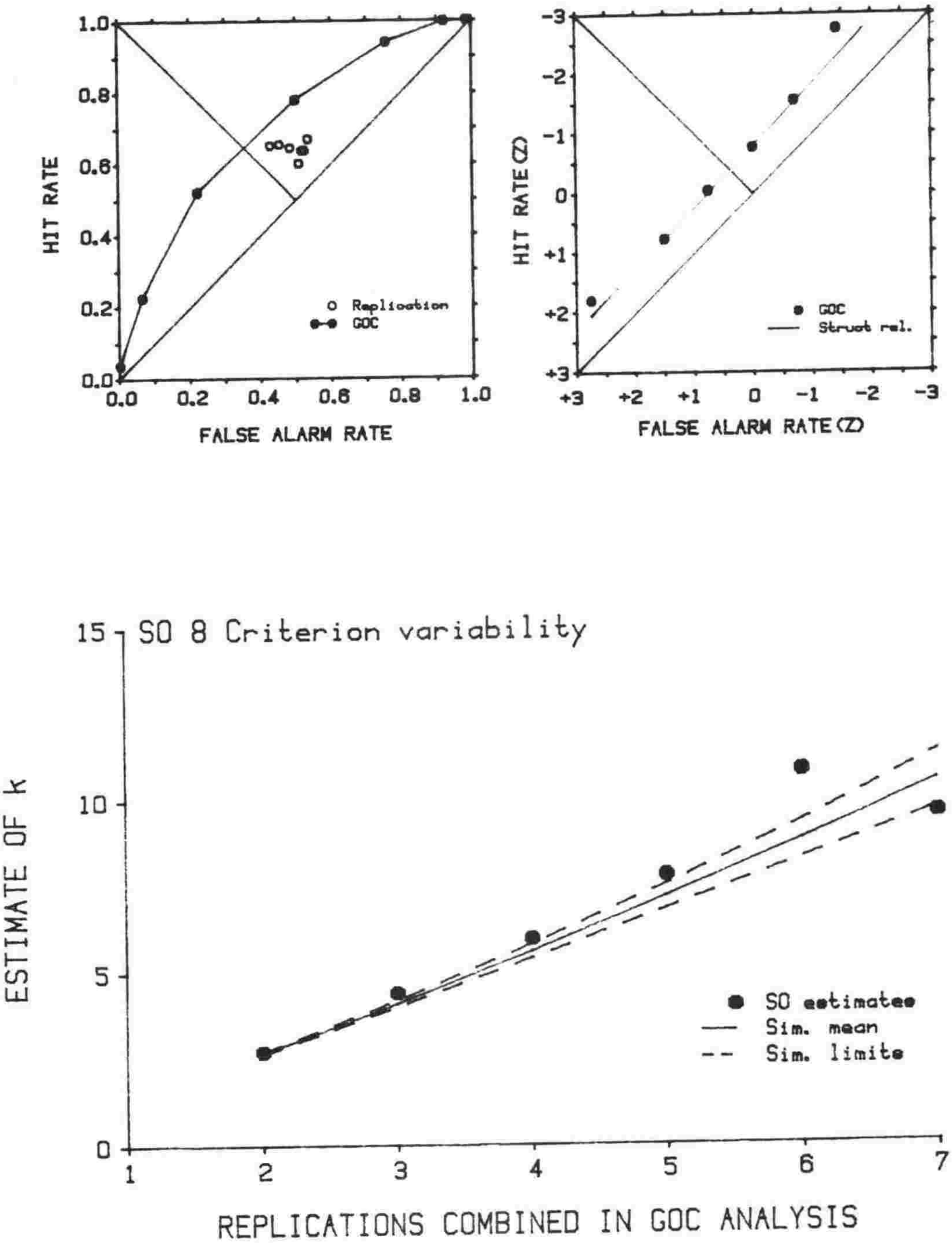


Figure 54. The GOC, individual replication results, and estimates of k for SO 8 in a hardware simulation involving criterion variability.

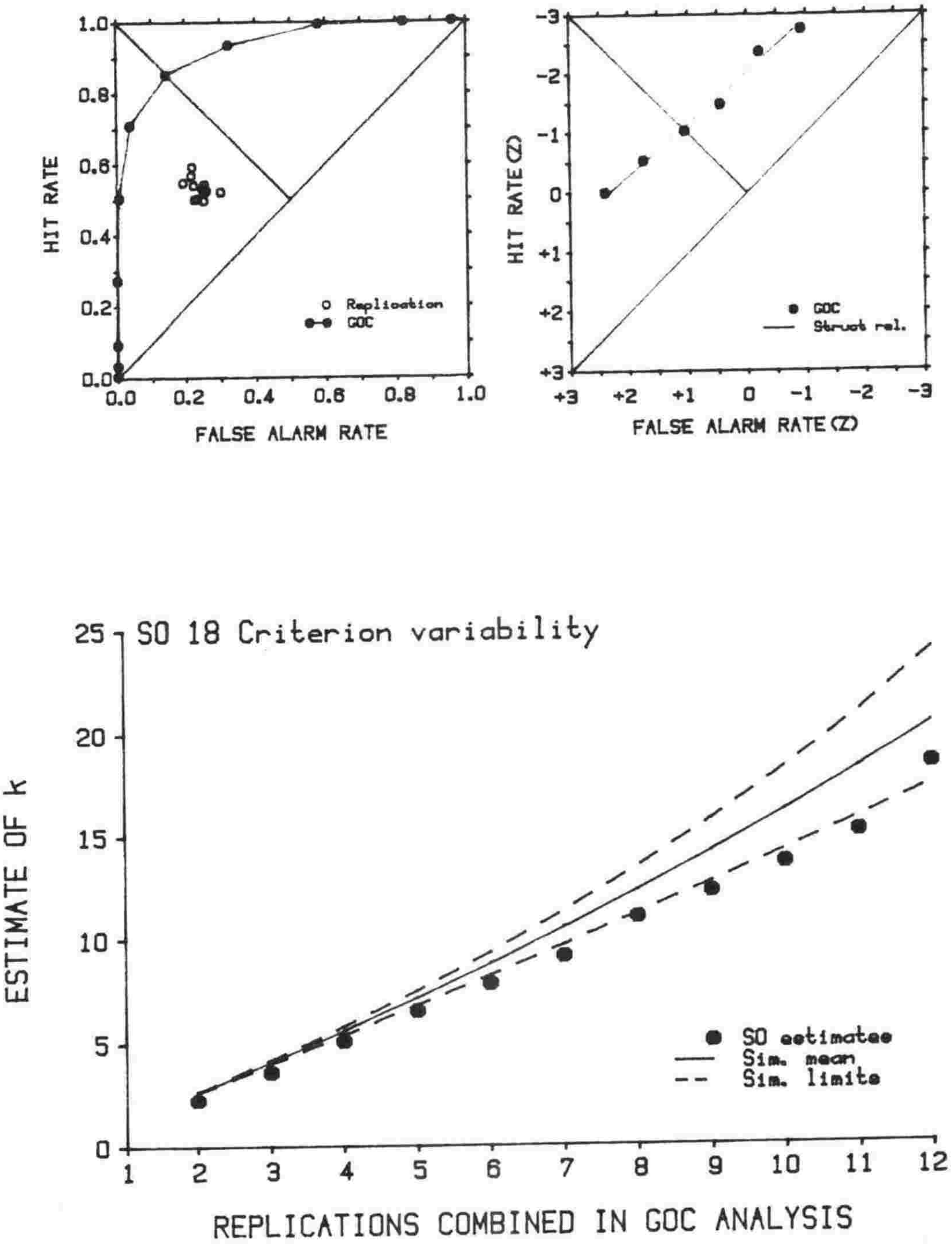


Figure 55. The GOC, individual replication results, and estimates of k for SO 18 in a hardware simulation involving criterion variability.

No explanation was found for the difference, and it must be tentatively attributed to error in the measurement of d_z. The percentage increases in d_z observed in the simulations were in the region where small variations in d_z led to very large changes in the estimate of $\underline{k}$ (Figure 22, Chapter 5).

Inattention. Figures 56 and 57 show the results obtained when the simulated observers ignored the stimulus. Though SO 8 ignored the stimulus on half of all trials and SO 18 on a third of trials, their performances were superior to those of the actual birds.

Most of the points representing the hit and false alarm rates for individual replications fell on the negative diagonal. Because of the large gap between the underlying distributions, the cutoff could vary considerably (in this case between replications rather than from trial to trial) without having much effect on the hit and false alarm rates. However, on four replications (1, 2, 4, and 5 in Table 11, Appendix C), the cutoff for SO 18 fell within the 650 Hz distribution. This led to a marked decline in the hit rate, but because there was no overlap between the distributions, the false alarm rate remained the same. Hence d_z fell markedly.

The estimates of $\underline{k}$ for both the simulated observers (especially those for SO 18) suggest departures from what would be expected if there was no common noise. However, once again $\underline{k}$ could be calculated directly, and in both cases was around 1000.

The direct calculation was possible because the means and standard deviations of the standard and comparison common noise distributions

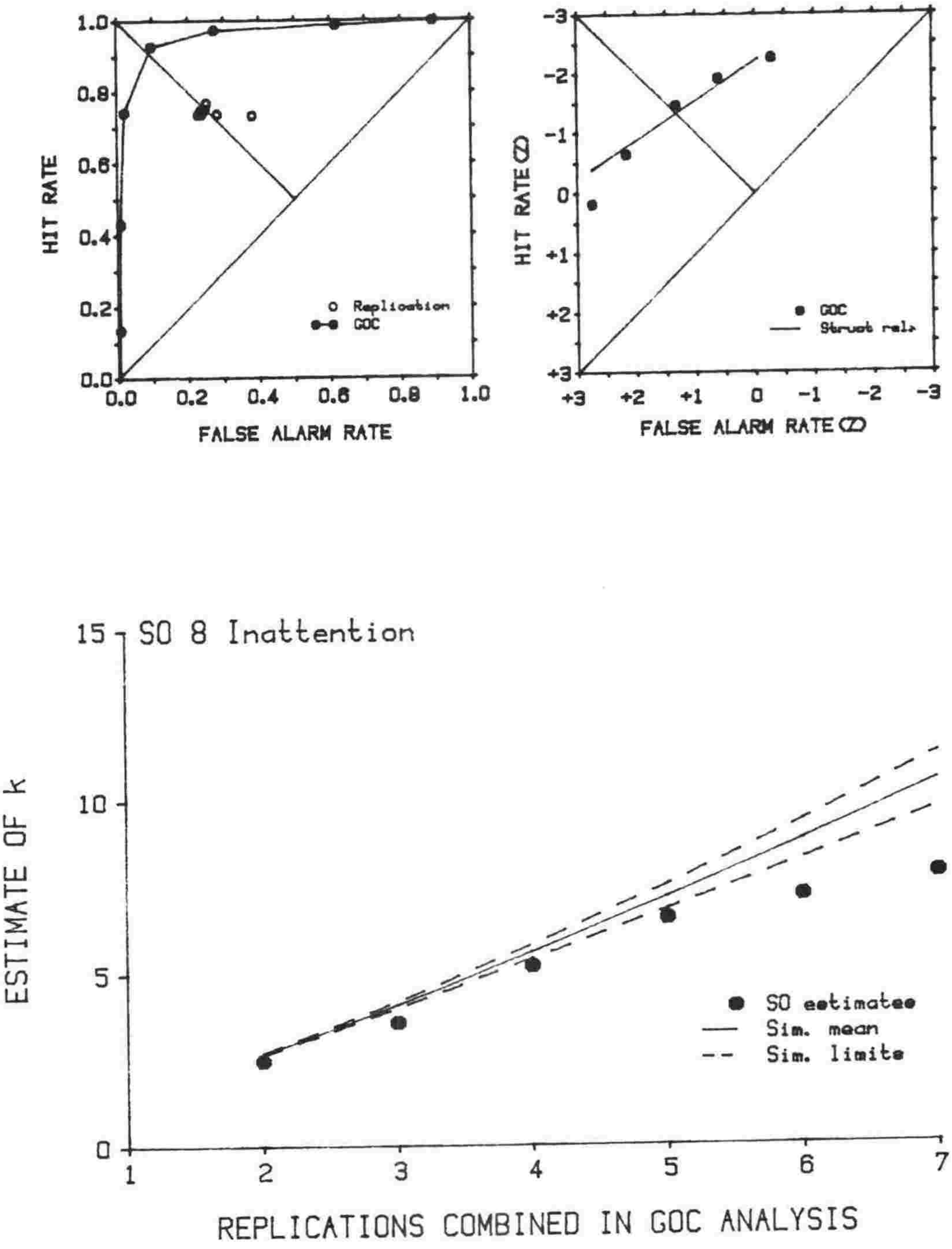


Figure 56. The GOC, individual replication results, and estimates of k for SO 8 in a hardware simulation involving inattention.

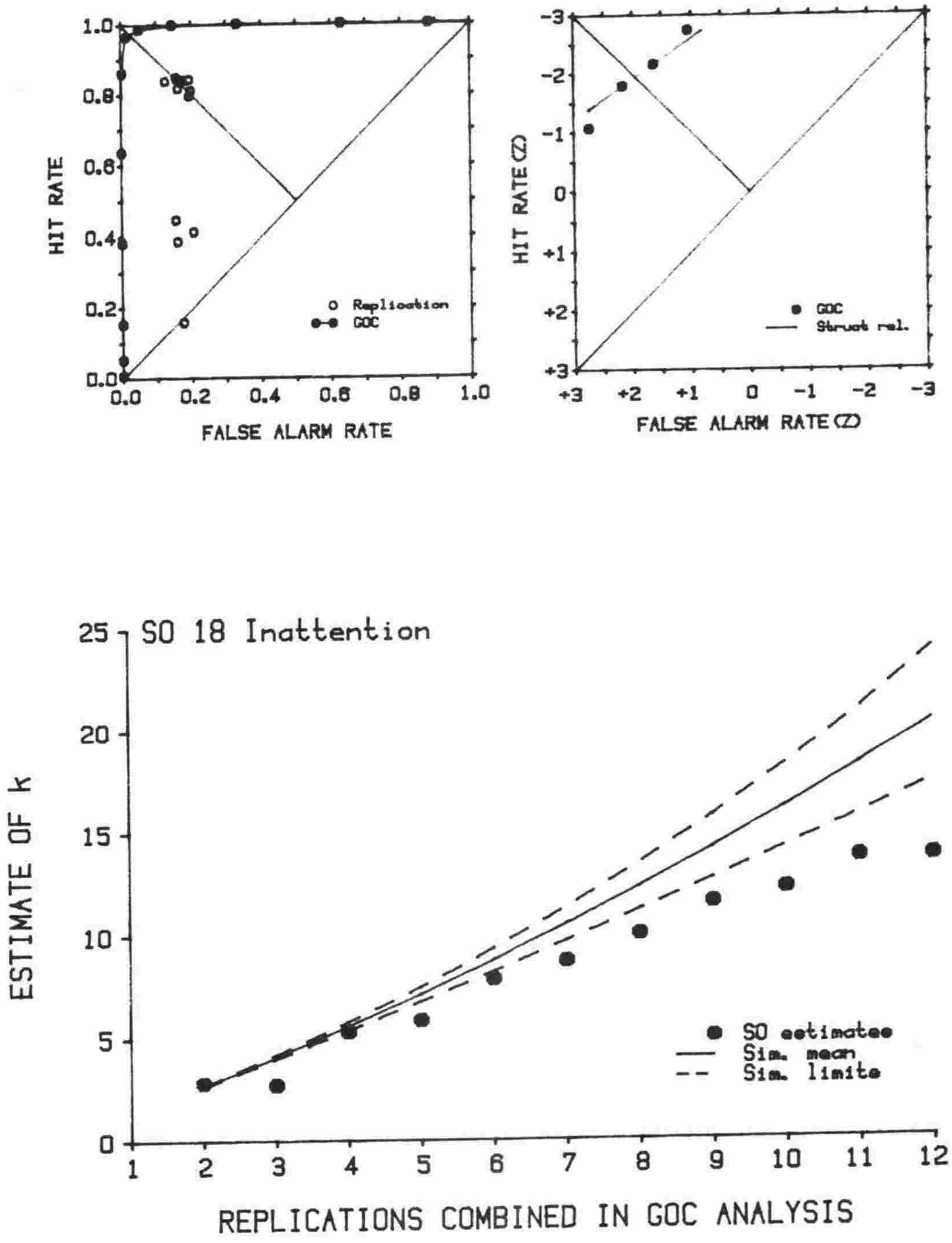


Figure 57. The GOC, individual replication results, and estimates of k for SO 18 in a hardware simulation involving inattention.

were known. From equation 5, Chapter 1:

$$d_z = (2^{1/2}m)/(\sigma_S^2 + \sigma_C^2)^{1/2}$$

$$= \frac{(2^{1/2}m)}{(\sigma_{SU}^2 + \sigma_{SC}^2 + \sigma_{CU}^2 + \sigma_{CC}^2)^{1/2}}$$

where the subscripts indicate the variance of the unique and common noise (U and C) making up the standard and comparison (S and C) distributions.

Since the values of d_z and of m, the difference between the means of the underlying distributions (34.4), were known for the simulation, the total variance could be calculated from a rearrangement of this expression, $2(m/d_z)^2$. The known variances of the underlying standard and comparison distributions could then be subtracted from the total variance to give the unique noise variance. The value of k was given by equation 6 (Chapter 4). For SO 8 k equalled 1182 and for SO 18 it was 1019.

Once again the departures of the results for SO 18 from the k estimates which would be expected are apparently misleading. Sampling error may be to blame, but in this case there are two other possible reasons. One is the drastic effect of cutoff variation for SO 18, discussed earlier. This would attenuate increases in d_z during GOC analysis. The other possibility is that estimates of d_z for GOCs based on larger numbers of replications were slightly attenuated. Because of the relatively high hit rates and low false alarm rates these GOCs were sometimes defined by fewer points than there were replications, the remaining points being lost because the values they represented were too large or too small to be shown on the normal-normal coordinates.

The fact that the simulated observers' performances were superior to those of the real birds despite the high levels of inattention suggests that if the birds did use temporal information in the same way as the hardware model, inattention alone was not responsible for degrading performance, a point that was in little doubt. However, the general outcome supports the idea that inattention is another source of unique noise, the effect of which will be removed in GOC analysis.

Filter Jitter. Variability in the setting of the "critical band" filter, unlike criterion variability and inattention, had an effect on the shape of the underlying distributions and even introduced some overlap. Nevertheless, because of the limits set by the possible range of the filter settings, the initial performances were still at a very high level, and a study of the effects of GOC analysis was not possible. In most cases, the straight lines necessary to calculate d_z could not be fitted to the GOC points since none, or only one, was defined by a hit and false alarm rate greater than zero and less than one. For example, d_z could only be estimated for 11 of the 78 possible two-replication combinations for SO 18, 64 of the 284 three-replication combinations and 110 of the 715 four-replication combinations. These results could have been quite misleading, and were not pursued further.

Discussion

A high level of degrading noise had to be added to the underlying distributions in order to produce discrimination performances similar to those of the actual birds. One possible conclusion is that the model is highly inadequate--that the temporal information available in the stimulus is not available, at least not in such discriminable form, to the birds themselves. However, consideration of the results obtained by Sinnott, Sachs and Hienz (1980) suggests that the model may not be entirely unrealistic.

At 500 Hz, Sinnott, Sachs and Hienz obtained frequency difference limens of around 10 Hz (Figure 37, Chapter 7). In the model used in the hardware simulation, signals of 500 Hz and 510 Hz would give rise to distributions of mean temporal intervals with means equal to $1/(500 \times 20^{-6})=100$ bins and $1/(510 \times 20^{-6})=98$ bins respectively. With this difference between the means of the standard and comparison distributions, and no masking noise (Sinnott, Sachs and Hienz did not use continuous or gated noise), ideal observers would discriminate perfectly. Since the pigeons in the Sinnott, Sachs and Hienz experiments did not do so, either the temporal model implemented here is inappropriate or noise from another source degraded performance. Assuming a d_z value of 1.28 (corresponding to a HR of 50% and a FAR of 10%) and a difference between the means of the distributions of $(100 - 98)=2$, the standard deviation of this noise would be 1.6. This seems small enough in comparison with the noise added in the hardware simulation to have some plausibility. Although Sinnott, Sachs and Hienz reduced the effects of noise due to memory, criterion variability and inattention, they could do little about unique noise from biological sources.

Pigeons do not appear to use temporal information as efficiently as the frequency discrimination model in the hardware simulation. However, the discrepancy between the predictions of the model and the pigeons' performances under reasonably favourable conditions is not so great as to rule the model out of consideration, given the presence of some unique noise.

If the model is realistic, the amount of unique noise added in the simulation may be similar to that actually present in the pigeon experiment with the 650 Hz signal. If so, the results would be similar to those expected in the absence of common noise.

According to this reasoning, the results obtained in both the pigeon experiments, which suggested that there was no common noise, are not incompatible with the presence of common noise which is obscured by high levels of unique noise.

The results of the simulation do not establish the sources of the degrading noise in the pigeon experiments.[22] However, they are consistent with the suggestion that decision noise--inattention and criterion variability due to imperfect memory for the standard (Jesteadt & Sims, 1975)--plays a relatively large part. The bandwidth of the filter system emulating the possible critical band in pigeons was much narrower than any suggested by the research on the budgerigar, yet the filter jitter had to be unrealistically extreme in order to have any effect on discrimination performance, as did the analogue of biological noise. This suggests the possibility that the critical band may be too large, as would be expected if unique noise affected the measures on which the estimates are based.

In sum, the hardware simulation showed how several types of noise could degrade performance. It also gave some support to suggestions that the results of the pigeon experiments were compatible with the presence of common noise in the pigeons' frequency discrimination performances. Thus the hardware simulation provided a demonstration of how such modelling might be used in conjunction with GOC analysis to test hypotheses about the nature of hearing mechanisms. This took a different form from that which would be expected if there was clear evidence of the presence of common noise.

---

[22] A source of noise not considered in the simulation was sequential dependency. Examination of the data for 450 Hz and 650 Hz trials in the pigeon experiment showed that both birds had a slight tendency to repeat responses. For Bird 8 $P[Yes(n)|Yes(n-1)]=0.589$ and $P[Yes(n)|No(n-1)]=0.513$; the equivalent figures for Bird 18 were 0.389 and 0.373. No other significant dependencies were found.

A final point raised by the results of the simulations is the difficulty of estimating k from the results of GOC analyses. In almost every case, the values indicated by the analyses differed from those calculated directly. This is probably partly because some of the assumptions underlying the direct calculations are not met, and partly a reflection of the high variability of k given small variations in d_z. The results also reinforce the point that comparison with the results of computer simulations is limited in that it may not be possible to incorporate factors affecting the true observer's performance, even if they are known, in the simulations. These simulations are based on assumptions in the same way as the Swets et al. model, though these are not as restrictive as those of the original model.

CHAPTER 10

SUMMARY AND DISCUSSION

This chapter consists of two main sections. The first is a summary of the previous nine chapters. It describes the concept of unique noise, outlines the four main areas of investigation, and gives the results obtained. The second part of the chapter discusses some of the main issues raised by the research.

Summary

Unique Noise

The concept of internal noise is a ubiquitous one in psychophysics. Identified as biological noise, it has been used to explain certain phenomena of binaural hearing and, more generally, to account for the variability of responses to the same physical stimulus in monaural hearing experiments. Attempts have been made to measure the magnitude and characteristics of internal noise, in particular by examining the consistency of responding with identical masking noises, any variability being attributed to the presence of additional noise.

Boven (1976) took up an approach described by Watson (1963) called Group-Operating-Characteristic (GOC) analysis. This technique is used to reduce the effects of unique noise, defined as a statistical concept referring to the idiosyncratic component of the total noise variation of an observer. By this definition unique noise can arise from biological sources and also from such things as criterion variability, faulty memory, inattention and sequential dependency. The complement of unique noise is common noise. In GOC analysis, responses made to the same signal on repeated trials (or by different observers on the same trial) are combined in a way which tends to average out the effects of unique noise. The analysis leads to a group equivalent of the ROC. The shape of this GOC, and the level of the performance it represents, are less affected by unique noise than those of the ROC.

Areas of Investigation

The overall aim of the research was to evaluate the effectiveness of GOC analysis in reducing the effects of unique noise on measures of sensitivity. The report covered four main topics within this general area. An outline of these topics will be followed by a brief description of the methods used to investigate them and finally a summary of the main outcomes.

The first topic followed on from work by Boven, which indicated that GOC analysis is less effective when the underlying distributions and the corresponding ROC are relatively complex. Boven suggested that when the ROC is thought to be complex, the number of trials in each replication should be increased. The present study looked at the effect of increasing the number of trials, the number of replications and the number of rating-scale categories.

The second main topic concerned the estimation of the relative variances of unique and common noise from the results of GOC analysis. Swets et al. (1959) and Watson (1963) gave formulae which can be used to estimate $\underline{k}$ from the changes in d' as replications are added in GOC analysis. However, in practice one or more of the assumptions underlying these formulae are unlikely to be met. In particular, when observers use rating scales with a finite number of categories, information is lost. As Boven pointed out, this will tend to lead to the underestimation of $\underline{k}$--in other words, to indicate that the relative variance of the unique noise is smaller than it in fact is. These points were followed up in the present study and an alternative method of estimating k was developed.

The third main area of interest lay in the use of GOC analysis to obtain ROC curves and measures of sensitivity relatively unaffected by unique noise. An important question in this sort of application is whether there is in fact any common noise. If identical noise which is presented to the ear on different occasions is not treated in the same way by the ear each time, all noise will be unique and GOC analysis will give rise to measures indicating perfect performance. In this investigation, GOC analysis was applied to the results of two experiments in which observers were trained to discriminate aural frequency.

The fourth area of interest was in the way theories of hearing might be studied by reproducing the results of GOC analysis with hardware and computer models. GOC analysis should provide a more clear-cut criterion against which to evaluate the outcomes of models than is usually available. In the absence of GOC analysis, allowance must be made for the presence of unique noise of unknown magnitude and characteristics.

Methods of Investigation

Four experiments were carried out with pigeons as observers. The birds were trained in a conventional operant chamber on a three-key discrete-trial procedure. Pecks at the centre key started each trial and thirty (not necessarily consecutive) pecks on one or other of the side keys led either to food or to a blackout, depending on the frequency of the auditory signal.

Other experiments involved simulated observers in models implemented on a computer. In most cases these simulations were entirely computer-based, with values of the evidence variable being generated by software. In a later simulation, however, signals previously presented to the pigeons were monitored and processed by a hardware system, and the simulated observer made decisions based on data transferred to the computer from the hardware.

One of the main advantages of the simulation (apart from the fact that the simulated observer or observers could use any number of rating-scale categories) was that the characteristics of both the common and unique noise could be nominated.

While there was no control over the unique noise of the pigeon observers, the first two experiments with pigeons were based on known distributions of common noise. The ROCs for these distributions provided a standard against which the results of GOC analysis were evaluated. The comparison was in one case all the more clear-cut because the distributions had shapes very different from those which might reasonably be expected for unique noise.

Results

This section describes the results obtained in each of the four main areas of investigation.

Multi-modal Distributions. The first experiment with pigeons was based on triangular probability mass functions. The ideal ROC for these functions was very similar to the general form of those obtained for normal equal-variance distributions. Thus, while the GOC curves for the birds approached the ideal ROC quite closely, unique noise, if approximately normally distributed, may have contributed to the apparent fit.

The GOC results obtained with bimodal distributions of common noise were, as expected, much less impressive, although those for all observers combined (a total of 18 replications) provided a reasonably unambiguous indication of the shape of the ideal ROC.

The computer simulation was used to demonstrate that if a larger number of rating-scale categories are used, a clearer picture of the ROC will emerge with a smaller number of replications. This may not be true for flesh-and-blood observers if, unlike the simulated observers, they cannot maintain multiple cutoffs consistently. Criterion variability could go a long way towards offsetting the advantages of extra rating-scale categories.

The computer simulation was also used to evaluate a suggestion by Boven that, when the ROC is thought to be complex, relatively large numbers of experimental trials should be used in order to reduce the effects of chance in the placement of points of the GOC curve. With the trimodal distributions developed by Boven, and a simulated observer using a four-category rating scale, better results were obtained by adding more replications than by adding more trials per replication. This conclusion may only apply in cases where the observers, like the simulated one, maintain consistent cutoffs, and the common noise is adequately sampled.

<u>The Estimation of $\underline{k}$.</u> The main difficulty in estimating $\underline{k}$, the ratio of the variances of the unique and common noise, was due to loss of information. This was especially serious with the pigeon observers, who used only two rating-scale categories. With their data, the formulae provided by Swets et al. (1959) led to gross underestimates of $\underline{k}$. Another difficulty arose when the slope of the ROC for individual replications could not be estimated because the two-category data produced only one point in the ROC space. This latter difficulty could be avoided by using $d_{z(2)}$ as the standard against which changes in d_z values due to GOC analysis could be measured. In other cases the slope

of the ROC for individual replications could be estimated by extrapolating a function fitted to the slopes of successive GOCs. Because of the all-combinations analysis, which balanced the changes in d_z due simply to the order in which replications were added, changes in GOC slope tended to be relatively orderly.

In the absence of an analytic theory dealing with the two-category case, $\underline{k}$ values for the pigeon experiment were estimated by comparing them with the results of computer simulations. In these, simulated observers used two-category rating scales, and because the parameters of the distributions of unique and common noise were known, $\underline{k}$ could be calculated directly.

This method was reasonably successful when applied to the results of the second pigeon experiment, which was based on normal unequal-variance distributions. However, the somewhat discrepant estimate for one bird, which was to some extent due to the use of discrete distributions, underlined the fact that the method has a number of limitations. The first is that the distributions of both unique and common noise are still assumed to be normal. Another is the assumption that all common noise is due to the reproducible noise in the signals; in fact, common noise may arise from such things as individual noise and sequential dependency. These would either be difficult to reproduce in a simulation or would give rise to so many free parameters as to make simulation impracticable.

A final problem is that small variations in d_z can produce large variations in the estimate of $\underline{k}$, especially when $\underline{k}$ is high. This leads to uncertainty when comparing the results of experiments and simulations, unless both are based on large numbers of noise samples, trials and replications. This problem is mentioned again in later sections.

Application of GOC Analysis in a Psychophysical Experiment. The results of the third and fourth pigeon experiments, in which the nature of the underlying distributions was unknown, indicated that *k* was very high. With such high levels of unique noise, the GOC curves, even that for Bird 18, which was based on thirteen replications, could reveal little about the underlying distributions of common noise. Instead the major question was whether there was any common noise at all.

In the third experiment each of the five comparison frequencies was represented by only a few trials per replication, and there were only five replications in all. The resulting variability in the estimates of *k* made the results inconclusive. This underlines the point made earlier about the effects of variability in d_z on variability in *k*.

In the fourth pigeon experiment, the GOCs were based on only one frequency, which was therefore better represented in each replication. Consequently, the estimates of *k* were much more orderly, and showed clearly that the results for the pigeons could not be distinguished from those obtained in simulations in which there was no common noise.

Modelling Pigeon Frequency Discrimination. The model of frequency discrimination implemented in the hardware simulation consisted of a bandpass filter and a device for measuring the interval between near-zero-axis crossings of the waveform. On each trial the observer used a parameter of the distribution of the intervals between positive-going axis crossings to make a decision.

The distributions of the parameters for standard and comparison signals did not overlap. The model predicted that, in the absence of unique noise, the birds would discriminate perfectly. High levels of unique noise had to be added to the model in order to produce levels of performance similar to those of the pigeon observers.

One interpretation of this finding is that the model implemented in the simulation is unrealistic: that the parameters are inappropriate or that pigeons do not use temporal information in the way suggested. The other possibility, which has some plausibility given the discrimination results reported by Sinnott, Sachs and Hienz (1980), is that the level of unique noise in the kind of pigeon experiment reported here is indeed very high.

If the levels of unique and common noise in the experiments were similar to those in the hardware simulation, the estimates of $\underline{k}$, like those obtained in the simulations, would be difficult to distinguish from those expected in the absence of common noise.

Discussion

The topics discussed in this section include the methodology used in the pigeon experiments, the estimation of $\underline{k}$, common and unique noise and finally GOC analysis. In the course of the discussion results of several experiments on internal or unique noise in human subjects, first referred to in Chapter 1, are described in some detail.

Methodology

The Trial Procedure and Signal Presentation. The procedure used with the pigeons was chosen because of its symmetry and the clear-cut nature of the choice responses. The birds were trained to respond to one key on standard trials and another on comparison trials, and "yes" and "no" responses were clearly defined. Originally only one response to either key was required. Clarity of definition was lost to some extent when the 30-response requirement (Krasnegor, 1971) was added, but the symmetry was retained.

The way in which the stimuli were presented was equivalent to the yes-no (YN) technique in human psychophysics. This technique has been previously used with pigeons (in conjunction with the above procedure) by Krasnegor and by Heinemann and his colleagues (e.g., Heinemann & Avin, 1973). Krasnegor's frequency discrimination results were very different from those obtained by investigators who had used some variant of the same-different (SD) procedure (Figure 37). The suggestion by Jesteadt and Sims (1975) that the inferior performances obtained with the YN procedure could be attributed to criterion variability indicated that this was an ideal procedure with which to investigate the GOC technique. However, in the third and fourth pigeon experiments, the YN method, as implemented in this study, may have led to levels of unique noise which overwhelmed any evidence of common noise.

For further experiments, a different method of stimulus presentation and perhaps a different procedure would need to be considered. The most extensive and favourable frequency discrimination results for pigeons were obtained by Sinnott, Sachs and Hienz (1980), who used a two-key procedure and a same-different format. They adopted the two-key procedure after pigeons trained on a three-key procedure (in which only one peck on either side-key ended each trial) failed to discriminate between the presence and absence of sinusoids in a sensitivity experiment (Hienz, Sinnott & Sachs, 1977). In the two-key procedure the birds pecked at an observation key to produce the stimuli and reported the presence of a signal on another key. In the frequency discrimination version of the procedure, they reported the alternation of different frequencies. The choice responses were not symmetrical, in that "yes" responses were registered by shifting to the report key while "no" responses were registered when a bird failed to shift during a certain time following occurrence of an observation interval.

The two-key procedure appears to have some advantages over the three-key procedure. However, good results might be obtained with the three-key procedure if the same-different method of presenting stimuli was coupled with the multiple-response requirement (i.e., 30 responses on a side-key) used in the present study. In this case the advantage of symmetrical choice responses could be retained.

The use of the same-different procedure would complicate things slightly in that the experimenter would have to decide how to pair samples of common noise associated with standard and comparison frequencies. For example, particular samples could always occur together, or all possible pairings could occur over the course of the experiment. If there was any doubt about whether the birds were integrating over a period longer than the duration of the signals, the first possibility would seem to be the most desirable.

The Number of Response Categories. The above discussion assumes that the birds may make only "yes" or "no" responses. As has been pointed out, the use of a greater number of categories would be desirable.

A number of experimenters have used measures such as response latency (e.g., Yager & Duncan, 1971; Green, Terman & Terman, 1979) and response rate (e.g., Blough, 1967) to construct ROCs for animal psychophysical data. Results based on these sorts of measures could be used in GOC analyses, and could be expected to lead to better-defined GOC curves because of the larger number of points. However, as Yager and Duncan (1971), and Emmerich, Gray, Watson and Tanis (1972) have noted for latencies, some measures may introduce variability or noise not present in simpler response measures. This would contribute to the unique noise in the experiment, and go some way towards negating the advantage of the extra categories.

Some preliminary work in the present study suggested that rating scales based on data collected in conjunction with two-category responses (e.g., the distribution of responses over the two keys and the number of changeovers), did not usually add anything to the final GOC results in terms of sensitivity.

As has been pointed out by Green, Terman and Terman (1979), ratings of the sort collected in animal experiments may be confounded with the "yes" and "no" choices. Thus, while the ratings may provide extra points which help to define a complex ROC more clearly, the sensitivity indicated by the ROC is unlikely to be significantly increased. Ratings independent of the "yes"-"no" choices could in principle be collected from one of the choice keys in the present procedure. However, the data would have been of limited value in these experiments, since the birds adopted patterns in which they nearly always made an initial burst of responses to the same key, changing over or remaining on that key according to the frequency of the stimulus.

While development of a rating scale superior to the two-category scales used by the birds might be possible, there is room for doubt that the resulting data would overcome the problems of information loss and subsequent under-estimation of _k_. Thus a strategy of the sort used in this research would still be necessary unless an appropriate analytical model were adopted.[23]

Stimuli. A study by Siegel (1979; cited in Spiegel & Green, 1981) was mentioned in Chapter 1. Internal noise was measured in a two-interval forced choice (2IFC) task in which observers attempted to detect a sinusoid in one of the intervals. On some trials the maskers

---

[23] Dr R. Davies of the Applied Mathematics Division of the DSIR kindly developed such a model. It assumed normally-distributed common noise and arcsin distributions of unique noise. While his model was not investigated as part of the present study, it could form the basis for further work involving two response-categories.

in the two intervals were identical, while in others they were different. The magnitude of internal noise was estimated from the difference between detection performance on the two different kinds of trials. The assumption was that when the maskers were identical the only noise variation was due to internal noise. When the maskers were different, both internal and external noise contributed to the noise variance. The smaller the superiority of performance with the identical maskers, the greater the relative variance of the internal noise was judged to be.

Siegel's results suggested a very high level of internal noise. The ratio of the standard deviations of the external and internal noise was much higher than those reported by Swets et al. (1959), Watson (1963) and Green (1964). In a later report Siegel (1981) suggested that the reason for his descrepant results was that, because his noise samples were relatively long (the duration of the signal was 256 msec), subjects were able to listen to different sub-intervals of identical maskers. They thus created independent maskers which were in effect equivalent to the truly independent maskers. Because there was little improvement in performance when nominally identical maskers were used, the variance of the internal noise was seen to be very large.

Although Siegel's later explanation was not directly supported by the results of a replication carried out by Spiegel and Green (1981)--their estimates of the ratio σ_I/σ_E fell between 1 and 2 for all observers--it is a possibility which must be considered. Thus, despite the fact that the acoustic transients used in the third and fourth pigeon experiments were relatively short (140 msecs total duration), and the intervals between them relatively long (360 msecs), the pigeons may have sampled within the stimuli, creating sets of independent maskers and raising the level of unique noise.

In the absence of clear evidence about integration times, the only solution is to reduce the duration of the noise samples as much as is feasible. In their replication of Green's (1964) study, first mentioned in Chapter 1, Spiegel and Green (1981) used observation intervals of 10 msec in a two-interval forced-choice experiment. Trials consisting of two independent samples of noise (no signal in either interval) were later repeated. The consistency of the observer's choice of the first or second interval as the one which contained the signal was used to infer the level of internal noise.

Estimates of internal noise were lowest when the observers were presented with only a 10 msec transient in each interval, and highest when reproducible noise began 250 msec before and ended 250 msec after each 10 msec observation interval. In the first case, there was no chance of the observer sampling outside the 10 msec interval; in the second the observation interval appears to have been marked only by the onset or offset of a light. The observers may have sampled outside the observation interval, and there may have been some variability in the part of the waveform actually sampled. This could account for the lower level of unique noise found when the observers could not sample outside the 10 msec interval.

Signal-to-Noise Ratio. The signal-to-noise ratio in the third and fourth pigeon experiments (40 dB) was high compared with that in most experiments with human subjects. If a lower ratio had been used, perhaps the results of the GOC analyses would have been less ambiguous. The reasoning here is that if the variance of the external noise had been greater, either $\underline{k}$ would not have been so high (given there was common noise), or the obtaining of $\underline{k}$ values indistinguishable from those expected without common noise would have pointed more strongly to the absence of such noise.

This reasoning does not take into account the possibility that internal noise may increase with increases in external noise (Swets et al., 1959). With the pigeons this could come about if a decrease in the signal-to-noise ratio led to a decline in performance due, not only to the greater masking of the sinusoids by reproducible noise, but also to an increase in such things as criterion variability and inattention. The effects of decreasing the signal-to-noise ratio, then, are not easily predicted.

The Estimation of *k*

If there is common noise, the GOC curve will, given enough replications, reflect the nature of the underlying distributions of common noise regardless of the number of response categories used. This is shown by the results of the first two pigeon experiments and some of the simulations. As has become obvious, however, results of GOC analysis may not be very useful in estimating the relative variances of unique and common noise. More accurate estimates of *k* may be obtained if observers use rating scales with large numbers of categories.

In the Swets et al. (1959) experiment, first mentioned in Chapter 1, the integration of information was carried out by the observers, rather than through the summation of responses. The subjects observed a set of four intervals five times on each trial. The signal always occurred in the same interval on a given trial, but the masking noise could be either constant or variable over each of the five sets of intervals in a trial. On the variable-noise trials, the observers could integrate over all the noise in the experiment (both external and internal) and the improvement in d' over the five observations within each trial was expected to follow that predicted by equation 1 (Chapter 1). The fact that, for the three observers used, the prediction was approximately correct suggested that the observers combined the information from the sets of intervals in more or less optimal fashion. This suggests that the estimates of k derived from the improvements in d' on trials with constant noise, in which the observers could only integrate over internal noise, were not significantly deflated by loss of information in combining observations.

The methods of Green (1964) and Siegel (1979) are not subject to problems caused by loss of information. However, as Spiegel and Green (1981) point out, the percentage agreement obtained in Green's procedure will be affected by bias towards one or other of the intervals in the 2IFC procedure. In both methods the error of estimate of k becomes high for certain values of the variable used to estimate it. These are a low percentage agreement in Green's procedure and a small percentage improvement in the d' obtained with identical as opposed to independent maskers in Siegel's procedure. As shown in Figure 22 in Chapter 5, the corresponding error in the estimate of k is large in GOC analysis when the percent improvement in d' over some base value is large.

In sum, there are difficulties involved in all of the above methods of estimating k. For human subjects, however, GOC analysis may have more drawbacks than other techniques, and if the estimation of k is the only goal it may not be the method of choice. For animals which cannot be trained to use the 2IFC method, there may be no alternative to GOC analysis.

Common Noise

The results of the present experiments do not answer the question of whether there is common noise in pigeon frequency discrimination. What they do show is that it can be very difficult to distinguish between the absence of common noise and high, but finite, levels of k.

Some of the studies of internal noise described here have led to results which clearly indicate the presence of common noise (e.g., Swets et al., 1959; Green, 1964), but all of these have involved the detection of signals in noise. The only available data on frequency discrimination are those of McAuley (1978), who used GOC analysis to facilitate comparisons between a model of frequency discrimination (referred to in Chapter 9) and results for human observers.

The experiments for which the results of GOC analysis are available involved masked sinusoids in a same-different (SD) procedure. Frequency differences ranged from 1 Hz to 6 Hz at 250 Hz, and there were two S/N ratios (23 dB and 30 dB). The highly trained observers used eight-category rating scales in order to reduce the loss of information.

In the GOC analysis, 10 replications for each of five observers were combined, making 50 replications in all. McAuley did not make estimates of k. For the present purpose, values of k were obtained by applying equation 4a. The lower value of d' was based on five, ten or 20 replications, while the higher value was based on ten, 20 or 50 replications.

In general, k was higher for higher S/N ratios and for smaller frequency differences. With a S/N ratio of 23 dB, k ranged from nine (at a frequency difference of 2 Hz) to five (at a difference of 6 Hz). With a S/N ratio of 30 dB, k ranged from 44 (at a frequency difference of 1 Hz) through 21 (at a difference of 2 Hz) to 14 (at a difference of 3 Hz).

These results support the suggestion made earlier in this chapter that lowering the S/N ratio will reduce the value of k. In other words, they suggest that the level of unique noise remained constant when the level of the reproducible masking noise was changed. The results thus conflict with the findings of Swets et al. (1959), which indicated that internal noise was proportional to external noise.

The variability of the estimates of k based on different numbers of replications suggested that some information was lost as responses were combined, and that k was underestimated. However, the loss of information does not appear to have been great and the estimates of k should be reasonably accurate. If the results are taken at face value, they show that there was common noise in McAuley's frequency discrimination experiments.

Unique Noise

The main aim of GOC analysis is to reveal something about the nature of the common noise in an experiment (insofar as this can be discerned from the GOC curve) and the sensitivity of the observers in the absence of unique noise. The characteristics of the unique noise--apart from its relative variance--while of interest, cannot be determined from the analysis.

For the purposes of the simulations and the analysis of the pigeon experiments, the noise was assumed to be normally distributed with zero mean, and the data have been consistent with this assumption. This is not surprising since when variance from such noise sources as criterion variability, lapses in concentration and physiological noise are added, the result will probably be approximately normally distributed. A consequence of the normal distribution is that ROC curves for underlying distributions which are not of equal variance will be systematically distorted so as to appear more like equal variance distributions than they in fact are.

GOC Analysis

The present study shows that GOC analysis is a demanding technique to implement: there must be some means of reproducing stimuli; signals must meet certain criteria if within- and between-transient sampling is to be avoided; there must be a large number of replications and a large number of trials per replication in order to reduce the sampling error (especially when k is high) and to represent the underlying distributions adequately; and the observers should use as many rating-scale categories as possible.

These requirements would be met more easily with human observers than with pigeons: human subjects can be trained with single very brief signals more readily than pigeons, they can get through trials more quickly and they can use a greater number of rating-scale categories.

For the reasons given above, further work with GOC analysis could probably be done most efficiently with human subjects, perhaps employing amplitude discrimination as well as frequency discrimination. With pigeons or other non-human observers, studies employing a less noisy psychophysical procedure could be contemplated. As with human observers, work could be done with both amplitude and frequency discrimination. With pigeons, at least, information on temporal integration, masking and critical bands would be very useful in this sort of study. The development of an effective method of responding with more than two categories would be another valuable preliminary to further applications of GOC analysis in animal psychophysics.

Summary of Findings and Conclusions

The results of GOC analysis gave a better indication of the nature of ideal ROCs for known underlying distributions than did results for individual replications.

GOC analysis was more effective (especially with complex ROCs) when simulated observers used rating scales with larger numbers of categories.

With tri-modal distributions, more favourable simulation results were obtained when more replications, rather than more trials per replication, were analyzed. This assumes that the analysis is based on sufficient independent trials.

When the rating scale used by the observer has a relatively small number of categories, the results of GOC analysis lead to underestimation of $\underline{k}$, the ratio of unique to common noise variance.

Large changes in d_z during GOC analysis indicate high values of $\underline{k}$. However, the errors in the estimation of $\underline{k}$ are greatest for large changes in d_z.

Some of the difficulties of estimating $\underline{k}$ were overcome in the present research by comparing experimental results with those obtained from simulations with known values of $\underline{k}$. Another solution would be to use analytical models developed specifically for the two-category rating scale.

The YN procedure used with the pigeons is a noisy one, and any common noise present in the third and fourth experiments was overwhelmed by unique noise.

GOC analysis is a demanding technique to implement. A large number of independent trials and replications are required, together with signals which do not allow observers to sample within them, and observers should use as many rating-scale categories as possible.

However, the technique allows an approach to be made to the problems posed by extraneous noise in psychophysical experiments which avoids pitfalls such as that pointed out by Green (1960).

The most profitable line for further research on GOC analysis would be to use human observers in amplitude as well as frequency discrimination tasks.

REFERENCES

Blough, D. S. Stimulus generalization as signal detection in pigeons. Science, 1967, 158, 940-941.

Blough, D., & Blough, P. Animal Psychophysics. In W. K. Honig & J. E. R. Stadden (Eds.), Handbook of operant behavior. New York: Appleton-Century-Crofts, 1977.

Boven, R. The use of multiple observers in signal detection theory: A method to remove the effects of unique noise from experimental data Unpublished Masters thesis, Victoria University of Wellington, 1976.

Bush, R. R., Galanter, E., & Luce, R. D. Characterization and classification of choice experiments. In R. D. Luce, R. R. Bush & E. Galanter (Eds.) Handbook of mathematical psychology (Vol. 1). New York: Wiley, 1963.

Coombs, C. H., Dawes, R. M., & Tversky, A. Mathematical psychology: An elementary introduction. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Creelman, C. D., & Macmillan, N. A. Auditory phase and frequency discrimination: A comparison of nine procedures. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 146-156.

Dalton, L. W. Conditioned suppression as a technique for determination of auditory sensitivity in pigeons. Journal of Auditory Research, 1967, 7, 25-29.

de Boer, E. Intensity discrimination of fluctuating signals. Journal of the Acoustical Society of America, 1966, 40, 552-560.

Delius, J. D., & Emmerton, J. Stimulus dependent asymmetry in classical and instrumental discrimination by pigeons. Psychological Record, 1978, 28, 425-434.

Delius, J. D., & Tarpy, R. M. Stimulus control of heart rate by auditory frequency and auditory pattern in pigeons. Journal of the Experimental Analysis of Behavior, 1974, 21, 297-306.

Diercks, K. J., & Jeffress, L. A. Interaural phase and absolute threshold for tone. Journal of the Acoustical Society of America, 1962, 34, 981-984.

Dooling, R. J. Temporal summation of pure tones in birds. Journal of the Acoustical Society of America, 1979, 65, 1058-1060.

Dooling, R. J., & Saunders, J. C. Hearing in the parakeet (Melopsittacus undulatus): Absolute thresholds, critical ratios, frequency difference limens, and vocalizations. Journal of Comparative and Physiological Psychology, 1975, 88, 1-20.

Egan, J. P. Signal detection and ROC analysis. New York: Academic Press, 1975.

Egan, J. P., Greenberg, G. Z., & Schulman, A. I. Interval of time uncertainty in auditory detection. Journal of the Acoustical Society of America, 1961, 33, 771-778.

Egan, J. P., Schulman, A. I., & Greenberg, G. Z. Operating characteristics determined by binary decisions and by ratings. Journal of Acoustical Society of America, 1959, 31, 768-773.

Eijkman, E., Thijssen, J. M., & Vendrik, A. J. H. Weber's law, power law and internal noise. Journal of the Acoustical Society of America, 1966, 40, 1164.

Emmerich, D. S., Gray, J. L., Watson, C. S., & Tanis, D. C. Response latency, confidence and ROCs in auditory signal detection. Perception and Psychophysics, 1972, 11, 65-72.

Evans, G. W., Wallace, G. F., & Sutherland, G. Simulation using digital computers. Englewood Cliffs, N. J.: Prentice-Hall, 1967.

Foree, D. D., & LoLordo, V. M. Attention in the pigeon: The differential effect of food getting vs. shock avoidance procedures. Journal of Comparative and Physiological Psychology, 1973, 85, 551-558.

Gilkey, R. H., Hanna, T. E., & Robinson, D. E. Estimates of the ratio of external to internal noise using reproducible samples of noise. Journal of the Acoustical Society of America, 1981, 69, Supplement 1, S23 (Abstract).

Green, D. M. Psychoacoustics and detection theory. Journal of the Acoustical Society of America, 1960, 32, 1189-1203.

Green, D. M. Consistency of auditory detection judgements. Psychological Review, 1964, 71, 392-407.

Green, D. M., & Luce, R. D. Counting and timing mechanisms in auditory discrimination and reaction time. In D. H. Krantz, R. O. Atkinson, R. D. Luce & P. Suppes (Eds.), Contemporary developments in mathematical psychology. San Francisco: Freeman, 1974.

Green, D. M., & Swets, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966.

Green, M., Terman, M., & Terman, J. S. Comparison of yes-no and latency measures of auditory intensity discrimination. Journal of the Experimental Analysis of Behavior, 1979, 32, 363-372.

Harrison, J. B., & Furomoto, L. Pigeon audiogram: Comparison of evoked potential and behavioral thresholds in individual birds. Journal of Auditory Research, 1971, 11, 33-42.

Heinemann, E. G., & Avin, E. On the development of stimulus control. Journal of the Experimental Analysis of Behavior, 1973, 20, 183-195.

Heinemann, E. G., Avin, E., Sullivan, M. A., & Chase, S. Analysis of stimulus generalization with a psychophysical method. Journal of Experimental Psychology, 1969, 80, 215-224.

Heise, G. A. Auditory thresholds in the pigeon. American Journal of Psychology, 1953, 66, 1-19.

Henning, G. B. Frequency discrimination in noise. Journal of the Acoustical Society of America, 1967a, 41, 774-777.

Henning, G. B. A model for auditory discrimination and detection. Journal of the Acoustical Society of America, 1967b, 42, 1325-1334.

Hienz, R. D., Sinnott, J. M., & Sachs, M. B. Auditory sensitivity in the redwing blackbird (Agelaius phoenicius) and the brown-headed cowbird (Molothrus ater). Journal of Comparative and Physiological Psychology, 1977, 91, 1365-1376.

Hirsch, I. J. Binaural summation: A century of investigation. Psychological Bulletin, 1948, 45, 193-206.

Isaac, P. D. Linear regression, structural relations, and measurement error. Psychological Bulletin, 1970, 74, 213-218.

Jeffress, L. A. Stimulus-oriented approach to detection. Journal of the Acoustical Society of America, 1964, 36, 766-774.

Jeffress, L. A. Stimulus-oriented approach to detection reexamined. Journal of the Acoustical Society of America, 1967, 41, 480-488.

Jeffress, L. A. Mathematical and electrical models of auditory detection. Journal of the Acoustical Society of America, 1968, 44, 187-203.

Jesteadt, W., & Sims, S. Decision processes in frequency discrimination. Journal of the Acoustical Society of America, 1975, 57, 1101-1168.

Knuth, D. E. The art of computer programming. Volume 2: Seminumerical algorithms. Reading, Massachusetts: Addison-Wesley, 1969.

Krasnegor, N. A. The effects of telencephalic lesions on auditory discriminations in pigeons (Doctoral dissertation, University of Maryland, 1970). Dissertation Abstracts International, 1971, 31, 5029B. (University Microfilms No. 71-04, 085).

Krasnegor, N. A., & Hodos, W. The evaluation and control of acoustical standing waves. Journal of the Experimental Analysis of Behavior, 1974, 21, 297-306.

Kriethen, M. L., & Quine, D. B. Infrasound detection by the homing pigeon: A behavioral audiogram. Journal of Comparative Physiology A, 1979, 129, 1-4.

McAuley, K. A temporal model of frequency discrimination. Unpublished doctoral dissertation, Victoria University of Wellington, 1978.

McFadden, D. Masking-level differences determined with and without interaural disparities in masker intensity. Journal of the Acoustical Society of America, 1968, 44, 212-223.

McNicol, D. A primer of signal detection theory. London: Allen & Unwin, 1972.

Madansky, A. The fitting of straight lines when both variables are subject to error. *Journal of the American Statistical Association*, 1959, 54, 173-205.

Moore, B. C. J. Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America*, 1973a, 54, 610-619.

Moore, B. C. J. Frequency difference limens for narrow bands of noise. *Journal of the Acoustical Society of America*, 1973b, 54, 888-896.

Osman, E. A correlation model of binaural masking level differences. *The Journal of the Acoustical Society of America*, 1971, 50, 1494-1511.

Pfafflin, S. M., & Mathews, M. V. Energy detection model for monaural auditory detection. *Journal of the Acoustical Society of America*, 1962, 34, 1842-1853.

Podd, J. V. *Type I and II ROC analysis*. Unpublished Masters thesis, Victoria University of Wellington, 1975.

Price, L. L., Dalton, L. W., & Smith, J. C. Frequency DL in the pigeon as determined by conditioned suppression. *Journal of Auditory Research*, 1967, 7, 229-239.

Quine, D. B., & Kriethen, M. L. Frequency shift discrimination: Can homing pigeons locate infrasounds by Doppler shifts. *Journal of Comparative Physiology A*, 1981, 141, 153-155.

Raab, D. H., & Goldberg, I. A. Auditory intensity discrimination with bursts of reproducible noise. *Journal of the Acoustical Society of America*, 1975, 57, 437-447.

Sabine, H. J. Acoustical materials. In C. M. Harris (Ed.), *Handbook of noise control*. New York: McGraw-Hill, 1957.

Saunders, J. C., Denny, R. M., & Bock, G. R. Critical bands in the parakeet (*Melopsittacus undulatus*). *Journal of Comparative Physiology A*, 1978, 125, 359-365.

Shaw, E. A., & Piercy, J. E. Physiological noise in relation to audiometry. *Journal of the Acoustical Society of America*, 1962, 34, 745 (Abstract).

Siebert, W. M. Frequency discrimination in the auditory system: Place or periodicity mechanisms. *Proceedings of the IEEE*, 1970, 58, 723-730.

Siegel, R. A. *Internal and external noise in auditory detection*. Unpublished Masters thesis, Massachusetts Institute of Technology, 1979. (Cited in Spiegel & Green, 1981.)

Siegel, R. A. Perspectives on additive internal noise. *Journal of the Acoustical Society of America*, 1981, 69, Supplement 1, S23 (Abstract).

Sinnott, J. M., Sachs, M. B., & Hienz, R. D. Aspects of frequency discrimination in passerine birds and pigeons. *Journal of Comparative and Physiological Psychology*, 1980, 94, 401-415.

Soderquist, D. R., & Lindsey, J. W. Physiological noise as a masker of lower frequencies: The cardiac cycle. Journal of the Acoustical Society of America, 1972, 52, 1216-1219.

Spiegel, M. F. & Green, D. M. Two procedures for estimating internal noise. Journal of the Acoustical Society of America, 1981, 70, 69-73.

Stebbins, W. C. (Ed.) Animal psychophysics. New York: Appleton-Century-Crofts, 1970a.

Stebbins, W. C. Studies of hearing and hearing loss in the monkey. In W. C. Stebbins (Ed.), Animal psychophysics. New York: Appleton-Century-Crofts, 1970b.

Swets, J. A. Is there a sensory threshold? Science, 1961, 134, 168-177.

Swets, J. A., Shipley, E. F., McKey, M. J., & Green, D. M. Multiple observations of signals in noise. Journal of the Acoustical Society of America, 1959, 31, 514-521.

Swets, J. A., & Pickett, R. M. Evaluation of diagnostic systems: Methods from signal detection theory. New York: Academic Press, 1982.

Thijssen, J. M., & Vendrik, A. J. H. Internal noise and transducer function in sensory detection experiments: Evaluation of psychometric curves and of ROC curves. Perception and Psychophysics, 1968, 3, 387-400.

Trainer, J. E. The auditory acuity of certain birds. Unpublished doctoral dissertation, Cornell University, 1946. (Data taken from Heise, 1953.)

Watson, C. S. Signal detection and certain physical characteristics of the stimulus during the observation interval. Unpublished doctoral dissertation, Indiana University, 1963.

Wilbanks, W. A., & Whitmore, J. Detection of monaural signals as a function of interaural noise correlation and signal frequency. Journal of the Acoustical Society of America, 1968, 43, 785-797.

Yager, D., & Duncan, T. Signal-detection analysis of luminance generalization in goldfish using latency as a graded response measure. Peception and Psychophysics, 1971, 9, 353-355.

Appendix A

TEST OF THE ATTENUATION OF AIRBORNE SOUND BY THE SOUND-ATTENUATED ROOM USED IN THE EXPERIMENTS

Tests were carried out by measuring low-pass filtered noise, generated outside the room, with a microphone located first outside the room and then inside it.

Apparatus

The apparatus is shown in block form in Figure 58. Gaussian noise, low-pass filtered at 2000 Hz (Butterworth, 24 dB/octave), was produced independently by two generators and fed, via electronic switches and further 2000 Hz low-pass filters, to the auxiliary inputs of a Kenwood audio amplifier. It was then output through a pair of three-speaker ensembles.

The output of the speakers was monitored by a B & K 25.4 mm condenser microphone, which fed, via a frequency analyser (B & K 2121), an attenuator, a 2000 Hz low-pass filter and an inverting amplifier, into a Hewlett-Packard 69422A 12-bit ADC card, housed in an HP 6940B multiprogrammer.

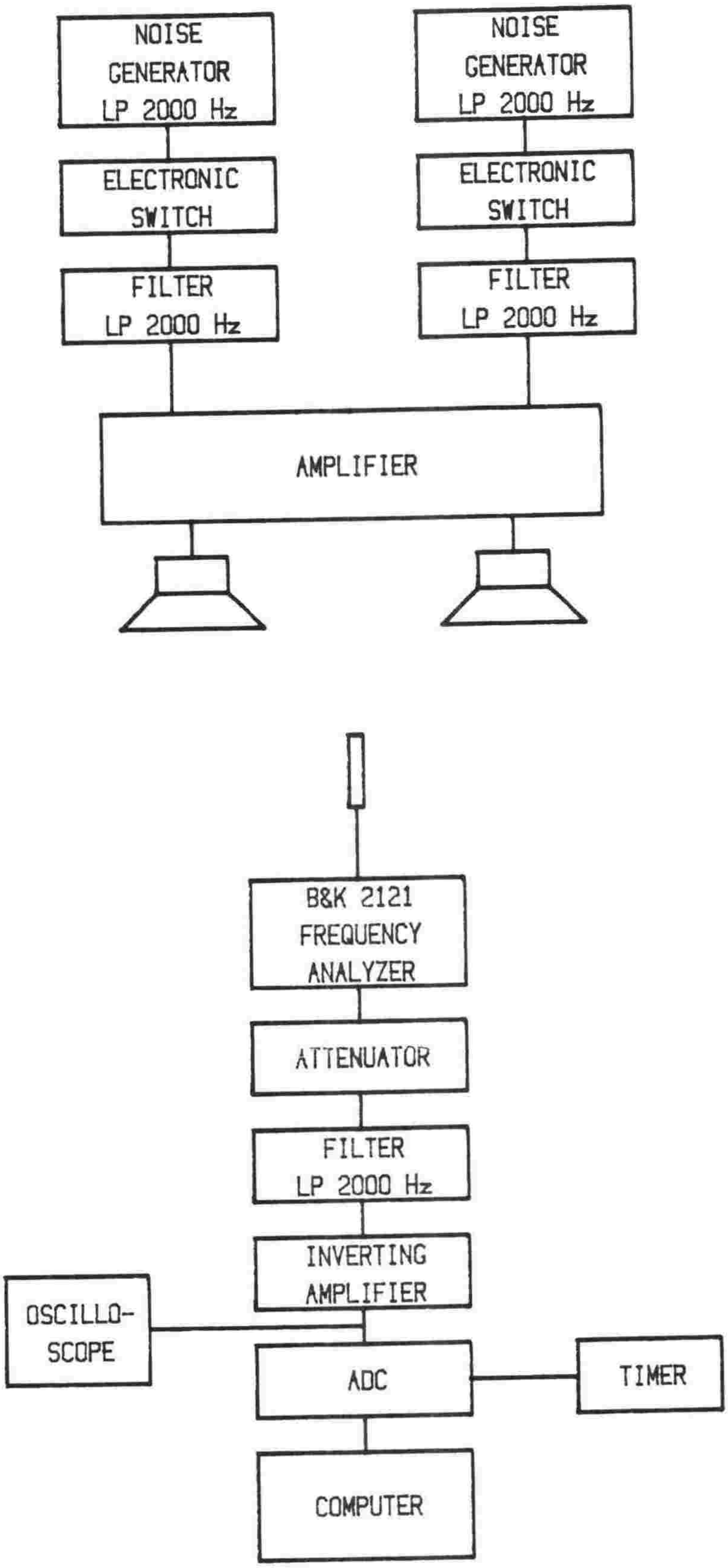


Figure 58. Block diagram of apparatus used to measure the attenuation of airborne sound by the sound-attenuating room.

Procedure

The microphone was first set up outside the room, about 30 cm from one of the walls, and the overall level of the noise was set at around 95 dB SPL.

Samples of the noise, which was switched on briefly, were taken by the ADC and fed to an HP 9825 computer, which calculated fast Fourier transforms (FFTs). Within each sample, points were obtained every 0.00021 seconds (a sampling rate of 4762 Hz, over twice the highest frequency in the filtered noise) and 1024 measurements were taken, giving a sample duration of 0.215 seconds. FFTs were calculated for a total of 32 samples and the results averaged. The above procedure was repeated with the microphone placed within the room.

Attenuation over a range of frequencies, down to the resolution achieved by the procedure (4.65 Hz), was then calculated by subtracting the results of the second group of FFTs from the results of the first. Allowance was made for the fact that the scale setting of the frequency analyser was 30 dB lower when the microphone was in the room than when it was located outside, near the speakers.

Results

Figure 59 shows a smooth attenuation function which was obtained by using the equation expressing the linear relation between attenuation in dB and log frequency from 28 Hz to 2000 Hz to generate estimates of attenuation over the linear frequency scale.

The attenuation at 28 Hz was about 15 dB, but it increased sharply to around 45 dB at 400 Hz and then more gradually to 65 dB at 2000 Hz.

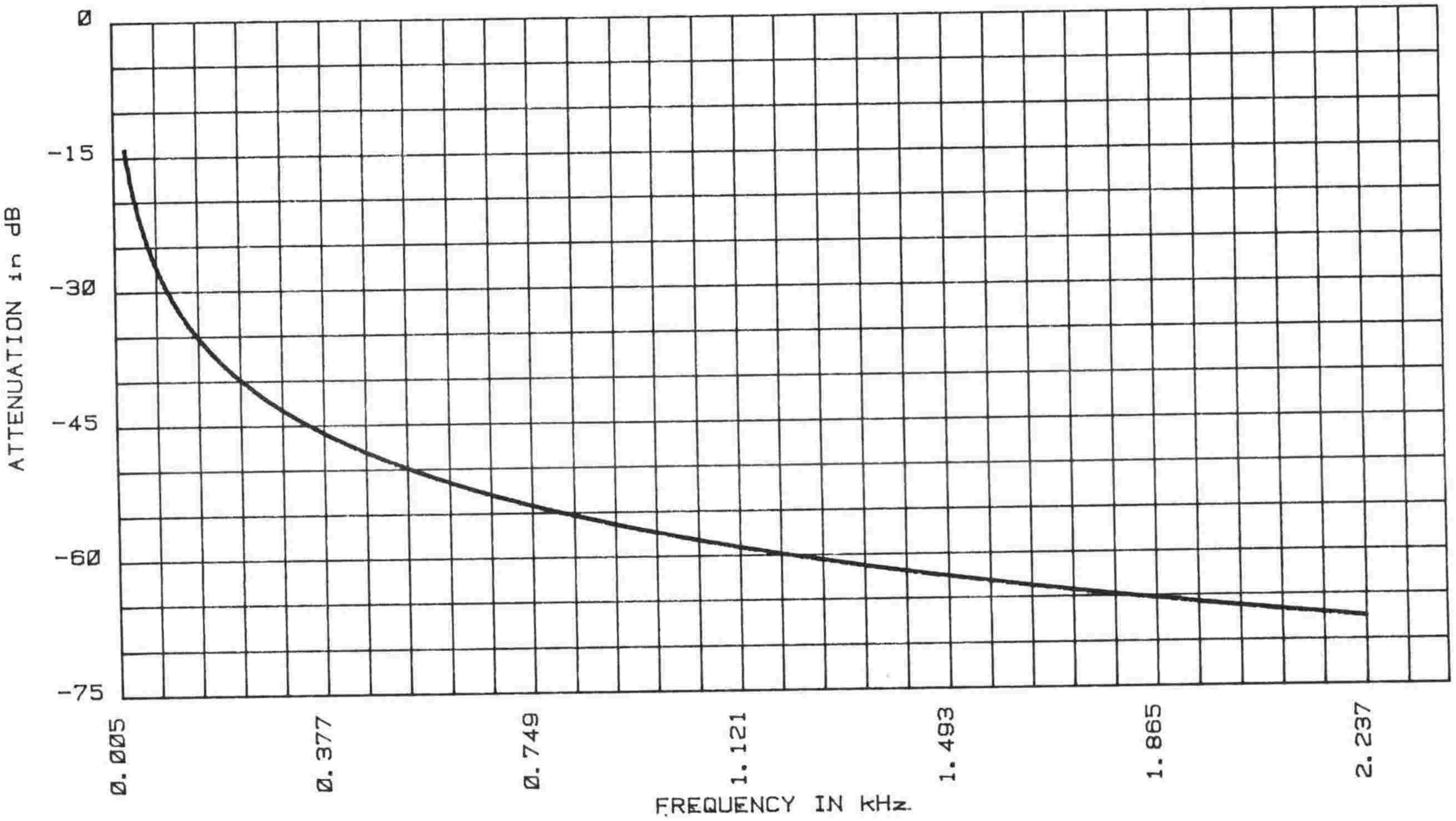


Figure 59. Attenuation of airborne sound by the sound-attenuating room.

Appendix B

THE OPERANT CHAMBER

The operant chamber was originally built for use in a free field. It consisted of a steel framework 540 mm long, 360 mm wide and 360 mm high, which housed a compartment for the birds which was 360 mm wide and long and 300 mm high. This compartment was surrounded by 15 mm wire netting and had a floor of 10 mm wire mesh made of 2 mm gauge wire.

At one end of the compartment three transparent response keys were fitted behind a slot 80 mm long and 115 mm high (Figure 7). The middle key, which was 25 mm wide, was hinged at the top, while the side keys were hinged at either side of the slot. A force of 0.19 N was required to operate the centre key. In the first two pigeon experiments a force of 0.15 N operated the side keys. Following recalibration, the minimum force on these keys rose to 0.20 N. Incandescent bulbs, which lit the ends of circular bezels 8 mm in diameter, were located behind the response keys.

The key slot was centred 225 mm above the floor of the compartment, and was surrounded by 10 mm thick polyurethane foam. This foam was added to break up sound reflections from the plate surrounding the keys. Below the keys a hole 50 mm square gave access to the food hopper. This hole was centred 85 mm above the floor of the compartment, and was surrounded by a 130 mm by 125 mm piece of 10 mm thick foam. The rest of the wall housing the keys was made of wire netting.

The food hopper was raised and lowered by a cam attached to a 12 volt windscreen-wiper motor.

A houselight was mounted in the ceiling of the chamber 80 mm back from the wall housing the keys.

When the sound-absorbing properties of the sound-attenuating room proved inadequate, a surround was developed for the previously open operant chamber. The shell consisted of (from the inside out) 100 mm thick polyurethane foam, acoustic tile and 1 mm aluminium sheet attached to a steel frame.

At the end of the chamber opposite the keys, the tile and aluminium were omitted, and a 180 mm diameter hole, centred 150 mm above the floor of the chamber, was cut in the foam (Figure 8). A speaker enclosure was mounted against the foam so that the 200 mm diameter speaker was located against the hole. A circular plug of 20 mm thick particle board was placed in the hole in the foam between the speaker and the chamber.

The chamber was ventilated by a Muffin Model WR2HI whisper fan located inside a 100 mm hole in a side wall of the chamber.

The operant chamber was mounted on a tubular steel frame 650 mm above the floor of the sound-attenuating room.

Appendix C

SUPPLEMENTARY TABLES AND FIGURES

TABLE 5

Hit and false alarm rates in the first pigeon experiment (triangular distributions)

	Bird					
	7		8		18	
	HR	FAR	HR	FAR	HR	FAR
Replication						
1	0.675	0.160	0.805	0.210	0.845	0.290
2	0.620	0.135	0.760	0.195	0.860	0.210
3	0.640	0.115	0.820	0.210	0.845	0.275
4	0.715	0.180	0.815	0.230	0.815	0.205
5	0.695	0.235	0.815	0.190	0.835	0.200
6	0.755	0.205	0.825	0.170	0.835	0.190

TABLE 6

Hit and false alarm rates in the first pigeon experiment (bimodal distributions)

	Bird					
	7		8		18	
Replication	HR	FAR	HR	FAR	HR	FAR
1	0.485	0.350	0.525	0.490	0.605	0.530
2	0.415	0.340	0.520	0.435	0.575	0.495
3	0.445	0.310	0.555	0.475	0.590	0.530
4	0.485	0.410	0.575	0.470	0.560	0.460
5	0.520	0.410	0.560	0.445	0.560	0.475
6	0.540	0.420	0.520	0.475	0.555	0.470

TABLE 7

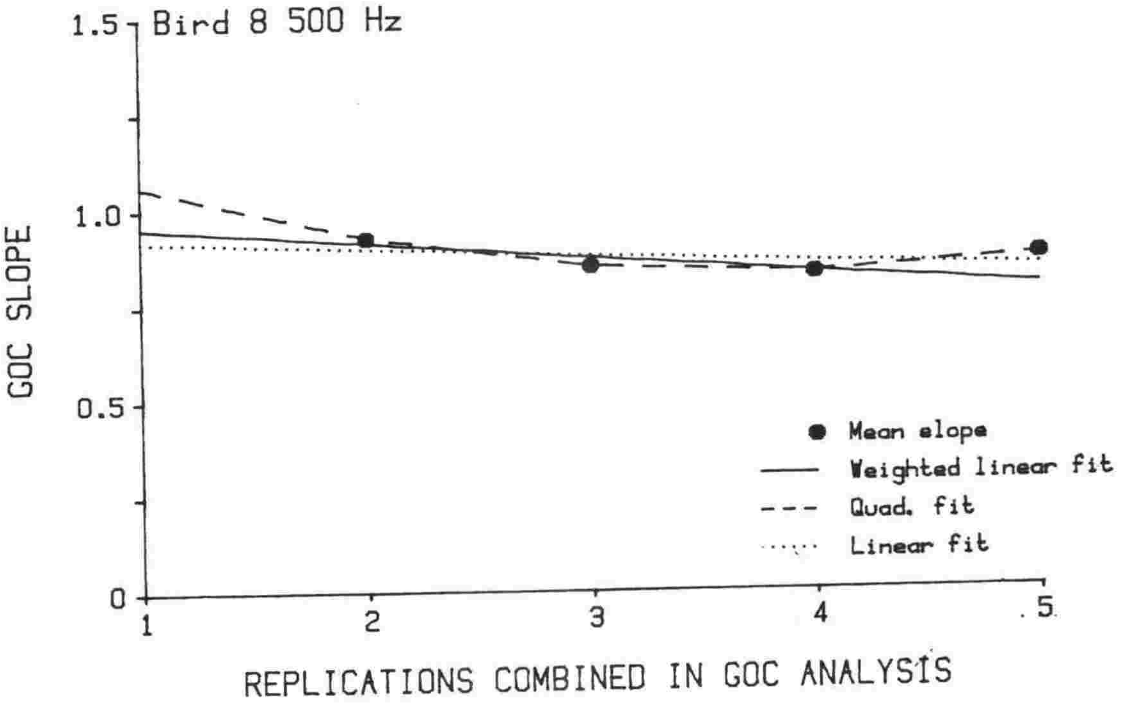
Hit and false alarm rates in the second pigeon experiment (normal unequal variance distributions)

	Bird			
	8		18	
Replication	HR	FAR	HR	FAR
1	0.752	0.574	0.901	0.525
2	0.718	0.436	0.817	0.505
3	0.619	0.436	0.941	0.480
4	0.738	0.450	0.886	0.446
5	0.782	0.485	0.921	0.446
6	0.871	0.639	0.916	0.470
7	0.842	0.569	0.901	0.455
8	0.797	0.480	0.941	0.475
9	0.733	0.485	0.946	0.455
Pooled	0.761	0.506	0.908	0.473

TABLE 8

Hit and false alarm rates in the third pigeon experiment (five comparison frequencies)

	FAR	HR				
Replication	450 Hz	500 Hz	600 Hz	700 Hz	800 Hz	900 Hz
Bird 8						
1	0.255	0.204	0.350	0.556	0.633	0.733
2	0.356	0.278	0.383	0.597	0.633	0.783
3	0.439	0.333	0.433	0.681	0.733	0.833
4	0.308	0.333	0.400	0.569	0.744	0.858
5	0.366	0.352	0.583	0.653	0.844	0.867
Pooled	0.345	0.300	0.430	0.611	0.718	0.815
Bird 18						
1	0.275	0.241	0.483	0.639	0.789	0.892
2	0.220	0.370	0.417	0.694	0.789	0.900
3	0.210	0.333	0.317	0.611	0.767	0.900
4	0.157	0.185	0.367	0.681	0.789	0.850
5	0.154	0.222	0.417	0.569	0.644	0.875
Pooled	0.203	0.270	0.400	0.639	0.756	0.883



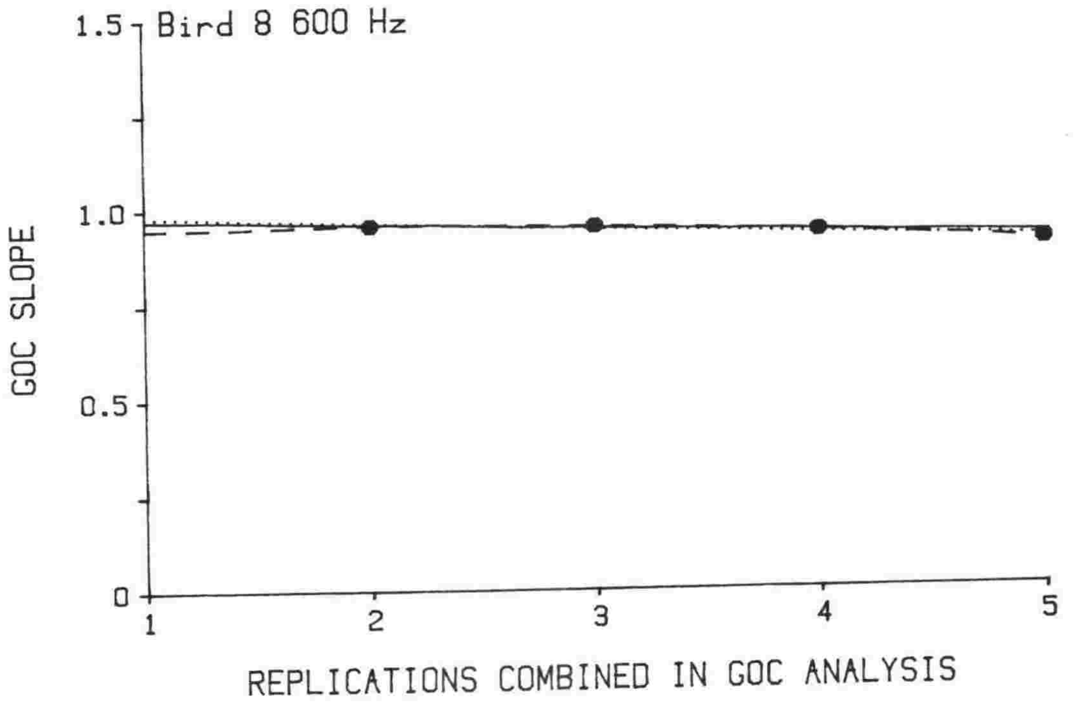
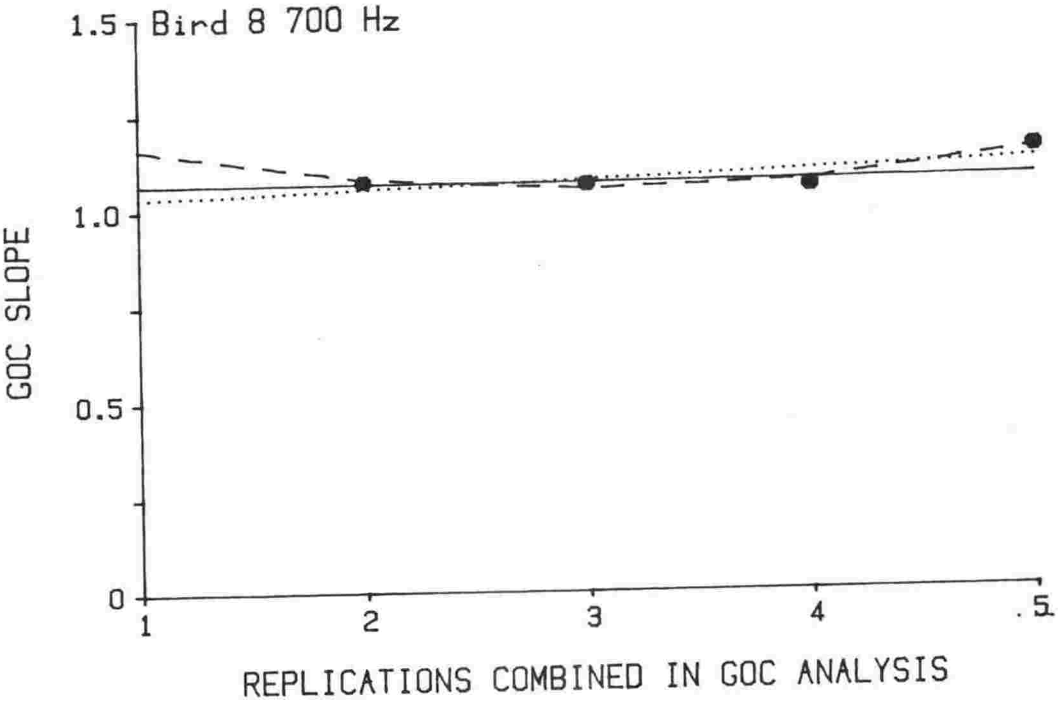


Figure 60. The mean slopes of GOCs based on different numbers of replications, obtained in the third pigeon experiment. The points are fitted with linear, weighted linear, and quadratic functions.



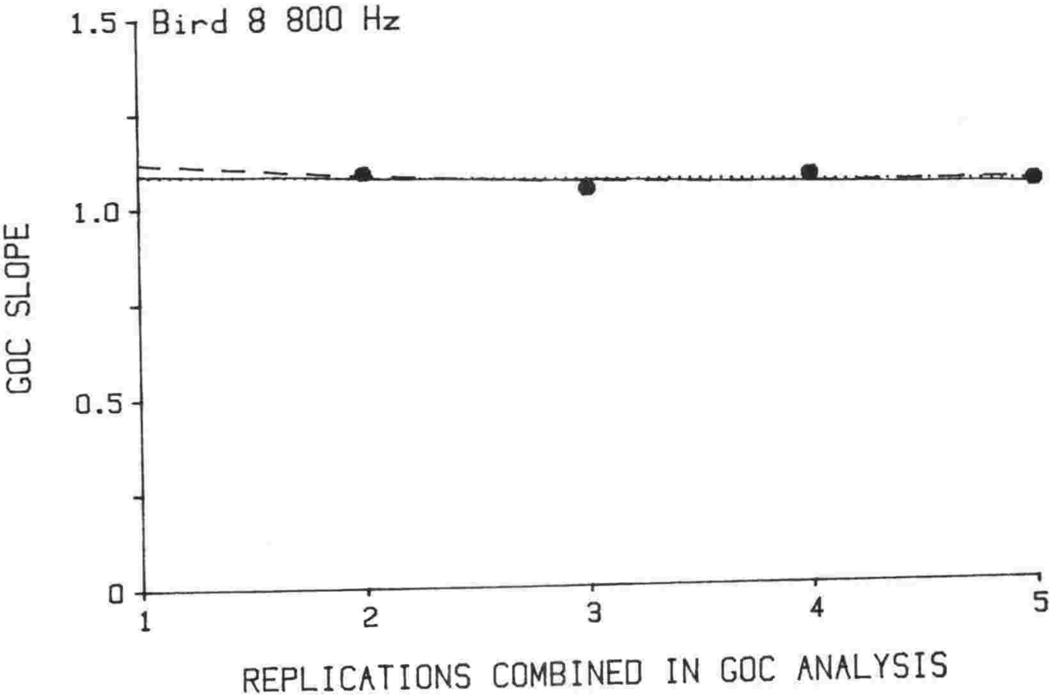


Figure 60 (continued).

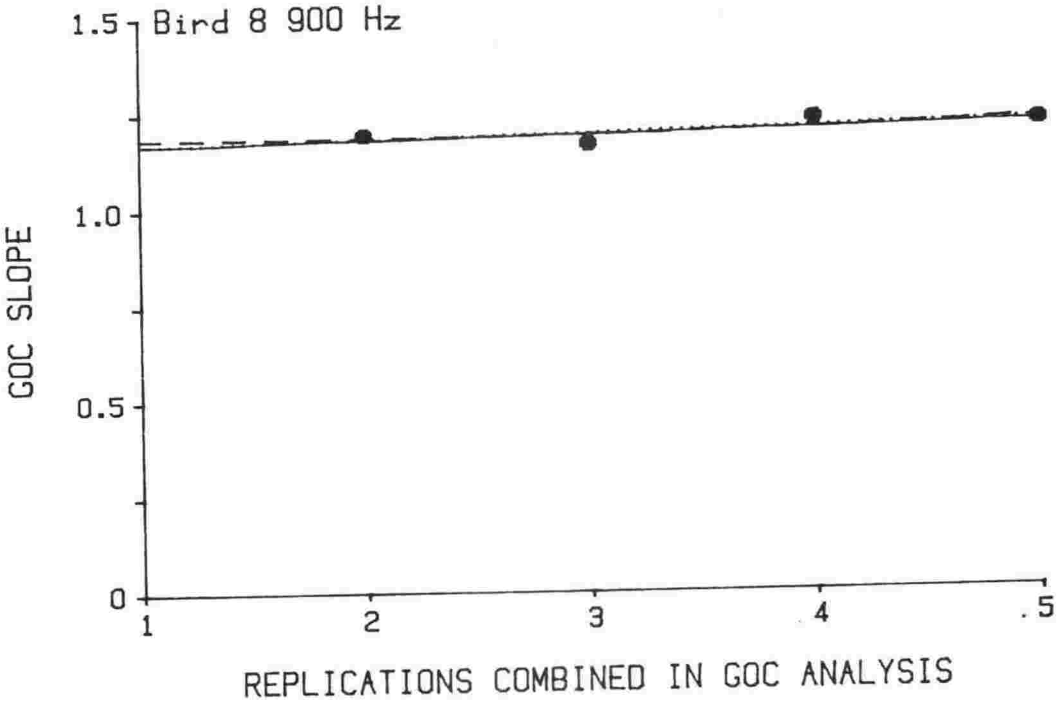
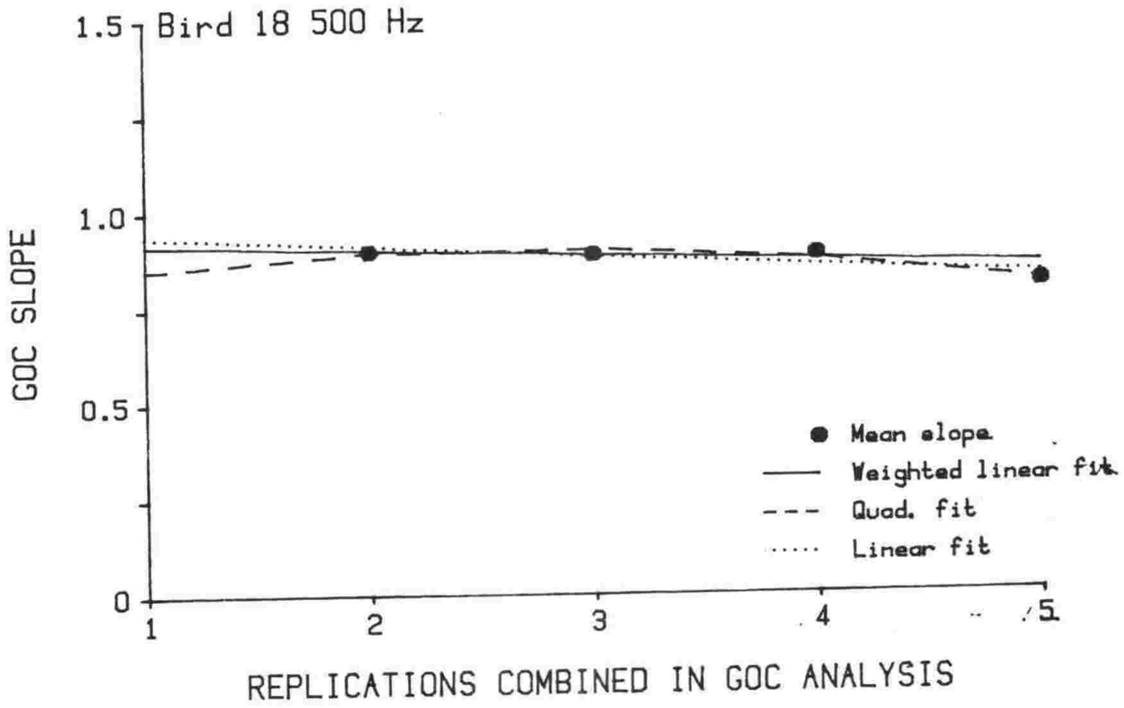


Figure 60 (continued).



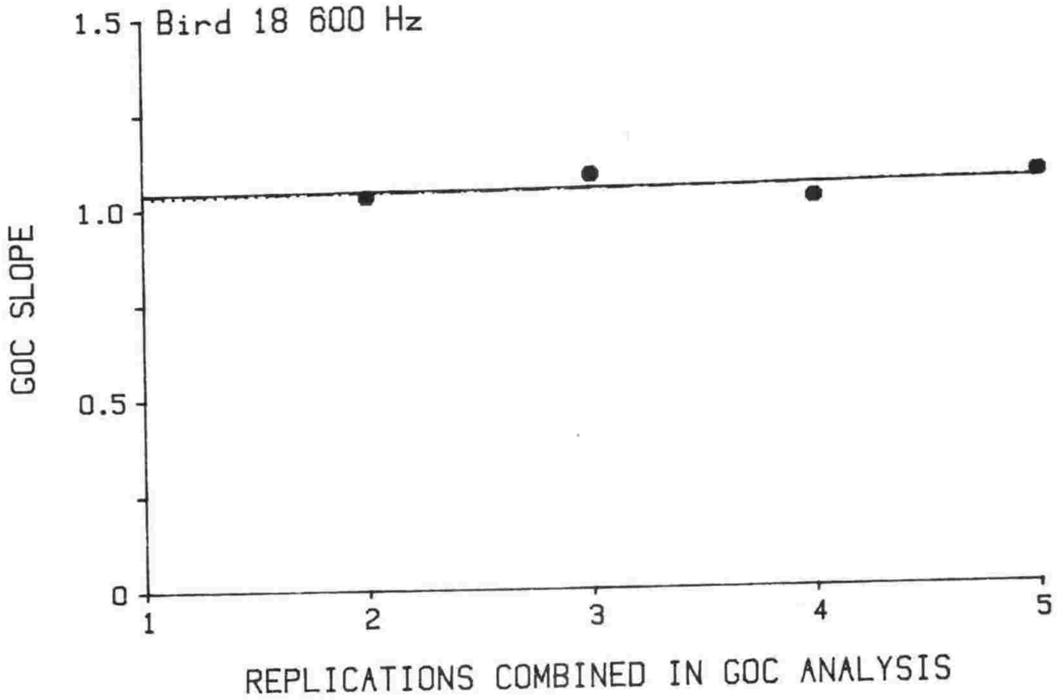
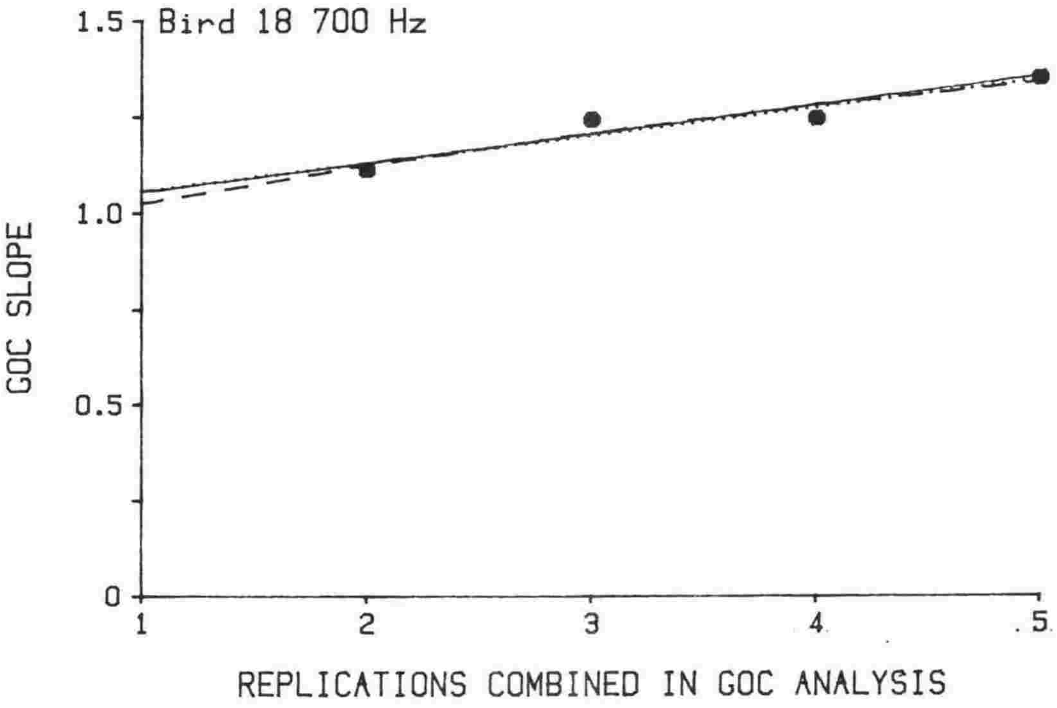


Figure 60 (continued).



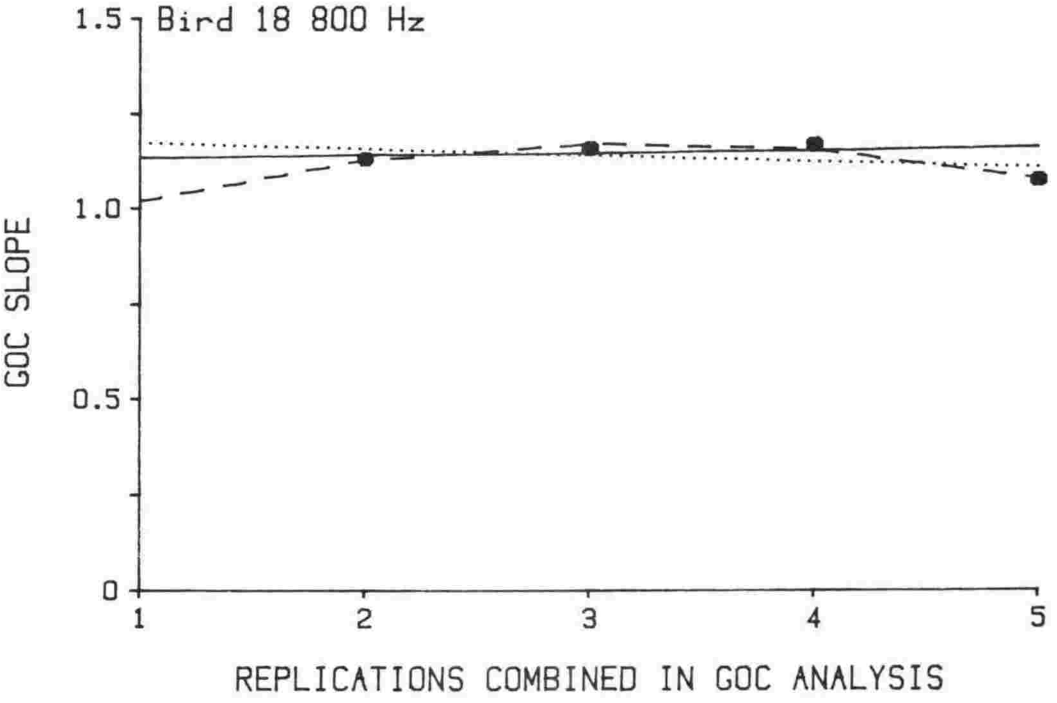


Figure 60 (continued).

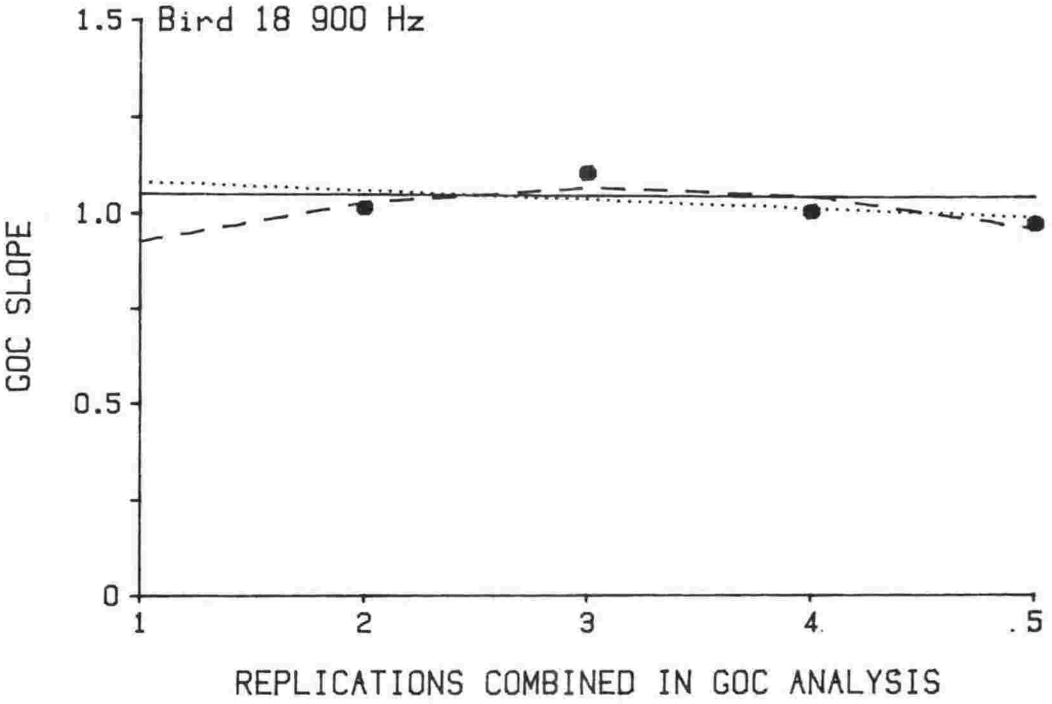


Figure 60 (continued).

TABLE 9

Hit and false alarm rates in the fourth pigeon experiment (two comparison frequencies)

	Bird					
	8			18		
	FAR	HR		FAR	HR	
	450 Hz	650 Hz	900 Hz	450 Hz	650 Hz	900 Hz
Replication						
1	0.396	0.574	0.867	0.242	0.500	0.850
2	0.482	0.646	0.917	0.227	0.497	0.850
3	0.447	0.574	0.783	0.225	0.533	0.933
4	0.427	0.598	0.733	0.278	0.542	0.883
5	0.467	0.690	0.933	0.199	0.536	0.917
6	0.510	0.631	0.800	0.194	0.464	0.750
7	0.561	0.720	0.800	0.227	0.536	0.950
8	0.485	0.708	0.833	0.237	0.518	0.967
9				0.235	0.589	0.950
10				0.202	0.515	0.967
11				0.255	0.568	0.983
12				0.255	0.643	0.967
13				0.275	0.607	0.983
Pooled	0.472	0.643	0.833	0.235	0.542	0.919

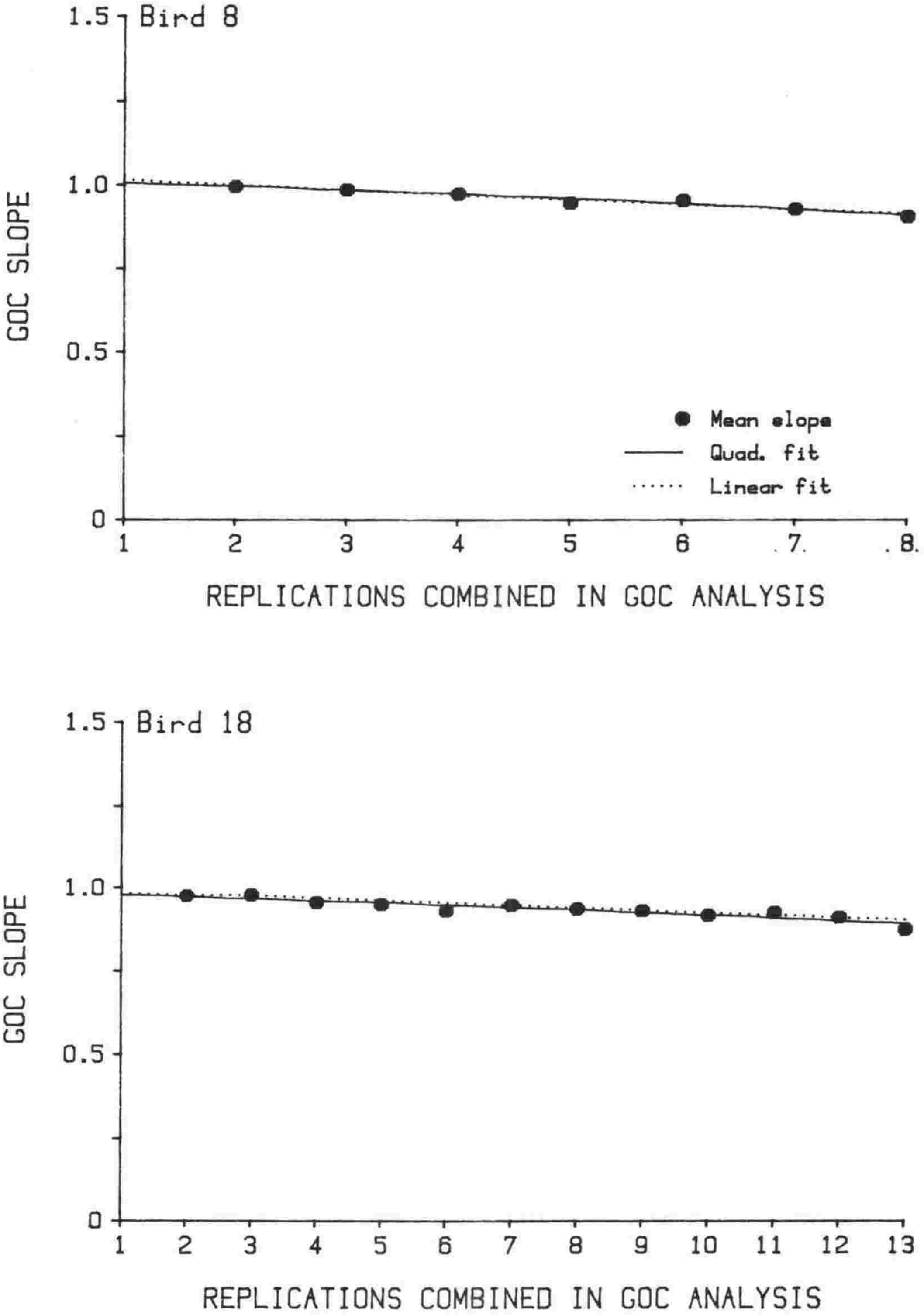


Figure 61. The mean slopes of GOCs based on different numbers of replications, obtained in the fourth pigeon experiment. The points are fitted with linear and quadratic functions.

TABLE 10

Parameters and hit and false alarm rates in the hardware simulation (criterion variability)

	Criterion Variability							
	SO 8				SO 18			
	Cutoff (bins)	s.d. (bins)	HR	FAR	Cutoff (bins)	s.d. (bins)	HR	FAR
Replication								
1	110	80	0.646	0.485	79	50	0.503	0.235
2	111	80	0.637	0.525	79	45	0.568	0.217
3	112	80	0.670	0.535	85	50	0.521	0.301
4	109	80	0.652	0.429	78	55	0.497	0.253
5	108	80	0.601	0.510	79	50	0.500	0.227
6	107	80	0.655	0.455	85	50	0.518	0.253
7	[105	80	0.563	0.492]a	85	40	0.592	0.220
8	113	80	0.637	0.518	[85	60	0.512	0.265]a
9					80	40	0.548	0.194
10					82	50	0.542	0.255
11					82	45	0.539	0.225
12					85	45	0.530	0.250
13					82	45	0.524	0.260
Pooled			[0.632	0.494]b			[0.530	0.243]b
			0.642	0.494			0.532	0.241

a Omitted in final analysis

b Pooled rates including omitted replications

TABLE 11

Parameters and hit and false alarm rates in the hardware simulation (inattention)

	Ignoring the Stimulus					
	SO 8			SO 18		
	P(ignore)=0.5 $P(\text{No}\mid\text{ignore})$=0.5			P(ignore)=0.333 $P(\text{No}\mid\text{ignore})$=0.5		
Replication	HR	FAR	cutoff (bins)	HR	FAR	cutoff (bins)
1	0.735	0.230	110	0.414	0.205	79
2	0.768	0.253	111	0.387	0.159	79
3	0.735	0.283	112	0.815	0.162	85
4	0.747	0.237	109	0.158	0.174	78
5	0.735	0.240	108	0.446	0.154	79
6	0.750	0.250	107	0.839	0.164	85
7	[0.735	0.237	105]a	0.810	0.199	85
8	0.732	0.381	113	[0.839	0.164	85]a
9				0.795	0.194	80
10				0.836	0.126	82
11				0.848	0.157	82
12				0.839	0.172	85
13				0.842	0.194	82
Pooled	[0.742	0.264]b		[0.682	0.171]b	
	0.743	0.268		0.669	0.172	

a Omitted in final analysis

b Pooled rates including omitted replications

Appendix D

METHODS USED TO PRODUCE RANDOM NORMAL DEVIATES IN THE SIMULATION

Normal Deviates

The Polar method is described by Knuth (1969, p. 104-105). The algorithm calculates two independent normally distributed variables X_1 and X_2, given two independent uniformly distributed variables, U_1 and U_2. Knuth gives the following four-step procedure.

a) Generate two independent random variables U_1 and U_2, uniformly distributed between zero and one. Create V_1 and V_2 as follows:

$$V_1 = 2U_1 - 1, \text{ and}$$
$$V_2 = 2U_2 - 1.$$

b) Compute $S = V_1^2 + V_2^2$.

c) If S is greater than or equal to one, return to the first step.

d) If S is less than one, create:

$$X_1 = V_1[(-2\underline{\ln}S)/S]^{1/2}, \text{ and}$$
$$X_2 = V_2[(-2\underline{\ln}S)/S]^{1/2}.$$

X_1 and X_2 are independent normally distributed variables. Knuth supplies a proof of the validity of the method (1969, p. 104-105).

Uniform Variable

Evans, Wallace and Sutherland (1967, Appendix B, p. 187-189) describe a method for producing pseudo-random numbers with a uniform distribution. Their method was implemented as follows. For numbers in the interval $0<r_i<1$,

$$r_{i+1} = [Cr_i],$$

where C is a constant multiplier, and the square brackets indicate that only the fractional part of the product is used. The value of C was set equal to:

$$C = (200A \pm B)$$

where A is any non-negative integer, and

B is one of the numbers in the set 3, 11, 13, 19, 21, 27, 29, 37, 53, 59, 61, 67, 69, 77, 83 and 91.

The starting value, r_0, was set equal to $R \times 10^{-11}$, where R was any integer not divisible by 2 or 5 and such that $0<R<10^{11}$. Candidates for R were chosen from a table of random numbers.

Evans, Wallace and Sutherland state that when the conditions they specify are adhered to, the generator will produce at least $5 \times 10^{p-2}$ numbers before repeating, where p is the number of decimal places used. Eleven digits were used in the simulations.

Appendix E

DERIVATION OF AN EXPRESSION WHICH ESTIMATES k FROM dz VALUES FOR PAIRS OF GOCS BASED ON ANY NUMBER OF REPLICATIONS

To (1) relate $d_{z(i)}$ and $d_{z(n)}$, where $d_{z(i)}$ and $d_{z(n)}$ are indices of sensitivity for GOC curves based on i and n replications respectively, and (2) produce an expression giving $\underline{k}$ for any pair of $d_{z(i)}$ and $d_{z(n)}$.

(1) Equation 3 shows that:

$$\frac{d_n}{d_1} = \frac{(1 + \underline{k})^{1/2}}{(1 + \underline{k}/n)^{1/2}} , \qquad (3)$$

where d_1 is the observer's sensitivity in the absence of GOC analysis (the subscript z has been omitted in the interests of clarity).

This can be solved for d_1, so that

$$d_1 = \frac{d_n(1 + \underline{k}/n)^{1/2}}{(1 + \underline{k})^{1/2}} .$$

This relation holds for i replications:

$$d_1 = \frac{d_i(1 + \underline{k}/i)^{1/2}}{(1 + \underline{k})^{1/2}} .$$

If the expressions involving d_n and d_i are combined and simplified

$$d_i(1 + \underline{k}/i)^{1/2} = d_n(1 + \underline{k}/n)^{1/2} \quad \text{, and} \qquad \{A\}$$

$$d_n = \frac{d_i(1 + \underline{k}/i)^{1/2}}{(1 + \underline{k}/n)^{1/2}} \quad .$$

Thus d_n can be estimated from d_i (and d_i from d_n) if $\underline{k}$ is known.

(2) Taking {A} above, and squaring both sides:

$$(1 + \underline{k}/i)(d_i)^2 = (1 + \underline{k}/n)(d_n)^2.$$

If the terms on both sides are multiplied,

$$(d_i)^2 + \frac{\underline{k}(d_i)^2}{i} = (d_n)^2 + \frac{\underline{k}(d_n)^2}{n} \quad .$$

Then

$$[\underline{k}(d_i)^2]/i - [\underline{k}(d_n)^2]/n = (d_n)^2 - (d_i)^2, \text{ and}$$

$$\underline{k}[(d_i)^2/i - (d_n)^2/n] = (d_n)^2 - (d_i)^2, \text{ so that}$$

$$\underline{k} = (d_n)^2 - (d_i)^2 \cdot \frac{1}{[n(d_i)^2 - i(d_n)^2]/in}$$

$$= \frac{in[(d_n)^2 - (d_i)^2]}{n(d_i)^2 - i(d_n)^2} \quad .$$

This expresses $\underline{k}$ in terms of any pair of d values d_i and d_n $(i<n)$.

Appendix F

DERIVATION OF EXPRESSIONS FOR ESTIMATING dz

(1) To obtain d_z given the slope and intercept of an ROC or GOC curve, plotted on normal-normal coordinates.

From equation 5

$$d_z = \frac{(2)^{1/2}m}{(\sigma_S^2 + \sigma_C^2)^{1/2}} ,$$

where σ_S^2 and σ_C^2 are the variances of the standard and comparison distributions respectively, and

m is the difference between the means of the two distributions.

(If σ_S is equal to σ_C, d_z is equal to d'.)

Since the slope, s, of the ROC curve is equal to σ_S/σ_C, σ_C^2 can be replaced, so

$$d_z = \frac{(2)^{1/2}m}{(\sigma_S^2 + \sigma_S^2/s^2)^{1/2}} .$$

This can be simplified and written as

$$d_z = \frac{(2)^{1/2}}{(1 + 1/s^2)^{1/2}} \cdot \frac{m}{\sigma_S} . \qquad \{A\}$$

The term m/σ_S is an expression for d' or, in the unequal variance case, delta-m. As McNicol (1972) points out, delta-m is equal to z(FAR) at the point where z(HR) is equal to zero. If these values are included in an equation relating hit rates and false alarm rates plotted on normal-normal coordinates, namely

$$z(HR) = s[z(FAR)] + c,$$

where c is the intercept, then

$$\text{delta-m} = -c/s \quad .$$

If this expression for delta-m is used in equation {A}

$$d_z = \frac{-c(2)^{1/2}}{s(1 + 1/s^2)^{1/2}} \text{, or}$$

$$d_z = \frac{-c(2)^{1/2}}{(s^2 + 1)^{1/2}} \quad .$$

(2) To obtain d_z given d_s

In the work reported here, d_z was usually calculated from d_s. This index is based on the hit and false alarm rate at the intersection of the ROC or GOC and the negative diagonal of an ROC space with normal-normal coordinates. At this point HR = -(FAR), so that application of the formula d' = z(FAR) - z(HR) gives $d_s = 2[z(FAR)]$. Egan, Greenberg and Schulman (1961) point out that if the prior probabilities of SN and N are equal, d_s represents the performance of an observer who says "yes" on half of all trials and whose hit rate and correct rejection rate are equal.

McNicol (1972) gives the following formula for d_s:

$$d_s = 2(m/\sigma_S)\ [s/(s + 1)] \quad .$$

This can be rearranged so that

$$m/\sigma_S = (d_s/2)\ [(s + 1)/s] \quad .$$

The right hand side of this expression can therefore be substituted for m/σ_S in equation $\{A\}$, so that

$$d_z = \frac{d_s(s + 1)(2)^{1/2}}{2s(1 + 1/s^2)^{1/2}}$$

$$= \frac{d_s(s + 1)(2)^{1/2}}{2(s^2 + 1)^{1/2}} \quad .$$

When the ROC slope is one, d_z is equal to d_s, but in other cases this is not so.

Appendix G

DERIVATION OF AN EXPRESSION FOR $\underline{k}$ WHEN THE COMMON NOISE DISTRIBUTIONS ARE KNOWN

An adaptation of equation (3) gives the following expression for $d_{z(n)}$:

$$d_{z(n)} = \frac{d_{z(1)}[1 + \underline{k}]^{1/2}}{(1 + k/n)^{1/2}} ,$$

where n is the number of replications on which $d_{z(n)}$ is based.

When n becomes very large, $(1 + \underline{k}/n)^{1/2} \to 1$ and $d_{z(n)} \to d_{z(1)}[1 + \underline{k}]^{1/2}$. If there is no common noise other than that actually provided by the known distributions, $\underline{k}$ can be estimated by letting $d_{z(n)}$ be the value for the known distributions, so that $\underline{k}$ can be estimated from this and $d_{z(1)}$, the performance prior to GOC analysis. To follow this reasoning through:

$$d_{z(n)} = d_{z(1)}[1 + \underline{k}]^{1/2}$$

$$\frac{d_{z(n)}}{d_{z(1)}} = (1 + \underline{k})^{1/2} \quad \text{, and}$$

$$\underline{k} = [d_{z(n)}/d_{z(1)}]^2 - 1 .$$

Appendix H

EXPERIMENTATION PRELIMINARY TO THE THIRD AND FOURTH PIGEON EXPERIMENTS (CHAPTERS 7 AND 8)

This work was carried out after a three month break following the second pigeon experiment. During the break the birds were on ad lib. food. Birds 7, 8, 18 and 16 were used. The first and last birds were not used in either of the final experiments.

Throughout preliminary experimentation all signals were generated by an analogue system. Sinusoids ranging in frequency from 400 Hz to 1000 Hz in 25 Hz steps (at an amplitude of 80 dB), and a continuous masking noise low-pass filtered at 8000 Hz, were produced as described previously. As before, the masking noise was set at an overall level of 55 dB, giving a spectrum level of 16 dB. Another masking noise, with bandwidth equal to 800 Hz (from 200 Hz to 1000 Hz), was generated by two noise sources in conjunction with two bandpass filters and a mixer. It was gated by an electronic switch, along with the sinusoid. This noise was later sampled and recorded digitally for use in the final experiments. More information about its generation and characteristics is given in Chapter 7.

Single Stimulus Presentations

In experiments with single stimuli, the gated masking noise, with a rise-fall time of 20 msec, was kept at a very low level, 44 dB, giving a spectrum level of 15 dB. Taking both continuous and gated noise together, the signal-to-noise ratio was about 58 dB.

The birds were trained to discriminate between the three highest frequencies (1000, 975 and 950 Hz) and the three lowest (400, 425 and

450 Hz). Sessions consisted of 54 trials, preceded by six warm-up trials.

Signals of limited duration were introduced following some preliminary training with signals which were turned off when 30 responses had been made to one key or the other, or when 30 seconds had elapsed since the beginning of the trial (the condition which existed in the first two experiments). If a bird's discrimination performance with a particular duration was above 90% correct during a session, a shorter duration was used in the next session. The durations used were 15, 10, 8, 6, 5, 4, 3 and 2 seconds. Two of the birds (7 and 8) never reached 90% correct for 3-second signals, but after five and two sessions respectively, they were run with signals two seconds long. Again, their performances did not reach 90% correct. Birds 16 and 18 achieved 90% correct for 3-second stimuli, but not for the 2-second stimuli. The general trend of the results is shown in Figure 62. Each point represents the average of the mean performances of each bird in all sessions run with a given stimulus duration. Birds 7, 8 and 16 contributed results to all the points, while Bird 18 contributed data at the 2- and 3-second durations.[24] The dashed lines give an indication of the spread of the mean trial lengths (the time taken to peck 30 times on one key or the other) and the mean latencies of the first pecks. The birds could perform well when, on some trials at least, the signal ended while they were responding. The birds did less well when the signal sometimes ended before they had started to respond.

Since in the final experiments the task was to be more demanding (stimuli less than 230 msec in duration, smaller differences between the standard and comparison frequencies, and a lower S/N ratio), these

---

[24] Bird 18 was run with longer durations, but with a lower S/N ratio than the other birds. The signal level was raised before the bird began with 3-second signals.

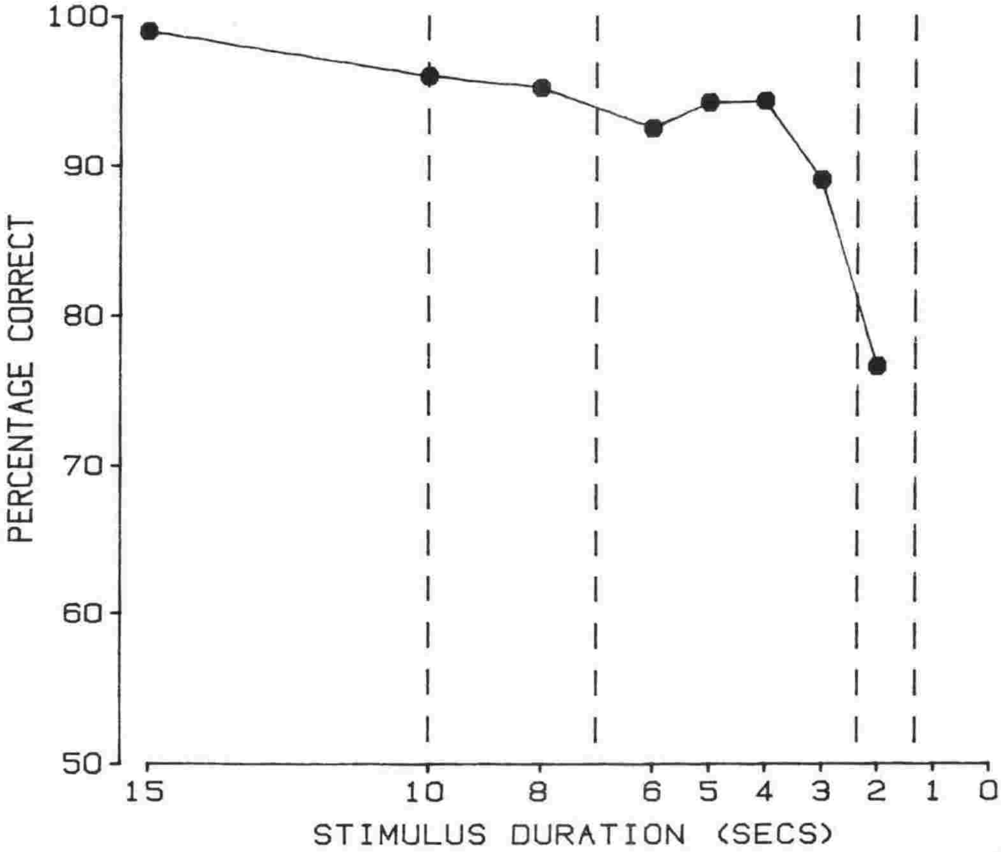


Figure 62. The mean percentage of correct responses as a function of stimulus duration, obtained during experimentation leading up to the third pigeon experiment. The pairs of dashed lines indicate the ranges of the mean trial lengths (left pair) and of the mean latency of the first peck (right pair).

results did not seem very promising, and work with single stimuli was abandoned. If time had allowed, the work would probably have been pushed a little further. There was no real hope that the birds could have coped with single stimuli of the type to be used, but the limits of their ability with relatively brief signals would be of interest.

Repeated Stimulus Presentations

All the birds which had been involved in single stimulus training were switched to the new condition. Initially the signal and noise conditions remained the same as before. The repeated signals were produced by switching the gated tone and noise off 100 msec after onset, and switching them on again 400 msec later. Because the rise-fall time was 20 msec, the signals were 120 msec in duration and were separated by 380 msec gaps.

Over the 82 sessions of preliminary training, various parameters of the auditory stimuli were changed in an attempt to find the most appropriate values for subsequent experimentation. Most importantly, the rise-fall time was increased to 40 msec and the S/N ratio was lowered. In addition, for Birds 7 and 16 the duration of the signals was increased (and the gap between them decreased) in order to raise their levels of performance.

The lowest S/N ratio used (on the 48th day of preliminary work) was 30 dB. At this stage the standard frequency was 450 Hz for all birds. Birds 8 and 18 had one comparison frequency--900 Hz--while Birds 7 and 16 had two comparisons, 900 and 925 Hz. In addition, the signal and gap durations used for Birds 8 and 18 were 140 msec and 360 msec respectively, while those for Birds 7 and 16 were 200 msec and 250 msec respectively. Despite their easier task, Birds 7 and 16 were performing near chance, while Birds 8 and 18 were averaging between 60% and 80%

correct. From this point on, the latter birds were prepared for the experiment proper, although the other birds were also run in case they were needed.

In the course of this latter lead-up to the experiment, the levels of the tones were adjusted slightly to obtain a constant S/N ratio over all frequencies. Given the slight increase in the spectrum level of the gated noise from 200 Hz to 1000 Hz (Figure 36), this entailed lowering the level of the 450 Hz signal by 4 dB, and the levels of the 500 and 600 Hz signals (which were introduced later) by 2.5 dB and 2 dB respectively. All other signals remained at 80 dB. Also, the signal-power-to-noise spectrum level ratio was raised to a final level of 40 dB.