AFOSR-TR-97-0705

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Hearing Research Center Department of Psychology 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NL
6c. ADDRESS (City, State and ZIP Code) University of Florida P. O. Box 112250 Gainesville, FL 32611-2250	7b. ADDRESS (City, State and ZIP Code) 110 Duncan Avenue Room B115 Bolling AFB, DC 20332-8080
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-93-1-0281
8c. ADDRESS (City, State and ZIP Code) 110 Duncan Ave, Room B115 Bolling AFB DC 20332-8050	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO.: 61102F PROJECT NO.: 2313 TASK NO.: BS WORK UNIT NO.
11. TITLE (Include Security Classification) Auditory Pattern Memory and Group Signal Detection	
12. PERSONAL AUTHOR(S) Robert D. Sorkin	
13a. TYPE OF REPORT Final Technical Report 13b. TIME COVERED FROM 93-9-1 TO 97-9-30	14. DATE OF REPORT (Yr., Mo., Day) 9-11-15 15. PAGE COUNT 72
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Auditory perception, temporal pattern discrimination, group decision making, group signal detection

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This project examined how human listeners discriminate temporal patterns and how groups of human observers detect signals presented on complex visual displays. The experiments with temporally-coded auditory patterns showed how listeners' attention is influenced by the position and the amount of information carried by different segments of the pattern. Analyses of group signal detection included mathematical analyses, computer simulations, and human experiments. These analyses specified the effects on performance of team member ability, team decision rule, correlation between member judgments, and type of member interaction. The results of this research may be useful for improving the design of auditory display systems and for optimizing the performance of decision making teams.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr John F. Tangney	22b. TELEPHONE NUMBER (Include Area Code) 202-767-8026 22c. OFFICE SYMBOL NL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

SECURITY CLASSIFICATION OF THIS PAGE

AUDITORY PATTERN PERCEPTION AND GROUP SIGNAL DETECTION

Robert D. Sorkin
Hearing Research Center
Department of Psychology
University of Florida
Gainesville, Florida 32611

15 November 1997

Final Technical Report for the Period 1 September 1993 to 30 September 1997

F49620-93-1-0281

Prepared for
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base, DC 20332-6448

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This project had two research components. The first component examined the discrimination of temporally-coded auditory patterns by trained listeners. The second component consisted of theoretical analyses, computer simulations, and human experiments, on group signal detection. The experiments revealed important principles about how team performance depends on the abilities of the team members, the team decision rule, and the constraints on interactions among members. The results of these two research efforts may be useful for improving the design of auditory display systems and for optimizing the performance of decision making teams.

Auditory Pattern Discrimination

The auditory experiments were addressed at understanding the mechanisms that underlie temporal pattern discrimination. The experimental paradigm had listeners discriminate between two sequences of tones, in which the times between tones were generated according to specific stochastic rules. The first series of experiments evaluated the effect on discrimination of the time between the two patterns to be discriminated. The results of this study provided support for a general statistical model of pattern discrimination. According to this model, a listener encodes the temporal pattern of times for each stimulus and then computes the statistical correlation between the two lists of intertone times. Memory for the temporal information contained in the first list is relatively insensitive to decay over the short time periods of the experiment.

A second series of auditory discrimination experiments was concerned with the distribution of listener attention to different pattern segments. It was suspected that the model's assumption of uniform emphasis on all temporal segments was incorrect. Earlier studies had indicated that an observer's attention is controlled by several factors, such as the relative duration or the variability of different pattern segments. The listener's attention to different segments was assessed by a technique that allowed calculation of the decision weights for each segment of the sequence. In one experiment, the statistical properties of the intertone time intervals were uniform across each sequence. In a second experiment, the mean or the variance of the duration of the intertone time intervals was varied within the sequence. The results may be summarized as follows:

(1) Listeners allocate more attention to the first and last occurring temporal positions of a stimulus pattern than to other positions. These segments normally play the dominant role in the listener's estimate of the difference between patterns.

(2) Giving one of the temporal segments a noticeably longer or shorter duration than the others does not affect the listeners' attentional strategy, as assessed by the decision weights given to the unique positions. Moreover, the absence of a correlation between the listeners' weights and the duration of the unique segment, indicates that the "proportion-of-total-duration" hypothesis is not a general rule in pattern discrimination.

(3) Giving one of the temporal segments a higher variance than the others results in a pronounced peak in the listener's attention to that position. In the temporal pattern discrimination task, higher variance segments are more diagnostic of whether the patterns are the same or different, and it appears that listeners utilize that information in their discrimination.

Group Signal Detection

The theoretical analysis of group signal detection behavior specifies how performance accuracy depends on the size of the group, the detection abilities of the members, the correlation between member judgments, the constraints on member interaction, and the group decision rule. The highest group performance is defined by the *Ideal Group*; this group uses an optimal statistical rule for combining the judgments of the individual members. Lower levels of performance are defined by *Condorcet Groups*; these groups base their decisions on the majority vote of their non-interacting members.

Computer simulations of Condorcet groups produced a surprising result: the accuracy of group performance decreased as the majority decision rule was made more stringent. Experiments with groups of human participants in a visual detection task closely matched the simulation results. Group performance was poorest in the unanimous condition and best in the simple majority condition. Little or no shifts in the team members' decision criterion were produced by different group decision rules.

Groups of from 2 to 12 members were also tested in freely interacting conditions; these were designed to see how near to Ideal performance could be obtained from human teams. Groups were tested in a visual detection task under different conditions of display difficulty and display correlation. The performance of the interacting groups was better than the Condorcet groups but worse than the Ideal predictions. Furthermore, group performance efficiency decreased with group size. The decrease in efficiency with size was not attributable to correlations between member judgments or to inappropriate weighting of judgments from individual members. Some of the decrease in efficience may be due to members reducing their individual detection effort as group size is increased. Additional experiments indicated that some of the decrease in efficiency was due to individual differences in how team members rate and communicate their estimates of signal likelihood to each other. When team member ratings of signal likelihood were ordered and displayed to the group via an optimized display, performance approached the theoretical ideal. Further improvements in group performance may be achieved by displaying member ratings that are normalized by the individual team member's rating function. These results are highly relevant to the performance and training of crews or teams assigned to perform decision-making tasks.

AUDITORY PERCEPTION OF TEMPORAL PATTERNS

I. INTRODUCTION

Which segments of a temporal pattern will receive the most listener attention? Are the attended segments the best ones for discriminating between two different patterns? Recent experiments have attempted to specify how a listener allocates attention to different spectral components of an auditory pattern (e.g., Green et al. 1983; Berg and Green, 1990; Kidd et al., 1991; Zara et al., 1993). In the present study, we apply similar techniques to determine the distribution of listener attention to the different temporal components of an auditory pattern.

In addition to improving our understanding of the basic mechanisms used to process an auditory stimulus, specifying how attention is concentrated on different portions of a temporal pattern may have practical implications for speech and music perception where the discrimination of temporal patterns plays an important role (Liberman et al., 1967; Stevens and House, 1972; Klatt, 1976; Steedman, 1977; Vos and Rasch, 1981). For example, Collins et al. (1994) showed that there is a correlation between the discrimination of random temporal patterns and performance on standard speech recognition tests. In music, several workers have argued that a temporal pattern-recognition scheme is probably used for the perception of the repetitive pattern of musical notes, both at a basic level and at more complex level of rhythmicity (Martin and Struges, 1974; Deutsch, 1979; Fraisse, 1982).

The importance of specific factors in the discrimination of temporal patterns has been demonstrated in a series of studies conducted by Watson and colleagues (e.g., Watson et al., 1975; Kidd and Watson, 1992). In one of the initial studies, Watson et al. (1975) found that the ordinal position of the information was very important. Listeners' performance was better when the change in the pattern occurred towards the end of the sequence (*recency* effect). A recent study (Kidd and Watson,1992; see also Kidd, 1995), reported that the relative target duration was also important. They found that performance improved with an increase in the ratio of the target's duration to the total pattern's duration. They proposed a rule, the "proportion-of-the-total-pattern-duration (PTD)" rule, to describe this result. The rule states that each individual component of an unfamiliar tone sequence is resolved with an accuracy that is based on its proportion of the total duration of the sequence. The basic assumption is that the listener evenly allocates attention over the pattern's duration, and therefore longer duration segments capture proportionally more of the listener's attention.

Lutfi (1993, 1995; also see Lutfi and Doherty, 1994) has suggested that Kidd and Watson's results can be accounted for by the Component Relative Entropy (CoRE) model. This model uses as its basis Shannon's (1948) definition of entropy, as a means of determining the amount of information contained in a pattern. The CoRE model has two basic premises. One is that listeners adopt an ideal decision rule that attempts maximum likelihood test for the task. The other is that listeners are incapable of ignoring any information in patterns that vary randomly from trial to trial. This premise is represented by multiplication of signal with a rectangular time

window. According to the model, "listeners are assumed to integrate information over a rectangular time-frequency window with bandwidth and duration equal to or exceeding the bandwidth and duration of possible signals". The CoRE model suggests that an important factor in these studies is the relative variance of the target tone and that performance can be predicted by a decision variable that is based on the weighted sum of the relative variances of each component's mean value. In the present experiment, we implement a further test of these different views. It should be noted that while the CoRE model makes predictions for d', the PTD rule does not.

The present experimental paradigm is an extension of one we have previously employed (Sorkin, 1990; Sorkin and Montgomery, 1991; Sorkin et al.,1994). The listener's task is to report whether two sequences of tones had the same or different temporal patterns. A sequence's temporal pattern is determined by the intertone time intervals between the tones within each sequence. On *same* trials, the set of intertone time intervals for the two sequences is identical (see figure 1). On *different* trials, the intertone time intervals are perturbed by a random process that results in intertone time intervals that are partially correlated (or uncorrelated) between the sequences. These studies showed that performance in this task depends on the correlation between the sequences on *different* trials and also on the (uniform) mean and variance of the intertone time intervals. These results are described by the Temporal Pattern Correlation Model, which assumes that the listener attempts to estimate the correlation between the sequences on each trial but is limited by a small, internal jitter in time estimation of between 10 and 20 ms (Sorkin, 1990; Sorkin and Montgomery, 1991).

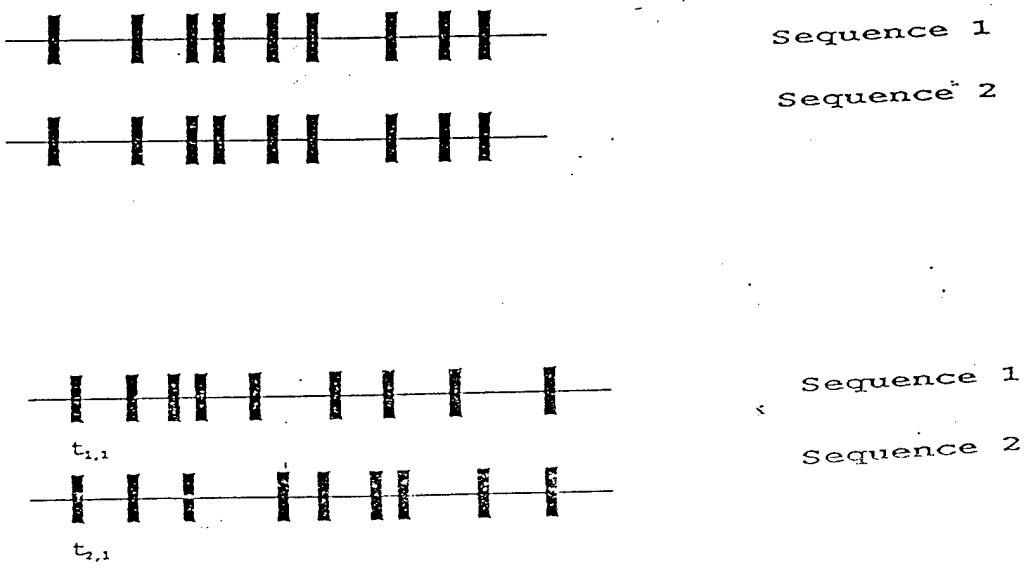

Figure 1. The top part of the figure illustrates a sample SAME trial for the two sequences of a pattern, and the bottom portion illustrates a sample DIFFERENT trial. The frequency and duration of each tone was 1000 Hz and 25 ms respectively.

One aspect of this experimental task is that information about the difference between the pair of sequences to be compared is distributed throughout the entire duration of the experimental stimuli. This is both a strength and potential weakness of the paradigm. The potential weakness arises when certain portions of the sequence play a critical role in discrimination that may not actually be evident in the experimental results. However, this uncertainty could be eliminated if one could determine the relative contribution made by individual segments of the sequence to the listener's discrimination performance. Such a technique is provided by Berg's (1990) weighting analysis or Conditional on Single Stimulus (COSS) technique and by Richards and Zhu's (1994) and Lutfi's (1995a) extension of the technique. The idea is to examine the strength of the relationship, over trials, between the listener's response and the value of each stimulus component. This technique enables the experimenter to calculate the relative contribution of different parts of the stimulus to the listener's decision. One then may infer from the resulting decision weights how much attention is given to different portions of the stimulus.

The main goal of the present study was to determine how attention is allocated to the different segments of each sequence in a temporal pattern discrimination task. In experiment 1, the statistical properties of the intertone time intervals were uniform within each sequence. That is, all intertone time intervals had the same mean duration and the same standard deviation in duration. We anticipated that the greatest listener weights would be given to the final temporal positions in the sequence, thereby replicating Watson et al.'s (1975) finding of best performance for a target that occupies the last temporal position in a sequence.

In experiment 2, we employed patterns with both uniform and non-uniform time durations. In the non-uniform cases, we made one segment of the patterns unique by manipulating either the mean or the variance of its intertone time interval duration. Increasing the mean duration of one of the sequence segments should make that segment more distinctive. We can think of two possible consequences of this manipulation: One possibility is that making a segment more distinctive would reduce the apparent randomness of the pattern, thereby improving the storage of information about the pattern. This could result in an improvement in the discriminability of the entire pattern. A second possible consequence of the manipulation is that the unique segment would draw the listener's attention to itself, and consequently less attention would be given to other segments. The resultant effect on performance could be good or bad depending on whether the unique segment carried more or less information about the task than the other segments.

The CoRE model and PTD hypothesis address these possibilities. Our interpretation of the PTD hypothesis is that the listener's attention will be spread evenly over the duration of the pattern to be discriminated, with the caveat that somewhat greater weighting may be given to the final portion of the pattern. It follows that increasing the mean duration of one segment will produce an effective increase in the listener's attention to that segment. If much of the task relevant information is contained in that segment, the result would be an increase in discrimination performance. However, the PTD hypothesis does not address the possible effect of manipulating the variance of an intertone time interval.

The CoRE model, on the other hand, is primarily concerned with the informational aspects of the patterns to be discriminated. This model makes no prediction about the effect of manipulating the mean duration of a segment on the listener's decision weights in the present task. The reason is that the intertone time intervals (t_i) are part of the statistic in this case. Therefore, the time-windowing does not contribute a further weight (i.e. it would be meaningless to write SUM [$t_{1i} * t_{1i} - t_{2i} * t_{2i}$] to reflect the effect of time-windowing on time weights). This is why, the CoRE model does not make any predictions about the proportional mean duration. In other words, the relative duration is the decision variable here and cannot be its own weight. In the next section we show that increasing the variance of a segment in our task makes that segment more diagnostic of whether the sequences are the same or different and therefore more appropriate for the assignment of a higher decision weight. This argument is consistent with the CoRE model, which would predict that manipulation of the variance of one segment can produce potentially large changes in the decision weights. Thus, experiment 2 should enable us to determine which model provides a better description of pattern discrimination.

A. Temporal Pattern Discrimination Mechanism

The sequence discrimination task requires the listener to compare the intertone time intervals for two sequences and then make a judgment about whether they were the same or different. The set of intertone time intervals for the first and second sequences are, respectively, $\{t_{1i}\}$ and $\{t_{2i}\}$; where the numbered index identifies the first and second sequence, respectively, and i identifies the ith intertone time position (e.g. i=1 to M for intertone time positions from 1 to M-1 tones). We assume that the listener's decision variable is based on the sum of the weighted (absolute value of the) difference between the pairs of corresponding intertone times, $\sum_{i=1}^{M} a_i g_i$, where the a_i is the decision weight for the ith temporal position, and g_i is a random variable equal to $|t_{1i} - t_{2i}|$, the absolute difference between the pair of intertone time intervals at position i in sequence 1 and 2. On *different* trials, the t_{ji} are independent, normal random variables with equal means and variances equal to σ^2_t. The variance, σ^2_t is the total variance in the observer's estimate of g_i and is given by $\sigma^2_t = \sigma^2_{internal} + \sigma^2_{exp}$ where $\sigma^2_{internal}$ is the variance due to the observer's jitter in estimating the time, and σ^2_{exp} is the variance in the intertone time intervals that is set by the experimental condition. On *same* trials, σ^2_{exp} is equal to 0, since all the intertone time intervals (t_{1i} and t_{2i}) are identical, which results in σ^2_t being equal to $\sigma^2_{internal}$. In other words, the pattern discrimination is only based on the differences between the intertone time intervals across the sequences in a trial. It should be noted that the sequences are independent across trials.

Since the g_i in different time positions are independent, the performance of the discrimination mechanism can be computed from knowledge of d_i', the individual discriminability of each intertone time interval pair. The segment discriminability, d_i', will depend on the mean and variance of g_i, given either the same or different sequences on a trial (see Sorkin, 1990 for d' calculation). It can be shown that E(g) is a function of σ_t^2. That is, an increase in the variance σ_{exp}^2 is analogous to an increase in signal strength. In other words, as the intertone time intervals

become more different from one another, they become more discriminable[1].

II. EXPERIMENT 1: EFFECT OF TEMPORAL POSITION ON LISTENER WEIGHTS

The goal of this experiment was to determine the effect of temporal position on listener attention when the statistics (mean and variance) of the intertone time intervals were constant across the different ordinal positions of the sequence. The correlational weight-estimation technique was used to determine the weight given to each ordinal position within the sequence. That is, at each intertone time interval position, we calculated the correlation (over *different* trials) between the listener's response and the absolute difference between the corresponding intertone time intervals in the two sequences, $|t_{1i} - t_{2i}|$. An assumption of the correlation-weight analysis is that the decision weight at each position is proportional to this normalized correlation. All weights were normalized to sum to unity ($\Sigma a_i = 1$).

A. METHOD

1. Listeners

One female and three male students from the University of Florida, with normal hearing (as determined from self report), participated in this experiment. One listener had prior experience with the task. All listeners were paid an hourly wage plus a bonus based on performance. Listeners were seated in a double-walled acoustically insulated chamber. The stimuli were presented monaurally via TDH-39 headphones.

2. Procedure

On a given trial, listeners were presented with two sequences of tones, each composed of nine 1000-Hz tones presented at 71-dB sound pressure level. The 25-ms tone bursts were generated by a T.T. Electronics system (precursor to Tucker Davis Technologies system), that sampled at 20000 samples per second. The tones were low-pass filtered at 7.5 kHz and gated with a TTE Cosine switch (set at cosine-squared shaping). These tones were shaped by a 4-ms linear rise and decay envelope. All times were defined from the 0-voltage point. The intertone time intervals between tones were generated by a process that enabled control of their mean and standard deviation (see Sorkin, 1990). The time between the tones (intertone time interval) had a mean duration of 50 ms and a standard deviation of 35 ms. The minimum intertone time interval was either 2 ms or mean intertone time interval minus 2.5 times the standard deviation (whichever was larger). The maximum intertone time interval was either 300 ms or mean intertone time interval plus 2.5 times the standard deviation of the intertone time (whichever was smaller). If an intertone time interval smaller or larger than the allowed values was drawn, then it was changed to the stated minimum and maximum.[2] An interval of 750-ms separated the two tone sequences. After listening to the pair of sequences presented on a trial, the subject indicated whether or not the temporal patterns of the tones were the same or different. Visual feedback about the correct responses was provided after each trial. The type of trial (*same* or *different*) was selected on a

random basis. On *same* trials, the temporal patterns were perfectly correlated (sequence pattern correlation =1.0). On *different* trials, the patterns were uncorrelated (sequence pattern correlation =0). Figure 1 displays a sample *same* and *different* trial. All aspects of this experiment, including the stimulus presentation and data collections were computer controlled.

Listeners participated in 2 or 3 experimental sessions per week. Each session consisted of 10 blocks of 100 trials. All experimental parameters (mean and standard deviation of intertone time intervals, etc.) were held constant within the block of 100 trials. Listeners participated in several practice sessions, which consisted of 700-1000 trials, before data collection was begun. There was no obvious improvement in discrimination performance, which was taken as lack of evidence for any practice effect.

Table 1. Summary of the experimental conditions and average performance (d') for the listeners in the experiments.

Experiment	unique position			other positions		obtained d'
	pos.	mean	σ_{exp}	mean	σ_{exp}	mean (std. error)
1		-none-		50	35	3.09 (0.22)
2		-none-		60	20	2.34 (0.23)
2a	2	20	20	60	20	2.80 (0.27)
	2	40	20	60	20	2.75 (0.29)
	2	80	20	60	20	2.70 (0.28)
	2	100	20	60	20	2.76 (0.40)
	6	20	20	60	20	2.83 (0.31)
	6	40	20	60	20	2.64 (0.37)
	6	80	20	60	20	2.62 (0.29)
	6	100	20	60	20	2.65 (0.17)
2b	2	60	40	60	20	2.67 (0.24)
	2	60	60	60	20	2.98 (0.23)
	2	100	100	100	20	2.52 (0.29)
	6	60	40	60	20	3.14 (0.29)
	6	60	60	60	20	3.51 (0.26)
	6	100	100	100	20	3.28 (0.28)

B. RESULTS

The average d' of the 4 listeners for a mean intertone time of 50 ms and a standard deviation of 35 ms was 3.09 (first row of table 1). The average weight data from all 4 listeners for experiment 1 are shown in Figure 2. The abscissa marks the temporal position of each intertone time interval and the ordinate indicates the relative weight given to each intertone time position. (The error bars indicate plus and minus one standard error of the mean.) It is clear that listeners allocated more weight to the first and last temporal positions. The individual data resembled the average plot, except that two of the listeners gave the highest weight to the first position, and two of the listeners gave essentially equal weight to the first and last positions. An ANOVA test

(repeated measures design with 4 listeners and 8 temporal positions) indicated a significant effect of temporal position [$F(3,7)= 5.27$, $p<0.05$]. Our model had not made any predictions regarding the effect of temporal positions.

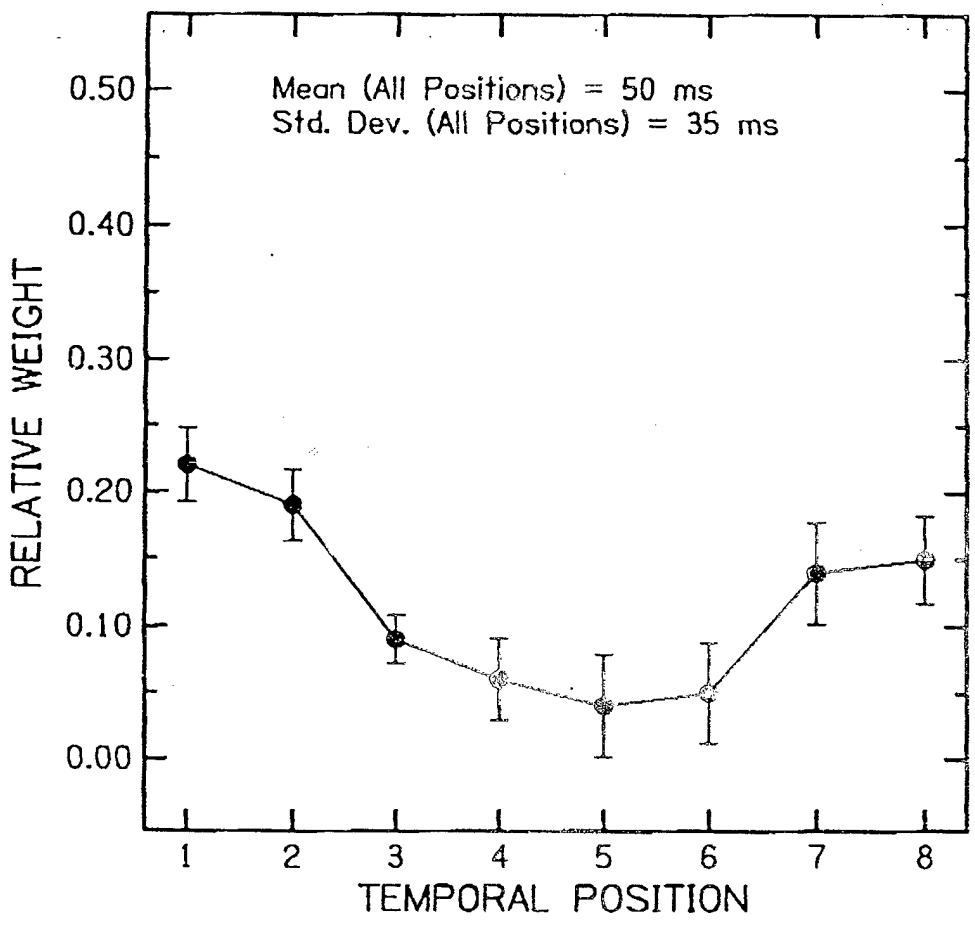

Figure 2. The relative weights obtained in experiment 1, averaged over 4 listeners, are plotted as a function of temporal position. (The error bars represent one standard error of the mean.)

III. EXPERIMENT 2: EFFECT OF SEGMENT MEAN AND VARIANCE ON LISTENER WEIGHTS

The results of experiment 1 indicated that listeners do not distribute their attention uniformly among intertone time intervals of equal duration, but rather give more weight to the early and late positions. In experiment 1, the statistics of the intertone time intervals were uniform across the sequence. Experiment 2 was designed to investigate whether a change in the properties of *one* of the intertone time intervals would be reflected by a change in the listener's weighting strategy. The statistics of one of the intertone time intervals was manipulated in two different ways: In part 1 of experiment 2, the average duration of one of the intervals was set to be either longer or shorter than all of the others (the others all had the same mean and standard deviation). Recall that the PTD rule predicts that increasing the mean duration of a component will produce greater listener attention to that segment and decreasing the mean duration will produce less attention to that segment. The CoRE model, on the other hand, makes no prediction about the manipulation of mean duration in this situation. In part 2 of experiment 2, the standard deviation of one of the intervals was set to be longer than the others. Our version of the CoRE hypothesis is that the relative variance of the segment will be an important factor in determining the listener's attention. This is because performance is assumed to depend on a decision variable based on the weighted sum of the detectability of each segment, and the detectability of a segment increases with the segment's variance. The PTD rule, on the other hand, makes no prediction about the effect of segment variances.

A. METHOD

1. Listeners

Four University of Florida students, two males and two females with normal hearing, participated in both parts of experiment 2. Two listeners had participated in the first study and had prior experience with the task. All listeners were paid an hourly wage plus a bonus based on performance. The same apparatus and stimuli as the first experiment were used in this experiment.

2. Procedure

The procedure, tones and task were the same as in experiment 1. Table 1 summarizes the parameters of the different conditions run in experiment 2. In the control condition, the mean and standard deviation of the intertone time intervals were uniform across all temporal positions, as in experiment 1. The mean intertone duration was 60 ms and the standard deviation of intertone duration was 20 ms. The reason for choosing these durations was to acquire a simple and systematic manipulation of the duration of the mean and standard deviation of the intertone time intervals. In part 1 of experiment 2 (2a), one of the intertone time intervals had either a higher or lower mean duration than the other positions. This intertone time interval occurred either at the 2nd position or the 6th position in the sequence. The different values of mean intertone-time tested ranged from 20 to 100 ms, in steps of 20 ms. This mean value was fixed during a block of

trials, as was the position of the unique intertone time (2nd or 6th). The mean and standard deviation for all the other positions within the sequences were kept at 60 and 20 ms respectively (see table 1). For part 2 of experiment 2 (2b), one of the intertone time intervals had a higher standard-deviation, which occurred at either the 2nd position or the 6th position in the sequence. The mean and standard deviation for all other positions were set to 60 and 20 ms respectively, except when the standard deviation value was 100 ms. In this case, the mean for all the positions was set to 100 ms. The different standard deviation values tested were 40, 60, 100 ms (see footnote 2). Listeners ran several hours of practice trials before the data collection began and no subsequent practice effects were observed.

B. RESULTS

The average d' for the four listeners and conditions is provided in table 1. Average performance in the uniform (control) condition yielded a d' of 2.34. Figure 3 depicts the average weight data from the four listeners in this condition. The first position is given the highest weight by these listeners. Note the similarity between the data in this figure and figure 2, of which this is essentially a replication. All the individual results showed the same pattern of results as figure 3. Average performance in experiment 2a, when the 2nd or the 6th position had a different mean value, are also shown in table 1. Performance was relatively constant over these 8 conditions, ranging between a d' of 2.62 and 2.83. The overall average d' for these conditions was 2.72, which is somewhat higher than the 2.34 value obtained in the uniform condition. Assuming an internal jitter of 10 ms, our simple pattern discrimination model predicted a d' value of 2.49 for both the uniform condition and these unique mean conditions.

The average weights from the 4 listeners for this case are plotted in figures 4 A-D. These figures represent the average data for the case where the 2nd position had a different mean value (filled circles) or the 6th position had a different mean value (unfilled circles). In this case, the standard deviations were all equal and fixed while the mean intertone time intervals of the 2nd or 6th position varied (20, 40, 80, 100 ms). It should be noted that the different mean durations of 20, 40, 80, 100 ms corresponds to 3, 5, 11, 16 percent of the total pattern duration, while 60 ms, for all the other intertone intervals corresponds to 8.5 percent. An ANOVA test (repeated measures design with 4 listeners and 5 mean values) showed no significant effect of a different mean intertone time interval at the 2nd position [$F(3,4)= 2.24$, $p<0.05$] or the 6th position [$F(3,4)= 1.00$, $p<0.05$]. However, the individual data in some cases showed a small tendency for a larger than usual weight given to the position with a different mean. Figure 5 is a plot of the average relative decision weight for the 2nd (filled circles) and 6th (unfilled circles) position as a function of the mean duration of that intertone time interval. The listeners' weights appear to be independent of the duration of the unique segment (or of the segment's proportion of the total pattern duration).

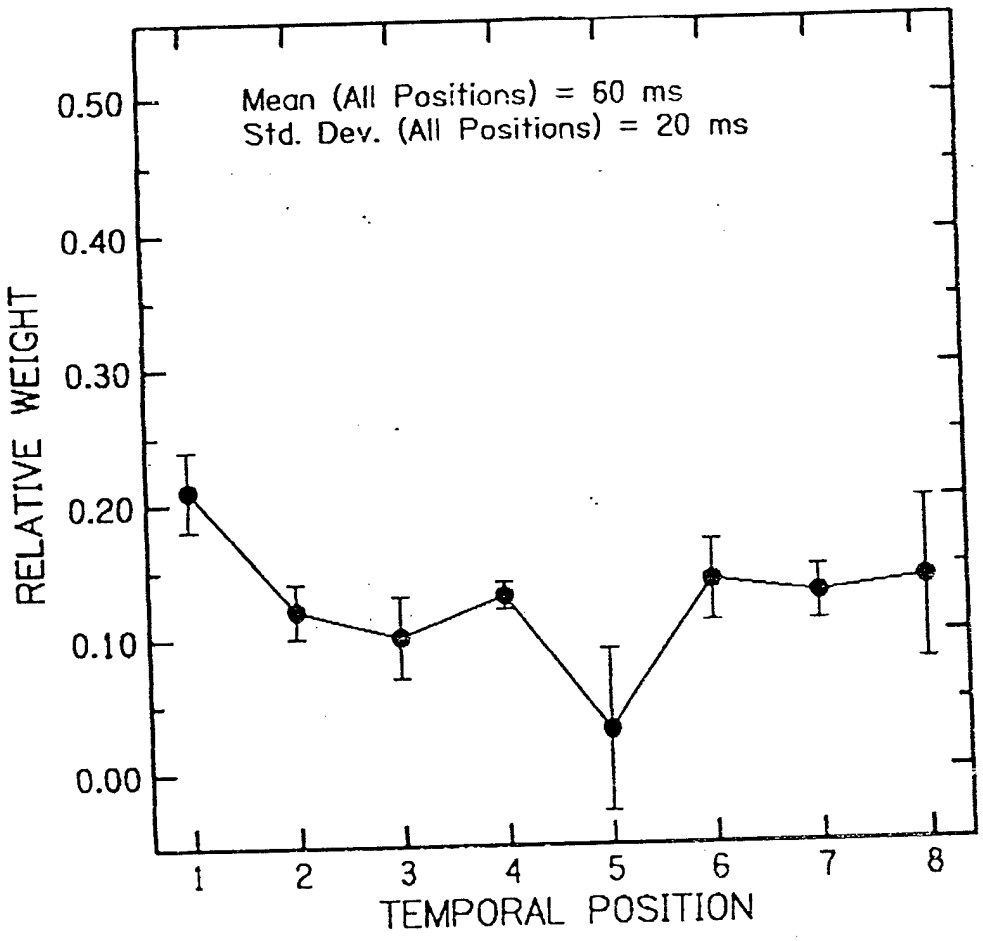

Figure 3. The average relative weights obtained in the uniform condition of experiment 2, averaged over 4 listeners, are plotted as a function of temporal position. (The error bars represent one standard error of the mean.)

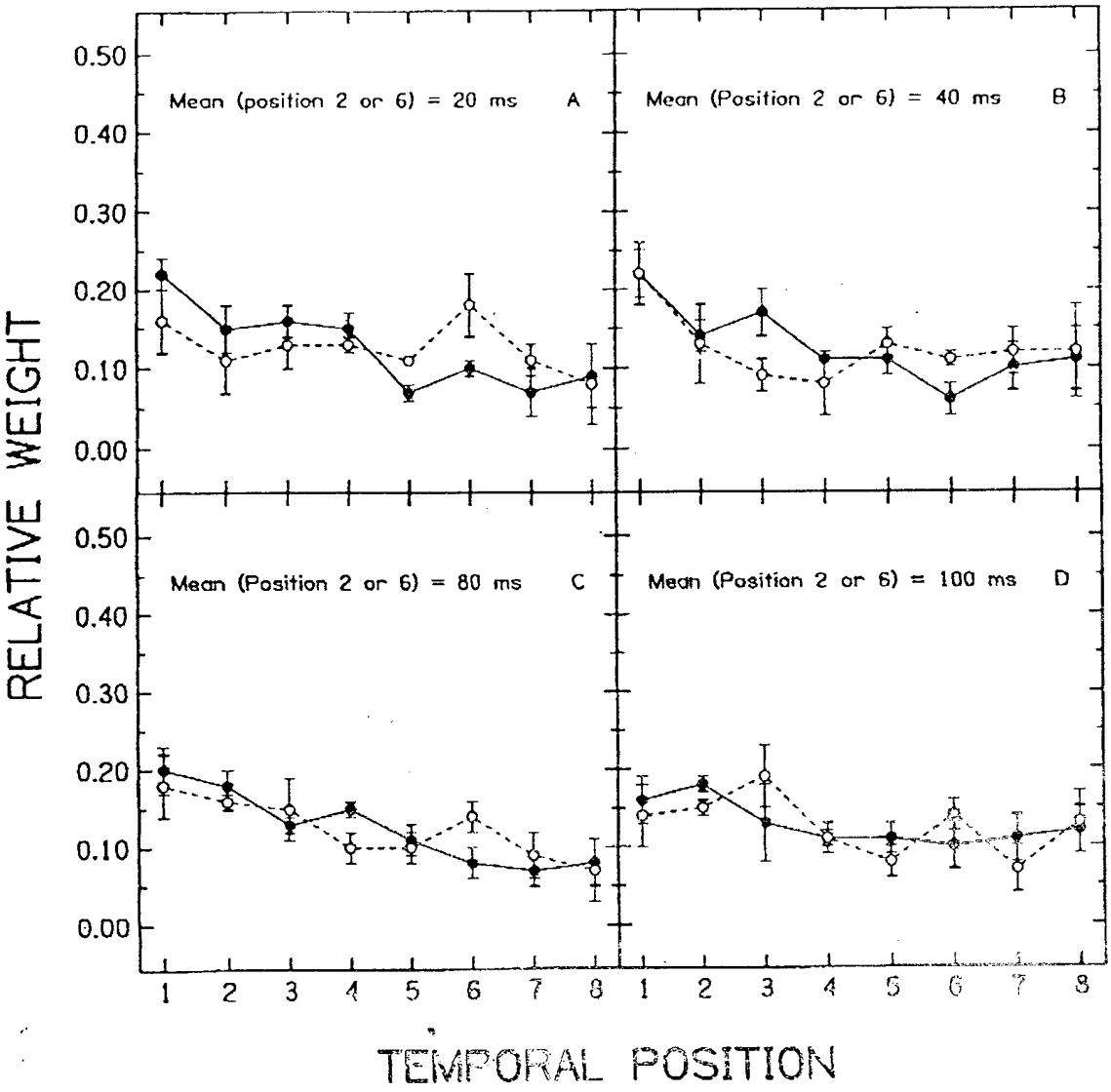

Figure 4A-D. The average relative weights obtained in experiment 2, averaged over all listeners, are plotted as a function of temporal position for the case when the longer or shorter duration segment was in position 2 (filled circles) or position 6 (unfilled circles). The mean duration for the unique position was 20 ms (panel A), 40 ms (panel B) , 80 ms (panel C), and 100 ms (panel D). (The error bars represent one standard error of the mean.)

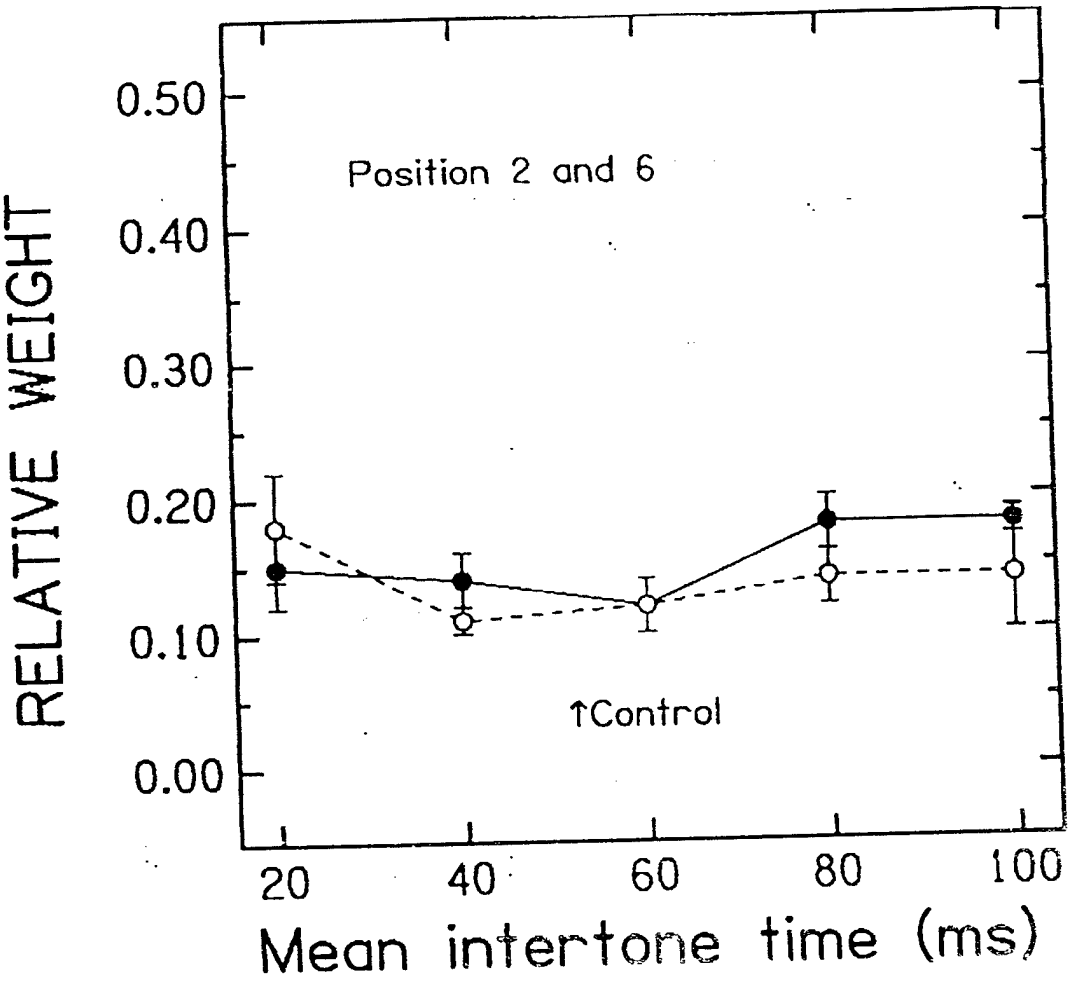

Figure 5. The average relative weight for the 2nd (filled circles) or 6th (unfilled circles) position is plotted as a function of the mean duration of the unique segment. (The error bars represent one standard error of the mean.)

Table 1 shows that the average d' in experiment 2b, when the 2nd or 6th time had a higher standard deviation, was somewhat higher than obtained in case 1, when the 2nd or 6th time had a different mean value. When the higher variance of intertone time interval was assigned to position 2, the obtained d' values were essentially the same as those obtained in case 1 (2.52 to 2.98). However, when the different standard deviation was assigned to position 6, performance was higher, ranging between 3.14 and 3.51. Our model predicts d' values of 2.77, 2.95, and 3.17 for the σ_{exp} values of 40, 60 and 100; however, the model makes no distinction between the effect of higher variance when the unique segment is in the 2nd or 6th position.

The effect of variance on weights can be seen in figures 6 and 7. Panel A, B, and C of figure 6 show average weight as a function of position for different values of the variance of the unique segment. Filled circles in each panel show the average weight when the unique segment is the 2nd position and the unfilled circles show the weight when the unique position is the 6th position. There is a clear peak in weight for the unique positions. This result was also true for the individual listeners; regardless of which position was assigned the higher standard deviation, all listeners consistently gave the highest weight to the unique position. Figure 7 shows the average weight as a function of the standard deviation of the unique position for the 2nd (filled circles) and 6th (unfilled circles) position. In general, the higher the value of the standard deviation of the unique position, the larger the weight is given to that position by the listeners. In this case, the weights increase in direct proportion to the standard deviation, which would be predicted by a maximum-likelihood decision maker. An ANOVA (repeated measures design with 4 listeners and 4 values for standard deviation) showed a significant effect of changing the standard deviation of intertone times at both positions; 2 [$F(3,3)= 15.7$, $p<0.05$] and 6 [$F(3,3)= 16.75$, $p<0.01$].

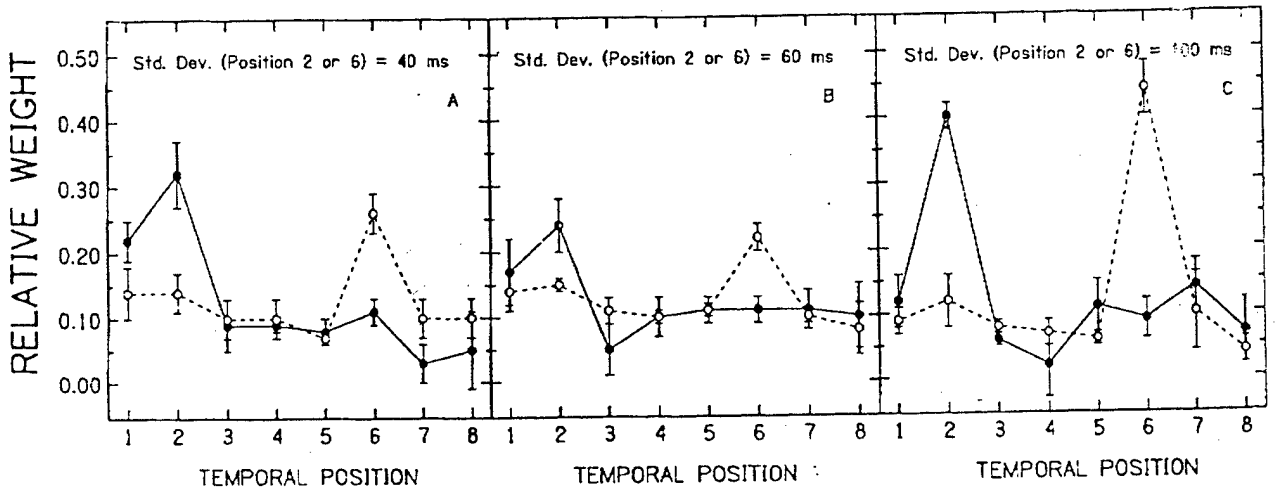

Figure 6 A-C. The average relative weights obtained in experiment 2, averaged over all listeners, are plotted as a function of temporal position for the case when the segment with the higher standard deviation was in position 2 (filled circles) or position 6 (unfilled circles). The standard deviation of the unique position was 40 ms (panel A), 60 ms (panel B) , and 100 ms (panel C). (The error bars represent one standard error of the mean.)

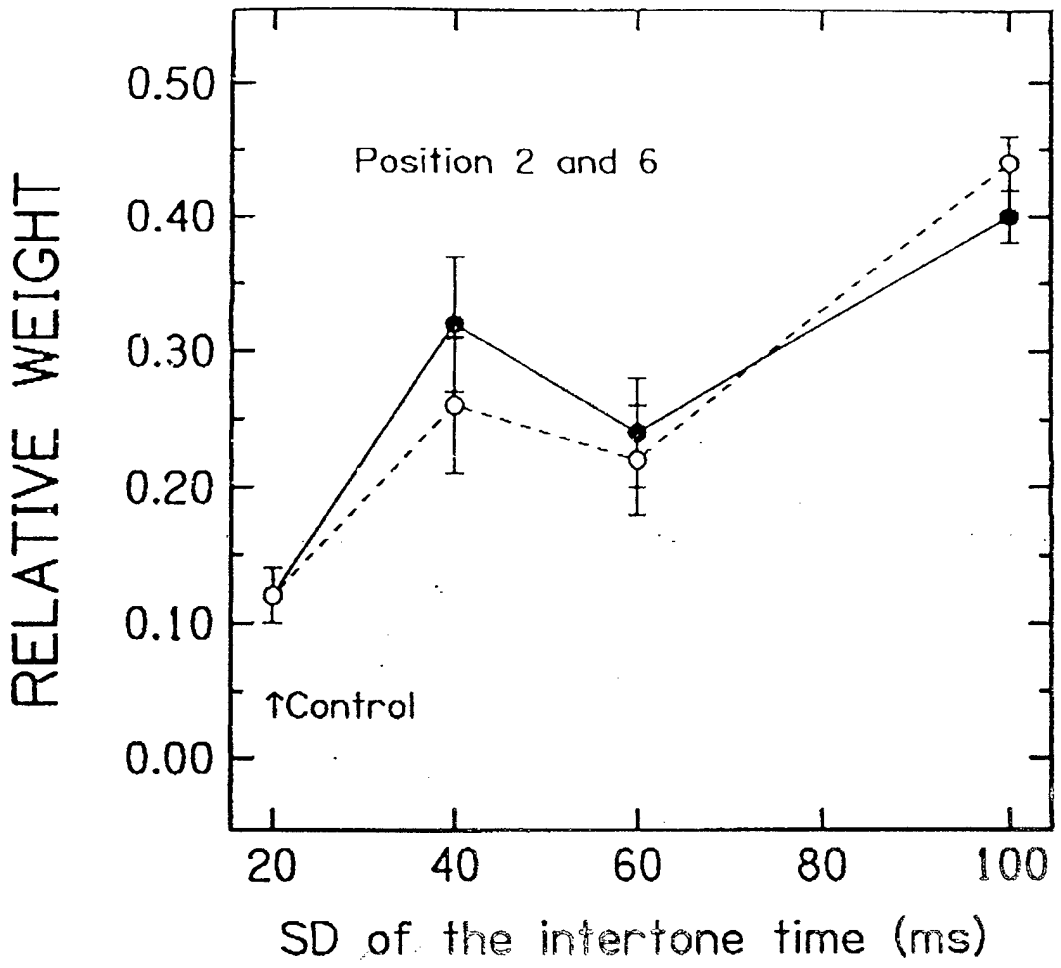

Figure 7. The average relative weight for the 2nd (filled circles) or 6th (unfilled circles) position is plotted as a function of the standard deviation of the unique segment. (The error bars represent one standard error of the mean.)

IV. GENERAL DISCUSSION

These experiments show that listeners can allocate more attention to certain temporal positions of a stimulus pattern than to others. Typically, most listener attention is given to the first and last occurring segments and the least listener attention is given to the middle occurring segments. These results have an important bearing on the Temporal Pattern Correlation and Proportion of Total Duration hypotheses, which each assume that a listener allocates approximately uniform attention over the duration of the stimulus. That assumption is not supported by the current results. Instead, the present results suggest that early and late-occurring segments normally play the dominant role in the listener's estimate of the difference between patterns, and that other segments may be given substantially more weight when appropriate for a specific task.

It is interesting to compare the present results with the Watson et al. (1975) study, which found that best performance was obtained when a signal occurred in the last temporal location of a pattern. There are several important differences between this experiment and the Watson et al. study (1975) that may account for the apparently divergent results. In the Watson et al. experiments, the sequences consisted of ten tones of 40-ms duration that ranged from 256-1500 Hz. The listener's task was to detect changes in the frequency of only one tonal component in the pattern, using a same-different task. The frequency of these tones varied randomly within a sequence. In some conditions, the observer knew which component was the relevant one, and in other conditions the observer was uncertain about which was the signal component. In either type of condition, Watson et al. found that the best frequency resolution was obtained when the signal component was the last one to occur. Watson and his colleagues also investigated conditions in which a silent interval was inserted after an early-occurring component of the sequence. When that gap was greater than about 70 ms, detection of the change in frequency of the component that occurred just before the gap was as good as when that component was the final component in the sequence. Apparently, the presence of a gap allowed the listener to retain information about a component that would otherwise have been disturbed by the occurrence of subsequent components.

We believe that limitations on memory encoding and memory capacity provide a reasonable explanation for the primacy and recency effects observed in the present experiment as well as for the differences in overall performance that we observed in different conditions. Suppose that the listener most hold detailed, i.e., analog, information about all the segments of a sequence for later comparison. This would impose a very large demand on the listener's memory capacity. This memory requirement could be eased if relevant information that was initially in an analog form could be transferred to a non-decaying, categorical store. This is essentially the two-process theory described by Durlach and Braida (1969) in their trace-context model of intensity discrimination (see also Sorkin, 1987). The theory assumes that data stored in the analog or trace mode requires a lot of storage capacity and degrades rapidly over time, whereas data that has been stored successfully in categorical or context mode does not decay. However, the process of encoding data in the context mode takes time and may be interfered with by the arrival of subsequent inputs (or the need for more detailed categorization). Separating the tones by gaps reduces the interference to this encoding process, as shown by the Watson et al. (1975) experiment.

In our experiment, the frequency of the tones was kept constant and the intertone time intervals between all the tones of the sequences was varied. All tones were separated by silent gaps of average duration equal to at least 60 ms. Thus, many segments in our sequences would have been effectively separated by gaps of sufficient length to allow some context-coding of the relevant time data, at least for early segments in the sequence. It is likely that the primacy effect observed in our experiment occurred because early-arriving information had been successfully stored in context mode. The data encoded from these early segments did not decay because it was successfully context-coded. As the sequences played themselves out, the demands on the context encoding mechanism increased to the point where only trace storage was viable. Moreover,

because the data in trace mode decays and is capacity limited, only the most recently stored trace data was available. As a consequence of these two memory processes, both primacy and recency effects were observed.

Experiment 2a showed that making one of the temporal positions noticeably longer or shorter than the others did not cause the observers to assign more decision weight to that position. The first temporal position still received the largest weight. However, there was a small gain in overall performance when one of the positions was assigned a different mean value. Apparently, making one of the temporal positions in the pattern different from the others can make the entire pattern more discriminable from other patterns, but is not a sufficient condition for changing the listener's attentional strategy across different portions of the stimulus pattern.

Changing a segment's duration in our task did not cause it to carry more information than other segments. This is because changing the mean duration of a component does not affect the characteristics of the distribution of the differences (or absolute value of the differences) between the paired time intervals in the two sequences. Rather, allotting more attention to the unique segment could cause an observer to neglect relevant information from the other segments. Since in this condition, (a) overall performance increased (albeit slightly), and (b) there was not a significant increase in weight to the unique segments, we must reject the PTD hypothesis. These results may suggest why the PTD hypothesis (Kidd & Watson, 1992) holds in some pattern-discrimination tasks. When one increases the mean duration of one segment of a pattern, the overall processing of that pattern may be improved because of the consequent change in the uniqueness of the entire pattern, not because more listener attention has been given to the unique segment.

Experiment 2b showed that when one of the temporal positions had a more variable intertone time interval than the others, listeners allocated substantially more attention to that position. Contrary to the previous manipulation of mean duration, this manipulation does increase the information available to the listener from the unique component. That is, increasing the variance of the intertone time interval provides the listener with additional diagnostic information about the pattern task because it increases that segment's signal-to-noise ratio. We observed a pronounced peak in the listener weight for the unique position, whether it was in the second or sixth position, a result fully consistent with the basic assumption of the CoRE model.

The particular position of the unique segment in case 2 of experiment 2 had an interesting effect on overall discrimination performance. Performance was higher when the unique segment occurred late in the pattern (at position 6). It appears that the listener can use more of the information from a particular segment when that segment occurs late rather than early in the stimulus. One possibility is that the information available from early segments is limited because it has been stored in context mode, and that more complete information is available from the final segments because that information has been stored in trace mode.

GROUP SIGNAL DETECTION

I. INTRODUCTION

How effectively can groups of people perform signal detection tasks, and how does performance depend on the abilities of the individual members and the constraints on team member interaction? We attempted to answer these questions by comparing the performance of human groups with the predictions of a signal detection analysis of group decision making. The signal detection analysis specifies how the accuracy of the group's performance should depend on the group's size, the detection abilities of the individual members, the correlation among member judgments, the constraints on member interaction, and the group decision rule. The analysis also allows specification of the absolute *efficiency* of group performance; that is, it yields a measure of how much the group's performance differs from that of statistically optimal groups. Our experiments with groups that performed visual detection tasks lead to several interesting conclusions about the efficiency of human decision making and to some specific suggestions for improving the performance of human groups.

Statistical arguments about the effects of size and member competence on group performance have existed for more than two hundred years, since Condorcet (1785) and more recently Einhorn, Hogarth, & Klempner (1977). According to the statistical argument, group performance should increase with group size, with the most rapid increase occurring when the competence of the group's members is high and when independent information is available to each member. These models assume that there is an effective way to combine the members' judgments. If the expertise of the members varies within the group, each member's input should be weighted proportionally by her competence at the task (Grofman, Feld & Owen, 1984; Grofman, Owen & Feld, 1983; Nitzan and Paroush, 1982, 1984; and Shapley and Grofman, 1984).

The empirical data on group performance indicates that human groups are generally less effective than would be predicted by models that claim optimal use of the members' information. In a fascinating sketch of 40 years of research on group decision making, Davis (1992), pointed out that most research has found group performance to be relatively inefficient. Group performance usually is superior to the average of individual performance, but less than the statistical expectation (see, also Hastie's 1986 review). Moreover, many studies have found that group performance either is insensitive to group size or that the advantage of size declines more rapidly than would be predicted from the statistical argument. All of these results can be attributed to inefficiency in group function such as might be caused by problems associated with member interaction or coordination, with reduced member motivation such as social loafing (Latané, Williams & Harkins, 1979; Shepperd, 1993) or with problems inherent in combining judgments from members using different scaling functions (Wallsten, Budescu, Erev, & Diederich, 1997; Myung, Ramaroorti, & Bailey, 1997).

Recent attempts to model group performance have employed signal detection theory (Erev, Gopher, Itkin, & Greenshpan, 1997; Metz and Shen, 1992; Pete, Pattipati, & Kleinman,

1993a,b; Sorkin and Dai, 1994; and Sorkin and Robinson, 1996). Metz and Shen analyzed the gains in the detection accuracy of reading X-ray images that resulted from replicated readings by the same or multiple readers. Erev *et al.* (1997) examined the strategic interaction between two observers in a signal detection task; i.e., when each observer's payoff structure was contingent on the outcome and the response of the other observer. Pete *et al.* (1993a) considered the case of multiple team members working in an uncertain, binary choice detection situation. They generalized the signal detection model to consider the individuals' as well as the group's, prior probability and payoff structure; that is, their model allowed joint optimization of the group aggregation rule and the individual decision rules of the group members.

Sorkin and Dai (1994) took a somewhat simpler approach to group signal detection than did Pete *et al.* (1993a). They assumed that each group member would provide a graded estimate of signal likelihood, and that the expertise of the members was known à priori to the group. These assumptions allowed them to sidestep the problem of how to aggregate binary responses from individuals who might have different biases toward the decision alternatives. They computed the performance that would result from the optimal aggregation of the members' likelihood estimates; this specified the performance of the *ideal group*. The ideal group analysis forms the basis for one class of detection models to be tested in the present study and provides an upper bound on the performance to be expected from any group of human participants.

Sorkin and Robinson (1996) also utilized detection theory to analyze the performance of classical Condorcet Groups. The members of these hypothetical groups do not exchange information and the group decision is determined by aggregation of the binary votes of the members[1]. The performance accuracy of these groups depends on the particular majority decision rule employed. Pete et al. (1993a) showed that a simple majority was the optimal rule, and Sorkin and Robinson showed how performance decreased as the majority rule shifted from a simple majority to more stringent majorities. The poorest performance was produced by a unanimous decision rule. The analysis of such statistically 'inefficient' groups is useful for modeling groups of human members. That is, a reasonable lower bound on the performance to be expected from a human group is provided by the (member-matched) Condorcet group whose decision is based on the majority vote of its non-interacting members.

In this report we first provide a description, in signal detection theory terms, of the group detection problem and of normative Ideal and Condorcet groups. Second, we describe several experiments which assess the performance of groups of human participants under different task conditions and different constraints on member interaction. Third, we compare the resulting behavior to the Ideal and Condorcet predictions. We will show that a signal detection analysis can account for much of the variance observed in the performance of groups in signal detection tasks. Finally, we address the question of why group performance efficiency decreases with group size and some possible ways to reduce that performance decrement.

The basic group detection task in all our experiments was to judge whether the stimulus input was due to a signal or non-signal event. On each trial, subjects were presented with

graphical visual displays which indicated either a *signal* or *non-signal* condition. Figure 8 is an example of the stimulus on a signal trial. The readings on the array of 9 gauges shown in the figure were determined by a statistical distribution. The task was to indicate whether the stimulus array had been generated from the signal or non-signal distributions; that is, on a signal trial, the readings on all of the gauges were drawn from the signal distribution. The signal distribution had a higher mean than the non-signal distribution, and both distributions had the same variance. The means and variance of the signal and non-signal distributions thus determined the difficulty of the task. After a group (or individual) decision was made, feedback about the correct answer was provided. A monetary payoff to the individual depended on the accuracy of the group's (or individual's) response. We assessed the performance of individuals and groups on this task under display conditions and different constraints on member communication and group decision rule. In the next sections, we present the group signal detection analyses of the Ideal and Condorcet groups.

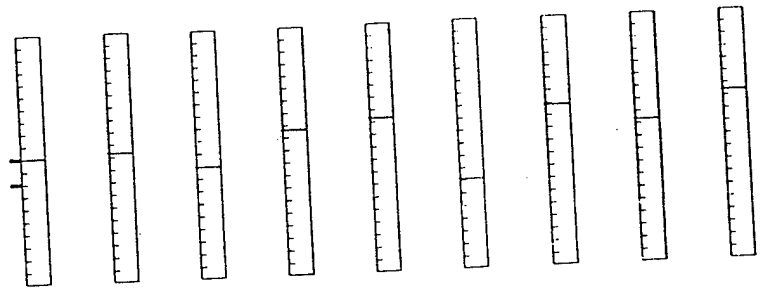

Figure 8. An example of the stimulus array presented to a subject on a trial of the experiment. The values displayed on the nine gauges were determined by sampling from one of two normal distributions: for signal, μ_s=5, and for non-signal μ_n=4 (bottom tic mark=0). The value of the common standard deviation, σ, determined the difficulty of the task.

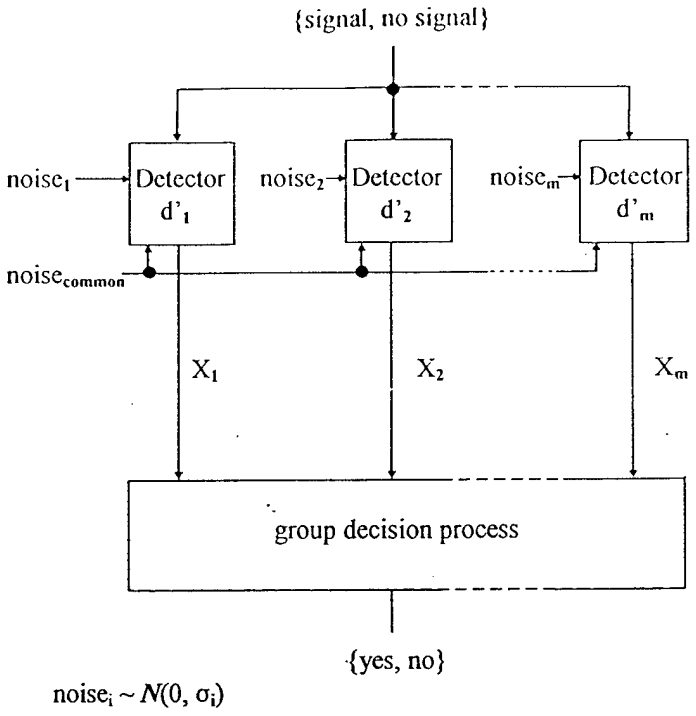

Figure 9. Diagram of a group signal detection system composed of m detectors (after Sorkin and Dai, 1994). Each detector is subjected to two sources of Gaussian noise, one unique and one common to the other detectors (see text).

A. GROUP SIGNAL DETECTION THEORY

A benefit of applying signal detection theory to a decision task is that it enables the experimenter to compute, from the obtained (group or individual) data, separate indices of performance *accuracy*, d', and *bias* (or criterion), c. The accuracy measure, d', which is expressed in standard deviate units, can vary between 0, for a chance level of performance, to approximately 4, for errorless performance. The criterion measure, c, is expressed in similar units; c equal to zero indicates that there is no preference toward a signal or non-signal response, and a positive value of c indicates that there is a preference for the "non-signal" response (Macmillan and Creelman, 1991). We used these measures to describe both individual and group performance in our experiments.

Figure 9 shows the general group signal detection paradigm. There are m group members. Each member has an index of detection accuracy, d'_i, and is subject to two sources of variance or noise: a source unique to that member, and a source common to all members. In a specific

decision situation, the members' judgments might be expressed as binary responses (yes, no), ratings of estimated signal likelihood, or other information. These member judgments are then combined in some manner in order to arrive at a group decision. This aggregation process might include the exchange of information among the members about member likelihood estimates, confidence, biases, etc.

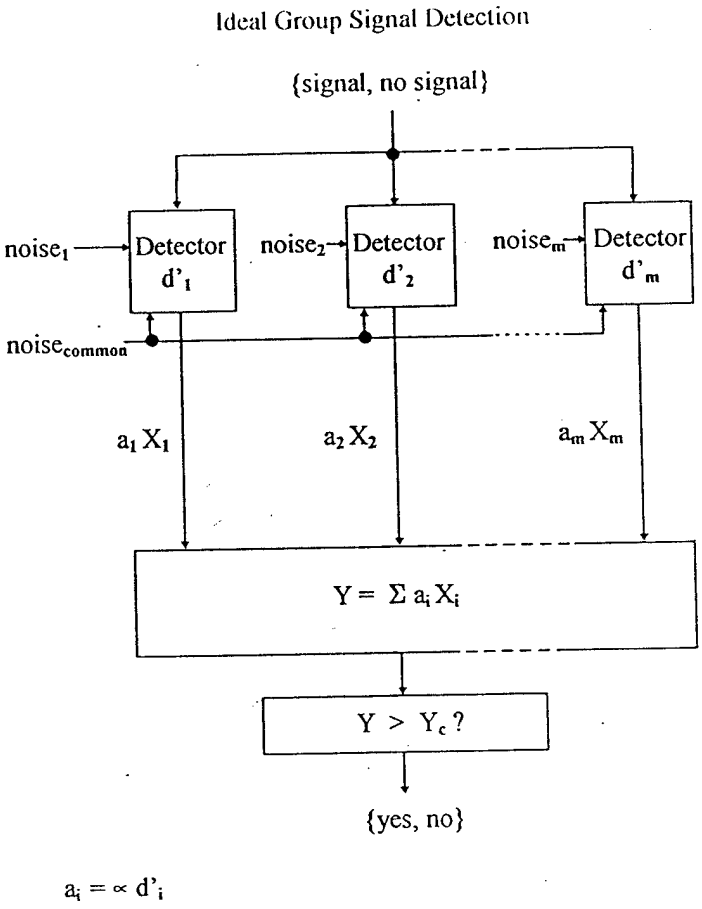

Figure 10. Diagram of the statistically optimal group signal detection system (after Sorkin and Dai, 1994). Each of the m detectors is subjected to two sources of Gaussian noise, one unique and one common to the other detectors (see text). The decision variable, Z, is formed from the weighted sum of the detector estimates.

1. Ideal Group

Another important benefit of detection theory is that it enables us to specify the behavior of the statistically optimal or ideal detection system (Green and Swets, 1966; Tanner and Birdsall, 1958). By definition, an ideal detection system employs an optimal decision rule (one based on a likelihood-ratio statistic) and suffers from no additional sources of noise or error. On average, an ideal detection system will produce the most accurate detection performance. The ideal analysis informs us about important task variables and provides us with an upper bound on the possible human performance.

In order to specify the performance of the ideal group, Sorkin and Dai (1994) extended Durlach, Braida, & Ito 's (1986) model of multiple channel auditory detection to the group detection problem. Figure 10 shows how our general group signal detection paradigm is modified to arrive at Sorkin and Dai's ideal detection system. On each trial, the array of m detectors is presented either with a signal input or a non-signal input, and the system must decide which was presented. On a signal trial, each detector receives an input, m_i, and on a non-signal trial, each detector receives a zero input. The task is made difficult by the presence of two Gaussian, zero-mean noise sources at each detector. The variance of these noise sources is, respectively, σ^2_{com}, which is the variance of a noise source that is common to all the detectors, and σ^2_i, which is the variance of a noise source that is unique to each detector. Although the unique noise source at each detector is independent of the noise at any other detector, the magnitude of the unique sources is constant across the array of detectors and is equal to σ^2_{ind}; that is: $\sigma^2_1=\sigma^2_2= \ldots =\sigma^2_{ind}$.

The output of each detector, X_i, will be normally distributed with a mean of μ_i or zero (respectively, depending on whether the trial was a signal or non-signal trial), and with a variance equal to the sum of the common and unique noise variances. The index of detection sensitivity, d'_i, for an individual detector is the difference between the means of the input on signal and non-signal trials, divided by the square root of the total noise variance,

$$d'_i = \mu_i / (\sigma^2_{com} + \sigma^2_{ind})^{1/2} \qquad (1)$$

By definition, the correlation between any pair of detectors is,

$$\rho = \sigma^2_{com} / (\sigma^2_{com} + \sigma^2_{ind})^{1/2} \qquad (2)$$

We simplify further by normalizing the total variance:

$$\sigma^2_{com} + \sigma^2_{ind} = 1 \qquad (3)$$

then,

$$d_i' = \mu_i \tag{4a}$$

$$\sigma^2_{com} = \rho \tag{4b}$$

$$\sigma^2_{ind} = 1 - \rho \tag{4c}$$

How should the detector outputs be combined to make the signal/non-signal decision? A decision statistic that is equivalent to a likelihood ratio statistic can be formed by linearly summing the weighted estimates of the individual detectors (see, e.g, Ashby and Maddox, 1992; Berg and Green, 1990; Green, 1992; Durlach *et al.*, 1986; Sorkin and Dai, 1994). The group decision statistic is

$$Z = \sum_{i=1}^{m} \hat{a}_i X_i \tag{5}$$

where the $\{\hat{a}_i\}$ are optimal decision weights applied to the estimates of the individual detectors. To arrive at the group response on a trial, the aggregate judgment, Z, is compared to a criterion value, Z_C. When $Z >= Z_C$, the group response is "signal", and when $Z < Z_C$, the response is "non-signal." The optimal weights $\{\hat{a}_i\}$ are specified (Durlach *et al.*, 1986; Sorkin and Dai, 1994) by:

$$\hat{a}_i = [1+\rho\,(m-1)]\,d_i' - \rho\, m\bar{d}' \tag{6}$$

Equation 6 shows that the optimal weights are proportional to the individual indices of detectability. The estimates of detectors having high d's should be afforded higher weights than detectors having small d's. Using the optimal weights yields the ideal performance (Sorkin and Dai, 1994):

$$d'_{ideal} = \left[\frac{m\mathrm{Var}(d_i')}{1-\rho} + \frac{m\bar{d}'^2}{1+\rho(m-1)} \right]^{1/2} \tag{7}$$

When the correlation is zero, equation 7 reduces to the familiar expression (Green and Swets, 1966):

$$d'_{ideal} = \left[\sum_{i=1}^{m}\left(d'_i\right)^2\right]^{1/2} \qquad (8)$$

Equation 7 specifies the performance of the ideal group; this is the maximum performance to be expected from a group of m detectors having a specified mean, variance, and correlation. The equation also suggests what to expect from groups whose performance is similar to, but less than the ideal's: (1) Group performance will increase when m increases; (2) performance will increase as $\sqrt{m}$ when $\rho=0$; (3) performance will increase when the variance in member ability increases; and (4) much of the advantage of group size will be lost when $\rho>>0.25$. The top curve of figure 11 shows ideal group performance as a function of group size, for a group of detectors with the parameters: $\rho=0$, mean d'=.69, and var(d')=0. This curve constitutes an upper bound on the performance of any group having these member statistics.

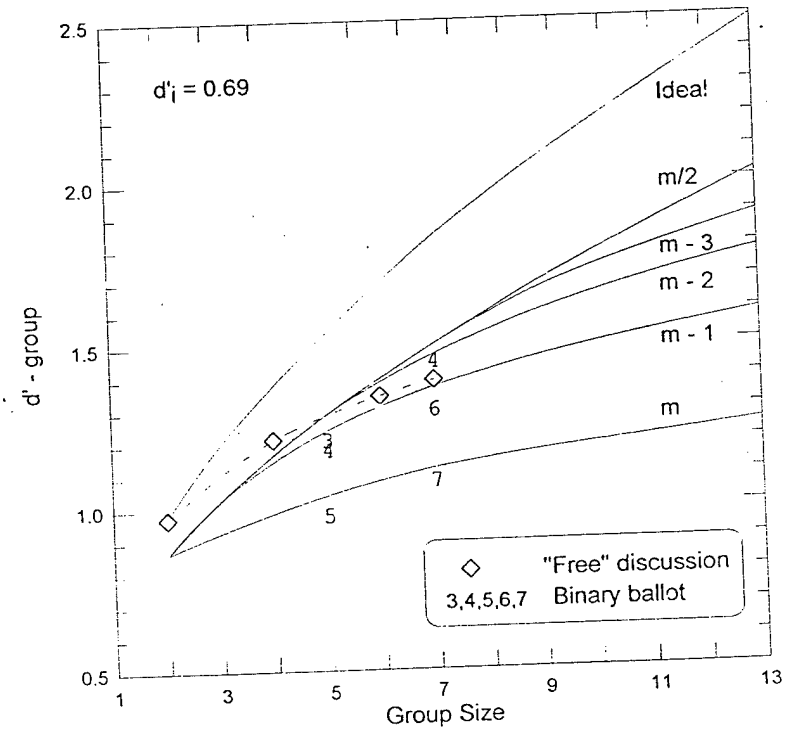

Figure 11. Performance of Ideal, Human and Condorcet Groups as a function of group size, m. The upper curve is the curve for the Ideal Group and the lower set of curves is for Condorcet Groups employing majority rules of m, m-1, m-2, m-3, and m/2 (see text). The numerical symbols are the data (scaled to d'=0.69) from experiment 1. The diamond symbols are the data from the DSNR=1, $\rho=0$ conditions of experiment 2.

What are the consequences of using non-optimal weights? If uniform weights are employed (i.e., if $a_i = 1/m$), performance is given by the right hand term of equation 7,

$$d'_{uniform} = (\bar{d}\sqrt{m}) / \sqrt{1+\rho(m-1)} \qquad (9)$$

Equation 9 provides the performance of ideal *Delphi* groups. The members of a Delphi group maintain their anonymity during deliberations. That is, they employ a uniform weighting strategy in spite of differences in the d' of individual team members. Delphi groups often have been used in the military and in industrial situations.

The performance of a group using an arbitrary set of weights, $\{a_i\}$, is given by d'_{weight}, where[3]:

$$d'_{weight} = \frac{\sum_{i=1}^{m} a_i d'_i}{\sqrt{(1-\rho)\sum_{i=1}^{m} a_i^2 + \rho(\sum_{i=1}^{m} a_i)^2}} \qquad (10)$$

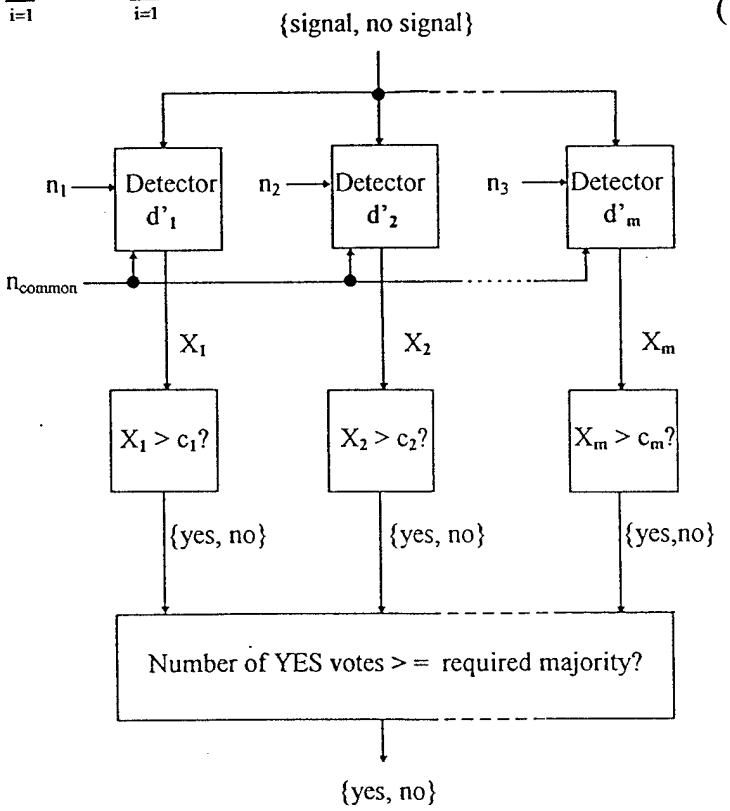

Figure 12. Block diagram of normative Condorcet signal detection system composed of group of m members (see text).

2. Condorcet Groups

A group whose decision is determined by a single binary ballot of the members is known as a Condorcet group, after the French mathematician, the Marquis de Condorcet (1785). Condorcet characterized the expertise of a group member by the single parameter, p_i, the probability that the member will vote for the 'correct' alternative. Condorcet's jury theorem states that if the (mean) p_i for the group is greater than 0.5, the probability of a correct majority vote will increase rapidly toward unity as the group size increases to infinity. Condorcet's theorem has been extended by researchers in many fields, such as economics, political science, electrical engineering, and psychology (Austin-Smith and Banks, 1996; Grofman, Feld, and Owen, 1984; Karotkin and Paroush, 1994; Miller, 1986; Pete, Pattipati, and Kleinman, 1993a, 1993b; Sorkin and Dai, 1994). One goal has been to define the best decision rule for obtaining a group decision, given a set of member competencies. For example, when the outcome matrix in the task is symmetric (there is an equal payoff for correctly choosing either alternative, an equal penalty for incorrectly choosing either alternative, and equal prior probabilities of each alternative), the optimal decision rule is a weighted majority rule, where individual members' votes are weighted by the logarithm of the odds of their making the correct choice (Nitzan and Paroush, 1982, 1984a,b,c; Grofman, Owen and Feld, 1983, Shapley and Grofman, 1984).

Contrary to the traditional analyses, the signal detection analysis recognizes that a group member's behavior cannot be characterized by the single competency parameter, p_i. In a signal detection framework each member's behavior is summarized by two parameters: the sensitivity, d'_i, and the response criterion, c_i. In an experiment these may be calculated from the obtained hit rate (the probability of responding yes given that the signal occurred), and false alarm rate (the probability of responding yes given that the signal did not occur). (See Green and Swets, 1966; Macmillan and Creelman, 1991; Swets and Pickett, 1982.)

Figure 12 shows a schematic array of a normative Condorcet group. As in the previous group models, the group is composed of m members, and each member is characterized by a detection sensitivity, d'_i. In this model, each member also has a response criterion, c_i. As before, each member is subject to two sources of noise, a unique source, and a common source. For simplicity, we now assume that the member's likelihood estimates are uncorrelated, i.e., $noise_{common} = 0$, but relaxing this assumption does not change the general results. Each member observes the stimulus input and then estimates the likelihood that the input on that trial was caused by a signal event. This estimate is compared to the member's response criterion, c_i. If the estimate exceeds c_i, the member votes yes. Only one ballot is taken and the group decision is determined by application of the majority rule, r, to the binary votes of the members. The specific majority rule is assumed to have no effect on the set of detection sensitivities, $\{d'_i\}$, or response criteria, $\{c_i\}$, of the individual members.

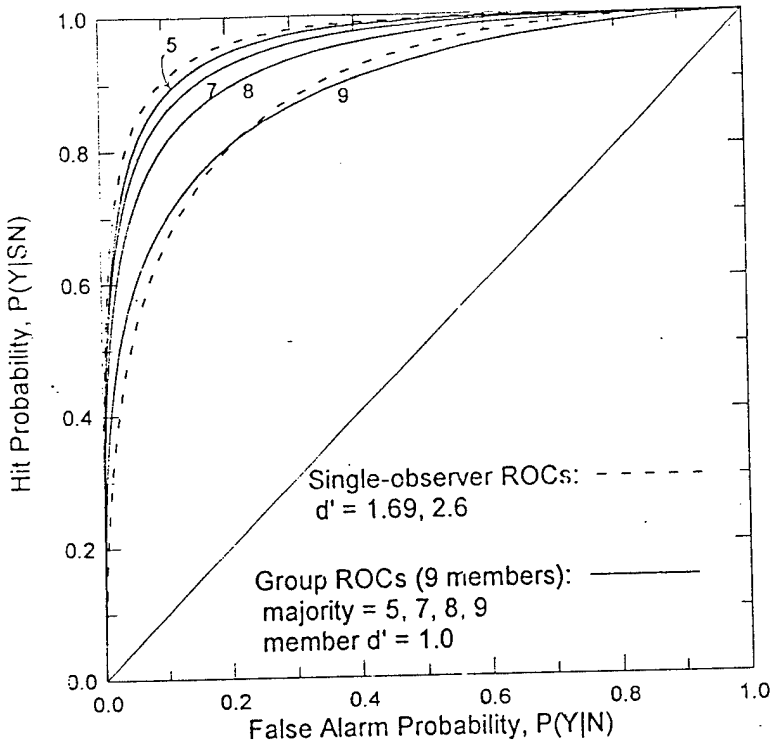

Figure 13 Receiver operating characteristics for single-observers and for 9-person group using majority rules of 5, 7, 8, and 9 yes votes.

a. Condorcet Receiver Operating Characteristic

An individual or a group's detection behavior over a set of trials can be summarized as a hit rate and a false alarm rate. This outcome is a point on a plot of hit rate versus false alarm rate as shown in figure 13. Signal detection models predict that as an observer shifts her response criterion from a very conservative to a very liberal bias, the resulting set of hit and false alarm rates will fall on a curve that increases monotonically from point (0,0) to point (1,1). This curve is called the observer's receiver operating characteristic (ROC). The two dashed lines in figure 13 show ROC curves for a single observer, generated under the usual assumptions of equal variance, normally distributed signal and non-signal hypothesis distributions, for d's of 1.69 (lower) and 2.6 (upper curve).

We wish to calculate the effect of a specific majority decision rule on a Condorcet group's performance. Suppose that we hold constant all the factors that affect the members' detection

sensitivities (task difficulty, signal level, etc.), but we cause the members to adopt new set of response criteria, $\{c_i\}$, more conservative and/or more liberal than the original set. The resulting group hit and false alarm rates would define the locus of ROC points obtainable at a constant member sensitivity and under a specific majority rule, i.e., the group ROC for that majority rule.

To assess the behavior of our normative Condorcet groups, we specified a fixed signal-to-noise level for the stimulus and assumed a set of member d'_i values. We then specified the majority rule, and calculated the probabilities that the group would vote for signal when there was a signal and the probability that the group would vote for signal when there was no signal. This calculation is straightforward when all the members have the same d' and the same c. Then we picked a new c value that was more liberal or more conservative that the initial value and repeated the calculation. Figure 13 shows the ROCs for a group of 9 members that resulted from using majority decision rules of 5, 7, 8, and 9. It is clear that the stricter majority rules lower the overall ROC and also slightly flatten the liberal portion (upper half) of the ROC.

To summarize the performance accuracy of a group's behavior, we used the equivalent normal-normal detectability, d'_e, derived from the area under the obtained ROC curve (Green and Swets, 1966; Macmillan and Creelman, 1991; Swets and Pickett, 1982; see also Swets, 1986a,b; Swets, 1988). The area under the ROC curve is equivalent to the percent correct in a two-alternative forced-choice version of the detection task. Other accuracy measures that we examined gave the same results because the group ROC curve closely resembles the normal-normal, single-observer ROC curve, as can be seen in figure 13. We shall refer to the area-derived-d'_e measure of group accuracy as d'_{group}.

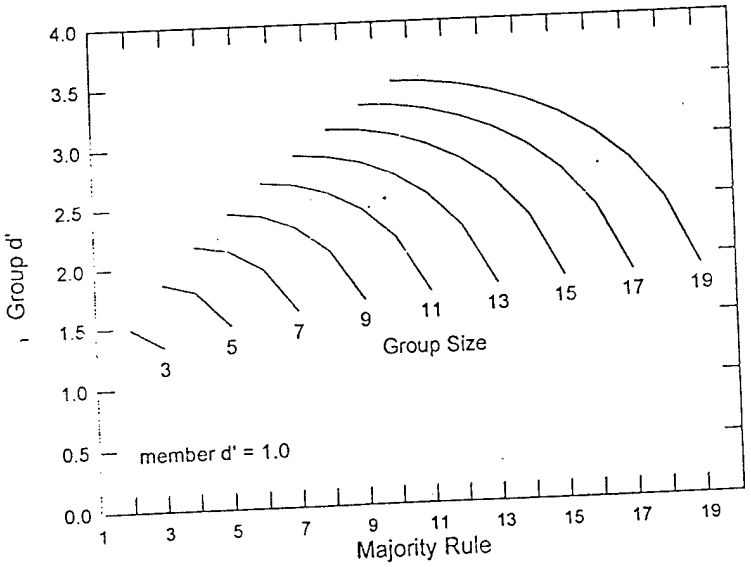

Figure 14. Group performance (d'_{group}) as a function of the majority rule for groups of different size.

Figure 14 shows the effect of different majority rules on the measure of group accuracy, d'_{group}, for a group whose member d'=1 . The performance level of different size groups is plotted as a function of the majority rule; the y-axis shows d'_{group} and the x-axis shows the number of votes required for a group signal decision. The parameter of the curves is the group size. Group detection accuracy decreases as the required majority becomes more stringent. For example, d'_{group} for a 9-person group using a simple majority rule (x = 5) is approximately 2.5. When that group uses a decision rule requiring 8 or 9 votes, d'_{group} drops to values of 2.0 and 1.7, respectively. Performance accuracy, d'_{group}, has the units of standard deviates, so a drop of 1 or even ½ is potentially important. For this 9-person group, there is about a 10% drop in percent correct from the simple majority to the unanimity cases.

When d' and c are not constant over the group's members, the computations of d'_{group} are more complicated but the result is the same. We ran Monte Carlo simulations of cases when the variance of the members' d' or c was non-zero: First, we generated sets of member sensitivities $\{d_i\}$ and criteria $\{c_i\}$ by assuming normally distributed distributions of d' and c and by specifying the set statistics: $\mu_{d'}$, $\sigma_{d'}$, μ_c, and σ_c. Then, we caused the group to shift to a *new* group response criterion, either more liberal or more conservative than the previous value by generating a new set of member criteria using the same σ_c as before, but with a new μ_c value. We repeated the assessment of group hit and false alarm rate, generating a range of hit and false alarm rate pairs for the cases when $\sigma_{d'}$ and σ_c exceeded 0.5. The group d' values that resulted from these ROCs were very close to those obtained in the zero variance cases.

Another way to view the normative results is to consider how group performance increases with group size for different decision rules. This is shown in the lower curves of figure 11. As the majority rule becomes more stringent, the performance advantage of having larger sized groups decreases.

3. Measures of Group Efficiency

It is useful to have a summary measure of how much the observed performance of a 'real' group of human observers differs from a hypothetical reference group such as the ideal group or the Condorcet group. In a hypothetical group situation, such as faced by the Condorcet, non-interacting, unanimous rule group, the detection theory analysis defines the highest level of performance possible for a real group of observers to attain in this situation. The degree to which the performance of the real group is less than this optimal level may be quantified the efficiency measure, η, (Tanner and Birdsall, 1958), where

$$\eta = (d'_{observed})^2/(d'_{ideal})^2 \qquad (11)$$

Efficiency is defined as the ratio of squared d's, because in many sensory situations $(d')^2$ is proportional to signal energy. In such cases, the efficiency calculation yields an energy ratio. An efficiency of 0.60 would mean that performance equal to the human detector's would be provided by an ideal detector that used 60% of the signal energy. Of course the definition of efficiency

depends on appropriate definition of the "ideal" comparison group, i.e., the particular constraints of the group detection situation.

4. Performance of Human Groups

Given the analyses of Condorcet and Ideal Groups, we can postulate five reasons why a group of human subjects might fail to reach the performance level of the ideal group (see Table 2). First, the judgments of individual subjects may be partially correlated; this would lead to performance much less than that of the zero-correlation ideal group. Second, the group decision might be made without communication among the members, based only on a ballot of the members' binary votes, as in a Condorcet group. Third, there might be inappropriate weighting of member judgments, as in the case where the judgments of certain members had an undue influence on the group decision. Fourth, additional sources of noise might be present in the members' or group's decision process such as: (a) a limit to the precision of aggregating the members' estimates of signal likelihood, or (b) variability in the scales employed by members in their judgments of signal likelihood. Finally, there might be motivational factors, such as social loafing, that would cause members to decrease their observational efforts as the group size increased. These possibilities were evaluated in a series of experiments with human groups. The first experiment tested hypotheses relating to the behavior of human in non-interacting groups. The second and third experiments attempted to obtain near ideal group performance.

Table 2. Hypotheses about the decrease in efficiency with group size for human groups.

1. Partially correlated judgments ($\rho>0$).
2. Incomplete member-member communication.
    a. Condorcet case - members lack knowledge of other members' judgments, expertise, or criteria.
    b. Delphi case - members lack knowledge of other members' expertise or criteria.
3. Inappropriate weighting of member judgments
    a. Delphi case - uniform weights
    b. Pathological case - the judgments of some members receive excessive or insufficient weight.
4. Additional noise.
    a. Limit to the precision of aggregate judgments.
    b. Variation in members' estimation scales.
5. Motivational factors - members reduce individual detection effort as group size increases.

II. EXPERIMENT 1. PERFORMANCE OF CONDORCET GROUPS

The signal detection Condorcet analysis raised several questions about the performance of groups of human observers in a signal detection task. First, would one see similar large decreases in the performance of groups of human participants as the majority rule is made more stringent? Second, does a more stringent rule produce a change in the individual's behavior? It is clear that a stringent rule will cause the group hit and false alarm rates to decrease, thereby resulting in a more conservative group decision criterion, c_{group}. The question is whether this will have an effect on member detection sensitivity or decision criteria. Specifically, will forcing the group decision to

be more conservative cause the member decisions to be more liberal? Experiment 1 was designed to address these questions by implementing the group detection situation under conditions where the group members could not communicate with each other and where the group decision was automatically determined by the majority rule of the binary votes of the group members.

A. METHOD

Observers performed a graphical signal detection task, either individually or in groups of 5 or 7 members. On each trial, observers were presented with a nine element visual display consisting of analog gauges similar to those shown in figure 8. Following the display, observers had to indicate whether the display indicated a signal or noise condition. The values displayed on the nine gauges were determined by sampling from one of two normal distributions: for signal, μ_s=5, and for non-signal μ_n=4. The value of the common standard deviation, σ, was set equal to 2.25. The detectability of a single element of the display is given by $d'_{\text{single element}} = (\mu_s - \mu_n)/\sigma =$ 0.44. The best detection performance, given the full display of nine independent display elements, would be $\sqrt{9}$ times the element detectability, giving a display signal-to-noise-level (DSNR) of 0.44 $\sqrt{9} =$ 1.33 (Green and Swets, 1966). In experiments using similar graphical materials, Sorkin, Mabry, Weldon, & Elvers (1991) and Montgomery and Sorkin (1996) showed that limitations on human processing, presumably caused by the display's brief duration and other factors, resulted in performance that was about 75 to 80% of the ideal values; this leads to a prediction of d'=1.0 for individual observers in the present experiment.

1. Subjects

University of Florida students, six men and two women, participated in the study. All of the subjects had normal or corrected-to-normal visual acuity. Subjects were paid $4.25 per hour plus an incentive bonus that was based on the accuracy of performance. In the individual conditions, the bonus depended on the accuracy of the individual's performance. In the group conditions, the bonus depended on the accuracy of the group's performance. The bonus averaged approximately $0.75 per person per hour.

2. Apparatus and Stimuli

Stimulus generation and presentation were done with Insight 4086 computers arranged in a local area network. The stimuli were displayed on 14 inch color monitors with the intensity set at approximately 100 cd/m^2 measured from a uniform white field. Subjects sat approximately 610 cm away from the monitor in a quiet, fluorescent lit laboratory room; the nine gauges subtended a visual angle of approximately 8° vertical by 22° horizontal. Responses were made via a standard computer keyboard. During the group phase of the experiment, subjects were seated close to each other but could see only their own monitor.

The display elements consisted of two parallel vertical lines with tick marks on the left line, dividing the gauge into 20 intervals (10 major tics and 10 minor tics). A value of 0 was represented by the tic at the bottom of the gauge, and a value of 10 was represented at the top. Two larger green ticks marked the mean of the noise and signal distributions. On a given trial, each of the gauges displayed, via a red horizontal line the width of the gauge, values that had been independently drawn from the same distribution. Half of the trials (randomly) were drawn from the signal, and half from the noise distribution. Stimulus duration was 167 ms. A white mask (200 ms) followed presentation of the stimulus. The majority required for a group 'yes' response was constant for a block of trials and was indicated before each trial. After all of the members had made their response, each member was provided feedback about the number of signal votes cast, the group decision, the trial outcome (correct or incorrect), and the cumulative number correct.

3. Procedure

Each subject had extensive experience in individual sessions of the task (at least 700 trials) before running in the group sessions. A session took approximately 1.5 hours and subjects were encouraged to take rest breaks after each block of 100 trials. After the individual session, subjects ran two sessions each in groups of either 5 or 7 members. Subject assignment to groups depended on schedule availability; 6 people were common to the two 7-member groups, and 4 people were common to the two 5-person groups. A trial block consisted of 100 trials at a fixed majority rule. The order of majority rules in the 5 person sessions were 5-3-4-5-3-4 and 3-5-3-5. The order in the 7-person groups were 4-7-6-4-7-6 and 4-6-7-7-4. Subjects were instructed not to talk during the experimental trials.

B. RESULTS

Figures 15 through 18 summarize group and individual performance as a function of the majority rule employed. Figures 15 and 16 are plots of the group detection index, d'_{group}, and the average detection index for the individual members, mean-d'_i, as a function of the majority rule imposed. Figure 15 shows the performance of two groups of 7 members and figure 16 shows the performance of two groups of 5 members. In all cases, there was a large drop in obtained d'_{group} as the majority rule increased from a simple majority to unaninity. The average drop in d'_{group} from the simple majority to the unanimous conditions was greater than 0.7 d' units. The worst-case standard error in d' for these conditions is approximately 0.2 d' units (200 trials). There were no significant differences in d'_i attributable to either the participant or the majority rule condition.

Figures 17 and 18 are plots of the group criterion, c_{group}, and the average criterion of individual members, mean-c_i, as a function of the majority rule condition. Figure 17 shows the criteria for two groups of 7 members and figure 18 shows the criteria for two groups of 5 members (absent one condition not run). As required by the majority rule manipulation, stricter majority rules produced smaller group hit and false alarm rates, and therefore large increases in c_{group}. We were interested in whether stricter rules also resulted in group members shifting to more

liberal individual criteria. The average c_i data indicated a tendency in this direction. Analyses of variance of the criterion data for the 5 and 7 person groups indicated significant effects of participant [$F(6, 38) = 4.84$, $p < 0.001$; $F(4, 40) = 4.09$, $p < 0.01$] but marginal effects of majority rule. A significant effect of majority rule was indicated for the 7-person groups [$F(2, 38) = 8.37$, $p<0.001$, but a marginal effect was indicated for the 5-person groups $F(2, 40) = 2.78$, $p < 0.10$]. An examination of the individual subject criteria indicated that three of the group members consistently shifted their criteria in a more liberal direction with stricter rules, while the other members shifted their criteria smaller amounts in either direction. After the experiment one of three consistent shifters remarked, "I get it; the trick is to respond yes more often when the majority rule is unanimous," or words to that effect.

How does the performance of these groups compare to the normative prediction? We used the mean and standard deviation of the obtained d'_i at each majority rule condition as the values for our Monte Carlo simulations of the normative Condorcet group. The resulting d'_{group} values are plotted as the circle symbols (and connected solid lines) in figure 15 and 16. These values provide a good fit to the performance of the human groups.

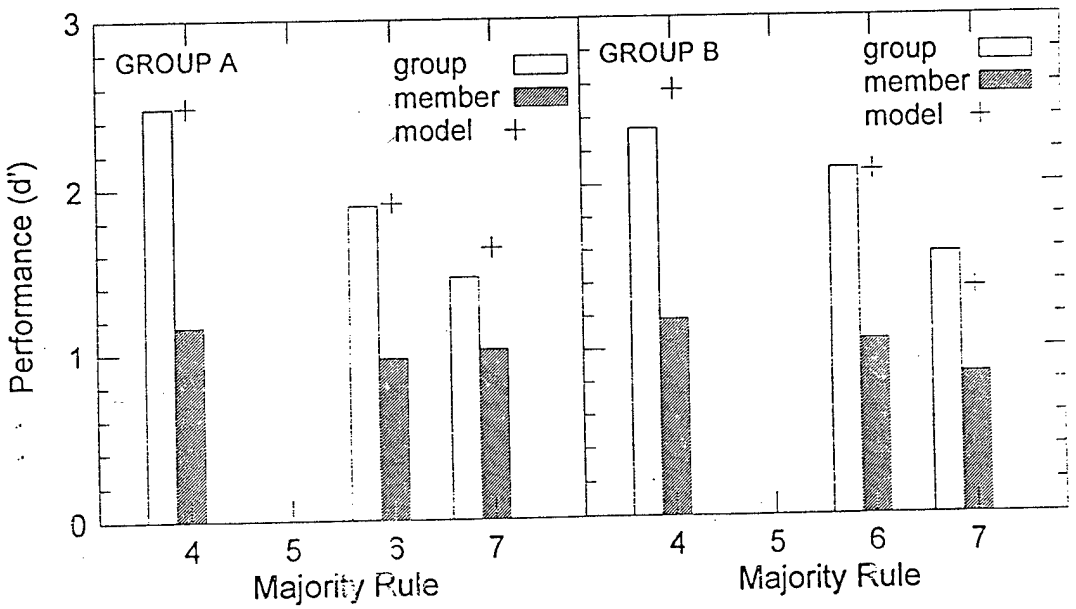

Figure 15. Group (open bars) and individual (shaded bars) performance (d') as a function of the majority rule for the 7-person groups. The plus symbols are the prediction of the normative model (see text).

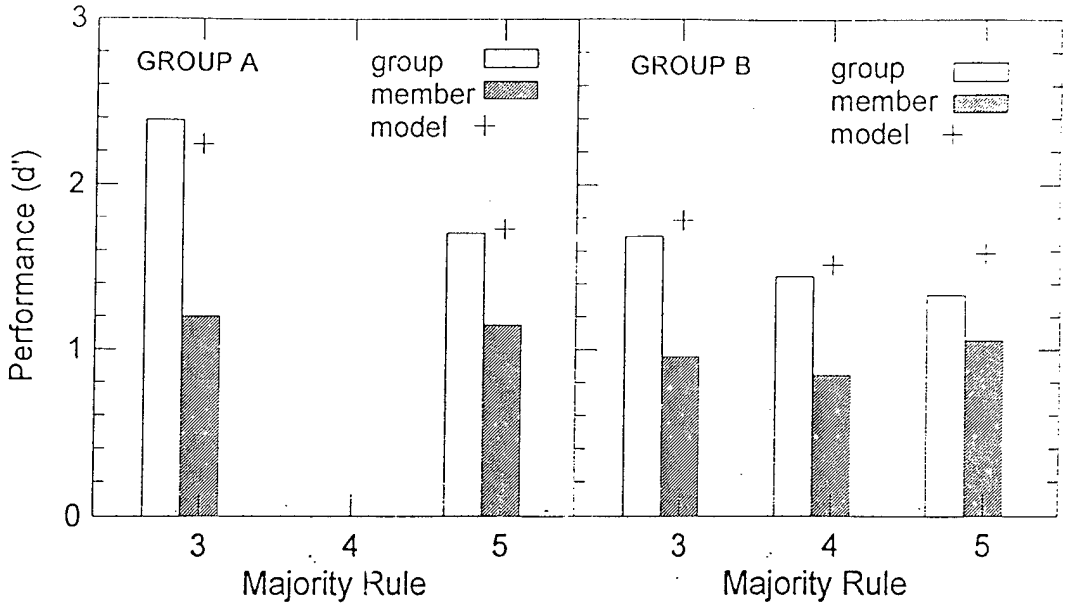

Figure 16. Group (open bars) and individual (shaded bars) performance (d') as a function of the majority rule for the 5-person groups. The plus symbols are the prediction of the normative model (see text).

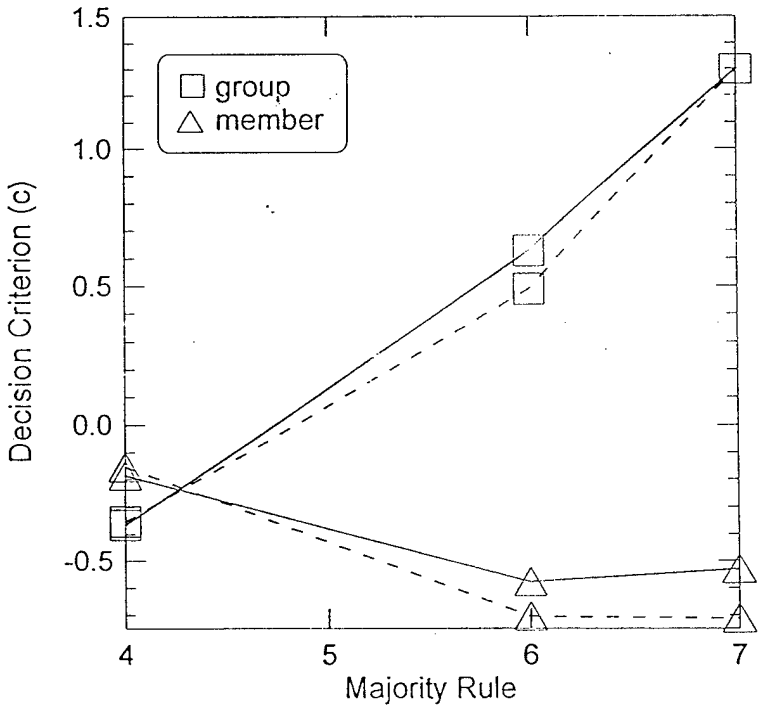

Figure 17. Group and individual criteria (c) as a function of the majority rule for the 7-person groups. The solid lines and symbols show the group results and the open lines and symbols show the average of the individual values.

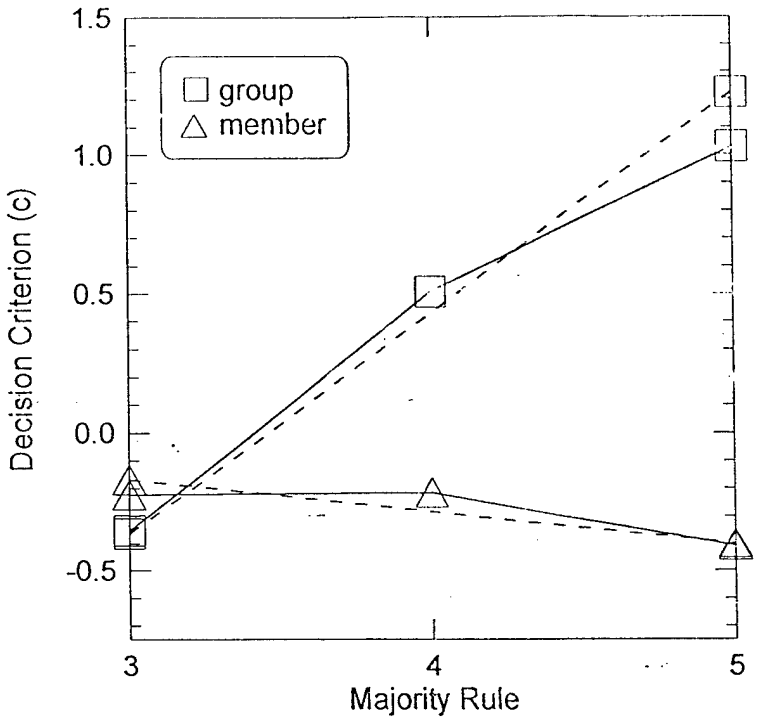

Figure 18. Group and individual criteria (c) as a function of the majority rule for the 5-person groups. The solid lines and symbols show the group results and the open lines and symbols show the average of the individual values.

C. DISCUSSION

Figure 11 showed the large differences in performance between the normative ideal group and the Condorcet simple majority group. These performance differences are due to the use of binary rather than continuous estimates of signal likelihood and to the weighting of member judgments without regard to individual expertise. The present analysis shows that there can be even further losses in performance accuracy when decision rules other than the simple majority are employed. For the purposes of comparison, we have scaled the group data shown in figures 15 and 16 to allow plotting on figure 11. The scaled data are plotted as numbered symbols that show the majority rule condition run. The data show that the Condorcet model provides an approximate description of the effect of majority rule in this condition.

We can conclued that, provided the detection sensitivities of a group's membership are constant, a more stringent majority rule will produce a decrease in the accuracy of the group's

performance. That is, the performance of a normative Condorcet group will decrease as the majority rule is changed from a simple majority to a two-thirds majority to a three-quarters majority to unanimity. When a stricter majority rule is imposed, group members can adjust their individual criteria to try to compensate for the more conservative group response criterion--but they cannot compensate for the loss in group performance accuracy that is produced by the strict rule. Furthermore, the loss in accuracy that is produced by a more stringent rule can be very large.

Given the normative result, it is not surprising that the performance of the human groups that we tested decreased with more stringent majority rules. In fact, the only way that a group of human participants could avoid having a decrease in their group performance with a more stringent majority rule would be if the participants *increased* their individual detection sensitivities under a more stringent rule. This would be a highly non-intuitive outcome, and in fact we found no evidence for changes in individual detection sensitivity with changes in majority rule. On the other hand, it does seem reasonable to expect participants to try to compensate for the bias shift imposed by a more stringent rule. We expected that, under stricter rules, members would shift toward more liberal individual criteria. The data favored this hypothesis but reached statistical significance only for the 7 person groups. A few subjects appeared to have adopted this strategy, but most did not vary their criteria with majority rule in any consistent way.

Can the performance of a voting group be improved by providing additional information to the members? We can imagine groups whose members, prior to voting, supply each other with information that allows each member to estimate the detection competence and response criterion (or graded likelihood estimate) of every other member. Certain formal constraints, such as those imposed on juries during important trials, may act to encourage such interactions. We would not expect to see the deleterious effect of strict majority rules on the performance of such groups. For example, if a more conservative rule produced an increase in the communication among group members, that rule would result in more accurate performance. Our simulations of such situations indicate that stricter majority rules increase the accuracy of group performance.

If a group must operate under Condorcet constraints on member interaction, is the simple majority always the preferred majority rule? We can say that it is usually best to operate on the receiver operating curve that is characterized by the highest d'. Therefore, if a conservative criterion is desired, it would be preferable to encourage the individual members to adopt more conservative response *criteria*, rather than to use a stricter majority rule. Such a strategy would shift the group's operating point to a new point on the simple-majority ROC, rather than to a point on the lower, strict-rule ROC. Finally, we remind the reader that our analysis applies only to cases when there is no sharing of information among members because allowing such communication might diminish or even reverse these majority rule effects. However, useful information sharing among members may not be feasible when the groups are very large. In those cases, the best rule probably is the simple majority rule.

III. EXPERIMENT 2. INTERACTING GROUPS

The constraints on observer inter-communication in experiment 1 produced a large effect on group performance. Suppose that these observers had been allowed to communicate freely about their detection judgments and had not been constrained to a specific majority decision rule? How much would performance have been improved? From figure 11, it is clear that there is a wide area of possible performance between that obtained with simple majority rules and that possible at an ideal level. Experiments 2 and 3 were designed to test whether the performance of human groups could be increased to levels approaching the ideal prediction. In addition, in order to test the applicability of equation 7, we examined the effects on performance of varying the inter-member correlation and the d' of individual members.

A. METHOD

Subjects again performed a graphical signal detection task, either individually or in groups of from 2 to 7. The values displayed on the nine gauges were determined by sampling from one of two Normal distributions: for signal, μ_s=5, and for non-signal μ_n=4. The value of the common standard deviation, σ, was set equal to 1.5, 2, 2.5 or 3 in different conditions of the experiment. Display element standard deviation values of 3, 2.5, 2, and 1.5, yielding DSNR conditions of 1, 1.2, 1.5, and 2, respectively, were used.

1. Subjects

Eight University of Florida students, seven women and one man, participated in the study. All of the subjects had normal or corrected-to-normal visual acuity. Subjects were paid $4.25 per hour plus a small incentive bonus that was based on the accuracy of performance. In the individual conditions, the bonus depended on the accuracy of the individual's performance. In the group conditions, the bonus depended on the accuracy of the group's performance. The bonus averaged approximately $0.40 per person per hour.

2. Apparatus and Stimuli

Stimulus generation and presentation were the same as in experiment 1 except that stimulus duration was 370 ms during practice sessions and 320 ms during all experimental trials and a centered cross (0.5") fixation stimulus (200 ms) preceded the stimulus, and a white masking screen (200 ms) followed presentation of the stimulus.

3. Procedure

Each subject first was tested alone in the individual detection sessions. The individual sessions were run before any group conditions, and then rerun again after all the group conditions. A trial block consisted of 125 trials at a given DSNR level. An experimental session consisted of 16 blocks, presented in sequences of flour blocks at a given DSNR. The DSNR levels were randomized both within and across sessions. After two practice sessions, subjects performed the task for 2000 trials at each of the four DSNR conditions. A session took approximately 1.5 hours and subjects were encouraged to take rest breaks after each block.

After the individual session, subjects were run in groups of 2, 4, 6, and 7 members. Subjects were randomly assigned to one group of 8-members, two groups of 4-members, and 2 groups of 2-members. The 2-member groups were randomly chosen from the 4-member groups. However, the male subject was purposely excluded from the 2-member groups, to minimize any chance that his presence would bias that group's decision (Clement & Schiereck, 1973). We had planned to use 8 subjects in the largest group. Due to absenteeism and scheduling problems, the large groups tested consisted of only 6 or 7, rather than 8, members.

In the group sessions, the trial blocks consisted of 100 trials, run in two, 4-block sets, randomized within and across sessions with respect to DSNR and correlation. The trial procedure was the same as for the individual sessions, except for the response sequence. In the individual sessions the subject had up to 1000 ms following the masking screen to make a response. In the group sessions, a 700 ms blank screen was presented after the masking screen . Then, one member of the group was randomly selected to receive a screen message telling her that she was to give the group's answer. There was no time limit for a response. Group members were encouraged to talk about their judgments both during the 700 ms blank period and the period after the group responder had been selected. Following the response, the entire group received feedback about the nature of the trial and the correctness of the response.

Two additional manipulations were made during the group sessions, in addition to changing DSNR. First, the distribution of DSNR was set either the same for all members of the group (the *equal DSNR* condition) or, for some 4-person groups, was intentionally varied so that the task difficulty for two members was twice that for the others (the *unequal DSNR* condition). Second, the stimulus displays were either independent ($r = 0$ condition) for all group members or were partially correlated ($r = 0.25$) between group members. Table 3 summarizes the conditions for the group sessions. (Some of the low difficulty conditions were omitted for the larger groups, because the group performance would have been too high to measure accurately.)

The correlation between group members was manipulated using a method used by Jeffress and Robinson (1962) and Sorkin (1990) in auditory experiments. The method can be understood by considering how the values were generated for element 1 (the left-most element) in subject A and subject B's displays. In the independent condition ($\rho=0$), each of the elements was drawn separately from a separate normal distribution as follows: For subject A, the value of element 1

was equal to x_a; for subject B, element 1 was equal to x_b; where x_a and x_b were normally distributed, independent, zero-mean, equal variance, random variables. However, in the correlated condition (ρ=0.25), the value for each subject's display element was generated as follows: For subject A, the value of element 1 was set equal to $0.87x_a+0.5x_c$, and for subject B, element 1 was set equal to $0.87x_b+0.5x_c$, where x_a, x_b and x_c were independent, normal, zero-mean, equal variance, random variables (i.e., the principle is the same as stated by equation 2). In the correlated condition, the corresponding elements in all pairs of subject displays were generated in a similar fashion.

Table 3. Experimental conditions for the group sessions.

Size m	Correlation r	DSNR	Group Membership	Number of Blocks
2	0	1	S1 S2	12
2	0	1	S5 S6	8
2	0	2	S1 S2	8
2	0	2	S5 S6	8
2	0.25	1	S1 S2	8
2	0.25	1	S5 S6	8
2	0.25	2	S1 S2	8
2	0.25	2	S5 S6	8
4	0	1	S1 S2 S3 S4	8
4	0	1	S5 S6 S7 S8	8
4	0	1.5	S1 S2 S3 S4	8
4	0	1.5	S5 S6 S7 S8	12
4	0.25	1	S1 S2 S3 S4	8
4	0.25	1	S5 S6 S7 S8	8
4	0.25	1.5	S1 S2 S3 S4	8
4	0.25	1.5	S5 S6 S7 S8	11
4U	0	1, 1, 2, 2	S2 S5 S7 S8	8
4U	0	1, 2, 1, 2	S2 S5 S7 S8	4
4U	0	2, 2, 1, 1	S2 S5 S7 S8	4
6	0	1	S1 S3 S5 S6 S7 S8	4
7	0	1	S2 S3 S4 S5 S6 S7 S8	4
6	0	1	S2 S3 S4 S6 S7 S8	4
6	0	1.2	S1 S3 S5 S6 S7 S8	4
7	0	1.2	S2 S3 S4 S5 S6 S7 S8	4
7	0	1.2	S1 S2 S3 S4 S5 S6 S7	4
7	0.25	1	S1 S2 S3 S5 S6 S7 S8	4
7	0.25	1	S1 S2 S3 S4 S5 S6 S7	4
7	0.25	1.2	S2 S3 S4 S5 S6 S7 S8	4
6	0.25	1.2	S2 S3 S4 S6 S7 S8	4

B. RESULTS AND DISCUSSION

We evaluated performance in all conditions of the experiment by calculating detection indices (d') and criterion (c) measures based on the obtained individual and group hit and false alarm rates (Macmillan and Creelman, 1991). The criterion measures were generally near zero,

indicating an absence of response bias, and they did not vary in a consistent way across conditions, subjects, or groups[4]. Therefore, our analysis will consider (a) the individual and group detection indices, and (b) the weighting strategies employed in the group conditions.

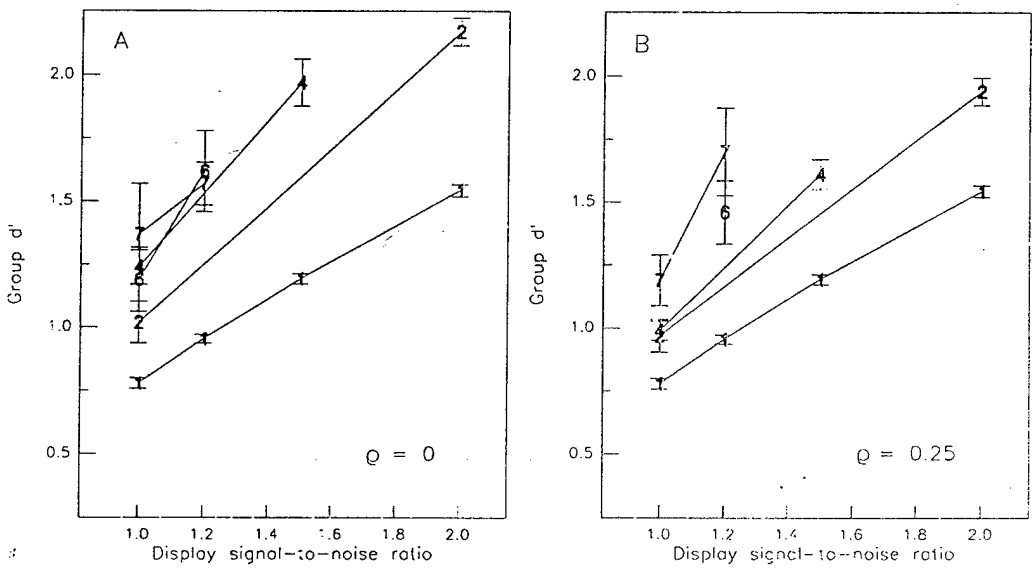

Figure 19. The obtained performance (d') of all groups, plotted as a function of the display signal-to-noise ratio. The left and right panels show, respectively, the data for the ρ=0 and ρ=0.25 correlation condition. The plotted symbols (*1, 2, 4, 6, 7*) indicate the group size. The data for the individual (*1*) condition is repeated in each panel. The brackets indicate plus and minus one standard error of the mean.

The lowest curve on each panel (*I*-symbols) in figure 19 shows the individual detection performance as a function of DSNR. The obtained detection performance was a linear function of DSNR, consistent with the predictions of (single-observer) detection theory. The level of performance was consistent with previous results using a similar task (Sorkin et al, 1991; Montgomery and Sorkin, in press). Average individual detectability at a DNSR equal to 1, was 0.77 with a standard deviation of 0.12. There was no significant difference between individual performance in the test (pre-group sessions) and retest (post-group sessions) conditions.

The indices of observer performance obtained in the individual conditions can be converted into measures of individual detection efficiency using equation 10, i.e., $h_i=(d'_i/DSNR)^2$. Observer efficiency in the individual sessions averaged 0.61, consistent with previous data. Efficiency was moderately consistent across subjects (the largest standard deviation in h_i across subjects at any DSNR was 0.17) and was highly consistent across display signal-to-noise ratios (the largest standard deviation in h_i across conditions for any subject was 0.1).

Figure 19 also shows the effect of DSNR on the performance of groups of size 2, 4, 6, and 7 (indicated by the plotted symbols *2, 4, 6, 7*). The left and right panels of the figure, respectively, show the results for the uncorrelated and correlated conditions. As in the individual case, performance increased with DSNR (the increase with DSNR for 6 and 7 member groups did not reach statistical significance). Group performance also increased with group size (for the $\rho=0$ condition, $F[4,51]=4.56$, $p<.003$). Consistent with the theory, the performance advantage of size was reduced in the $\rho=0.25$ conditions.

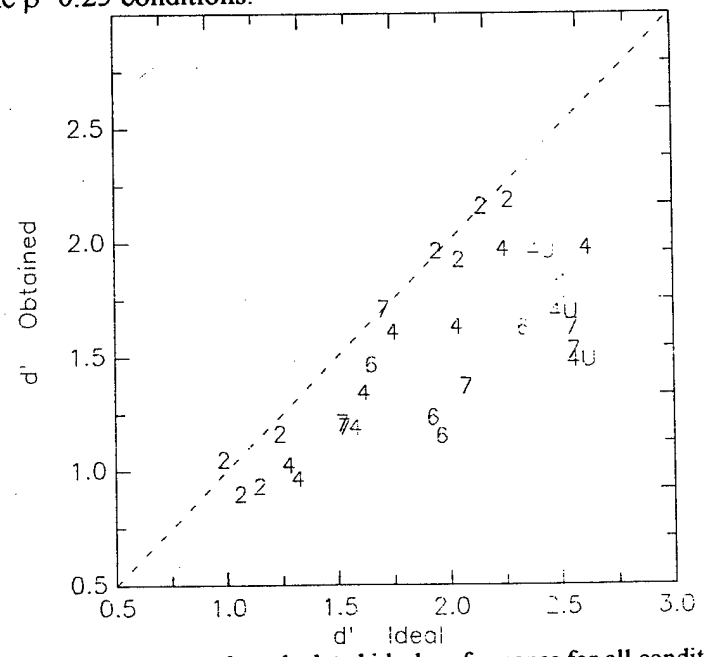

Figure 20. A plot of the obtained versus the calculated ideal performance for all conditions in the experiment. The ideal d' predictions were generated from the data in the individual conditions. The data symbol shows the group size; U indicates the unequal difficulty (DSNR) conditions.

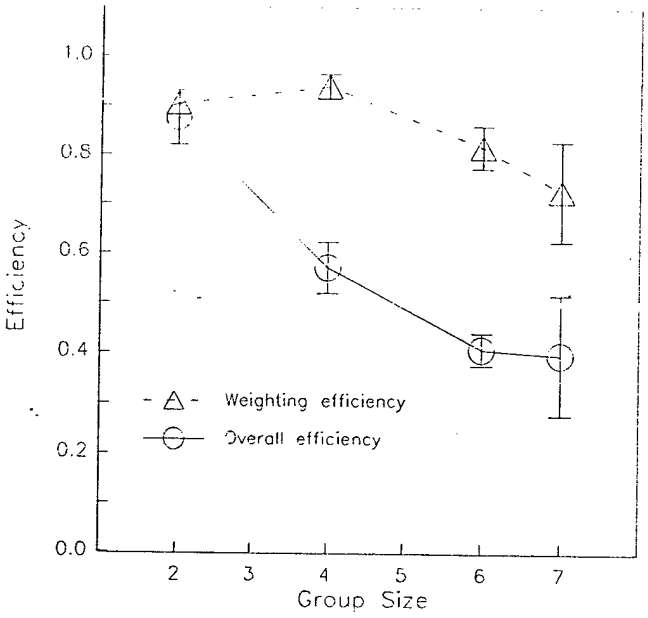

Figure 21. A plot of performance efficiency as a function of group size, for the ρ=0 conditions (circle symbols). The triangle symbols shows the weighting efficiency, the group's ability to weight a member's contribution by the member's expertise.

1. Analysis of Group Efficiency

Theoretical predictions of performance in the group conditions were calculated in the following way. For each group, the d' values obtained in the members' individual sessions were taken to specify the values for those members in the group conditions. For example, in order to predict the performance of the 4-person group that consisted of subjects S1, S2, S5, and S6, the d' values that had been obtained in the individual sessions for these subjects were used to evaluate equation 7. Figure 20 is a plot of the obtained vs. predicted d' values for all the group conditions.

An interesting result is evident in figure 20; the discrepancy between obtained and ideal performance increased with group size. For example, the performance of the 2-person groups was very close to the prediction, while the performance of the 6 and 7 person groups was generally much less than the predicted ideal level. The points marked with a U symbol, indicate the 4-

person, *unequal difficulty*. groups in which we had arranged that two subjects had high and two subjects had low DSNRs. The unequal difficulty conditions are considered further in the section on decision weights.

The obtained and ideal d' values shown in figure 20 were used to compute the efficiency, η, of group performance. As in the individual conditions, efficiency in the group conditions was highly consistent across different DSNR levels and somewhat consistent across groups of similar size. Figure 21 shows (circle symbols) the efficiency of group performance as a function of group size, for the ρ=0 conditions. The figure shows more clearly the result suggested by the previous figure; efficiency started out a very high level, 90% for the 2-person groups, but fell rapidly as group size was increased. This result also can be seen in figure 11. The diamond symbols show the data for the DSNR=1, ρ=0 conditions of experiment 2. Group performance increases with group size but at a slower rate than the ideal model. In fact, the function appears to fall into the Condorcet region when m=5 members. In the next section, we consider the possibility that the decrease in efficiency was due to the use of a non-optimal weighting strategy.

a. Analysis of Decision Weights

A correlational technique developed by Lutfi (1995) and Richards and Zhu (1994) was used to determine the relative weights assigned to each member during the response deliberation process. This technique is based on Berg's (1989,1990) Conditional on Single-Stimulus (COSS) analysis of decision weights in individual sensory tasks. Sorkin *et al.* (1991) employed the COSS technique in a study of visual displays using stimuli similar to those in the present experiment. The basic idea of the correlational analysis is to compute the point bi-serial correlation between the stimulus presented to an observer and the group's response, over trials. This correlation (scaled by the variance of the stimulus data, and using partial correlations when r>0) provides a measure of the relative impact of that observer's stimulus on the group's decision. In previous applications of this method, the goal was to assess the relative influence of *different components* of the stimulus on the response of a *single observer*. In the present experiment, we were interested in assessing the relative influence of *different observers* on the response of the *group*[5].

Figures 22 through 25 show the weights that were obtained from the correlational analysis. Figures 22 and 23 show weights obtained in representative equal DSNR conditions. In all the figures, the shaded bars show the ideal weights and the open bars show the obtained weights (the average of the weights computed separately on signal and non-signal trials; the individual data for these trials are shown as the plotted *S* and *N* symbols). In some of the equal DSNR conditions, the magnitude of the obtained decision weights had approximately the same ordering as the ideal weights. That is, the largest weights were given to the most sensitive observers. This is evident to some extent in figure 22a (6-person, ρ=0, DSNR=1.2) and 22b (7-person, ρ=0.25, DSNR=1) and to a greater extent in figure 23a (4-person, ρ=0, DSNR= 1) and 23b (4-person, ρ=0.25, DSNR= 1.5). In other conditions, such as figures 22c (7-person, ρ=0.25, DSNR=1.2), 7d (7-person, ρ=0, DSNR=1.2), 23c (4-person, ρ=0, DSNR= 1.5), and 23d (4-

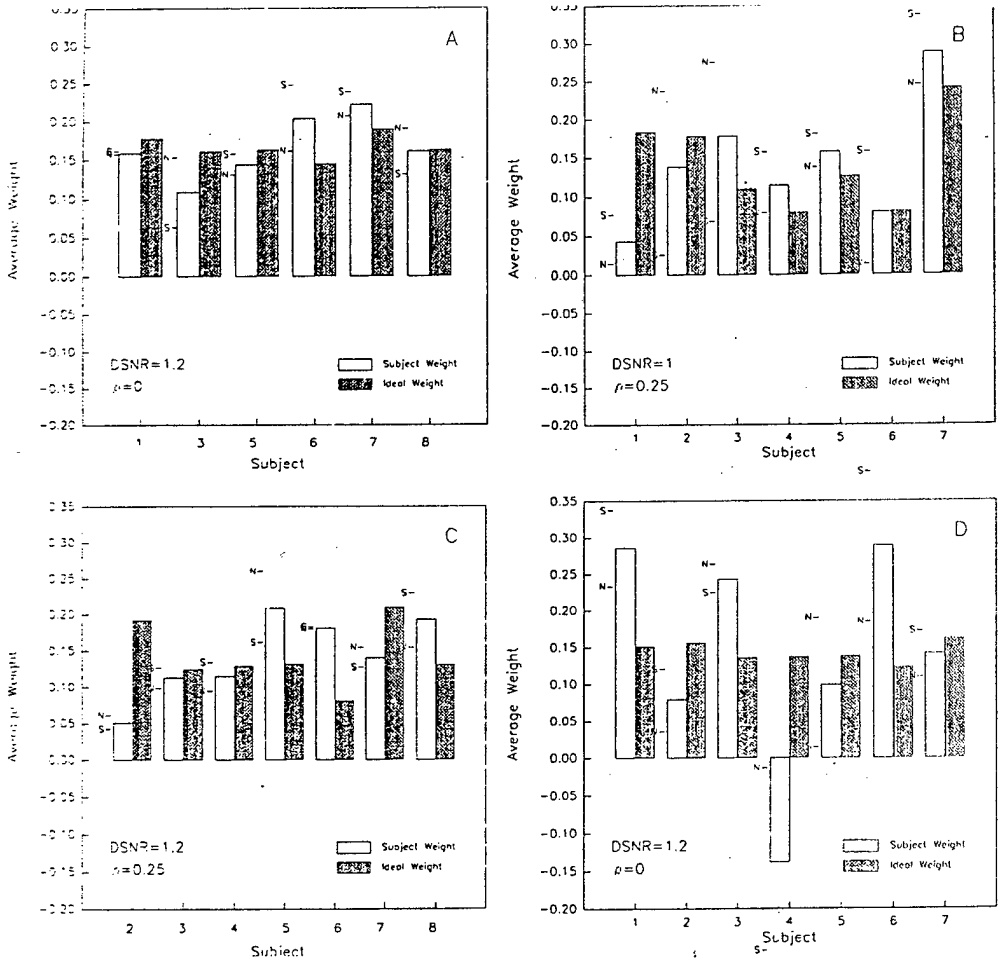

Figure 22. The average relative weights obtained for each member of some 6 and 7 member groups for different conditions of display correlation and signal-to-noise ratio. Signal-to-noise ratio was the same for all members of the group. The *S* and *N* symbols show the weights calculated separately on signal and non-signal trials. In panels A and B, the obtained weights appear to be partially correlated with the ideal weights. Note the negative weight given to subject 4 in panel D. This subject had irritated the other members of the group by arriving late.

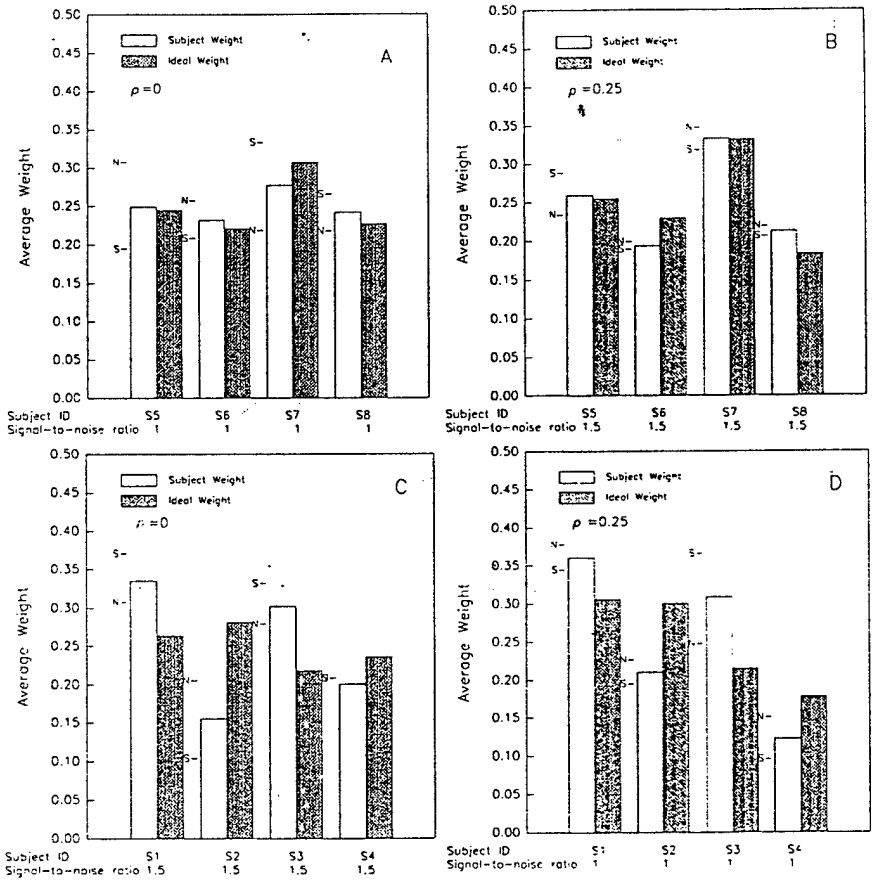

Figure 23. The average relative weights obtained for each member of some 4-member groups for different conditions of display correlation and signal-to-noise ratio. The *S* and *N* symbols show the weights calculated separately on signal and non-signal trials. In panels A and B, the obtained weights appear to be correlated with the ideal weights; in panels C and D, the weights appear to be random.

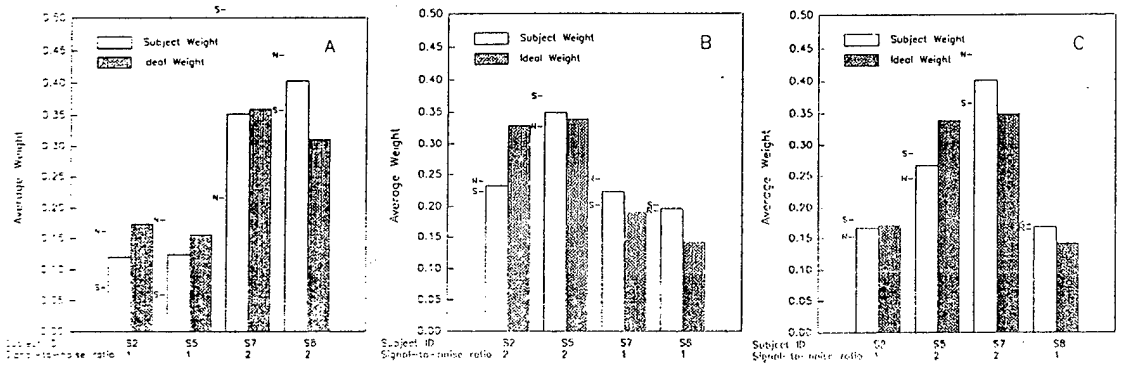

Figure 24. The average relative weights for each member of the 4 member groups in the unequal difficulty conditions. The *S* and *N* symbols show the weights calculated separately on signal and non-signal trials. The ordering of obtained weights is close to the ordering of ideal weights.

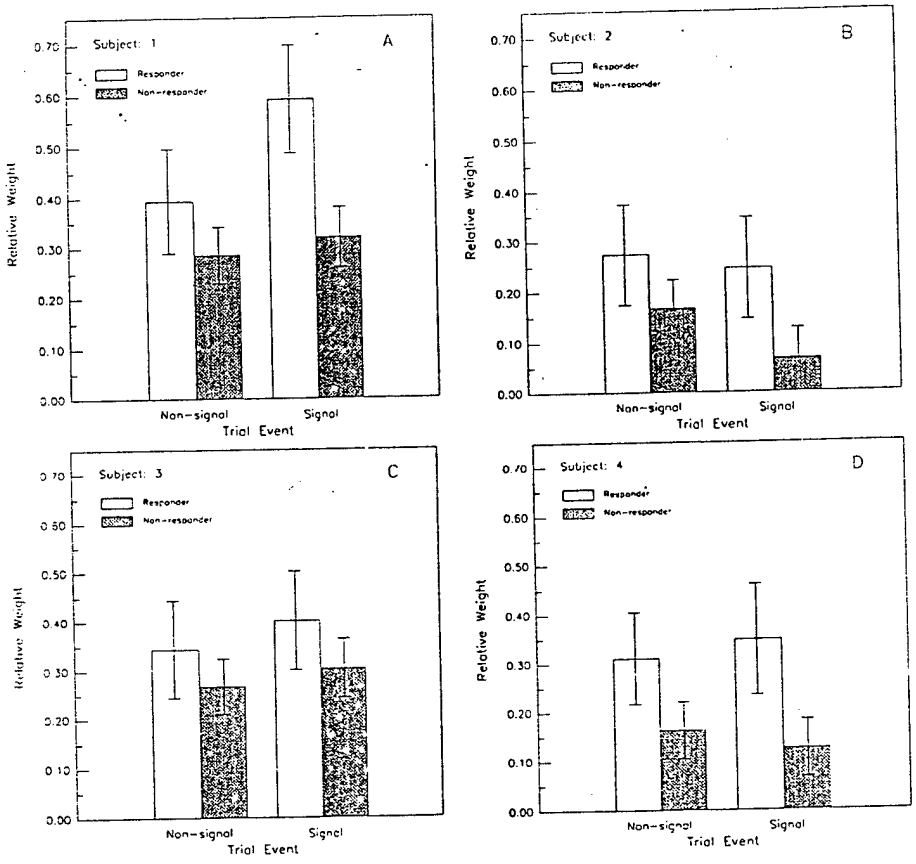

Figure 25. The relative weights obtained for members of a 4-person group (DSNR=1.5, ρ=0) when the person was chosen to respond for the group (open bar) and when the person was not chosen to respond (hashed bar). The responder weights are based on approximately 100 trials and the non-responder weights are based on approximately 300 trials; the brackets indicate plus and minus one (estimated) standard error of the mean.

person, ρ=0.25, DSNR=1), the ordering appears random. In most cases, the variation in the obtained decision weights for a condition was approximately the same as the variation in ideal weights for that condition. Figure 22d shows the only case when a negative weight was given to an observer. This particular observer was late for two consecutive experimental sessions, thereby forcing the group to wait before being able to start the session. Apparently, this behavior resulted in her being given the negative weight. Figure 24 shows the more consistent weighting pattern that was obtained in the unequal DSNR conditions (4-person, ρ=0). Here, the weights corresponded more closely to the ideal values. In all three conditions, the two highest weights were given to the two members with the highest d's (and highest DSNRs).

Based on a statistical argument and a computer simulation, we estimated that the standard deviation, $\sigma_{\hat{a}i}$, of the obtained weights was approximately 0.04. Using that estimate, 5 of the 29 conditions tested produced weights that were outside of a 99% confidence interval around $\hat{a}_i$; that is, they differed significantly from the ideal. This result is not surprising, given the small variation in the ideal weights for the equal DSNR conditions. Recall that this variation was due entirely to the variance in the subjects' d's; at a constant DSNR=1, $\sigma_{d'}$=0.12. In the unequal DSNR condition, where $\sigma_{d'}$ exceeded 0.45, none of the unequal DSNR conditions produced weights that were significantly different from the ideal.

Berg (1990) coined the term, *weighting efficiency* (h_{weight}), to describe how accurately weights were assigned in a detection task. Weighting efficiency is obtained using the following equation:

$$\eta_{weight} = \left[d'_{weight} / d'_{ideal} \right]^2 \qquad (12)$$

where d'_{weight} is the index of detectability that would have been obtained if the group's inefficiency were entirely due to using the obtained, rather than the ideal, weights. It is calculated with equation 11, using the d'_i indices obtained in the individual sessions, and the weights derived from the correlational analysis of the group session. Berg suggested that overall efficiency can be factored into two components, the weighting efficiency and the *noise efficiency*, where the latter relates to all inefficiencies *other* than those attributable to the weighting strategy, i.e.,

$$\eta = \eta_{noise} \times \eta_{weight} \qquad (13)$$

We have plotted weighting efficiency on figure 21 (triangle symbols), to allow comparison with overall efficiency (circle symbols). Although there is a drop in weighting efficiency with group size, the drop in weighting efficiency cannot account for the larger decrease in overall efficiency with size, especially at m=4.

The preceding analysis may not have revealed one non-optimal weighting effect that could have decreased efficiency. Suppose that (1) the person designated to make the group response consistently gave a higher weight to her own judgment, and (2) every member followed that same strategy to an equal degree. Since the choice of responder was random and the higher weight for each responder would be averaged across every member who responded, this higher-weight-to-responder strategy would be transparent to our weighting analysis. We tested for this strategy by comparing the correlation between a member's stimulus and the group's response, when that member was the responder and when that member was not the responder.

The data from the four-member equal DSNR (ρ=0) condition was partitioned into those trials when a member was the designated responder, and those trials when that member was not the designated responder. We then calculated the correlation between the group response and that member's stimulus (separately) for the two sets of partitioned trials. Two drawbacks of this analysis are the reduced number of trials on which the weight is computed and the necessity to compare weights computed from different sets of trials. However, the results were unambiguous: weights were consistently higher when a subject was responder than when she was not. Figure 25 shows the results; the open columns are weights when responder, the hashed columns are when non-responder. The error bars represent plus and minus one (estimated) standard error of the mean, $[1/n\text{-}3]^{1/2}$. The size of the responder effect, in terms of average difference in decision weight, was approximately 0.15.

How much of a drop in efficiency would be produced by a responder effect of this magnitude? Using equation 11, we estimated the drop in performance that would result from a consistent weight increment of 0.15 given to the responder over the other members. The uniform weight case was used as a comparison, because the effect was assumed to be averaged across all group members. The estimated decrease in efficiency (for equal DSNR and ρ=0) turned out to be less than 0.10 in all cases. This decrement is much less than the observed drop in efficiency with group size.

We considered whether the observed decrease in group efficiency could be attributed to changes in the motivation of individual members. A frequent observation about group behavior is that individuals reduce their effort when in a group (e.g., Kerr, 1993; Latané *et al.*, 1979, Shepperd, 1993). We incorporated this observation into a model via the following assumption: Each observer reduces her individual sensitivity (d'_i) by an amount proportional to the number of other members in the group. Then, individual detection sensitivity is a function of m such that

$$d'_i(m) = d'_i(1 - \delta_m m) \qquad (13)$$

where d'_i is the detectability measured in the individual condition and δ_m is a constant between 0 and 1. From equation 7, performance will be given by:

$$d'_{group} = (1 - \delta_m m)\left[\frac{m\mathrm{Var}(d'_i)}{1-\rho} + \frac{m\bar{d}'^2}{1+\rho(m-1)}\right]^{1/2} \quad (14)$$

We calculated a least-squares fit of equation 14 to the group data. The value of δ_m obtained from the fit was 0.056. According to this model, a four-person group would suffer a drop of 0.22 in each member's d' (standard deviate units). The magnitude of the predicted drop for group performance is the same as that for individual performance. The slope of the predicted function was too small at small m and too great at large m[6].

IV. EXPERIMENT 3. OPTIMIZED GROUPS

One problem with the experimental procedure of experiment 2, was that it did not require the individual members to make formal responses on each trial. The reason that we did not require individual responses from the members was because we did not want members to have a commitment to a particular decision prior to the group's deliberation on that trial. In fact, this was a weakness in the experiment's design, because the index of detectability for each member had to be inferred from the member's performance in separately run (individual) conditions. If the member's detection effort changed as a consequence of the group test parameters (such as group size), we could not detect that effect in the group condition. Thus, this could have led to variability in the calculated efficiency of group performance in the group conditions. Experiment 3 was designed to remedy this limitation by requiring each member to make a (public) numerical estimate of signal likelihood on each trial. In addition, we (1) recorded the mean of the actual stimulus displayed to each member on each trial, and (2) provided an additional decision aid to the group on every trial. This aid was designed to optimize group performance by enabling the group to use all of the relevant information about the group members' estimates of signal likelihood.

Our measure of the group's index of detection accuracy was based on the group hit and false alarm rate. In addition, we used the individual estimates of signal likelihood to calculate the detection sensitivities of the individual members. The calculation of group efficiency then would be based on the observed detection performance of the group members *in that condition*. Thus, if an individual member reduced his or her detection effort, that would not produce an observed decrease in the calculated group detection efficiency. In addition, we thought that providing each group member with the formal numerical estimates of all group members would improve the efficacy of member communication and interaction and would produce improvements in the efficiency of the group's detection performance. In fact, we expected group performance to be very near to the ideal group prediction.

In addition to these measures of individual rating and group performance, we recorded the mean of the actual stimuli displayed to each subject on each trial. These mean values provide an additional estimate of the possible performance over a set of trials, since they incorporate the effects of statistical and experimental variability in the signal and noise parameters over trials. A group d' based on the summed mean subject stimulus display was calculated for this purpose.

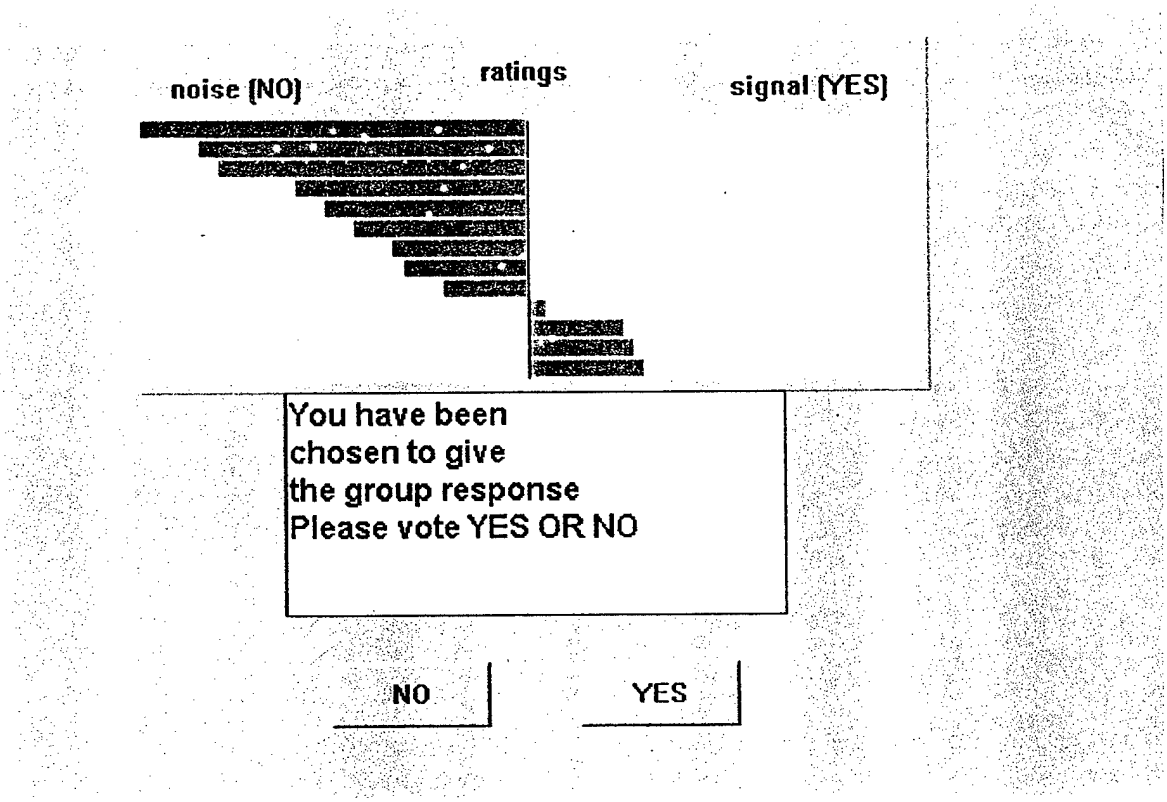

Figure 26. Example of the member rating display.

A. METHOD

1. Subjects, Apparatus, and Stimuli

Four groups of University of Florida students participated in the study. Two groups were composed of 3 students and the two were composed of 5 students. All of the subjects had normal or corrected-to-normal visual acuity. Subjects were paid $4.25 per hour plus an incentive bonus that was based on the accuracy of the group's performance. The bonus averaged approximately $0.75 per person per hour. The apparatus and stimuli were as described for experiment 1.

2. Procedure

The basic task was the same as the previous experiment. After presentation of the stimulus display, the group members were required to provide a numerical estimate, via a response slider, on a scale of 0 to 100, of the judged likelihood of signal occurrence on that trial. After all of a group's members had provided a numerical estimate of signal likelihood, the estimates were rank ordered and graphically displayed to all members of the group. Then, as in experiment 2, one of the members was randomly selected to respond with the group decision. Also as in experiment 2, the monetary payoff to the individual members was determined by the accuracy of the group's decision. Conversation among group members was allowed after the likelihood estimates were displayed and until the presentation of the stimulus on the next trial. Figure 26 shows a sample display of the ranked rating information.

Table 4. Measures employed in experiment 3.

measure	how obtained	notes
d_i , $d_{ideal\text{-}1}$	Based on individual ratings, dichotomized to 'yes' and 'no' responses.	Estimate of group d_{ideal} $d_{ideal\text{-}1} = \sqrt{\Sigma(d_i')^2}$
$d_{ideal\text{-}2}$	Based on sum of actual stimuli presented to each observer, dichotomized to group yes and no response.	Hypothetical group response based on ΣX_i summed mean display.
$d_{observed}$	Obtained group d'	Based on actual group responses.
$d_{sum\text{-}R}$	Based on summed individual ratings, dichotomized to yes and no responses.	Hypothetical group response based on ΣR_i summed individual ratings.
$d_{normal\text{-}R}$	Based on summed normalized individual ratings (see text).	Hypothetical group response based on $\Sigma X_i'$ summed predicted display values based on regression of individual ratings on display values.

B. RESULTS AND DISCUSSION

On each trial of the experiment, we recorded each member's numerical estimate of signal likelihood, R_i, as well as the identity of the respondent and the group's (respondent's) decision. As in the previous experiments, the group decision was used to calculate the group hit and false alarm rate, and hence the group d', indicated as $d_{observed}$ in table 4. In order to estimate the individual detection sensitivity of each member, we compared each member's numerical estimate on each trial to an assumed criterion value. That comparison yielded a hypothetical yes or no response for that member on that trial. Rather than constructing (and integrating) a complete

ROC curve for each member, we chose a range of hypothetical criteria (around R=50) and averaged the resulting d'_i values for each observer. These values were then used as one estimate of the ideal group d' for that condition, $d_{ideal-1}$. This estimate ignores decreases in group performance due to inefficiencies in individual observer performance. A second estimate of the ideal group d' was provided by the mean of the actual display values, X_i, presented to each observer on each trial. These values were summed and then dichotomized to form hypothetical yes and no responses for the group and to a calculation of $d_{ideal-2}$. This d' value assumes that individual member efficiency is 100%. Finally, we calculated two additional measures of hypothetical performance: d_{sum-R} and $d_{normal-R}$. The first of these measures simply took the sum of the member ratings on each trial and dichotomized that sum to yes and no responses. The second measure examined each member's function relating the rating, R_i, to the displayed input, X_i, and computed a regression of R on X. The individual regression equations were then used to transform each observer's rating to a best estimate of the input, X'. These estimated (equivalent to normalized ratings) were then summed and dichotomized to yes and no responses to calculate $d_{normal-R}$.

Table 5. Detection indices obtained in experiment 3.

group	measure					individual d's				
	$d_{observed}$	$d_{ideal-1}$	$d_{ideal-2}$	d_{sum-R}	$d_{normal-R}$	d_1	d_2	d_3	d_4	d_5
3A	**1.43**	1.63	1.92	1.69	1.83	0.99	0.95	0.88		
3B	**1.63**	2.39	1.91	1.92	1.48	0.81	1.40	1.75		
5A	**1.95**	2.22	3.13	1.99	2.16	1.06	1.26	0.85	1.01	0.67
5B	**1.63**	1.77	2.48	1.78	1.92	0.94	1.02	0.77	0.48	0.62

Table 5 shows the group detection indices observed in experiment 3 as a function of group size, for four groups. In all cases: (1) $d_{ideal-1}$ was below $d_{ideal-2}$ indicating that there was some loss in efficiency due to the detection behavior of members acting individually, rather than as a group; and (2) obtained performance, $d_{observed}$, was below either of the two measures of ideal group performance. The difference between the observed performance and the measure based on the individual observer indices, $d_{ideal-1}$, was within statistical error. (The worst-case standard error in d' for d'=1.8 based on a block of 100 trials is approximately 0.3 d' units). This indicates that there was little or no loss in efficiency in aggregating member judgments and in making the group response. Comparisons of the indices based on summed ratings also indicates that the group members were combining member rating information in a highly efficient fashion. In all cases but one there was a benefit to using the normalized or predicted display reading, rather than the actual rating. This may indicate a possible way to further improve the display of member judgments, i.e., by computing a best estimate of each member's displayed input given on the member's rating, prior to the group decision. This is equivalent to normalizing each member's rating to have a standard slope and intercept.

V. GENERAL DISCUSSION

A theory of group decision making should be able to predict how performance depends both on the properties of the decision task and the abilities of the group members. Group signal detection theory promises to meet those requirements. The theory allows us to quantify the accuracy of group performance and to specify the relationship between performance, the level and distribution of member expertise, and some formal aspects of member interaction and decision rule. Moreover, empirical tests of the theory lead naturally to hypotheses about the causes of inefficiencies in the behavior of human groups.

According to the theory, the accuracy of a group's performance must fall between two extremes: the zero-interaction binary-voting group and the statistically ideal group. In general, group performance will be limited by the number and ability of the members and the correlation between member judgments. In addition, the ideal group is assumed to weight each member's graded judgment in proportion to the member's expertise. The present experiments were designed to give groups of human subjects the opportunity to reach their maximum level of performance, via training, feedback, monetary payoff, and minimal constraints on member interaction. The resulting performance was generally consistent with the detection theory of the ideal group, in that: (1) the level of group performance was high and increased with group size, and the advantage of size decreased when member judgments were correlated; and (2) decision weights were assigned in accordance with member expertise when there was sufficient variability in the ability of individual members. However, in experiment 2, the efficiency of group performance decreased with group size, possibly due to members reducing their individual efforts when working in a group or to problems in combining judgments from different group members.

The present results are qualitatively consistent with studies that have found group decision performance to be less than the statistical optimum (see Davis, 1992). In the present study, we quantified the effects on performance of inter-member correlation and member ability, and we showed that groups can effectively weight the judgments of their members. Libby, Trotman, and Zimmer (1987) showed that the variance in expertise within a group and the group's ability to recognize the relative expertise of its members, are crucial in determining group performance when member interaction is allowed. Our results are consistent with their observations. Our weighting results also are consistent with studies that show that groups can recognize the relative expertise of group members. For example, Henry (1993), demonstrated that groups can estimate the ability of members, even when no specific feedback about the correctness of judgments is provided.

The observed decrease in efficiency with group size might be attributed to a number of different causes, including inter-member correlation, binary voting without member interaction, inappropriate weighting, individual differences in likelihood functions, and extraneous noise (see table 2). We now discuss the likelihood that each of these possibilities acted in the present experiment, beginning with the hypothesis of inter-member correlation.

A. INTERMEMBER-CORRELATION

Implicit in our signal detection model is the assumption that each member make an observation and that this observation leads to a graded estimate of signal likelihood. One member's estimate could be correlated with another member's estimate as a consequence of common genetics, background, or experience. However, the correlation between members is confined by assumption to the individual's *perceptual* stage of processing; i.e., it can arise only in the first stage of the system shown in figure 10. Conversely, the influence one member has on another is confined to the *decision* stage of the system, specifically, the setting of decision weights on each member's contribution to the group decision (e.g., Kriss, Kinchla & Darley, 1977; Robinson and Sorkin, 1985).

We attempted to control the inter-member correlation in experiment 2 by manipulating the stimulus information that was presented to each member. Our manipulation was designed to produce member judgments that were either independent or at a set correlation of 0.25. An advantage of signal detection theory is that it allows one to quantify the performance decrement that will be produced by a correlation among member judgments. We observed a large drop in performance as a consequence of our experimentally *increasing* the correlation from $\rho=0$ to $\rho=0.25$. If the member judgments had been correlated to a significant degree *prior* to our manipulation, the effect of the increase in correlation would have been much smaller (equation 7 or 9). The high levels of performance that we obtained also make it unlikely that subjects' estimates were correlated at levels above of the experimental values.

B. NON-INTERACTIVE BINARY VOTING

The second possible cause of decreased efficiency is that the subjects used a non-interactive, binary voting strategy such as modeled by our simulation of a Condorcet group. It is difficult to test this hypothesis directly. We monitored the interactions of our groups, and observed that many groups in experiment 2 did take binary ballots during their deliberations. However, it was apparent that members communicated graded likelihood information when they conveyed their binary votes, i.e., they varied the tone of their voices from tentative to emphatic, and included descriptive phrases, such as, "I think", "definitely a signal", "not sure", etc. Even so, the efficiency of these groups, for group sizes greater than about m=4, was comparable to that of Condorcet groups. Furthermore, the efficiency of the normative Condorcet group was essentially independent of group size; that is, its performance is below and parallel to the ideal group. This insensitivity to size is inconsistent with the result obtained in experiment 2 with real groups. Therefore, even if zero-interaction binary voting were the dominant mode of group "interaction" in experiment 2, it could not explain the observed decrease in efficiency as a function of size.

C. INAPPROPRIATE DECISION WEIGHTS

The third possible cause of the obtained inefficiency is that groups used inappropriate weights. There are at least two reasons for rejecting this hypothesis. First, our analysis of

weighting efficiency indicated that the groups were effective at weighting the judgments of individual members according to the members' detection competence. Appropriate weighting by member d' was very clear in the unequal DSNR conditions of experiment 2, but also was evident in many of the equal difficulty conditions, where there was no strong monetary incentive to do so. This result is not surprising. It is not necessary for the group (or each member) to make an accurate estimate of the d'_i of every other member, because on the average, a group member's response can convey that information. The group member given the high-difficulty display will find it difficult to make signal/no-signal discriminations; if she is honest, she will consistently indicate her uncertainty to the group. Furthermore, the group soon would lose confidence in, and hence lower the weight for, a subject whose estimates were consistently at odds with the trial-by-trial feedback.

The second reason for rejecting the weight hypothesis is that the variance in member d' (in the equal DSNR conditions of experiment 2) was too small to allow for the observed inefficiencies to occur, assuming a simple weighting strategy. Suppose that the group could not differentiate small differences in member ability and instead used a uniform weighting strategy. That would have produced a resultant drop in efficiency of less than 3%. A noisy or random weighting strategy would have produced approximately the same size drop. An even less likely possibility is that the group used a complex and inappropriate weighting strategy. For example, a group could give a positive weight to the worst member and negative or zero weights to the best members. Such a pathological strategy would have produced efficiencies that decreased with group size, but the resulting performance would have been much worse than any observed in the experiment.

D. INDIVIDUAL DIFFERENCES AND ADDED NOISE

The fourth possible cause of group inefficiency is the presence of additional sources of noise. This noise could be present in each member's decision process or could be generated when the member estimates are aggregated into the group decision. Suppose that there were individual differences in the scales used by members in converting their judgments of signal likelihood to numerical estimates or ratings. The problem of aggregating such estimates has begun to receive theoretical attention by decision scientists (see, e.g., Myung, Ramamoorti, and Bailey, 1996; Wallsten, Budescu, Erev, and Diederich, 1997). Variability in the scales used by different members could lower the efficiency of group performance, and our simulations indicate that these inefficiencies would increase as a function of group size.

The results of experiment 3 indicated that, when the members were forced to provide numerical ratings of signal likelihood, and these ratings were ranked and displayed to all members, performance approximated the ideal performance level predicted by the actual member d's obtained in that condition. Inefficiencies of group performance were almost entirely attributable to inefficiencies in the individual performance of members. Hence, the decreased efficiency observed in experiment 2 relative to the (separately assessed) individual performance of the members was not observed. We conclude that the inefficiencies associated with increased group size observed in

experiment 2 may be attributed to either (1) inefficient aggregation of member judgments or (2) factors relating to changes in the detection efficiency of individual members. The former inefficiency was ameliorated in experiment 3 by the requirement for (and display of) numerical ratings of member likelihood estimates.

E. MOTIVATIONAL FACTORS

Kerr (1993), Shepperd (1993), and others have discussed the reasons why subjects may reduce their individual efforts in a group situation, i.e., indulge in "social loafing." If some version of the social loafing hypothesis were true in our experiments, we would conclude that the monetary payoff to each subject, as determined by the level of group performance, was not sufficient to completely dominate incentives to reduce individual effort or participation. Perhaps subjects will "loaf" if they cannot see the statistical benefit of their contribution to the group's performance (i.e., to their own payoff), and if they can do so anonymously. Harkins and Petty (1982) found that subjects who viewed their contributions as non-essential reduced their individual efforts more than subjects who viewed their efforts as significant. Following that reasoning, groups with high inter-member correlation should show greater decreases in efficiency with size, because the benefit of each member's contribution will be reduced. We did not find a greater drop in efficiency with size in the correlated condition, perhaps because the level of correlation between member stimuli (0.25) was not high enough to be obvious to group members.

We tried to model the possibility of as an incremental decrease in each member's d', that increased in proportion to the group size in experiment 2. For our data, the magnitude of the constant reduction per member was 0.056 (standard deviate units), which would result in a drop in d' of 0.22 for an individual in a four-person group. A group member could be confident that this much reduction in performance accuracy would not be apparent to another group member. The drop would be proportionally larger in large groups, but an individual might feel protected by a greater sense of anonymity in a larger group. Although the model fit was not impressive, that may have been due to the presence in experiment 2, of additional inefficiencies relating to the aggregation of member estimates (and ameliorated in experiment 3). Experiment 3 also provided some evidence for members reducing their individual effort as the group size was increased. Further experiments, combining the techniques of experiments 2 and 3 and perhaps utilizing additional methods to control individual detection effort, may be able to further isolate this factor.

NOTES

1. The expected value of g cannot be written in non-integral form, as a result d_i' was evaluated by Monte Carlo simulation. Receiver Operating Characteristic (ROC) curves were generated using as the noise distribution, g_n, the absolute value of the difference of two Gaussian random variables with zero means and a standard deviation of 10 ms. This value was based on previous work (Sorkin, 1990). The signal-plus-noise distribution, g_{s+n}, was calculated as the absolute value of the difference of two Gaussian variables with zero mean and a standard deviation equal to $(10^2+\sigma_{exp}^2)^{1/2}$, where σ_{exp}^2 is the variance introduced by the experimenter. The area under the ROC at each σ_{exp} was calculated and converted to an equivalent d' value. An approximation for d_i' in terms of the perturbation σ_{exp} and an assumed value for $\sigma_{internal}$ of 10 ms is given by:

$$d_i' = 0.2313 + 0.0362\sigma_{exp} - 0.00017\sigma_{exp}^2$$

If the sequences consisted of 8 segments with uniform statistics, and the listener assigned equal weights to each segment, it follows that d' would be $d'_i\sqrt{8}$. If the sequence has one unique segment, d' would be $[(d'_i\sqrt{7})^2 + (d'_{unique})^2]^{1/2}$. A necessary condition for this case is that a proportionately higher weight would be applied to the time difference obtained from the unique segment. This model makes no assumption about the effect of mean segment duration on performance.

2. These restrictions affected our distributions in the following manner. The instances when the intertone time interval exceeded the maximum allowed value occurred only rarely. It should be noted that while the distribution was "normal" appearing for values above 2 ms (minimum selected value), this was not the case for values at or less than 2 ms. In these instances, there was a second peak at 2 ms. This means that there were more tone bursts that were very close together than would be expected based on a normal distribution of intertone time intervals. In cases where this occurred, the actual mean and standard deviation values were respectively slightly elevated and reduced than the assigned values. However, we do not believe this had a significant effect on the obtained results.

3. The performance, d', of a system that employs the Z statistic and weights, a_i, is given by the expected value (over trials) of Z given signal, minus the expected value of Z given noise, all divided by the standard deviation of Z. Let $X_{i,j}$ be the value of detector i's estimate on the j^{th} trial. The expected value of Z given noise is zero, so the numerator for d' is just the expected value of Z given signal. That is,

$$E[Z|_{signal}] - E[Z|_{noise}] = E[\sum_{i=1}^{m} a_i(X_{i,j}|_{signal})] = \sum_{i=1}^{m} a_i d_i'$$

The variance of Z over trials is the same given signal or noise and since the independent and common components of X are independent,

$$\mathrm{Var}[Z] = \mathrm{Var}\left[\sum_i a_i X_{i,j}\right] = \mathrm{Var}\left[\sum_i a_i X_{IND,j} + \sum_i a_i X_{COM,j}\right]$$

$$= \mathrm{Var}\left[\sum_i a_i X_{IND,j}\right] + \mathrm{Var}\left[\sum_i a_i X_{COM,j}\right]$$

$$= \sum_i \left[\mathrm{Var}(a_i X_{IND,j})\right] + \mathrm{Var}\left[\left(\sum_i a_i\right) X_{COM,j}\right]$$

$$= \sum_i \left[a_i^2\, \mathrm{Var}(X_{IND,j})\right] + \left(\sum_i a_i\right)^2 \mathrm{Var}[X_{COM,j}]$$

$$= \sigma^2_{IND} \sum_i a_i^2 + \sigma^2_{COM} \left(\sum_i a_i\right)^2$$

$$= (1-\rho) \sum_i a_i^2 + \rho \left(\sum_i a_i\right)^2$$

and,

$$d'_{weight} = \frac{\sum_{i=1}^{m} a_i d'_i}{\sqrt{(1-\rho)\sum_{i=1}^{m} a_i^2 + \rho\left(\sum_{i=1}^{m} a_i\right)^2}}$$

Note, that our simplification of the multiple channel model requires that $\rho \neq 1$.

4. The average criterion in the group conditions was 0.021 with a standard deviation of 0.17. The lack of criterion effects was consistent with our previously reported data on individual subjects in similar tasks (Sorkin *et al.*, 1991) and generally consistent with results reported by Pete *et al.*, (1993b) in their study of distributed detection by 3-person groups. The criteria used by Pete *et al.*'s observers (operating points) reflected a relatively neutral bias and were somewhat insensitive to experimental manipulation of event probability and cost structure. However, the direction of the observers' criterion shifts was in the direction of the optimal criterion. In all of our experimental conditions, the optimal group criterion was zero (c=0) .

5. If the observers had made individual judgments about the likelihood of signal on each trial, those judgments could have been used to calculate the group decision weights. In that case, we could have computed the correlation between each observer's judgment and the group decision. Such correlations would have provided a more accurate estimate of the weight given to each observer's judgment. However, we had decided not to require any overt trial-by-trial response by individual observers. We wished to avoid any chance that either the individual and/or the group behavior would be affected by the requirement for individual responses.

6. A similar model contains the assumption that each observer suffers from an additional, uncorrelated noise that is proportional to the size of the group. The performance of this model is given by:

$$d'_{group} = \left[\frac{1}{1+m\sigma_m^2}\right]^{1/2} \cdot \left[\frac{mVar(d'_i)}{1-\rho} + \frac{m\bar{d}'^2}{1+r(m-1)}\right]^{1/2}$$

where $\sigma_m{}^2$ is the noise added per member. The fitted value of $\sigma_m{}^2$ was 0.1787. This model behaves in a similar way to the d'-decrement model, but there is less drop predicted for the performance of large groups (m>8). The model fits to the data (m<8) were equivalent to the d'-decrement model.

REFERENCES

Ashby, F. G. & Maddox, T. W. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance*, **18**, 50-71.

Austin-Smith, D., and Banks, J.S. (1996). Information aggregation, rationality, and the Condorcet jury theorem. *American Political Science Review*, **90**, 34-46

Batchelder, W. H. and Romney, A. K. (1986). The statistical analysis of a general Condorcet model for dichotomous choice situations. In B. Grofman and G. Owen (Eds.), *Information pooling and group decision making: Proceedings of the second University of California, Irvine, conference on political economy*. Greenwich: CT: JAI Press.

Berend, D., and Harmse, J. E. (1993). Expert rule versus majority rule under partial information. *Theory and Decision*, **35**, 179-197.

Berg, B. G. (1989). Analysis of weights in multiple observation tasks. *Journal of the Acoustical Society of America*, **86**, 1743-1745.

Berg, B. G. (1990). Observer efficiency and weights in a multiple observation task. *Journal of the Acoustical Society of America*, **88**, 149-158.

Berg, B. G. and Green, D. M. (1990). Spectral weights in profile listening. *Journal of the Acoustical Society of America*, **88**, 758-766.

Clement, D. E. & Schiereck, J. J. (1973). Sex composition and group performance in a visual signal detection task. *Memory & Cognition*, **1**, 251-255.

Collins, L. M., Wakefield, G. H., and Feinman, G. R. (1994). "Temporal pattern discrimination and speech recognition under electrical stimulation," *Journal of the Acoustical Society of America*, **96**, 2731-2737.

Condorcet, Jean-Antoine-Nicholas de Caritat, marquis de, (1785). *Essai sur l'application de l'analyse a la probabilite des descisions redues a la pluralite de voix.* Paris: De l'Imprimerie royale.

Davis, J. H. (1992). Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: Selected examples, 1950-1990. *Organizational Behavior and Human Decision Processes*, **52**, 3-38.

Deutsch, D. (1979). "Binaural integration of melodic patterns," Percept.& Psychophys. 25, 399-405.

Durlach, N. I. and Braida (1969). "Intensity perception. I. Preliminary theory of intensity resolution," *Journal of the Acoustical Society of America*, **46**, 372-383.

Durlach, N. I., Braida, L. D., & Ito, Y. (1986). Towards a model for discrimination of broadband signals. *Journal of the Acoustical Society of America*, **80**, 60-72.

Einhorn, H. J., Hogarth, R. M., & Klempner, E. (1977). Quality of group judgment. *Psychological Bulletin*, **84**, 158-172.

Erev, I., Gopher, D., Itkin, R., & Greenshpan, Y. (in press). Toward a generalization of signal detection theory to n-person games: The example of two-person safety problem. *Journal of Mathematical Psychology*.

Fraisse, P. (1982). "Rhythm and Tempo," in *The Psychology of Music*, edited by D. Deutsch (Academic Press Inc., New York).

Green, D. M. (1992). The number of components in profile analysis tasks. *Journal of the Acoustical Society of America*, **91**, 1616-1623.

Green, D. M., Kidd, G., and Picardi, M. C. (1983). "Successive versus simultaneous comparison in auditory intensity discrimination," *Journal of the Acoustical Society of America*, **73**, 639-643.

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

Grofman, B., Feld, S. L., & Owen, G. (1984). Group size and the performance of a composite group majority: Statistical truths and empirical results. *Organizational Behavior and Human Performance*, **33**, 350-359.

Grofman, B., Owen, G., and Feld, S. (1983). Thirteen theorems in search of the truth. *Theory and Decision*, **15**, 261-278.

Harkins, S. G., and Petty, R. E. (1982). Effects of task difficulty and task uniqueness on social loafing. *Journal of Personality and Social Psychology*, **43**, 1214-1229.

Hastie, R. (1986). Experimental evidence on group accuracy. In B. Grofman & G. Owen (Eds.), *Decision research* (Vol. 2). Greenwich, CT: JAI Press.

Henry, R. A. (1993). Group judgment accuracy: Reliability and validity of postdiscussion confidence judgments. *Organizational Behavior and Human Decision Processes*, **56**, 11-27.

Hillman, B. J., Hessel, S. J., Swensson, R. G., and Herman, P.G. (1977). Improving diagnostic accuracy: A comparison of interactive and Delphi consultations. *Investigative Radiology*, **12**, 112-115.

Hinsz, V. B. (1990). Cognitive and consensus processes in group recognition memory performance. *Journal of Personality and Social Psychology*, **59**, 705-718.

Jeffress, L. A. and Robinson, D. E. (1962). Formulas for the coefficient of interaural correlation for noise, *Journal of the Acoustical Society of America*, **34**, 1658-1659.

Karotkin, D. and Paroush, J. (1994). Variability of decisional ability and the essential order of decision rules. *Journal of Economic Behavior and Organization*, **23**, 343-354.

Karotkin, D., Nitzal, S., and Paroush, J. (1988). The essential ranking of decision rules in small panels of experts. *Theory and Decision*, **24**, 253-268.

Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis *Journal of Personality and Social Psychology*, **45**, 819-828.

Kidd, G., Mason, C. R., Uchanski, R. M. Brantley, M. A., and Shah, P. (1991). "Evaluation of simple models of auditory profile analysis using random reference spectra," *Journal of the Acoustical Society of America*, **90**, 1340-1354.

Kidd, G. R. (**1995**). "Proportional duration and proportional variance as factors in auditory pattern discrimination," *Journal of the Acoustical Society of America*, **97**, 1335-1338.

Kidd, G. R., and Watson, C. S. (1992). "The "proportion-of-the-total-duration-rule" for the discrimination of auditory patterns," *Journal of the Acoustical Society of America*, **92**, 3109-3118.

Klatt, D. H. (1976). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208-1221.

Kriss, M., Kinchla, R. A., & Darley, J. M. (1977). A mathematical model for social influences on perceptual judgments. *Journal of Experimental Social Psychology*, **13**, 403-420.

Latané, B. (1991). The psychology of social impact. *American Psychologist*, **36**, 343-356.

Latané, B., Williams, K. & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, **37**, 822-832.

Libby, R., Trotman, K.T., and Zimmer, I. (1987). Member variation, recognition of expertise, and group performance. *Journal of Applied Psychology*, **72**, 81-87.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of speech code," Psych. Review. **74**, 431-461.

Lutfi, R. A. (1993). "A model of auditory pattern analysis on component relative-entropy *Journal of the Acoustical Society of America*, **94**, 748-758.

Lutfi, R. A. (1995a). Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks. *Journal of the Acoustical Society of America*, **97**, 1333-1334.

Lutfi, R. A. (1995b). "Further comments on proportional duration and proportional variance as factors in auditory pattern discrimination," *Journal of the Acoustical Society of America*, **97**, 1339-1340.

Lutfi, R. A. and Doherty, K. A. (1994). "Effect of component-relative-entropy on the discrimination of simultaneous tone complexes *Journal of the Acoustical Society of America*, **96**, 3443-3450.

Macmillan, N. A., and Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Martin, J. G., and Struges, P. T. (1974). "Rhythmic structures in auditory temporal pattern perception and immediate memory," J. Exp. Psych. **102**, 377-383.

Metz, C. E., and Shen, J. (1992). Gains in accuracy from replicated readings of diagnostic images: Prediction and assessment in terms of ROC analysis. *Medical Decision Making*, **12**, 60-75.

Miller, N. R. (1986). Information, electorates, and democracy: Some extensions and interpretations of the Condorcet jury theorem. In B. Grofman and G. Owen (Eds.),

Information pooling and group decision making: Proceedings of the second Unaiversity of California, Irvine, conference on political economy. Greenwich: CT: JAI Press.

Montgomery, D. A. and Sorkin R. D. (1996). Observer sensitivity to element reliability in a multi-element visual display. *Human Factors*. **38**, 484-494.

Myung, I.J., Ramamoorti, S., and Bailey, A.D. (1996). Maximum entropy aggregation of expert outcome predictions. *Management Science*, **42**, 1420-1436.

Nitzan, S., and Paroush, J. (1982). Optimal decision rules in uncertain dichotomous choice situations. *International Economic Review*, **23**, 289-297.

Nitzan, S., and Paroush, J. (1984a). The significance of independent decisions in uncertain dichotomous choice situations, *Theory and Decision*, **17**, 47-60.

Nitzan, S., and Paroush, J. (1984b). A general theorem and eight corollaries in search of a correct decision, *Theory and Decision*, **17**, 211-220.

Nitzan, S., and Paroush, J. (1984c). Partial information on decisional competencies and the desirability of the expert rule in uncertain dichotomous choice situations, *Theory and Decision*, **17**, 275-286.

Pete, A., Pattipati, K.R., and Kleinman, D. L. (1993a). Optimal team and individual decision rules in uncertain dichotomous situations. *Public Choice*, **75**, 205-230.

Pete, A., Pattipati, K.R., and Kleinman, D. L. (1993b). Distributed detection in teams with partial information: A normative-descriptive model. *IEEE Transactions on systems, man, and cybernetics*, **23**, 1626-1647.

Pollack, I. & Madans, A. B. (1964). On the performance of a combination of detectors. *Human Factors*, **6**, 523-531.

Richards, V. M. & Zhu, S. (1994). Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients. *Journal of the Acoustical Society of America* , **95**, 423-434.

Robinson, D. E. & Sorkin, R. D. (1985). A Contingent Criterion Model of Computer Assisted Detection. In R. Eberts & C. G. Eberts (Eds.) Trends in Ergonomics/Human Factors, Vol. II. North-Holland: Amsterdam, 75-82.

Shannon, C. E. (1948). "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379-423.

Shapley, L. and Grofman, B. (1984) Optimizing group judgmental accuracy. *Public Choice*, **43**, 329-343.

Shepperd, J. A. (1993). Productivity loss in performance groups: A motivation analysis. *Psychological Bulletin*, **113**, 67-81.

Sorkin, R. D. (1987). "Temporal factors in the discrimination of tonal sequences *Journal of the Acoustical Society of America*, **82**, 1218-1226.

Sorkin, R. D. (1990). Perception of temporal patterns defined by tonal sequences. *Journal of the Acoustical Society of America*, **87**, 1695-1701.

Sorkin, R. D. & Dai, H. (1993). Psychoacoustic models of group signal detection. *Journal of the Acoustical Society of America*, **93**, 2366.

Sorkin, R. D. & Dai, H. (1994). Signal detection analysis of the ideal group. *Journal of Organizational Behavior and Human Decision Processes*, **60**, 1-13.

Sorkin, R. D. & Montgomery, D. A. (1991). Effect of time compression and expansion on the discrimination of tonal patterns. *Journal of the Acoustical Society of America*, 90, 846-857.

Sorkin, R. D., Montgomery, D. A., and Sadralodabai, T. (1994). Effect of sequence delay on the discrimination of tonal patterns. *Journal of the Acoustical Society of America,* **96,** 2148-2155.

Sorkin, R. D., Mabry, T. R., Weldon, M., & Elvers, G. (1991). Integration of information from multiple element display. *Organizational Behavior and Human Decision Processes*, **49,** 167-187.

Steedman, M. J. (1977). "The perception of musical rhythm and meter," Perception, **6,** 555-569.

Stevens, K. N., and House, A. S. (1972). "Speech perception," in *Foundations of modern auditory theory,* edited by J. T. Tobias (Academic Press, New York) .

Swets, J. A. (1973). The relative operating characteristic in psychology. *Science*, **182**, 990-1000.

Swets, J. A. (1986a). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin,* **99**, 100-117.

Swets, J. A. (1986b). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance.*Psychological Bulletin*. **99**, 181-198.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285-1294.

Swets, J. A., and Pickett, R. M. (1982). *Evaluation of diagnostic systems. Methods from Signal Detection Theory*. New York: Academic Press.

Tanner , W. P. & Birdsall, T. G. (1958). Definitions of d' and h as psychophysical measures. *Journal of the Acoustical Society of America*, **30**, 922-928.

Voss, J., and Rasch, R. (1981). "The perceptual of musical tones," Percept. Psychophys. **29,** 323-335.

Wallsten, T.S., Budescu, D. V., Erev, I., and Diederich, A. (1997). Evaluating and combining subjective probability estimates. *Journal of Behavioral Decision Making*, **10**, 243-268.

Watson, C. S., Wroton, H. W., Kelly, W. J., and Benbasset, C. A. (1975). "Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals," *Journal of the Acoustical Society of America,* **57,** 1175-1185.

Zara, J., Onsan, Z. A., and Nguyen, Q. T. (1993). "Auditory profile analysis of harmonic signals," *Journal of the Acoustical Society of America,* **93,** 3431-3441.

PROFESSIONAL PERSONNEL ASSOCIATED WITH THE RESEARCH PROJECT

Hays, Christopher J. Graduate Student, Department of Psychology, University of Florida. Mr. Hays was on assignment from the Air Force to study for a M.S. in Psychology. He completed worked on the group detection project and completed his Masters Degree in 1995.

Montgomery, D. A. Assistant Professor Psychology, Bradley University. Dr. Montgomery completed her Ph.D. at the University of Florida in 1993.

Robinson, Donald E. Professor, Department of Psychology, Indiana University. Dr. Robinson assisted with the computer simulation and analysis of Condorcet groups.

Sadralodabai, Toktam. Postdoctoral Research Fellow, Boys Town National Research Hospital, Omaha, NE. Dr. Sadralodabai completed her Ph.D. at the University of Florida in 1996.

Sorkin, R. D. Principal Investigator, Professor of Psychology, University of Florida.

West, Ryan. Graduate Student, Department of Psychology, University of Florida.

PUBLICATIONS AND DISSERTATIONS

Sadraladabai, T., and Sorkin, R. D. (1993). Serial position effects in temporal pattern discrimination. *Journal of the Acoustical Society of America*, **93**, 2385-2386 (abstract).

Sorkin, R. D. and Dai, H. (1993). Psychoacoustic models of group signal detection. *Journal of the Acoustical Society of America*, **93**, 2366 (abstract).

Montgomery, D. A. and Sorkin, R. D. (1993). The effects of display code and its relation to the optimal decision statistic in visual signal detection. *Proceedings of the Human Factors and Ergonomics Society*, **2**, 1325-1329.

Barfield, W., Salvendy, G., and Sorkin, R. D. (1994). Judgments of orientation for computer-generated images as a function of type of rotation and level of figure complexity. *Studia Psychologica*, **36**, 283-299.

Sorkin, R. D., Montgomery, D. A., and Sadralodabai, T. (1994). Effect of sequence delay on the discrimination of tonal patterns. *Journal of the Acoustical Society of America*, **96**, 2148-2155.

Sorkin, R. D. and Dai, H. (1994). Signal detection analysis of the ideal group. *Organizational Behavior and Human Decision Processes*, **60**, 1-13.

Sadraladabai, T., and Sorkin, R. D. (1995). The effect of different component means and variances on the discrimination of temporal patterns. *Journal of the Acoustical Society of America*, **97**, 3278 (abstract).

Montgomery, D. A. and Sorkin, R. D. (1996). Observer sensitivity to element reliability in a multi-element visual display. *Human Factors*, **38**, 484-494.

Sadraladabai, T. *The discrimination of auditory temporal patterns: Effect of temporal position, channel, and intertone time statistics*. University of Florida Doctoral Dissertation, 1996.

Sorkin, R. D., Hays, C. J., and R. West. (under revision). A signal detection theory of group decision making. *Psychological Review*.

Sadraladabai, T., and Sorkin, R. D. (under revision). The discrimination of auditory temporal patterns: Effect of temporal position and intertone time statistics. *Journal of the Acoustical Society of America.*

Sorkin, R. D., West, R., and Robinson, D. E. (submitted.) Group performance depends on the majority rule. *Psychological Science*.

Sorkin, R. D. (submitted). Spreadsheet Signal Detection. *Behavior Research Methods, Instruments, & Computers*.