NASA CR-173,296

NASA-CR-173296
19840009790

Annual Progress Report

January-December 1983

OBJECTIVE MEASURES OF PILOT WORKLOAD

Cooperative Agreement NCC 2-235

between

Purdue University
Department of Psychological Sciences
West Lafayette, IN 47907
Barry H. Kantowitz, Principal Investigator

and

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Ames Research Center

BEST

AVAILABLE

COPY

Annual Progress Report

January-December 1983

OBJECTIVE MEASURES OF PILOT WORKLOAD

Cooperative Agreement NCC 2-235

between

Purdue University
Department of Psychological Sciences
West Lafayette, IN 47907
Barry H. Kantowitz, Principal Investigator

and

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Ames Research Center

This is the second progress report for Cooperative Agreement NCC 2-235. The NASA Technical Officer is S.G. Hart, Ames Research Center, Man-Vehicle Systems Research Division.

Three experiments have been completed during this year. The first, conducted by R.J. Shively, investigated timesharing behavior in a data-entry task, similar to a pilot entering navigation data into an on-board computer. The second, conducted by MarySue Weldon and Patricia Casper, examined auditory reaction time as a function of stimulus information and dimensionality. This study has direct implications for stimulus selection for secondary tasks used in the GAT flight simulator at Ames Research Center. The third experiment, conducted by Charles Caldwell, studied attention effects of heat and cold stress in a psychological refractory period task. The focus of interest is the general effects of stress on attention rather than upon specific temperature related phenomena. Since the Human Information Processing Laboratory at Purdue has a special environmental chamber, temperature is a convenient way of stressing the operator. Brief descriptions of each experiment follow. The Appendix contains a complete report of work accomplished in each of the three experiments.

Research conducted by R. J. Shively

In this experiment operators were required to enter data into a computer. Data were digits and the operators either entered the digit itself or transformed the digit by adding one to it. Two data entry devices were used: keyboard or light pen. Thus, there were four experimental conditions (2 levels of data entry complexity crossed with 2 levels of data entry device). Auditory choice-reactions were used as the secondary task. Two different tone frequencies (High = 4900 Hz; Low = 2900 Hz) were presented in two pulses to obtain four possible tone combinations: High-High, High-Low, Low-High, Low-Low. These pairs were chosen to mimic similar pairs of Beeps and Buzzes used in previous research at Ames Research Center.

At present, only the data entry task results have been completely analyzed. While task complexity influenced speed of data entry, no significant differences were found between keyboard and light pen devices. Imposition of the secondary task did not significantly increase data-entry time and no speed-accuracy trade-off was found. These results indicate that the assumptions necessary to interpret secondary-task data have been met. Analysis of secondary-task data is in progress. When this is completed, the mathematical network model described in the original proposal will be fit to all the data. A complete report should be ready before Mr. Shively begins his internship at Ames this coming summer.

Research conducted by MarySue Weldon and Patricia Casper

This experiment compared reaction times to tones generated by an Apple II microcomputer versus Beeps and Buzzes (multi-dimensional stimuli) generated by a Cyborg ISAAC Model 91A. Both one and two bits of stimulus uncertainty were used, mapped to 2 or 4 responses to maintain unique 1:1 stimulus-response mappings. Previous research at Ames has found professional pilots to be unable to complete this reaction-time task using tones with two bits of uncertainty (4 equi-probable tones) although they were able to accomplish the task when multi-dimensional stimuli consisting of Beep-Buzz pairs were used to create two bits of uncertainty. The Purdue undergraduates in this experiment were able to learn either tones or multi-dimensional stimuli to the same criterion with no significant difference in number of acquisition trials. The most important result was an interaction between stimulus information and stimulus dimensionality. For the multi-dimensional stimuli, rate of transmitted information increased with stimulus information. However, for the tones rate of transmitted information actually decreased as stimulus information increased from one to two bits. Since professional pilots probably have worse hearing than college students, it is likely that a replication of this study using pilots would show an even more marked interaction. This has implications both for the design of auditory alerting signals in the flight deck and for selection of auditory stimuli for secondary-tasks in simulated flight.

Research conducted by Charles Caldwell

A major goal of workload research is to establish how pilot workload is affected by stress. While such stress is most often imposed by the procedures and equipment required to maintain satisfactory flight, exogenous factors such as rapidly deteriorating weather conditions can also impose stress. This experiment determines whether or not changes in ambient temperature in a laboratory setting are sufficient to induce stress that will alter attentional processes. The manipulation of temperature is thus a method for inducing psychological stress. Thus, no biological or physiological indices were recorded.

Three ambient temperatures were used: 20° C, 5°, and 35°. A four-choice psychological refractory period paradigm with two neon lamps mounted close together and two more mounted further away to form a straight horizontal line was used. Reaction time was higher for lamps at the extreme spatial positions, for shorter inter-stimulus intervals, and for normal room temperature. A significant interaction between inter-stimulus interval and spatial position showed that attention was narrowed into the center (foveal) spatial positions more at shorter inter-stimulus intervals. These results, based upon 18 male subjects, are currently being replicated with 18 female subjects.

APPENDIX

1. SHIVELY. MENTAL WORKLOAD IMPOSED BY A DATA-ENTRY TASK.

2. WELDON, CASPER, & KANTOWITZ. AUDITORY CHOICE-REACTION TIME AS A FUNCTION OF STIMULUS INFORMATION AND DIMENSIONALITY.

3. CALDWELL. THE EFFECTS OF HEAT AND COLD ON ATTENTION.

Mental Workload Imposed by a Data-Entry Task

Robert J. Shively

Purdue University

Advances within a field are frequently the result of borrowing techniques or paradigms from other disiplines. Cognitive psychology, especially, has benefitted from interactions with several diverse areas. Shannon(1948), a researcher in communications, developed an information metric that allows quantification of the information available in a stimulus and a response, and the information transmitted from the stimulus to the response. Shannon's metric was widely applied in its early usage(Miller,1953, Atteneave,1954). Many thought that this metric held the answer to quantifing most stimulus-response relationships. However, limitations have been found to the utility of the information metric in some situations. For example, while quantifing the information contained in a array of lights may be an easy task, to do the same for a prose passage, however, may be difficult if not impossible. Although some limitaions have been found to the information metric, its ability to express reaction time and error perfoemance as a single metric (Bits/Sec) has maintained its usefulness.

The field of computer science has also impacted on pyschology. In addition to such uses as computer simulations and automation of laboratories, the use of an analogy based on computer processing of information has become widespread(Broadbent,1958, Sternberg,1969). These models allow cognitie pyschologists to clearly depict how they think information is processed by the human. This approach has come to typify the modern human information processing(HIP) approach to cognitive psychology.

The benefits to be gained from collaboration between two fields does not need to be a one-way street, benefits may be derived by both participants. Two fields that are ripe for such an intersection are human information processing and human factors(HF). Kantowitz(1981), has made the arguement for correspondence between these two areas. They both stand to reap benefits, HF from gaining theory that can guide research and organize experimental results and HIP from seeing theories put to applied uses and from the new areas of interest to venture into. Human factors is a vast disipline which touches almost all facets of human life. One area of HF which could benefit greatly from HIP theory is the human factors of computer systems.

While the computer began to prosper in the 1960s and 1970s, the 1980s have become and will continue to be the age of the computer. The use of computers in todays society is staggering. Examples of computer influence in our lives are everywhere from video games to word processors to robots used in industrial manufacturing. The majority of these computers are controlled by Visual Display Terminals (VDT) operated by a keyboard. In the United States alone, it is estimated that 10 million people now operate VDTs and that by 1990 that figure may rise to 25 million(Salvendy,1982). This level of usage by all segments of society has propelled the study of Human-Computer interaction to the forefront of Ergonomics. The study of Human-Computer interaction can be divided into two areas, the first of which is concerned with software development. Researchers in this area are concerned with programmer productivity, structured programming, specific commands in a given language, and any other aspects of software design(Shniederman,1981). The second area places emphasis on non-programmming aspects of computer usage. Work in this field is often concerned with the use of editors, word-processors and other interactive systems(Moran,1981). It is this area that is of concern here. Any interactive system depends on the mode of entry that is used to convey commands and input information from the user to the computer. The traditional QWERTY keyboard has been joined by a host of other input devices such as light-pens, joysticks, roller balls, mice, speech recognition systems and others. Most of these devices were designed to move a curser about a video screen and thus studies have looked at the speed and accuracy of these curser controllers(e.g. English, Engelbert & Berman, 1967, Earl & Goff,1965).

A problem that is common throughout the area of Human-Computer interaction is the lack of a theoretical base to guide experimental design. Faced with a decision between two text-editors, a choice based soley on empirical data may be sufficient. However, a disregard for the theoretical basis of why one text-editor may result in better performance than the other provides no information for comparing two other text-editors. The development of a comprehensive theoretical base will allow prediction in many similar circumstances. Theory may also allow seemingly disparate results to be organized into a coherent data base. Therefore, this experiment is a comparison of data entry devices to computers couched within the framework of HIP theory. As mentioned earlier, HIP theorists attempt to map the flow of information through the human. One method that has proven useful is the secondary task paradigm.

The secondary task paradigm involves performance of two tasks simultaneously. The subject is given instructions that emphasize the importance of the primary task. That is, the subject is to perform the primary task at the highest possible level, even if this performance causes a decrement in performance of the secondary task. The reasons for choosing one of the tasks as primary differ among experiments; the decision is often made on the basis of theoretical or applied considerations. The basic premise of the secondary task paradigm is that when a subject performs two tasks and those tasks require less capacity than the total (human) system has available, then those tasks may be performed without decrement. If however, the capacity required to perform the two tasks exceeds the capacity that is available, then perforamce degrades and this is usually manifested in the performance of the secondary task. It should be noted that the total capcity demanded for performance of the two tasks is not necessarily the sum of the capacity demanded by performing each task in isolation. Following the additive factors logic of Sternberg(1969), the two tasks may require processing in the same stage (which has a limited capacity). If this interaction occurs, the demand for capacity to perform both tasks will be greater than the sum of the capacity demanded for the two tasks performed in isolation. In addition, there may be capacity demanded by the requirment of performing two tasks at once, Navon and Gopher(1979) call this a concurrence cost. The secondary task may occur independently of the primary task; this asynchronized arrangement yields less information than a synchronized task due to the lack of knowledge of the temporal relationships of the stimuli. The synchronized arrangment does give information about the temporal relationships, i.e. the stimulus for the secondary task follows the stimulus for the primary task at some specific interval, the Inter-Stimulus Interval(ISI). Thus, the models to be discussed here address the synchronous situation and were developed to explain results garnered from this experimental paradigm.

There are two methods that can be used to increase the processing capacity demanded by the two tasks. The first method is to increase the amount of information contained in the stimulus of each task. This can be accomplished by increasing the number of alternatives or changing the frequency of occurence of the alternatives. The second method is to reduce the amount of time between presentation of the stimuli for each of the tasks, the ISI. The period of time after presentaion of the first stimulus, in which presentaion of a second stimulus will result in a delay of response to one or both of the stimuli, has been called the psychological refractory period(PRP, Telford,1931). The effect was so named due to an apparent correspondence with the neural refractory period, however, it was later shown not to be mediated by the same mechanism(Kantowitz,1974a). The theories that evolved to explain this phenomenom centered around a single-channel limited capacity model(e.g. Broadbent,1958, Welford,1959). Although differing on the locus of processign capacity limitations, the theories agreed that at some point in the processing, the second stimulus must wait until processing for the first stimulus is complete(i.e. a processing bottleneck). This explanation of the single-channel model does not, however, explain the findings of a delay of the response to the first stimulus, especially when no response is required to the second stimulus. In an effort to explain these results, models were proposed that extended the single channel model.

Moray(1967) and Kahnamen(1973) proposed models which did not have a limited capacity, but a dynamic supply of capacity. These variable-allocation capacity models do not explicitly retain the processing bottleneck from the single channel model. Instead they hypothesize a variable supply of capacity which increases with respect to the capacity demanded by a task, regardless of the intentions of the subject. An apparent bottleneck occurs because the capacity of the system increases at a slower rate than it is demanded by the performance of the task. The variable-allocation model can explain a delay in the response to the first stimulus, if the second stimulus is assumed to require capacity, even if a response to the second stimulus is not required. If this is true, then the total system capacity, although increasing, is exceeded by the total demand of capacity by the two stimuli. However, it is not clear why the second stimulus would require capacity if it is not to be responded to.

Another model proposed to explain this pattern of results is the hybrid modle proposed by Kantowitz & Knight(1976a). Unlike Kahneman(1973), Kantowitz & Knight retain the limited capacity of the singel channel model. This model, shown in figure 1, extends the response-conflict model of Kantowitz(1974b). The hybrid model proposes at least two parallel stages, S1 and S2, and a response organization and executon stage, S3. The limited capacity feeds the response stage as well as the earlier stages. The allocation of capacity between the earlier stages is determined by the relative importance placed on the primary and secondary tasks(Kantowitz & Knight,1976a). In addition to traditional stage models, network models may prove useful.

Critical path analysis extends the basic logic of Sternberg's additive factors to allow analysis of parallel processes or stages(Scweickert,1978). The basic premise of critical path analysis is that the total reaction time is the sum of the durations of all processes on the longest path of the network, i.e. the critical path. The critical path of the network represented in figure 2 is the process labelled 'E'. The duration of this path is 15 units, the longest in the network. The method requires that certain conditions exist for the analysis to apply to a reaction time distribution. The first is that the processes must be able to be represented in a directed acyclical network. This is illustrated in figure 2, none of the processes feedback to any process that is on the same path as itself. Secondly, no process can begin until those precediing it on a path have finished. The third condition is that every process has a duration. The processes can be either sequential, joined on a directed path, for example, processes 'A' and 'B' in figure 2. If a process 'A' precedes a process 'C' as in figure 2, and if some other process 'B' is executed in parallel with 'A' and takes longer than 'A', then the time when 'A' is completed will not be crucial in determining when 'C' will start. The amount of time that 'A' can be delayed without delaying the start of 'C' is the slack for 'A' with respect to 'C' and is written S(AC). In figure 2, 'A' can be delayed for 3 units before it delays the start of 'C', thus S(AC)=3 units. The amount of time that a process 'A' can be delayed without delaying the response is the total slack for that process and is written S(Ar). In the example in figure 2, S(Ar)=6 units. The coupled slack for 'A' and 'C' is equal to the total slack for 'A' minus the slack for 'A' with respect to 'C' and is denoted as K(AC)=S(Ar)-S(AC).

Use of critical path analysis will allow determination of wether the processes are executed in series or in parallel. If the processes are in parallel and the prolongations of 'X' and 'Y' are sufficently long to prolong the response if delayed seperately then equation 1 will hold.

$$T(X, Y)=MAX(T(X,0)\ T(0, Y)) \quad (1)$$

Thus the total time that the response is delayed is equal to the maximum delay in the response caused by delaying either 'X' or 'Y' seperately. If however, 'X' and 'Y' are arranged sequentially and the prolongations are large enough to make both 'X' and 'Y' critical, then equaton 2 should hold.

$$T(X, Y)=T(X,0) + T(0, Y) + K(XY) \quad (2)$$

The couple slack in equation 2 should be constant for all prolongations of 'X' and 'Y' which are long enough to make 'X' and 'Y' critical. For example, in figure 2 the coupled slack of 'A' and 'C' should always be 3 units as determined above. Thus, if 'A' is prolonged by 11 units, the resulting delay in response time is 5 units. If 'C' is seperately prolonged by 6 units, then the resulting increase in response time is 3 units. If both processes are prolonged by those amounts at the same time then the response is delayed by 11 units. Thus, when these results are placed in equation 2, the value yielded for K(AC) is 3 units. Therefore the equation holds for these prolongations. If this relationship does not hold, then the reation time distribution does not lend itself to critical path analysis.

These models relate especially well to the evaluation of mental workload. This area attempts to determine the level of mental workload over a wide variety of tasks. This relates directly to the amount of capacity demanded by the task. However, as the models point out, this demand does not exist in isolation. Instead the task demand interacts with allocation of capacity and availability of capacity to determine task performance. These models explicitly address those issues and thus are well suited for extension to the evaluation of mental workload. However, these performance measures represent only one of three major thrusts of research in the field of mental workload. The other two areas are subjective evaluation and physiological measures.

Subjective evaluation is, as the name indicates, essentially asking the subject how much mental workload is associated with a task just after the task has been performed. These evaluations are structured into card sorting procedures or rating scales. One of the most fully developed evaluations is the subjective workload assessment technique (SWAT). The development and validation of this technique are discused by Reid, Shingledecker, Nygren & Eggemeier(1981) and will not be restated here. The obvious advantages of subjective evaluations are that they are cheap and easy to administer. However, the ease of use of this technique ends after administration. The results may be very difficult to interpret. For example, the SWAT system rates three dimensions of an event: 1) time load, 2) mental effort load and 3) psychological stress load. Each of these three dimensions are rated as 1,2 or 3 for each event, with three representing greater workload. These three values then define a cell of a previously developed sacle which yields a single numerical value. These numerical values are the workload values associatied with that particular event. Applications of SWAT have yielded accurate discriminations between workload of two events that are intuitivly associated with different workload levels such as landing an aircraft in adverse weather and landing in good weather. While this technique yields differences between the two events, it does not give us information about how much more workload one event causes than the other. The scale must certainly be ordinal and thus does not contain the information about how much more workload a task is associated with as compared to another. For example, the we cannot tell if a value of 200 represents twice as much workload as a value of 100 or if the change from 100 to 200 is the same as the change from 200 to 300. An additional drawback to subjective evaluations is that they must be collected after completion of the task to be evaluated, not 'on-line' during the task. Even given these difficulties, the ease of administration and lack of expense will assure that subjective ratings will continue to be a major technique for evaluating mental workload.

Another area that has spurred a great deal of interest is the use of physiological measures to evaluate workload. These measures have the advantage of being unobtrusive and can be collected 'on-line' as the task is being performed. Some of the more popular measures include electroencephalogram, galvanic skin response, and sinus arrhythmia. While these measures are promising, they share a major problem with the subjective evaluations, i.e. a difficulty in interpreting the data. Wierwille & Conner(1983) compared twenty physiological measures of mental workload and concluded that few, if any, are proven to the degree that they could be widely applied as a measurement technique of mental workload. O'Donnell(1981), while discussing a battery of nine measures designed for the Air Force, suggests that the major uncertainties about the measures lie in the sensitivity to changes in workload. Thus, our measures may have not progressed enough to be useful on a wide range of tasks and people.

At this time, therefore, behavioral secondary-task techniques seem to be the avenue that will yield the most information about mental workload. This is not to discount subjective evaluations and physiological measures as unimportant. Indeed, as Eggemeir & O'Donnell(1982) argue, any single measure of mental workload is doomed to failure. However, a fusion of the three areas is well beyond the current state of research and thus development of the best single measure seems an appropriate place to begin.

Method

Subjects. Subjects were sixteen volunteers from the pyschological sciences pool of subjects at Purdue University. The subjects participated to fulfill a requirement of an introductory psychology course. Participation in this experiment was limited to subjects having little or no computer experience. This was defined as having taken no computer courses or having any other associations with computer usage such as having a home computer. The subjects were divided randomly into two groups defined by mode of entry. Each subject participated in two one hour sessions on succesive days.

Apparatus. The visual display consisted of a virtual keyboard containing the digits from zero to nine. This arrangement is displayed in figure 3. The display was produced on a Sony televison monitor model number KV-1206, driven by an Apple II microcomputer model number AA110408. The modes of entry were the standard Apple keyboard and Symtec light-pen. Reaction times and responses to the digit entry task were also recorded by the Apple II. An Automated Data Systems (ADS) 1800E microcomputer was employed to produce the tone pairs which represented the secondary task. This was accomplished by driving two sonalerts of 2900 Hz and 4900 Hz. The subjects classified tone pairs by depressing the appropriate key of a four key keyboard. The ADS also recorded the reaction times and responses to these tone pairs.

Procedure. The subjects were greeted by the experimenter and read instructions emphasizing the speed of the data entry task. The use of data entry as the primary task allowed evaluation of two data entry devices, a keyboard and a light-pen. These two input modes defined the two between-subjects groups so that every subject used just one entry device. The data entry task consisted of two levels of complexity, entry of the digit itself (N), and entry of the digit plus one (N + 1). In the latter case, if the stimulus digit was a nine, the subjects were instructed to input a zero. The secondary task was a tone classification task. Two tones, 2900 Hz and 4900 Hz, were factorially combined to produce four arrangements: high-high, high-low, low-high, and low-low. Each one had a duration of 50 msec and the interval between tones was 50 msec. The tone pairs were presented following the digit stimulus for the first task, following an ISI of 0,100 or 200 msec. The ISIs were randomized within each block. The subject responded with a key press of the key corresponding to the appropriate tone pair. The secondary tone classification task involved two levels of difficulty. The first level was a two alternative choice-reaction time task, the tone pairs were high-high or low-low. Thus, the presentation of the stimulus produced one bit of information. The second level consisted of classifing all four pairs of tones, a four alternative choice-reaction time task, therefore the stimulus information was two bits. The first session, considered practice, consisted only of single stimulation blocks. This was to allow subjects to establish proficiency on the stimulus-response mappings prior to the addition of a second task. The second day consisted of all four single stimulation blocks and all four of the double stimulation blocks. Each block consisted of 66 trials, the first three and last three of which were deleted from analysis. The digits were randomized within each block under the constraint that each occur equally often. One double stimulation trial would proceed as follows. A warning beep was sounded 500 msec prior to digit presentation and coincided with a fixation point on the screen. Following an ISI of 0,100 or 200 msec, the tone pair was presented. The subject was intructed to respond to the digit first and then to the tone pair. In addition to verbal emphasis on the digit entry task as primary, the subjects recieved visual feedback of performance on this task. The mean reaction time and errors were displayed on the screen following each block of trials. The purpose of this feedback was to produce primary focus on the digit entry task.

Design. Two between-subjects groups were defined by mode of entry; kyboard and light-pen. Eight subjects were randomly assigned to each group. Each subject participated for one hour on two succesive days. On day one, subjects participated in all four single stimulation conditions and were given an additional block of practice on the more difficult variant of each task for a total of six blocks. The second day consisted of participation in all four double stimulation single stimulation blocks and all four double stimulation blocks. A Latin square was used to define eight orders of presentation. Thus the experimental design was a 2(mode of entry) X 2(levls of digit entry) X 2(levels of tone classification) X 2(single vs. double stimulation). The dependent variables were reaction times and errors for both the digit entry and the tone classification.

Results

The results presented here are a preliminary analysis of the data. Complete analysis will be presented in a paper at a later date. The data presented are the primary task results of sixteen subjects, eight in each of the groups defined by the data entry device. The primary task manipulation, inputting the digit plus one, increased the mean RT for both entry groups above the mean RT level of entering the digit itself. The mean RT for entering the digit itself was 1,125 msec. and for the digit plus one, it was 1,380 msec. This comparison was significant $F(1,14)=7.454$, $p=.015$. The light-pen group did show faster entry times, 1,155 msec. than did the keyboard group, 1,350 msec. However, this comparison failed to reach significance. The mean RT increased as the level of the secondary task increased. Thus for level 0, i.e. single stimulation, the mean RT was 1,165 msec., for one bit of information the mean RT was 1,246 msec. and for 2 bits mean RT was 1,345 msec. Again, this comparison was not significant. In addition, a Dunnett's t comparison showed no differences between conditions. These results are displayed in figure 3. Analysis of the error data revealed no significant differences, thus ruling out a speed-accuracy trade-off.

Discussion

The preliminary results reported here are only from the primary task. Any discussion of performance in a dual task situation must be in terms of both tasks. However, those results obtained are in accordance with the predictions made. That is, a significant difference between N and N+1 entry and increases in

secondary task level produce increases in mean RT but this effect does not reach significance. A complete discussion will be provided pending completion of analysis.

References

Attenave, F. Some informational aspects of visual perception. Psychological Review, 1954, 61, 3, 183-193.

Broadbent, D.E. Perception and Communicatons. London: Pergamon Press, 1958.

Eggemier, F.T. & O'Donnell, R.D. A conceptual framework for development of a workload assessment methodology, Workshop mental worload assessment, Human Factors Society Annual Meeting, Norfolk, Virgina, October, 1984.

English, W.K., Engelbart, D.C. & Berman, M.L., Display selection techniques for text manipulation. IEEE transactions on Human Factos in Electronics HFE-8, 5-15, 1967.

Kahneman, D. Attention and Effort, Englewood Cliffs, N.J.: Prentice Hall, 1973.

Kantowitz, B.H. Double Stimulation. In B.H. Kantowitz (Ed.), Human Information Processing: Tutorials in Performance and Cognition. Potomac, Md.: Erlbaum, 1974(b).

Kantowitz, B.H. Interfacing Human Information Processing and Engineering Psychology. In W.C. Howell and E.A. Fleishman(Eds.), Human Performance and Productivity. Volume 2, Hillsdale, N.J.: Erlbaum, 1981.

Kantowitz, B.H. & Knight, J.L. Testing Tapping Timesharing. II. Acta Psychologica, 1976, 40, 343-362.

Miller, G.A. What is information measurement ? American Psychologist, 8, 1, 1953, 3-11.

Moran, T.P. An Applied Psychology of the User, Computing Surveys, 13, 1, 1981, 1-11.

Moray, N. Where is Capacity Limited ? A Survey and a Model. Acta Psychologica, 1967, 27, 84-92.

Navon, D. & Gopher, D. On the Economy of the Human-Processing System. Psychological Review, 86, 3, 1979, 214-256.

O'Donnell, R.D. Physiological Measures, Proceedings of the International congress on Cybernetics and Society, IEEE, Atlanta, Ga., 1981.

Reid, G.B., Shingledecker, C.A. & Eggiemier, F.T. Application of Conjoint Measurement to Workload Scale Development, Proceedings of the Human Factors Society 25th Annual Meeting, Rochester, N.Y., 522-525.

Salvendy,G. Human-computer Communications with Special Reference to Technological Developments, Occupational Stress and Educational Needs, Ergonomics, 1982, 25, 6, 435-447.

Schweickert, R. A Critical Path Generalization of the Additive Factor Method: Analysis of a Stroop Task, Journal of Mathematical Psychology, 1978, 18, 2, 105-139.

Shannon, C.E. A Mathematical theory of Communications, Bell Systems Technical Journal,1948, 7, 379-423.

Shneiderman, B. Software Psychology, Cambidge, Mass.:Winthrop, 1980.

Sternberg, S. The Discovery of Processing Stages: Extensions of Donders Method, Acta Psychologica,1969, 30, 276-315.

Telford, C.W. The Refractory Phase of Voluntary and Associative Responses, Journal of Experimental Psychology, 1931, 14, 1-36.

Welford, A.T. The "Psychological Refractory Period" and the time of High Speed Performance: a Review and a Theory. British Journal of Psychology, 1952, 43,2-19.

Wierwille, W.W. & Conner, S.A. Evaluation of 20 Worload Measures Using a Psychomotor Task in a Moving-Base Aircraft Simulator, Human Factors,25, 1, 1983, 1-16.

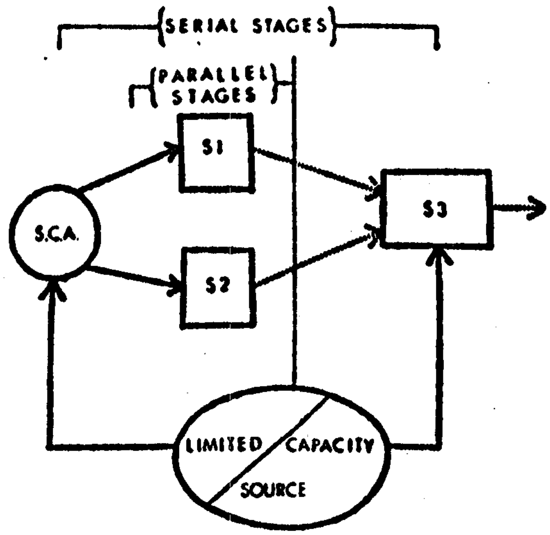


Figure 1 The hybrid model of Kantowitz and Knight (1976a). Solid lines represent capacity allocation to stages, whereas dotted lines represent information flow between stages. The information flow inputs to stages 1 and 2 (S1 and S2) have been deleted to improve the clarity of the diagram. A limited-capacity source dynamically feeds both S3 and a static-capacity allocator (S.C.A.), which partitions capacity between S1 and S2. S1 and S2 operate in parallel and S3 operates in serial with both preceding stages. Each stage can potentially be broken down into smaller stages depending on the level of analysis. For example, S3 is a molar representation of response selection, execution, and control processes. [From Kantowitz and Knight (1976a).

Figure 2

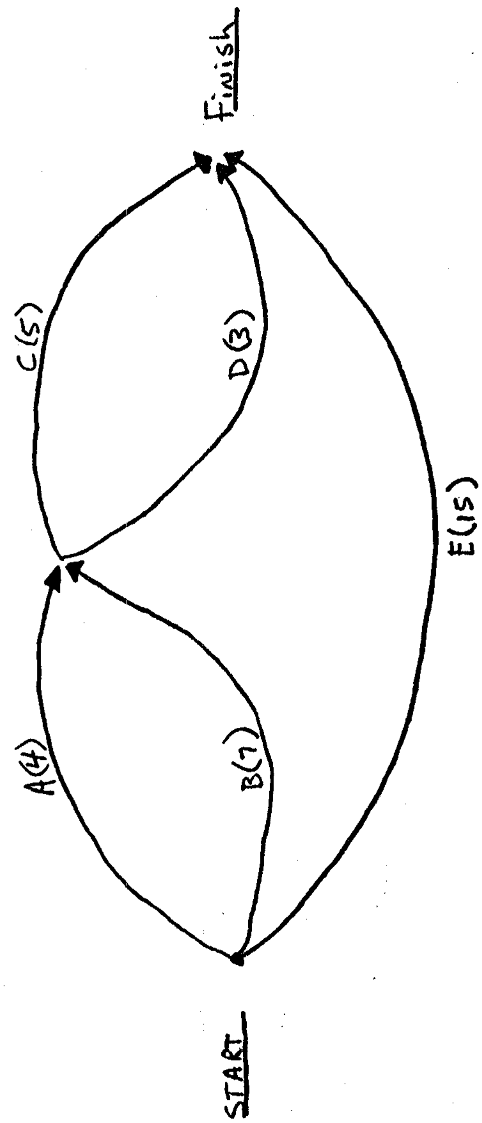
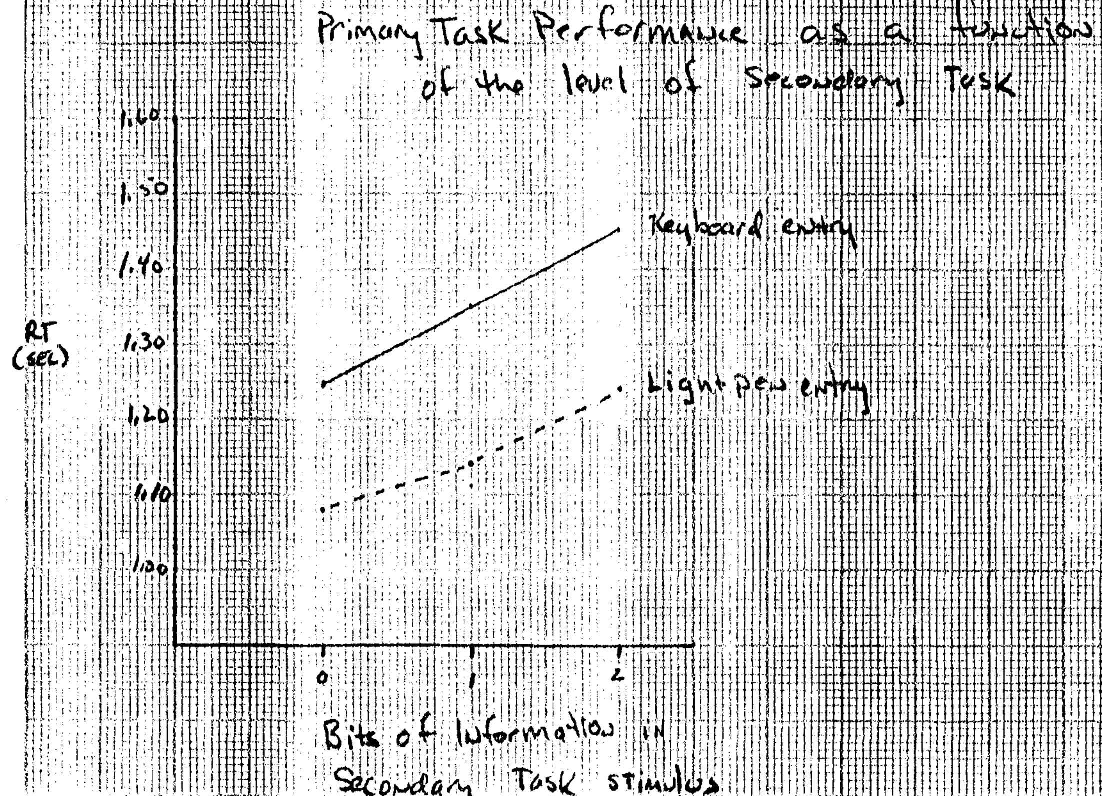
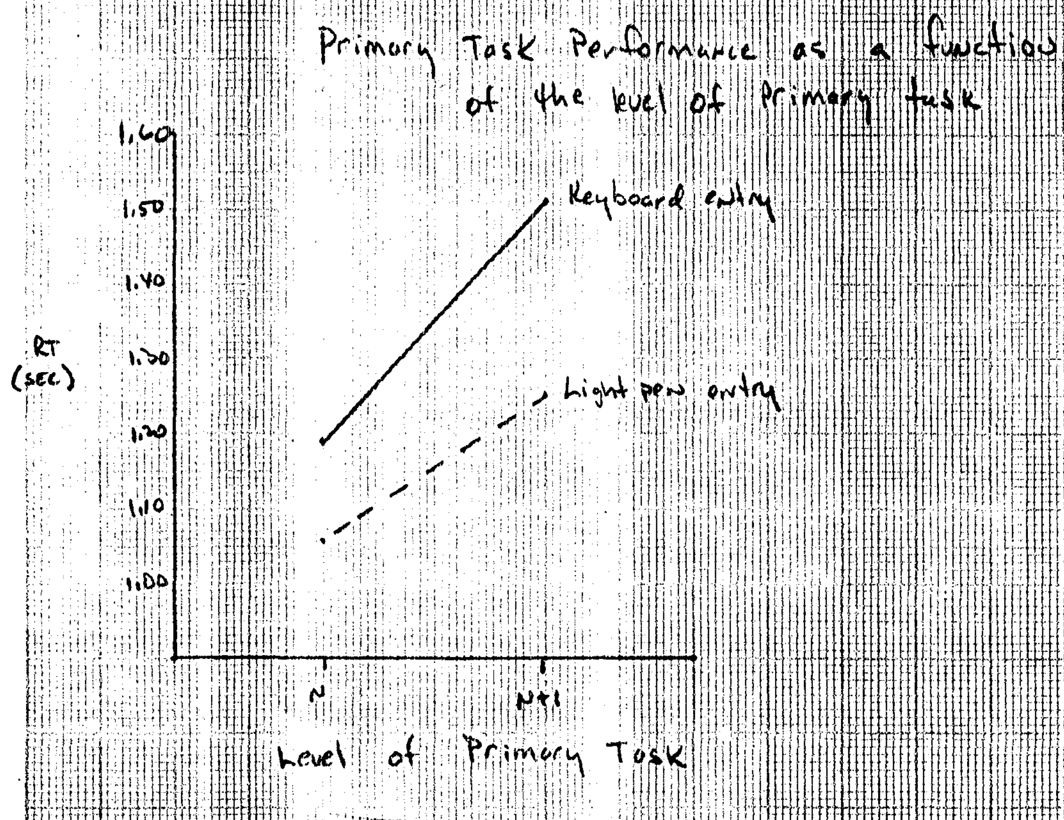


Figure 3


Primary Task Performance as a function of the level of Secondary Task
1.60
1.50
1.40
1.30
1.20
1.10
1.00
RT (sec)
Keyboard entry
Light pen entry
0
1
2
Bits of Information in Secondary Task stimulus


Primary Task Performance as a function of the level of Primary task
1.60
1.50
1.40
1.30
1.20
1.10
1.00
RT (sec)
Keyboard entry
Light pen entry
N
N+1
Level of Primary Task

Auditory choice-reaction time as a function of stimulus information and dimensionality

MarySue Weldon, Patricia Casper, and Barry H. Kantowitz
Purdue University

Many recent studies of aircraft pilot workload have utilized a secondary-task technique to measure the workload demands of simulated flight. Very often a probe reaction-time task has been selected as the secondary task. While a simple reaction task has been widely used, there are both practical and theoretical difficulties associated with using simple Donders' A-reaction probes as secondary tasks (Kantowitz, 1984; Klapp, 1977). For example, obtaining a null result with a simple probe has been interpreted as revealing no workload effects. But, an alternative explanation that a simple probe task presents too low a secondary-task load to reveal possible attentional effects cannot be dismissed without investigating more complex or more difficult secondary tasks. Furthermore, theoretical models of timesharing behavior (Kantowitz & Knight, 1976) suggest that in many instances a simple probe task will be insensitive to small variations in primary-task attentional demands. Such considerations have lead to increased use of choice-reaction secondary tasks in place of simple probe tasks.

Since a pilot flying under IFR conditions must constantly monitor visual cockpit displays, the auditory modality is an obvious choice for the secondary task. However, auditory stimuli can vary across several dimensions including frequency, timbre and other factors related to the stimulus waveform. While pure tones are most often used in laboratory research, it becomes more difficult to perceptually discriminate among pure tones as the number of possible tones increases. This was not a problem for earlier research where only a single tone was used as a probe, but now this difficulty emerges when choice-reaction secondary-tasks are utilized. Using multi-dimensional auditory stimuli reduces perceptual confusion. Furthermore, it is possible that a set of multi-dimensional stimuli will be learned faster than an equivalent set of pure tones. Finally, airplane pilots may have poorer hearing than the general population since FAA licensing criteria do not stress auditory processing and indeed, there are many anecdotal instances of pilots whose hearing is poor enough to require some sort of hearing aid having little trouble renewing their licenses. Thus, a population of professional pilots might have more than average difficulty perceiving pure tones in a secondary-task paradigm.

The present study evaluates an auditory-reaction task using subjects drawn from a population of college undergraduates. If results show that this relatively young population cannot perform as well with tones as with multi-dimensional auditory stimuli, then replication of this experiment using subjects drawn from a population of professional pilots is in order. Furthermore, the tendency to increase the number of auditory alerts in newer planes which have from 14 to 17 alerting signals (FAA, 1977), despite recommendations that only four or five signals should be used (Cooper, 1977) also argues for more detailed study of reactions to auditory stimuli.

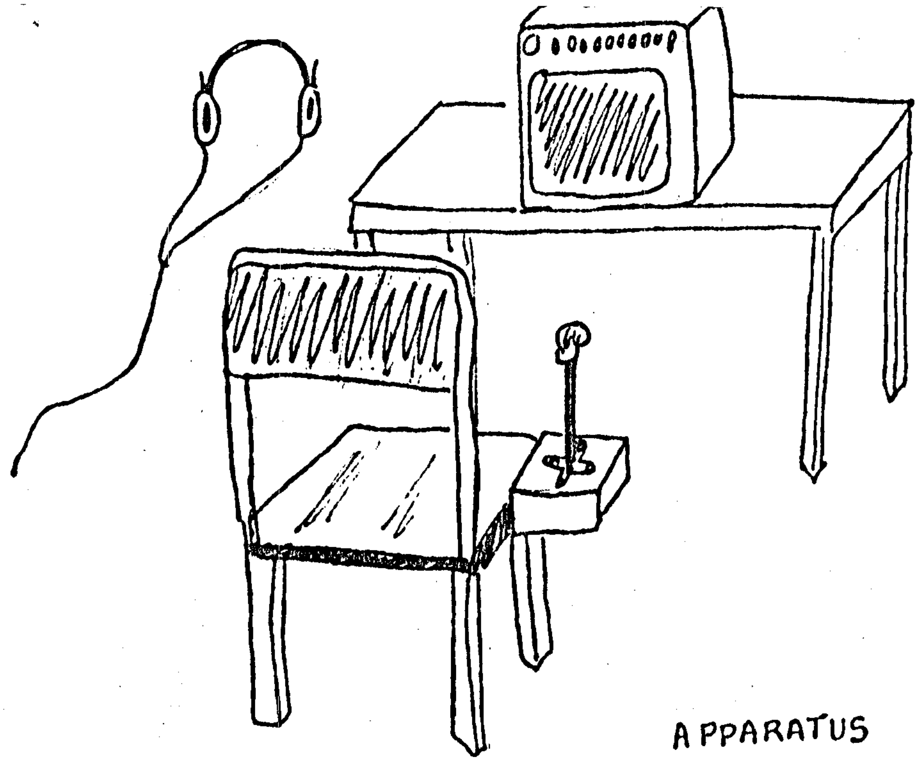
METHOD

Materials and Design

Twenty-four male students from an undergraduate introductory psychology course participated in this experiment. The subjects each received one hour of credit for their participation, partially fulfilling a class requirement for research involvement. All of the subjects were right handed (a requirement due to the nature of the apparatus) and had no hearing deficits.

The between-subjects variable was the type of stimulus presented. Half of the subjects were given pure-tone stimuli and half of the subjects heard stimuli consisting of beeps and buzzes. All of the subjects were given two- and four-choice reaction time tasks. Due to restrictions imposed by the equipment, all of the subjects in the pure-tones condition were run prior to the subjects in the beep-buzz condition. Subjects in both conditions, however, were randomly assigned to one of six orthogonal orders of presentation of the four blocks of two and four-choice reaction time tests: 2244,2424,2442,4422,4242,and 4224.

An Apple II computer was used to randomly generate 4 tones of 741,1176,1961 and 4000 Hz, each of which was presented continuously for 300 milliseconds on a tone trial. To produce the tones, a program was used which accessed certain memory locations in the computer which contained clicks. The frequencies at which these clicks were emitted determined the pitch of the tones. A Cyborg model 91A ISAAC computer interface generated random presentations of 4 beep and buzz combinations: beep-beep, beep-buzz, buzz-beep, and buzz-buzz. The first half of a beep-buzz stimulus was presented for 100 milliseconds, followed by a 100 msec silent pause, followed by the second half of the beep-buzz pair which also lasted for 100 msec. With respect to response positions, the beep-beeps were equivalent to the high tones, the buzz-buzzes to the low tones, the beep-buzzes to the middle-high tones and the buzz-beeps to the middle-low tones. A Realistic brand (model SA-10) solid state stereo amplifier was employed to amplify the sounds produced by the computer. Intensity levels for the "low" tones and the buzzes were measured at 63 dB(A) SPL using a General Radio brand 1565-D decibel meter with a 1560-P83 earphone coupler (9A type) and a non-standard cushion. The loudest tones were measured at 70 dB(A) SPL. The intertrial interval was five seconds for both types of stimuli. Stimuli were presented to subjects over Grason Stadler model TDH39-300Z headphones and visual feedback was provided on a Sony Trinitron color television serving as a video display terminal. The joystick device was mounted on the right-hand side of a chair so that the subject could sit comfortably and activate the lever. The lever had a round plastic knob on the top and could be moved in four directions: forward, backward, left and right. See diagram of apparatus below.



APPARATUS

Procedure

Upon arrival at the experiment, subjects were told that they would receive training on a reaction-time task to tones, and would later be tested on the task. They were also informed that their performance on the training task would not be recorded. In order to fix the trade-off between speed and acuracy, the subjects were told to work towards a 95 to 98% accuracy criterion on the two-choice reaction time task. The subjects were instructed to move the joystick into the forward position if a high tone was heard, and into the backward position if a low tone was heard. A diagram with flashing arrows indicating the correct response position was presented on the screen for the first block of 24 training trials in both the two-and four-choice task in order to help the subjects learn the correct responses. Before the first block of trials (both two- and four-choice) the stimuli were demonstrated for the subjects three consecutive times. Thereafter the stimuli were presented once before each block of 24 trials. After each response the subject's reaction time in milliseconds and their accuracy was presented on the screen in front of the subject. Subjects were given three seconds to respond to each stimulus, after which the response was scored as an omission error by the computer. After each block of 24 training trials the subject's mean reaction time and accuracy rate was presented on their screen, along with a message advising them to respond faster if their accuracy was greater than 98%, or a message advising them to try and be more accurate if their accuracy was below 95%. If a subject achieved the 95-98% accuracy rate, the message "Well Done!" appeared on the screen.

Termination of training on the two-choice task occurred when a subject reached criterion on a block of 24 trials without the arrow prompts, or after two consecutive blocks with 100% accuracy had occurred. During four-choice training the subjects were instructed to continue responding as before to the high and low tones, and two new stimuli were introduced: a "low-middle" and a "high-middle" tone, to which they were to respond to the right and to the left, respectively. Subjects were again told to work towards a 95-98% accuracy level. Training on the four-choice task continued as detailed above for the two-choice task, and ended when the 95-98% criterion was met on a trial without arrow prompts. Subjects were allowed a two-minute break after reaching criterion in the four-choice training task. Testing consisted of four blocks of 48 trials each.

```
                    BEEP-BEEP or
                      HIGH

                        ^
BEEP-BUZZ or            ^           BUZZ-BEEP or
MID-HIGH          <<<< >>>>         MID-LOW
                        V
                        V

                    BUZZ-BUZZ or
                      LOW
```

STIMULI AND RESPONSE DIAGRAM

Results

The dependent variable of primary interest in the data analysis is bits per second of transmitted information (H(T)/sec) computed from the reaction time and accuracy data. This measure is appropriate because it accounts for both reaction time and error data in computing the amount of information the subject is processing accurately per unit time. Therefore, the two- and four- choice tasks are more directly comparable. Other dependent variables of interest include mean accuracy and reaction time, and trials to criterion.

Trials to criterion. In order to determine whether the four-choice task is more easily learned using the beep-buzz (BB) or the pure tone (PT) stimuli, a comparison was made of the trials required to reach four-choice criterion in the two stimulus conditions. An analysis of variance revealed no significant difference between the BB (x = 3.92) and PT (x = 6.75) conditions, $F(1,22)=3.79$, $p > .05$. Because the intertrial interval was constant (5 seconds), subjects in the two different stimulus conditions required approximately the same amount of time to reach criterion. Therefore, significant differences between the two stimuli in the test trials cannot be attributed to differential amounts of time spent learning the reaction time task.

Block and order effects. Preliminary analyses of variance were performed to test for order and block effects. Order refers to the order in which the two-and four-choice tasks were presented during testing. An order effect might indicate differential carryover as a function of different presentation orders. Blocks are the the four sequential test periods; each block tested either a two- or four- choice task, depending on the order to which the subject was assigned. Block effects would reflect a change in performance as a function of time, practice, or fatigue.

Transmitted information for the orders and blocks are presented in Table 1. No significant differences in H(T)/sec were found among orders, $F(5,12) = .88$, $p > .05$, or among blocks, $F(3,36) = .88$, $p > .05$. Furthermore, no significant interaction was obtained between these two factors on this measure $F(15,36) = .90$, $p > .05$.

Reaction time and accuracy data are presented in Table 2 and Table 3, respectively. These data reveal a slightly different pattern of results than that obtained for transmitted information. With respect to the order factor, no main effects of reaction time or accuracy were found. This suggests that the order in which the two- and four-choice tasks were presented did not affect overall performance. With respect to the block factor, however, a significant main effect was found in the accuracy of performance, $F(3,36) = 3.21$, $p < .05$. Accuracy increased slightly over the first three blocks (Block 1: x = 94.79%; Block 2: x = 95.14%; Block 3: x = 96.09%), then declined during the fourth block (x = 92.8%). The decline in accuracy probably reflects fatigue or disinterest. Since it did not interact with the stimulus type, however, it affected performance equally under both stimulus conditions. The decline in accuracy does not reflect a major decrement in performance; recall that H(T)/sec did not decline significantly across blocks. Therefore, block effects appear to be negligible, and do not pose a threat to straightforward interpretations of the stimulus and task difficulty effects, which are ofprimary interest.

The block by order interaction was significant for the accuracy measure, $F(15,36) = 6.61$, $p < .001$, and also for the reaction time measure, $F(15,36) = 48.50$, $p < .001$. This interaction is an artifact of the counterbalancing of the two- and four-choice tasks across blocks, as inspection of Table 2 and Table 3 reveal. The magnitude of a measure in a particular block within a particular order condition depends on whether that order requires a two- or four- choice task in that block. In other words, task difficulty is confounded with the block by order interaction. Thus, this interaction does not reflect a true carryover effect. This assertion is further supported by the lack of a block by order interaction in the H(T)/sec variable. Recall that H(T)/sec makes the two- and four- choice tasks directly comparable, eliminating the inherent differences in reaction time and accuracy.

For the reasons cited above, there appears to be no evidence for carryover effects. Therefore, the primary analysis was conducted in a straightforward manner.

Stimulus and task difficulty effects. Table 4 displays the average H(T)/sec for the two auditory stimuli and two levels of task difficulty. Neither the main effect of stimulus type $F(1,12) = 1.58$, $p > .05$, nor of task difficulty $F(1,12) = .24$, $p > .05$, are significant . The interaction between stimulus type and task difficulty, which is the outcome of interest, is significant, $F(1,12) = 8.69$, $p < .05$. Notice that in the beep-buzz condition the rate of information transmission increases as task difficulty increases. In the pure tone condition, however, the rate of information transmission decreases as task difficulty increases.

Inspection of the reaction time and accuracy data in Table 4 indicate that the differences in performance reflected in H(T)/sec are primarily due to differences in accuracy among the conditions, and not to differences in reaction time. With respect to reaction time, only the main effect of task difficulty was significant, $F(1,12)= 469.62$, $p < .001$. As expected, the two-choice tasks ($x = 467$ msec) were performed more quickly than the four-choice tasks ($x = 802$ msec). It is well-known that choice reaction time increases as the number of alternative choices increases. No other main effects or interactions revealed significant differences in reaction times as a function of treatment conditions.

The treatment conditions have their major effects on the accuracy of performance. There was a significant main effect of auditory stimulus type, $F(1,12) = 7.36$, $p < .05$. Performance was more accurate with the beep-buzz stimuli ($x = 96.18$) than with the pure tones ($x = 93.23\%$). With respect to task difficulty, performance was significantly more accurate on the two-choice task ($x = 98.4\%$) than on the four-choice task ($x = 90.9\%$), $F(1,12) = 55.65$, $p < .001$. Most importantly, the interaction between stimulus type and task difficulty was significant, $F(1,12) = 9.01$, $p < .05$. Inspecting the means in Table 4, it can be seen that accuracy is almost identical in the two-choice task for both types of stimuli. In moving to the four-choice task, however, there is a larger decrement in accuracy in the pure tone condition than in the beep-buzz condition.

Recall that reaction time showed no differential increase in the four-choice task as a function of stimulus type. Thus it appears that decreases in accuracy account for the inferiority of the pure tones in the four-choice task.

Effects of Loudness. Due to equipment limitations, it was not possible to equate the loudness of all auditory stimuli on the dB(A) scale. In order to determine whether these differences affected the experimental outcome, the low tone and buzz-buzz tones (ie., the back position on the joystick) were equated at 63 dB(A). Data from this one position were then analyzed.

Table 5 displays mean reaction times and accuracy as a function of stimulus type and task difficulty. (H(T)/sec cannot be computed with only one stimulus in the stimulus set.) It can be seen that the pattern of results is identical to that of the complete stimulus set (Table 4). Furthermore, analyses of variance revealed nearly equivalent results. With respect to reaction time, only the main effect of task difficulty was significant, $F(1,12) = 227.36, p < .001$. With respect to accuracy, the main effect of task difficulty was significant $F(1,12) = 18.55$, $p < .001$, as was the main effect of stimulus type, $F(1,12) = 6.16$, $p < .05$. Therefore, the only difference between the results for the complete and equal-loudness stimulus sets was the lack of a significant interaction between stimulus type and task difficulty with respect to accuracy. (In fact, this interaction approached significance in the equal-loudness set, $F(1,12) = 4.34$, $p < .06$.)

These results indicate that the effects of stimulus type and task difficulty were not artifacts of the small differences in loudness among the auditory stimuli.

Table 1

Transmitted Information (H(T)/sec) as a Function of
Block and Stimulus Presentation Order

	Block			
Order*	1	2	3	4
1 (2244)	2.05	2.12	2.18	2.07
2 (2424)	1.91	2.10	1.86	1.76
3 (2442)	2.13	1.84	1.98	2.02
4 (4422)	1.87	1.81	1.99	1.89
5 (4242)	2.21	2.26	2.48	2.00
6 (4224)	1.89	1.73	1.82	1.86

* Level of task difficulty for each block is indicated. 2 = 2-choice task; 4 = 4-choice task.

Table 2
Reaction Time as a Function of
Block and Stimulus Presentation Order

	Block			
Order *	1	2	3	4
1 (2244)	460	465	742	765
2 (2424)	467	784	494	838
3 (2442)	471	856	840	485
4 (4422)	808	809	434	420
5 (4242)	731	412	673	420
6 (4224)	907	543	536	874

* Level of task difficulty for each block is indicated. 2 = 2-choice task; 4 = 4-choice task.

Table 3
Accuracy (percent correct) as a Function of
Block and Stimulus Presentation Order

	Block			
Order *	1	2	3	4
1 (2244)	97.9	100.0	92.7	87.5
2 (2424)	98.4	92.1	98.4	87.5
3 (2442)	100.0	91.2	94.3	98.4
4 (4422)	89.6	89.6	98.4	95.8
5 (4242)	90.6	99.0	93.2	96.9
6 (4224)	92.2	99.0	99.5	90.6

* Level of task difficulty for each order is indicated. 2 = 2-choice task; 4 = 4-choice task.

Table 4
Effects of Auditory Stimulus Type and
Task Difficulty for the Complete Stimulus Set

Task Difficulty	Auditory Stimulus: Beep-Buzz			Pure Tone		
	H(T)/sec	RT(msec)	ACC(%)	H(T)/sec	RT(msec)	ACC(%
Two-choice	1.98	461	98.4	1.97	473	98.5
Four-choice	2.16	789	93.9	1.84	814	87.9

Table 5
Effect of Auditory Stimulus Type and Task Difficulty for the Equal-Loudness Stimuli

Task Difficulty	Auditory Stimulus: Beep-Buzz		Pure Tone	
	RT(msec)	ACC(%)	RT(msec)	ACC(%)
Two-choice	461	98.8	472	98.3
Four-choice	721	95.8	748	89.8

Discussion

The results of this experiment reveal an interaction between task difficulty and auditory stimulus type. In the easy, two-choice reaction time task, performance is the same with both types of auditory stimule. In the more difficult four-choice reaction time task, however, performance is superior when beep-buzz patterns rather than pure tones, are used as auditory stimuli. This pattern of results suggests that differential performance on the middle-range tones in the four-choice task may be the source of the interaction, since performance is the same on the extreme tones which are present in the two-choice task.

The superior performance with beep-buzz stimuli in the four-choice task is probably due to the number of stimulus dimensions on which they can be discriminated. Beep-buzzes provide at least three dimensions, including frequency, timbre, and change in tonal characteristics between the first and last segments of the stimulus onset trial (eg. a "beep" then "buzz"). By contrast, pure tones can only be discriminated along the frequency dimension. When task demands are low, as in the two-choice task, this additional information may not be as useful as when the task requires more difficult discriminations, as in the four-choice task.

The interaction between task difficulty and auditory stimulus type with respect to information processing rate is of particular interest (see Table 4). Note that in the the beep-buzz condition, the rate of information transmission ((H)T/sec) increased from the two-choice to the four-choice task. In the pure tone condition, however, the rate of information transmission decreased between the two-choice and four-choice task. This suggests that although the four-choice task imposes increased processing demands, subjects were better able to meet the demands when responding to beep-buzz stimuli rather than pure tone stimuli. Again,beep-buzz tones possess more discriminable stimulus dimensions than do pure tones, providing more information on which to base choice reaction time decisions. Thus, subjects are able to process information at a higher rate.

On the basis of these results, the authors recommend using beep-buzz sounds as auditory stimuli when a four-choice reaction time task is used as a secondary task. Pure tones are difficult to discriminate in the middle ranges, and discriminating between middle-range tone imposes greater processing demands than discriminating extreme high or low pure tones. The resultant inconsistency in task demands and performance across stimulus values is undesirable in a secondary task. Ideally, a secondary task should impose a constant processing demand at all times, so that changes in performance on the secondary task reflect changes in primary task demands, and not changes in secondary task demands.

Of course, the beep-buzz stimuli probably do not provide completely consistent performance across all four stimulus values. Extreme values are inherently more discriminable than middle-range values, and the beep-beep and buzz-buzz tones are probably more discriminable than the beep-buzz and buzz-beep tones. Nevertheless, as shown in Table 4, overall four-choice task performance is better with the beep-buzz stimuli, and this is probably attributable to better performance on the middle-range stimulus values of the beep-buzz stimuli (since two-choice performance is the same for both stimulus types). It is likely that beep-buzz stimuli provide more consistent processing demands across the four stimulus values than do pure tones. Therefore, beep-buzz tones are more appropriate auditory stimuli for a secondary four-choice reaction time task.

The effects of heat and cold on attention

Charles D. Caldwell
Purdue University

Running head: Attention and extreme temperatures

Abstract

Hot (95° F) and cold (41° F) temperature conditions were compared to a medium (68° F) temperature to examine how temperature affects attention; when a narrowing of attention occurs, is it lost of funnelled? Eighteen males were exposed to each temperature for 1 hour; testing began after 30 minutes. Stimuli consisted of four lights positioned horizontally either near or far from the point of fixation and were presented singly or in pairs with either a 60 msec or a 240 msec interstimulus interval. Reaction times and error rates were recorded. There were significant main effects for position of stimulus, interstimulus interval, and temperature ($p < .001$); no significant interactions were consistently found ($p > .05$). The data supported the arousal theory; that heat and cold increase activation and arousal. The data lent support to the theory that attention was affected differently by heat than by cold; cold aroused attention while heat aroused and funnelled attention.

The effects of heat and cold on attention

There has been a wealth of material published in the past concerning the effects of hot and cold environments. Early studies were designed to assess physical work limits and exhaustion under stressful environmental conditions (Haldane, 1905, for example). The next phase of research examined the effects of stressful temperatures on less physically exhausting tasks, such as a study done by Mackworth (1946) who found that the coding of messages was subject to thermal decrement. This task required very little muscular effort, which in this case was not affectd by heat; however, the skill required was. Mackworth also commented on concentration as well; this, too, was negatively affected by heat. A third phase of research is the investigation of how cognitive processes are affected by hot and cold environments; the experiment presented in this paper is of this phase.

Bell (1975) asked subjects to estimate the length of varying time intervals while being exposed to hot or neutral conditions. The results were not statistically significant; subjects could accurately estimate time intervals in any of the temperature conditions. Hence Bell concluded that a human's internal clock was not temperature dependent.

Past research in the area of environmental stress and its effect on cognitive processes includes a study done by Bateman (1981). His experiment was based on two opposing theories: first, that mental tasks that are complex are less tolerant to heat stress than mental tasks that are simple (Hendler, 1963). The second theory (Wing, 1965) stated that the effects of heat stress are the same for all kinds of activity. Bateman tested subjects at six different levels of task difficulty, from simple to complex, during exposure to different levels of temperature (temperatures up to 32.2° C; no beginning temperature was mentioned). In general, the results were that, as temperature increased, performance was degraded for simple tasks but not for complex tasks, and in some cases performance improved for complex tasks.

Bateman commented on both theories. The results of neither Hendler (1963) nor Wing (1965) were duplicated here. The effects of heat stress were not the same for all kinds of activity. complex tasks were tolerant (in some cases they were more tolerant) to heat stress; performance during exposure to high temperatures was statistically equal to or better than performance during exposure to lower temperatures. Bateman's results support the idea that states of arousal are dependent upon the complexity of the cognitive task required as well as the environmental conditions under which they are performed.

Ellis (1982) exposed his subjects to cold conditions (-12° C) in order to examine the effects of such a temperature on serial choice reaction time (classification of a series of digits as "even" or "odd"), simple reaction time (a response was required each time a stimulus was presented), verbal reasoning (the test suggested by Baddeley, 1968), and a Stroop Word Color Test (Stroop, 1935). For the choice reaction time, errors were greater during cold exposure and lags (reaction times greater than Ellis' standard) were not significantly different. Mean simple reaction times were greater during cold exposure. Mean verbal reasoning performance (accuracy) was better during cold exposure than during the pre-exposure condition. Mean time for color naming during cold exposure was not statistically higher than pre- or post-exposure conditions. Ellis attempted to lend support to one of two hypotheses concerning cold stress; an arousal hypothesis (cold increases activation and arousal) and a distraction hypothesis (cold leads to discomfort, which acts as a distraction to ongoing processing). Unfortunately, both hypotheses explained the results equally well; in some cases performance was better during cold exposure (arousal) and in some cases it was worse (distraction). The arousal hypothesis explains why complex tasks are more durable in stressful environments and the distraction hypothesis explains why easier tasks are less durable in stressful environments.

The remainder of this report concerns the present experiment; it was thought that the results would support one of these hypotheses. In the experiment presented in this report, subjects were presented with a cognitive task in a hot, medium, and cold environment. It was predicted that performance would be affected by extreme environmental conditions. Easterbrook (1959) discussed the generalization that cue utilization in any situation tends to become smaller with an increase in emotion or "noxious" external stimulation. On some tasks, a reduction in or narrowing of the range of cue utilization may improve performance if irrelevant cues are excluded. In other tasks, if responding demands a wide range of cues, performance may be hindered when relevant cues are excluded. Based on the findings of Easterbrook (1959) that stress narrows one's focus of attention, an environment that induces stress will lead a subject to narrow his/her field of visual attention to preclude the conscious processing of items on the periphery. Given the total amount of attention possible (TOT) for a subject to devote to a cognitive task under nonstressful conditions, the following formula arises:

$$\text{TOT} = \text{CENTER} + \text{PERIPHERAL} + \text{SPARE} \qquad (1)$$

where CENTER is the amount of attention devoted to stimuli that are well within the focus of attention, PERIPHERAL is the attention devoted to those items in the periphery, and SPARE is the attention devoted to other items or unused. Under stressful conditions, the following equation arises from a narrowed focus of attention:

$$\text{TOT} = (\text{CENTER} + \text{TRANS}) + (\text{PERIPHERAL} - \text{LOST}) + \text{SPARE} \qquad (2)$$

where LOST is the amount of attention lost from the peripheral stimuli and TRANS is either all of or a subset of LOST that is transferred or funnelled toward and gained by the central stimuli. This idea of "funnelled attention" in stressful environments has been reported by McCormick and Sanders (1982) using stress created by high levels of noise, and Bursill (1958) using stress created by hot and humid conditions.

Bursill (1958) used six neon bulbs placed symmetrically in front of his subjects; 20, 50, and 80 degrees to the left and right of the subject's point of fixation, all approximately 73 cm from his (all subjects were male) eyes. The subject was asked to press a corresponding response key when a bulb was illuminated. In addition to this task, the subject had to perform a pursuit motor task. The experimental subjects performed all of this in 41° C environment, while the control subjects had been in a 30° C environment. The reader is directed to the Bursill (1958) report for an in-depth data analysis, but, in general, reliable differences were found between the two temperature conditions in the number of stimuli missed (he did not report reaction times) in relation to their physical position. Thus the phenomenon of "narrowed attention" did occur under hot conditions.

However, although the processing of items in the periphery can be affected by a change in attention, equation 2 does not represent the only possibility for a narrowing of attention. If a "funnelling" effect occurs, then there would be greater attention devoted to the items still within the narrowed focus; an inhibition effect for peripheral stimuli and a facilitation effect for central stimuli. The other alternative is that the attention devoted to those items in the periphery is non-transferable to those items in the center. The following formula arises:

TOT = CENTRAL + (PERIPHERAL - LOST) + (SPARE + TRANS) (3)

where, again TOT is total attention, CENTRAL is attention devoted to the central items, PERIPHERAL is attention devoted to peripherally located stimuli, LOST is attention lost from peripheral items due to stress, and SPARE is attention devoted to neither items in the periphery nor the items in the center. However, equation 3 differs from equation 2 in that none of the attention lost from the peripheral items is transferred to the central items; it is instead added to the OTHER category. Perhaps a "funnelling" effect does not occur; perhaps the attention is simply "lost". In this case there would be an inhibition effect for peripheral stimuli, but no facilitation effect for central stimuli. This experiment will examine the two possibilities of altered attention, whether it is "lost" or "funnelled".

Although the variables used in this experiment will be covered in the methods section, one is discussed here because of its theoretical nature. This experiment provided both single and double stimulation (stimuli were visual). This was done to determine the effects of temperature extremes on the psychological refractory period (Telford, 1931 discusses this PRP), which is a brief period after reacting to the first stimulus during which time a subject's information processing mechanism is "being reset". S/he is unable to react to a second stimulus until the end of this PRP. Kantowitz (1974) discusses what can be inferred about information processing by using an S1-R1, S2-R2 paradigm and varying the interstimulus interval (ISI), or the time between the offset of the first stimulus and the onset of the second. In this experiment, there were two interstimulus intervals used, 60 MSEC and 240 MSEC. Short interstimulus intervals cause an inhibition effect for reaction times as compared to long intervals (Kantowitz, 1974). It was hypothesized that hot and cold environments would cause a further inhibition effect for reaction times as compared to a neutral temperature. Any change in the S2-R2 reaction time at each ISI between temperatures would be due to the effect of the temperature on the PRP.

The independent variables (stimuli location, interstimulus interval, and temperature) and the dependent variables (reaction time and accuracy) are covered in the next section. To summarize this section, each hypothesis will now be formally stated.

Temperature would have an effect on attention in that subjects in hot or cold conditions would have significantly better performance than in a medium temperature condition. Either lower reaction times, lower error rates, or both were predicted; no speed-accuracy tradeoff would result. This would be due to extreme temperatures having an arousal effect on attention, rather than a distracting effect.

The position of the stimulus would affect performance.

Peripherally located stimuli would have significantly higher reaction times or error rates than centrally located stimuli. This would be due simply to the extra distance the eye must move to reach the outer stimuli.

There would be an interaction between temperature and position of stimulus. At extreme temperatures, reaction times, error rates, or both would be greater for stimuli in outer or peripheral positions than for stimuli in inner or central positions. This would be discussed either in terms of a "funnelling" or a "loss" theory for attention.

There would be an effect for interstimulus interval. Performance would be significantly better for a long ISI than a short ISI.

It was hypothesized that the psychological refractory period would be affected by extreme temperatures. If this were to happen, an interaction would emerge between temperature and the interstimulus interval.

A short interstimulus interval itself can be thought of as a factor that would affect attention. In such a case there would be an interaction between the spatial position of the stimulus and the interstimulus interval.

If performance on a complex task is more durable in extreme temperatures than performance on a simple task (the results of Ellis, 1982, stated earlier), and if a short interstimulus can alter attention, then a short interstimulus interval in an extreme temperature would have an effect on performance that was something more than additive factors. This would be demonstrated by a three-way interaction between the spatial position of the stimulus, temperature, and the interstimulus interval.

Method

Subjects

The 18 subjects were all male undergraduate students. Fifteen of the subjects participated in order to fulfill a class requirement for Introductory Psychology. The remaining subjects had responded to an advertisement in the campus newspaper and were paid for their participation at a rate of $3.00/hour. Subjects participated for one hour on each of three days for a total of three hours.

Apparatus

The experiment took place in an environmental chamber manufactured by the Kysor Industrial Corporation. The chamber allowed for controlling humidity, temperature, and luminance, however in this experiment neither humidity nor luminance was a variable. Humidity was kept constant (the humidity control was set to zero and thus did not vary throughout the experiment) and luminance was controlled by the use of one 150W incandescent bulb in the socket located in the chamber. In the chamber were a chair and a table; upon the table was the stimulus/response panel (see figures 1a-c). The panel consisted of four neon stimulus lights (each 1 cm square) symmetrically arranged horizontally across the panel, with distances from the center of 1.5 cm for the inner lights and 22.5 cm for the outer lights. With the subject's eyes 55 cm from the panel and the stimuli 22 cm below the subject's horizontal line of sight, the visual angles from the point of fixation to the inner and outer lights were 1.45 and 20.80 degrees, respectively. Responses to each light were made by depressing one of four keys, each similar in operation and design to a piano key. However, the keys were not separated as the lights were, but instead were placed in the center of the panel so as to be operated without lateral arm movement. The keys corresponded to the light in a one-to-one manner; the left-most key was the correct response for the left-most stimulus, etc. The stimuli were numbered for easy reference; the outer left light was #1, the inner left was #2, the inner right was #3, and the outer right was #4. All stimuli were computer-generated and all response choices and reaction times (precise to one msec) were computer-evaluated, using an ADS 1800E computer.

Procedure

The experimental design was completely within subjects. The three temperature conditions were cold (5°C or 41°F), medium (20°C or 68°F), and hot (35°C or 95°F), resulting in six (3!) different orders. Three interstimulus intervals (60 msec, 240 msec and single stimulation) also resulted in six different orders. The combination of these two variables resulted in 36 different orders (see appendix A) and the experiment was to have one subject in each possible order using 18 males and 18 females. As of this writing, the design had been partially completed; the data for the 18 male subjects had been collected. The completed design would control for any possible order effect. Each subject was tested individually.

The procedure was similar for each of the three sessions, the only change being the temperature of the chamber. Upon entering the laboratory, the subject was shown the experimental chamber and asked to sign a consent form. Although the subject had been told ahead of time the nature of the experiment, this was verbally reiterated along with the fact that, should the subject feel uncomfortable, attrition (leaving the experiment before it was finished) was possible with no loss of credit or compensation (after the first day no consent form was signed; the experimenter repeated the attrition allowance nevertheless). No subjects declined participation at any time.

The subjects were then directed to a dressing room to change into shorts and a t-shirt (each subject was asked to supply a pair of shorts to be worn during each of the three sessions; these articles of clothing, socks, and shoes were all the subjects wore). To allow for temperature adaption, the subject entered and remained in the chamber for 30 minutes and was allowed to read magazines during this time.

Following the adaption period, the experimenter entered the chamber, positioned the subject's chair in front of the panel, and read the instructions to the subject. Subjects were told to sit upright in the chair and to place their index and middle finger of each hand upon the response keys. An explanation was given on how the keys corresponded to the lights. The subject was then ready to begin the four groups of trials. Upon exiting the chamber, the experimenter checked the lighting, humidity, and temperature, and initiated a block of 33 trials. The subject was alone during testing; the experimenter entered only at the end of a block of trials to explain the specific instructions for the next block.

The first block of trials consisted of single stimulation (S1-R1) and was used as practice. Here subjects were told to react as quickly as possible by depressing the key that corresponded to an illuminated light. Lights were illuminated, in somewhat random order, one at a time with a six second lag between each trial. To alert the subject to an oncoming stimulus, a warning tone was sounded 500 msec before each trial; this tone continued until the onset of the stimulus. Upon hearing this tone the subject had been instructed to focus his eyes at the center of the stimulus panel. The subject's reaction time and accuracy were recorded for each trial.

Following this first block of trials were three blocks each consisting of a single interstimulus interval; these were presented in the order determined by the experimental design. The instructions for the SS condition (single stimulation) were the same as those for the practice trials. For the 60 MSEC ISI and the 240 MSEC ISI the instructions included those for the practice trial as well as additional information about the second stimulus. The second stimulus always would appear on the side opposite to the first stimulus (if S1 were on the right, then S2 would be on the left). This removed one bit of uncertainty; subjects then had to make a choice between two stimuli instead of four (the exact instructions can be found in appendix B). In each condition, the subject's response and reaction time for each stimulus were recorded. The 33 trials in each block included one practice trial and eight of each type of the possible S1, S2 combinations of the four stimulus lights with the limitations resulting from the advance information (the combinations did not include the following pairs: 1 and 1; 1 and 2; 2 and 2; 2 and 1; 3 and 3; 3 and 4; 4 and 3; or 4 and 4).

Each block took approximately 6-7 minutes to complete. Finally, the subject left the chamber and changed clothes. At the end of the first and second day of the experiment, the subject was reminded to bring the same clothing for subsequent sessions.

Results

The descriptive statistics calculated here were the appropriate mean reaction times for all correct responses and the percent accuracy of all responses. Nie et. al. (1975) covered this procedure. For the inferential statistics, the decision to use the residual mean square as the error term for corresponding F-ratios resulted from the "never pool" rule (Winer, 1971).

Plotted in figure 2a are the mean reaction times of response 1 (RT_1) for the spatial position of the stimulus (STIM) by the interstimulus interval (ISI) interaction. As expected, RT_1 was higher for the 60 msec ISI than the 240 msec ISI, which is turn was higher than RT_1 for single stimulation. RT_1 was also higher for the outer stimuli (numbers 1 and 4) than for the inner stimuli (numbers 2 and 3). There was a significant main effect for STIM ($MS = 230478$, $F = 8.99$, $p < .001$) and for ISI ($MS = .264 \times 10^8$, $F = 1030.92$, $p < .001$). The interaction was also significant ($MS = 224744$, $F = 8.77$, $p < .001$).

To create compatibility between the graphs, accuracy (percent correct) was converted to error rate (percent error) in all cases, and the error rates (ACC_1) corresponding to the data in figure 2a are plotted in figure 2b. Quite noticable is the extremely high percent error for stimulus 3 during single stimulation. Theoretically, this whould have been one of the easiest stimuli to respond to; it is not followed by a second stimulus and is not in an outer position. Presently this result is unexplained.

The percent error for 60 msec ISI responses were higher for the outer stimuli than the inner stimuli, while the reverse is true for 240 msec ISI and single stimulation. The main effect for STIM was significant ($MS = 4.31$, $F = 59.74$, $p < .001$), as was the main effect for ISI ($MS = 3.24$, $F = 44.93$, $p < .001$) and the interaction ($MS = 2.67$, $F = 37.03$, $p < .001$). The reader may notice that performance was worse for the 60 msec ISI for outer stimuli than for inner stimuli (both RT_1 and ACC_1 were higher for outer stimuli), but there was a speed-accuracy trade-off for responses during the 240 msec ISI (higfher RT_1 and lower ACC_1 for outer stimuli as compared to inner).

Plotted in figure 3 are the mean reaction times (RT_2) and percent error (ACC_2) for the second response of the STIM by ISI interaction. In figure 3a, RT_2 during the 60 msec ISI was higher than during the 240 msec ISI (significant main effect; $MS= .18 \times 10^8$, $F= 548.97$, $p< .001$). There was also a main effect for STIM ($MS= 101817$, $F= 3.09$, $p<.05$). The interaction was not significant ($MS= .21$, $F= 1.99$, $p> .05$). The reader may notice that RT_2 was somewhat higher for outer stimuli than inner, but ACC_2 was somewhat lower for outer stimuli than inner; a speed-accuracy trade-off for both levels of the ISI.

Plotted in figure 4 are the responses for the STIM by temperature (TEMP) interaction. In figure 4a, performance was better (RT_1 was lower) for both the "HOT" and "COLD" conditions than for the "MED" temperature condition. The main effect for temperature for RT_1 was significant ($MS = 212413$, $F= 8.29$, $p< .001$). Here RT_1 was slightly lower for inner stimuli than outer stimuli, but the temperature did not differentially affect RT_1 (no significant interaction; $MS= 13365$, $F= .52$, $p> .05$). Figure 4b represents the corresponding percent error. One would expect somewhat symmetrical error rates for the spatial position of the stimuli, yet stimulus 2 was quite different from 3, as was 1 from 4. However, ACC_1 was not significantly different for each temperature ($MS= .06$, $F= .76$, $p> .05$). Errors were less for the HOT and COLD temperatures than MED, and for outer stimuli rather than inner, yet there was no significant interaction (MS= .06, F= .84, p> .05). Although there was a speed-accuracy trade-off for STIM (for inner stimuli, lower RT_1 and higher ACC_1 than outer), performance (both RT_1 and ACC_1) during HOT and COLD temperatures was either better than or equal to performance during the MED temperature.

Plotted in figure 5 are the performance measures for the second response of the STIM by TEMP interaction. In figure 5a, RT_2 was lower for the HOT and COLD temperatures than for the MED temperature, and this main effect was significant (MS= 113945, F= 3.46, p< .05). However, there was no differential effect for TEMP at each of the spatial positions of the stimuli (no significant main effect; MS= 31890, F= .97, p>.05).

For the error data plotted in figure 5b, ACC_2 was lower during the HOT and COLD temperatures than during the MED temperature, and this main effect was significant (MS= .72, F= 6.81, p< .001). The percent error was higher for the iner stimuli than the outer, yet the interaction between STIM and TEMP was not significant (MS= .03, F= .29, p> .05). For this second response, performance was better (lower RT_2 and ACC_2) for extreme temperatures. When the spatial position of the stimulus was taken into account, there was a speed-accuracy trade-off (RT_2 increased while ACC_2 decreased).

Plotted in figure 6 are the data for response 1 of the ISI by TEMP interaction. As may be seen in figure 6a, RT_1 was lower for the HOT and COLD temperatures than the MED temperature (teh significant main effect was mentioned previously), and RT_1 was lower for the 240 msec ISI than the 60 mesc ISI (this sig. main effect was also mentioned previously). The interaciton between ISI and TEMP was not significant ($MS= 9497$, $F= .37$, $p> .05$). Similar results were found in figure 6b. For ACC_1, the extreme temperatures were associated with lower percent errors for both levels of the ISI (but not single stimulation). The interaction was not significant ($MS= .14$, $F= 1.98$, $p> .05$).

Of special note in figure 6 was that no speed-accuracy trade-off occurred; conditions with low reaction times also had low error rates. Performance during an ISI of 60 msec was definitely worse than during 240 msec; the same was true for the MED temperature when compared to either extreme.

Plotted in figure 7 are the data from response 2 of the ISI by TEMP interaction. The results were similar to response 1. The ISI of 60 msec had better performance scores associated with it (both RT_2 and ACC_2) than the ISI of 240 msec, for both TEMP and ISI (the main effects were mentioned previously). The two factors, though, had no differential effect (no significant interaction) for either RT_2 ($MS= 11646$, $F= .35$, $p> .05$) or ACC_2 ($MS= .08$, $F= .74$, $p> .05$).

Plotted in figures 8 and 9 are the 3-way interaction between STIM, TEMP, and ISI for responses 1 and 2, respectively. In each case (although difficult to discern from figures 8b and 9b), performance was better (lower scores) for the HOT and COLD conditions than the MED condition. However, for the 3-way interaction, neither RT_1

(MS= 4970, F= .19, p> .05), ACC_1 (MS= .08, F= 1.05, p> .05), RT_2 (MS= 23225, F= .71, p> .05), nor ACC_2 (MS= .06, F= .57, p> .05)were significant.

Following Figure 9 is a summary table of the analyses of variance; Table 1a includes data from response 1 (RT_1, ACC_1) and Table 1b includes data from response 2 (RT_2, ACC_2).

Figures 10 through 15 will not be covered in detail as were Figures 2 through 9. These last six figures were the same data collapsed across the spatial position of the stimuli. These recoded figures present the data in a clear manner and give a good indication of just how performance changes in each level of the variables. The inner stimuli were relabeled "1" and the outer stimuli were relabeled "2".

From the data and statistics presented here, the following conclusions were drawn concerning the hypotheses stated previously:

1. Temperature had an effect on performance.
2. The spatial position of the stimulus had an effect on performance.
3. There was no interaction between temperature and position of stimulus.
4. The interstimulus interval had an effect on performance.
5. There was no interaction between temperature and the interstimulus interval.
6. There was an interaction between the interstimulus interval and the spatial position of the stimulus, but only for response 1.

7. There was no 3-way interaction (TEMP by ISI by STIM).

Discussion

The first result that emerged from the data was that performance was almost always better during HOT or COLD temperature conditions. This result was similar to that found by Bateman (1981), that complex tasks such as the one used in this study are durable in stressful environments. This result also lends support to the arousal hypothesis that cold and heat increase activation and arousal and therefore lead to better performance.

In terms of the Easterbrook finding (1959) that "noxious" external stimulation decreases cue utilization, the results found here can be explained by considering that perhaps irrelevant cues were excluded during extreme temperatures from those utilized by the subjects in the medium temperature condition. A significant interaction would have supported the hypothesis that the peripherally located lights would be associated with higher reaction times due to their becoming irrelevant cues.

No interaction arose between temperature and position of stimulus; this may be due to the outer lights not being considered irrelevant (hence lower reaction times) when observed in a cold or hot environment than in the medium temperature condition. However, the fact that there was a main effect for the spatial position of the stimulus lends support to the idea that attention is not equally distributed in any condition; that cue utilization differs in the visual field.

The longer interstimulus interval was indeed associated with better performance. This was clearly predicted by Kantowitz (1974). However, it was predicted in this paper that there would be an interaction between the interstimulus interval and temperature. This was not the case. A possible explanation may be that, just as a human's ability to judge time intervals was not temperature dependent (Bell, 1975), perhaps the psychological refractory period was not temperature dependent.

With the arousal hypothesis supported, a different aspect of the "funnel" vs. "loss" controversy was examined. If performance was better for stressful conditions (Sanders, 1983), then where did the improvement come from? With the results of this experiment, some portion of the SPARE category (refer to equation 1) of attention was devoted to both the CENTER and PERIPHERAL components of attention. This supported the hypothesis that attention can be altered.

To determine whether attention was lost or funnelled, the relative values of CENTER and PERIPHERAL at each level of stress must be examined. In figure 12a, the observation of the lines repersenting the medium and the cold temperatures appear somewhat parallel, indicating that attention changed across the visual field equally. For the hot temperature, it appeared that more attention was gained in the inner positions of the stimuli than the outer positions; in fact, it appears that no attention was gained for the outer position. Perhaps this was due to attention being altered differently by the hot condition as compared to the cold condition. Perhaps there were both an arousal effect and a funnelling effect for heat, but simply an arousal effect for cold. This explanation is made with caution; there was no interaction between the spatial position of the stimulus and temperature. It may be that while heat and cold both increase arousal, they affect attention differently.

At the present time, data are being collected for female subjects. It is hoped that most of the results found here are duplicated; yet that speed-accuracy trade-offs are avoided and a significant interaction between temperature and stimulus position emerges; any discussion of how attention was focused was also limited by the significant main effects for error rates. Any differences found between these data and those that follow will be due to a gender effect, although Paolone, Wells, and Kelley (1978) found that there are no gender differences in abilities during temperature extremes except when the task requires much physically; the tasks in this experiment may or may not turn out to be equally demanding on males as females.

Requiring that subjects perform to a criterion in terms of being accurate in their responses would certainly help in further experiments. In this way a better explanation can be developed of just how attention is affected by stress.

References

Baddeley, A. D. (1968). A 3 minute reasoning test based on grammatical transformation. Psychonomic Science, 10, 341-342.

Bateman, R. P. (1981). The effects of increased ambient temperature on dual task performance. Proceedings of the Human Factors Society 25th Annual Meeting, 379-383.

Bell, C. R. (1975). Effects of lowered temperature on time estimation. Quarterly Journal of Experimental Psychology, 27, 531-538.

Bursill, A. E. (1958). The restriction of peripheral vision during exposure to hot and humid conditions. Quarterly Journal of Experimental Psychology, 10, 113-129.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. Psychological Review, 66, 183-201.

Ellis, H. D. (1982). The effects of cold on the performance of serial choice reaction time and various discrete tasks. Human Factors, 24, 589-598.

Haldane, J. S. (1905). The influence of high air temperature. Journal of Hygiene, Cambridge, 5, 494.

Kantowitz, B. H. (1974). Double stimulation. In Human Information Processing: Tutorials in Performance and Cognition. (Kantowitz, B. H, ed.) Hillsdale: Lawrence Erlbaum Associates.

Mackworth, N. H. (1946). The effect of heat on wireless operators hearing and recording morse code messages. British Journal of Industrial Medicine, 3, 143.

McCormick, E. J. & Sanders, M. S. (1982). Human Factors in Engineering and Design. New York: McGraw-Hill.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. (1975). Statistical Package for the Social Sciences (2nd ed.). New York: McGraw-Hill.

Paolone, A. M., Wells, C. L., & Kelley, G. T. (1978). Sexual variations in thermoregulation during heat stress. Aviation, Space, and Environmental Medicine, 49, 715-719.

Sanders, A. F. (1983). Towards a model of stress and human performance. Acta Psychologica, 53, 61-97.

Stroop, J. R. (1935). Studies of interference in serial verbal reaction. Journal of Experimental Psychology, 18, 643-662.

Winer, B. J. (1971). Statistical Principles in Experimental Design. New York: McGraw-Hill.

APPENDIX A

TREATMENT ORDERS FOR MALE SUBJECTS

SUBJECT	TEMPERATURE			INTERSTIMULUS INTERVAL (ISI)*		
NUMBER	DAY 1	DAY 2	DAY 3	BLOCK 2	BLOCK 3	BLOCK 4
1	41	95	68	0	60	240
2	41	95	68	0	240	60
3	41	95	68	60	0	240
7	41	68	95	0	60	240
8	41	68	95	0	240	60
9	41	68	95	60	0	240
13	95	68	41	0	60	240
14	95	68	41	0	240	60
15	95	68	41	60	0	240
19	95	41	68	0	60	240
20	95	41	68	0	240	60
21	95	41	68	60	0	240
25	68	95	41	0	60	240
26	68	95	41	0	240	60
27	68	95	41	60	0	240
31	68	41	95	0	60	[illegible]
32	68	41	95	0	240	[illegible]
33	68	41	95	60	0	240

*AN ISI OF "0" DENOTES SINGLE STIMULATION.

POTENTIAL
TREATMENT ORDERS FOR FEMALE SUBJECTS

SUBJECT	TEMPERATURE			INTERSTIMULUS INTERVAL (ISI)*		
NUMBER	DAY 1	DAY 2	DAY 3	BLOCK 2	BLOCK 3	BLOCK 4
4	41	95	68	60	240	0
5	41	95	68	240	0	60
6	41	95	68	240	60	0
10	41	68	95	60	240	0
11	41	68	95	240	0	60
12	41	68	95	240	60	0
16	95	68	41	60	240	0
17	95	68	41	240	0	60
18	95	68	41	240	60	0
22	95	41	68	60	240	0
23	95	41	68	240	0	60
24	95	41	68	240	60	0
28	68	95	41	60	240	0
29	68	95	41	240	0	60
30	68	95	41	240	60	0
34	68	41	95	60	240	0
35	68	41	95	240	0	60
36	68	41	95	240	60	0

* AN ISI OF "0" DENOTES SINGLE STIMULATION.

Appendix B

Instructions

Please remain seated where the chair has been placed, and sit upright with your back against the backrest. Also, keep all four legs of the chair on the floor. In front of you are four response keys. I want you to place the index and middle fingers of each hand on the keys.
(demonstration)

The keys correspond to the four lights in a one-to-one manner. The left-most light corresponds to the left-most key, etc. Do you have any questions about the light-key arrangement? There will be four groups of trials in all. I will give you specific instructions before each of the groups of trials that pertain to that specific group of trials.

Single Stimulation When one of the lights comes on, respond as quickly as possible by depressing the key corresponding to that light. When you press the correct key the light will go off; you are to press only one key on each trial. There will be six seconds between each trial, and shortly before each trial a warning tone will sound to alert you to be ready to respond to a light. When you hear the sound, focus your eyes here at the center of the board. Do you have any questions?

Double Stimulation Once again, a tone will signal the beginning of a trial. Again focus your eyes at the center of the panel and respond to the first light that comes on. Now, however, a second light will also come on. Respond to it in the same way as the first. Please respond to the first light first and the second light second. Also, the second light will always be on the other side of the display from the first. Thus, if one of the two right-most lights comes on first, the second light will be one of the left-most lights and vice versa. Is this arrangement clear? Any questions?

Figure 1

STIMULUS PANEL

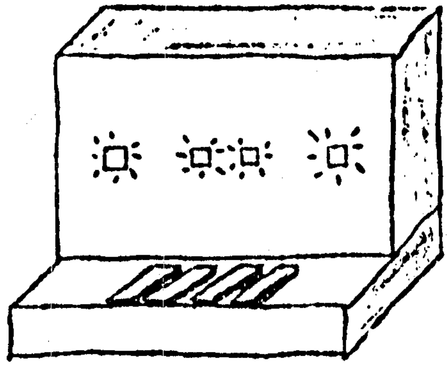


FIGURE 1a: FRONT VIEW OF STIMULUS PANEL. THE INNER LIGHTS ARE 1.5 cm FROM THE CENTER OF THE PANEL; THE OUTER LIGHTS ARE 22.5 cm FROM THE CENTER. THEY ARE NUMBERED 1-4 FROM LEFT TO RIGHT.

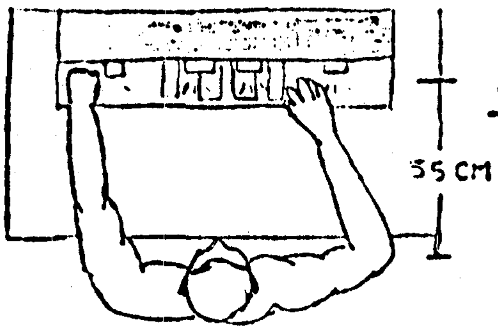


FIGURE 1b: TOP VIEW OF STIMULUS PANEL. THE SUBJECT IS APPROXIMATELY 55 cm FROM THE PANEL.

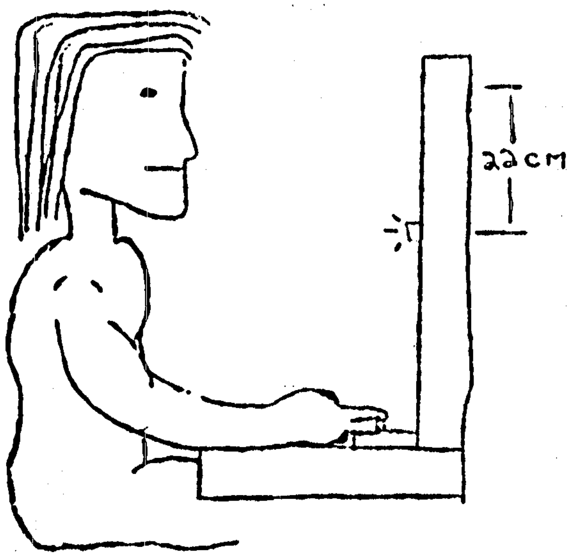
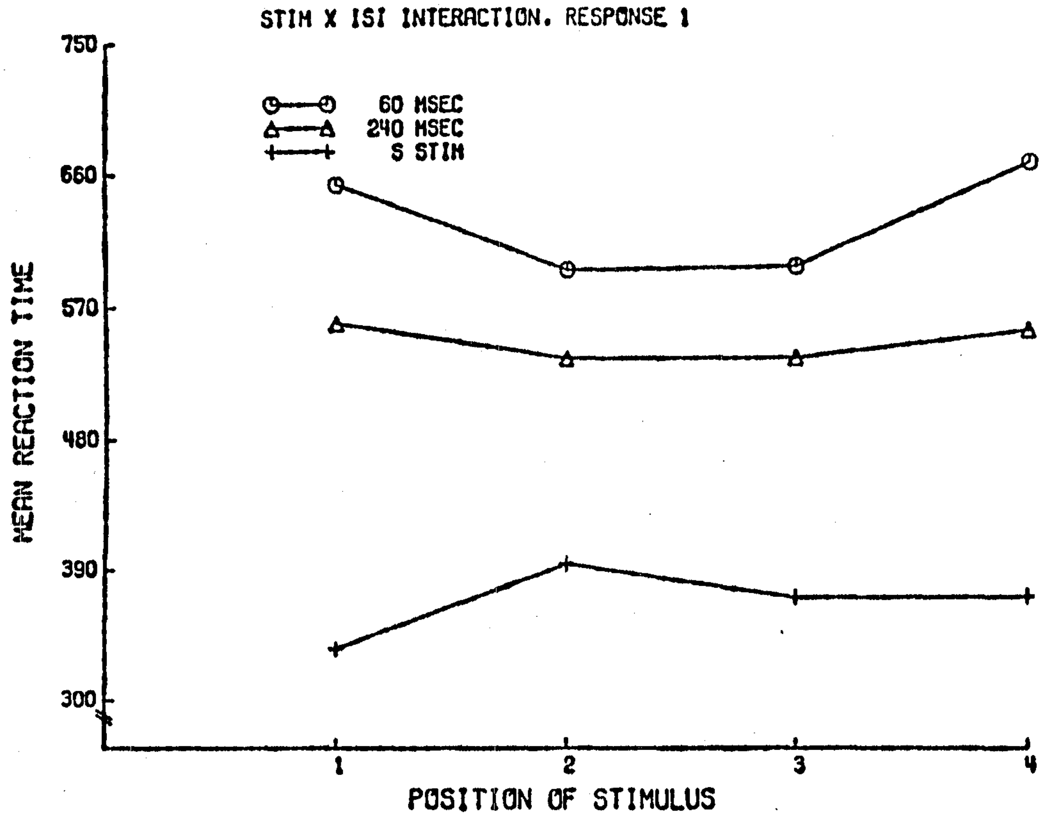


FIGURE 1c: SIDE VIEW OF STIMULUS PANEL. THE STIMULI ARE APPROXIMATELY 22 cm BELOW THE SUBJECT'S HORIZONTAL LINE OF SIGHT.

Figure 2


STIM X ISI INTERACTION. RESPONSE 1
750
660
570
480
390
300
MEAN REACTION TIME
60 MSEC
240 MSEC
S STIM
1
2
3
4
POSITION OF STIMULUS

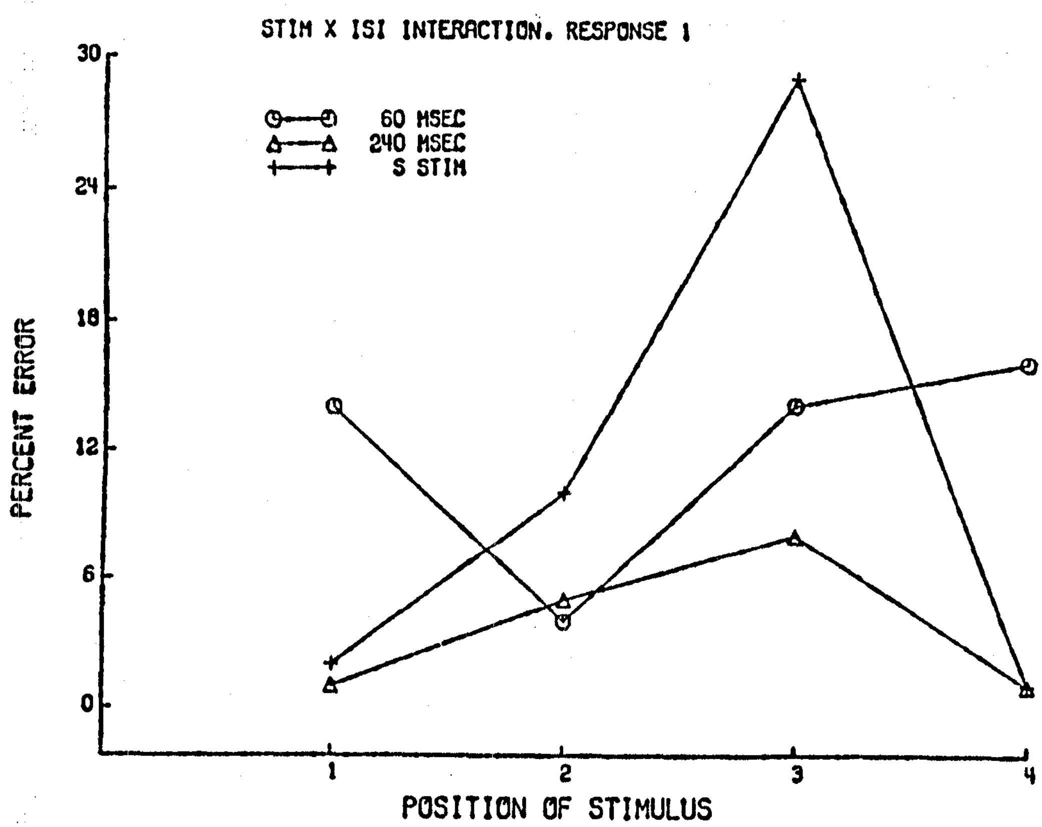
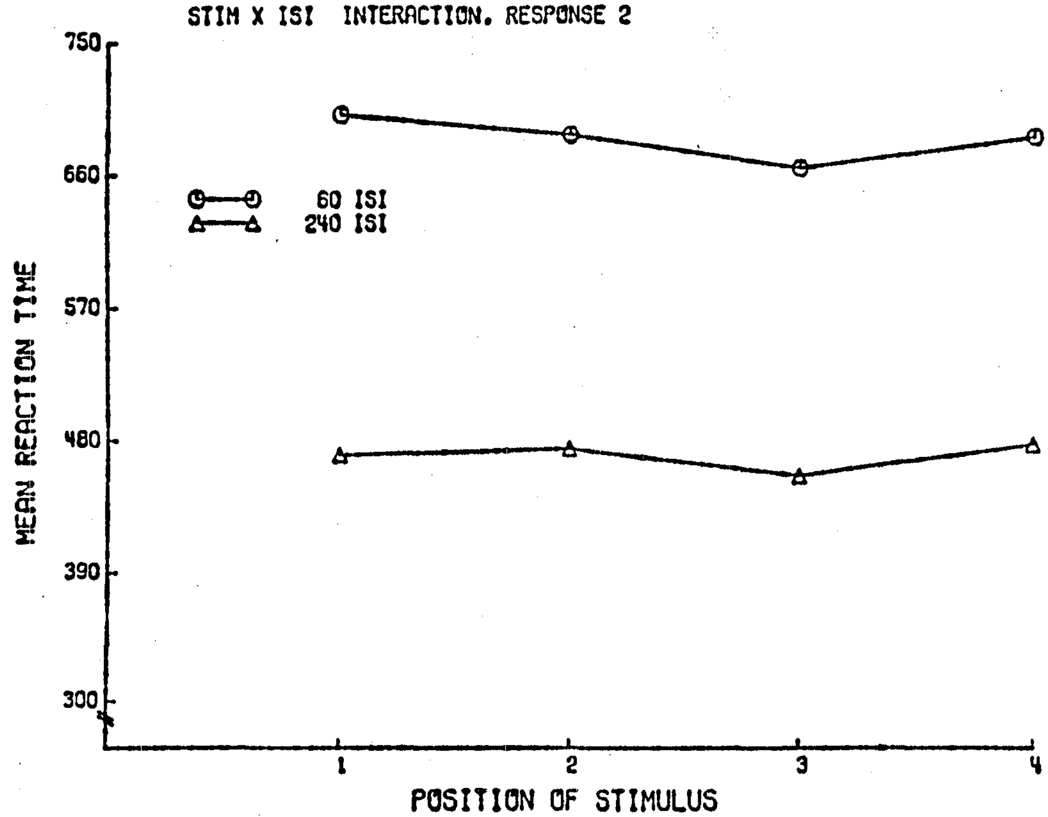

STIM X ISI INTERACTION. RESPONSE 1
30
24
18
12
6
0
PERCENT ERROR
60 MSEC
240 MSEC
S STIM
1
2
3
4
POSITION OF STIMULUS

Figure 3


STIM X ISI INTERACTION. RESPONSE 2
750
660
570
480
390
300
60 ISI
240 ISI
MEAN REACTION TIME
1
2
3
4
POSITION OF STIMULUS

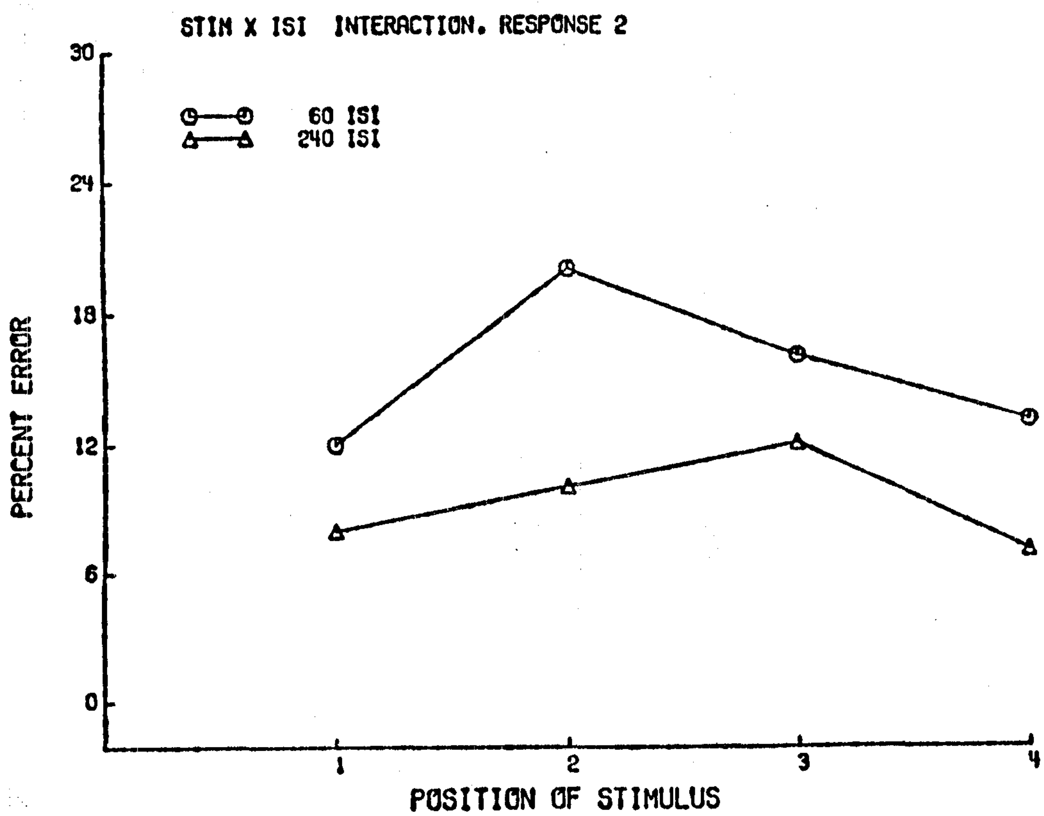
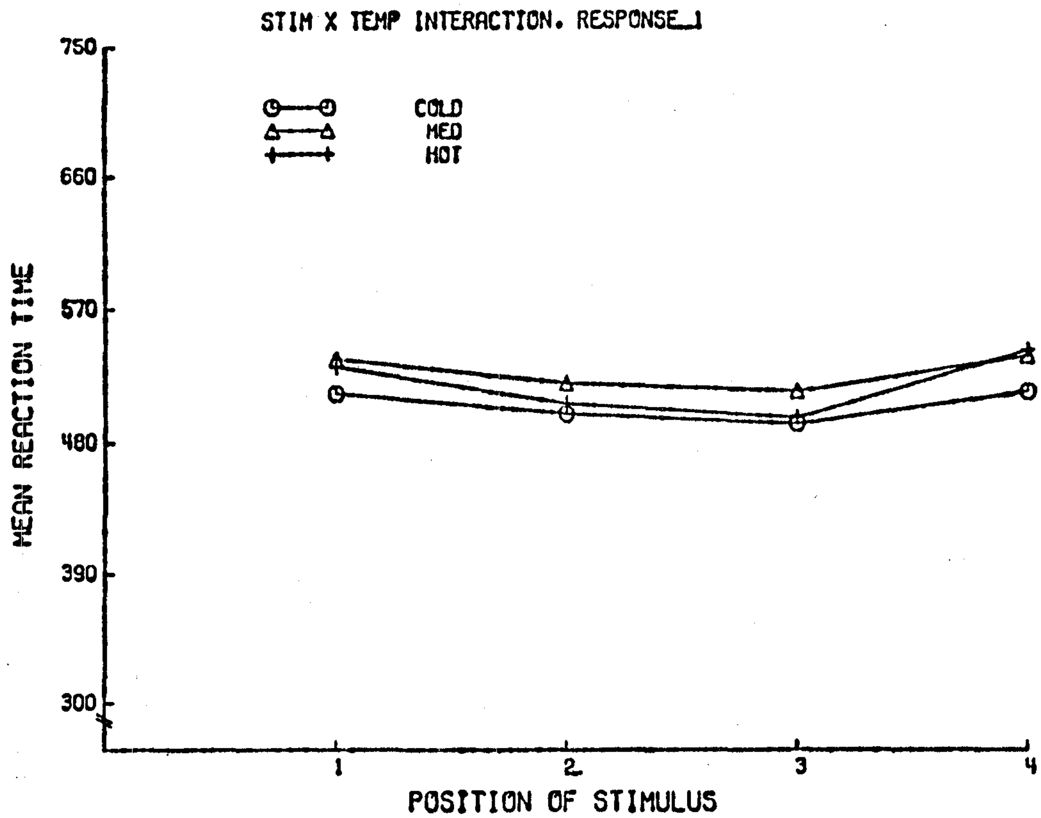

STIM X ISI INTERACTION. RESPONSE 2
30
24
18
12
6
0
60 ISI
240 ISI
PERCENT ERROR
1
2
3
4
POSITION OF STIMULUS

Figure 4


STIM X TEMP INTERACTION. RESPONSE_1
750
660
570
480
390
300
MEAN REACTION TIME
COLD
MED
HOT
1
2
3
4
POSITION OF STIMULUS

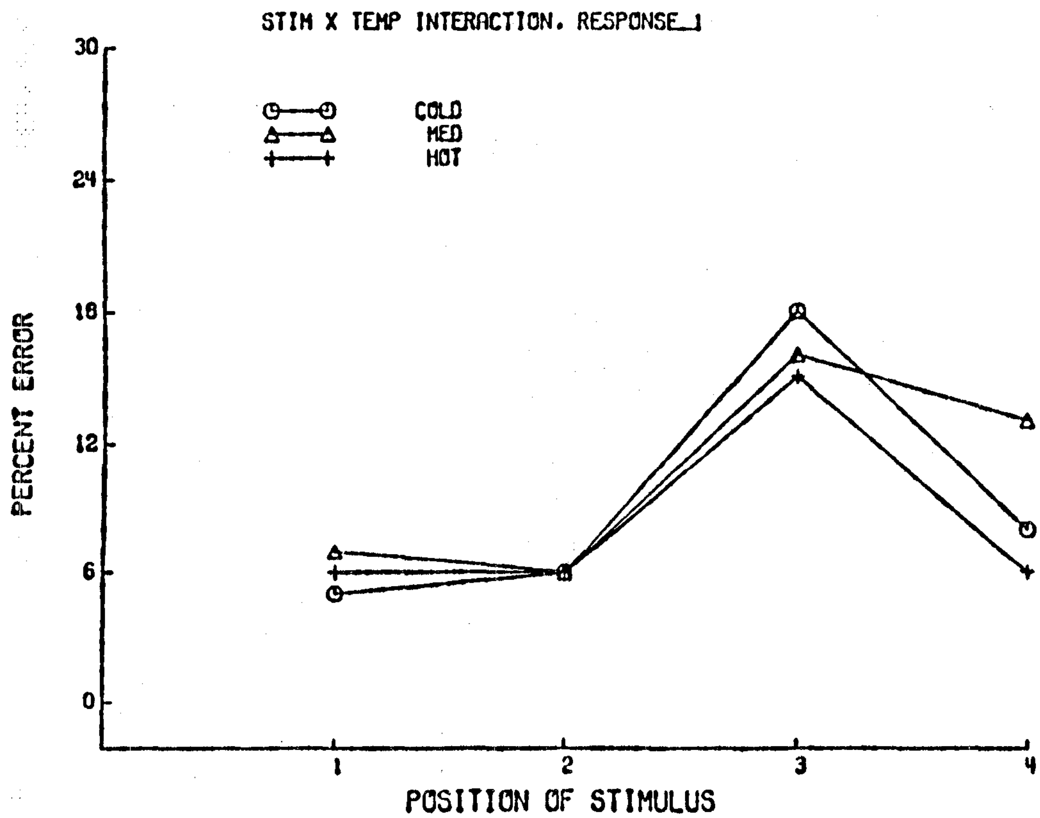
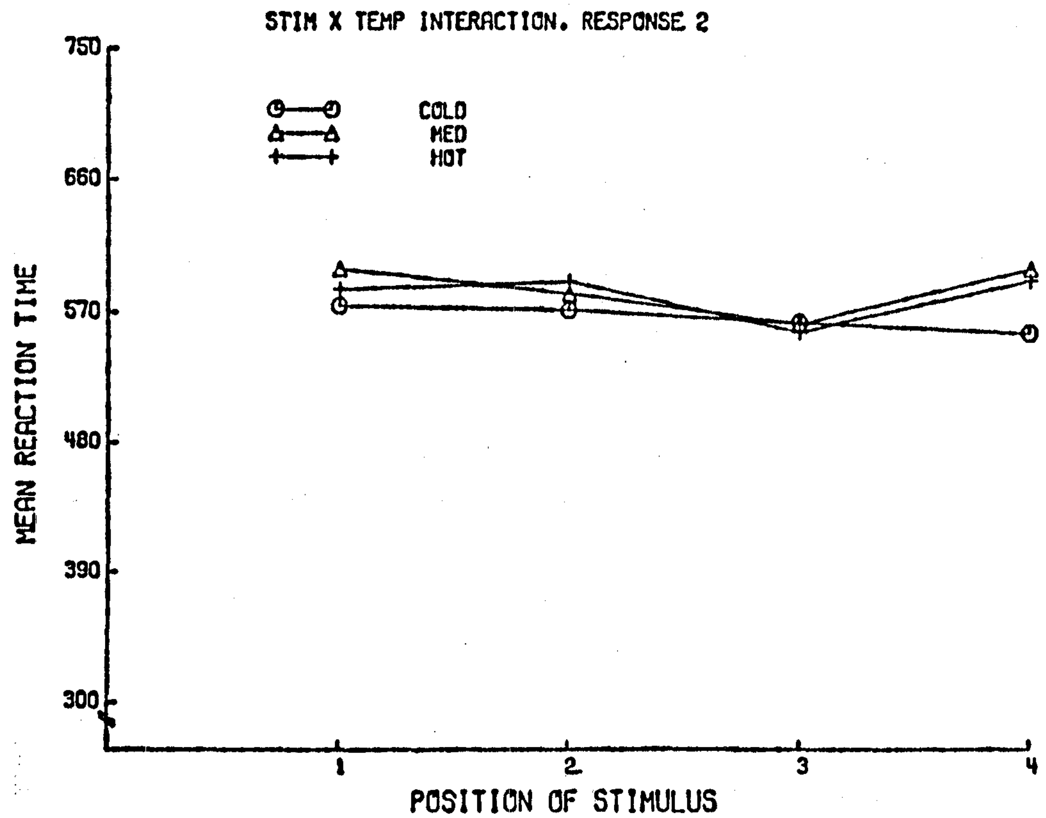

STIM X TEMP INTERACTION. RESPONSE_1
30
24
18
12
6
0
PERCENT ERROR
COLD
MED
HOT
1
2
3
4
POSITION OF STIMULUS

Figure 5


STIM X TEMP INTERACTION, RESPONSE 2
750
660
570
480
390
300
MEAN REACTION TIME
COLD
MED
HOT
1
2
3
4
POSITION OF STIMULUS

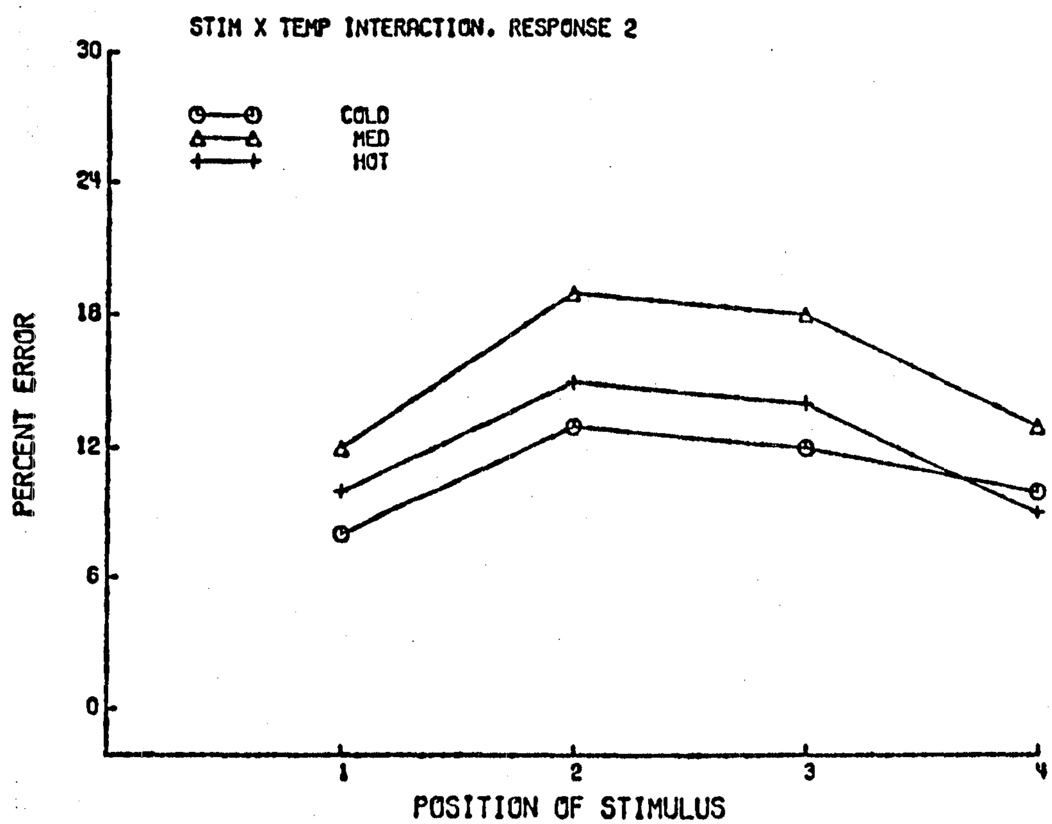
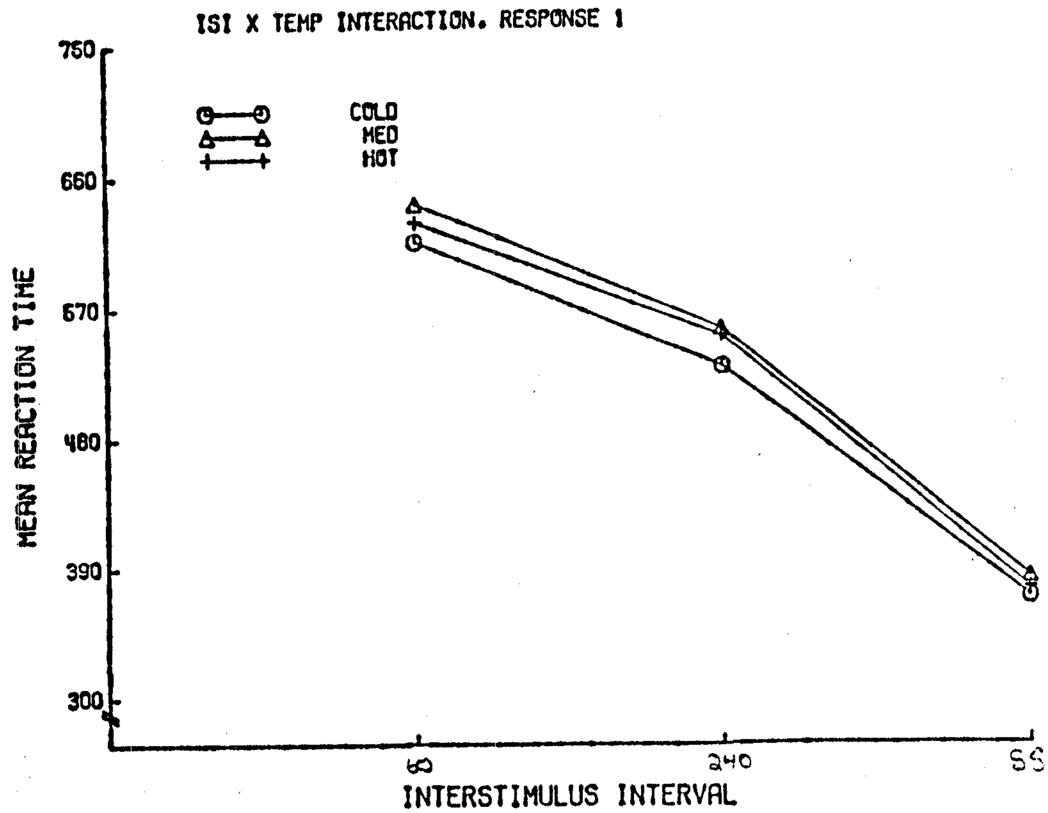

STIM X TEMP INTERACTION, RESPONSE 2
30
24
18
12
6
0
PERCENT ERROR
COLD
MED
HOT
1
2
3
4
POSITION OF STIMULUS

Figure 6


ISI X TEMP INTERACTION, RESPONSE 1
750
660
570
480
390
300
COLD
MED
HOT
MEAN REACTION TIME
60
240
SS
INTERSTIMULUS INTERVAL

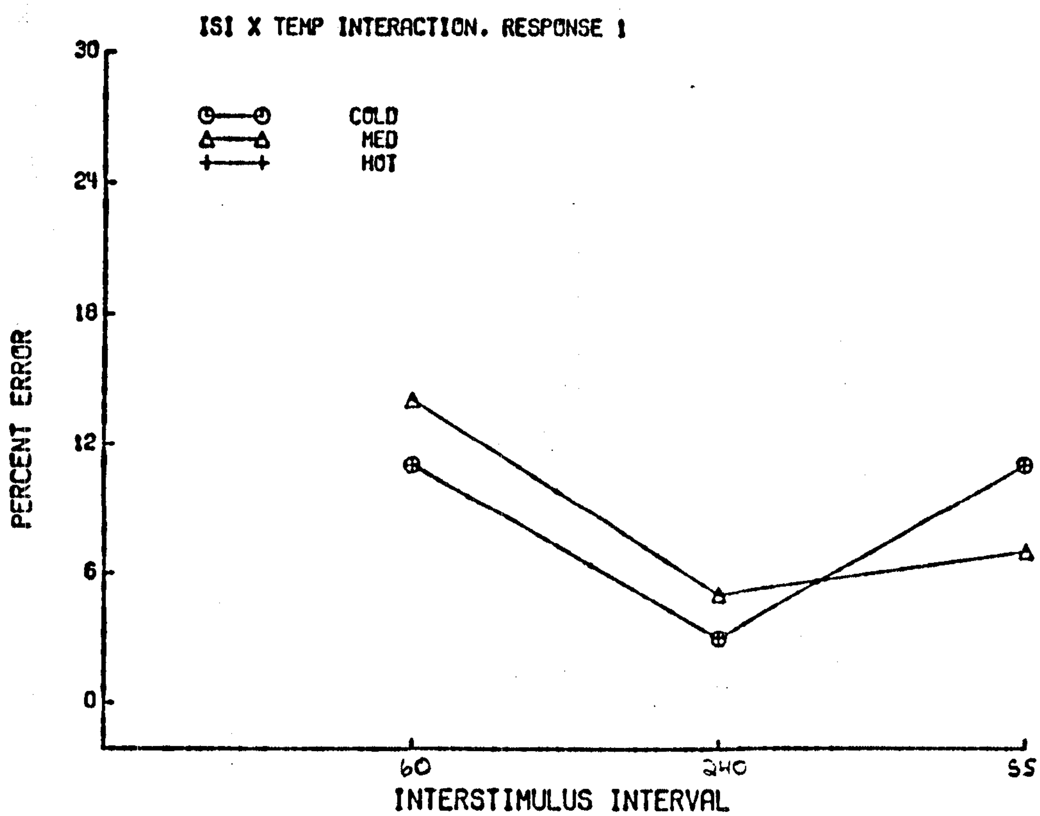
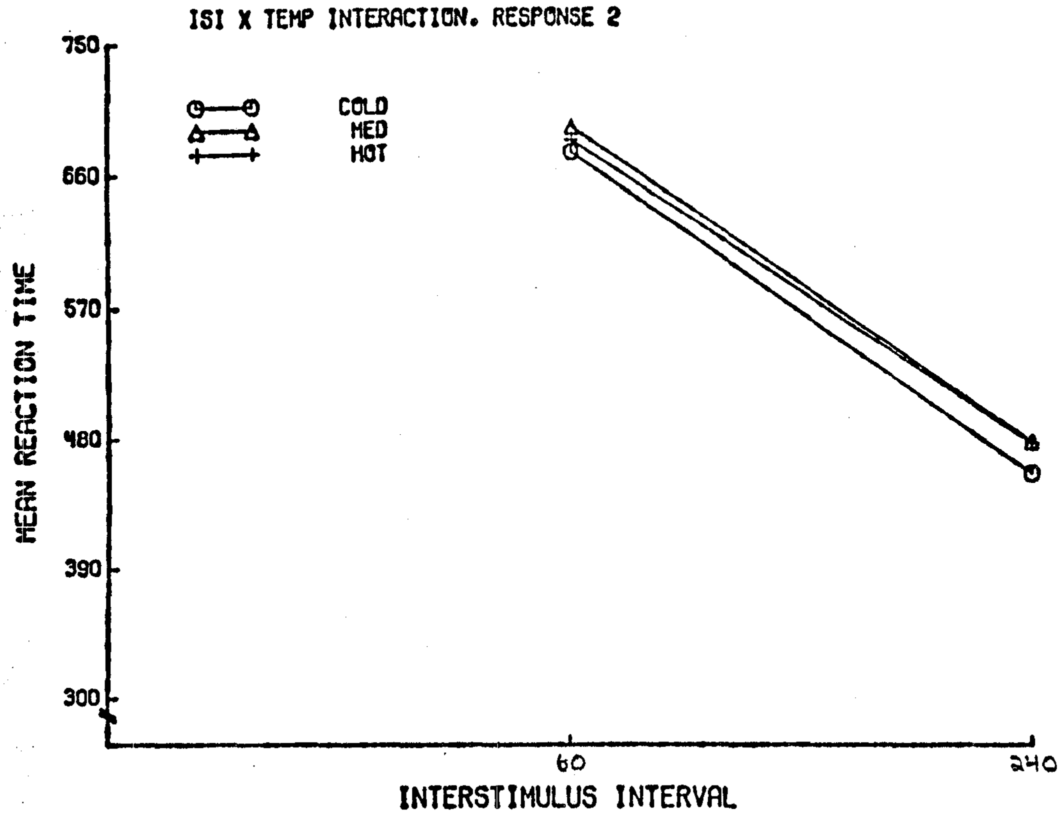

ISI X TEMP INTERACTION, RESPONSE 1
30
24
18
12
6
0
COLD
MED
HOT
PERCENT ERROR
60
240
SS
INTERSTIMULUS INTERVAL

Figure 7


ISI X TEMP INTERACTION. RESPONSE 2
COLD
MED
HOT
MEAN REACTION TIME
750
660
570
480
390
300
60
240
INTERSTIMULUS INTERVAL

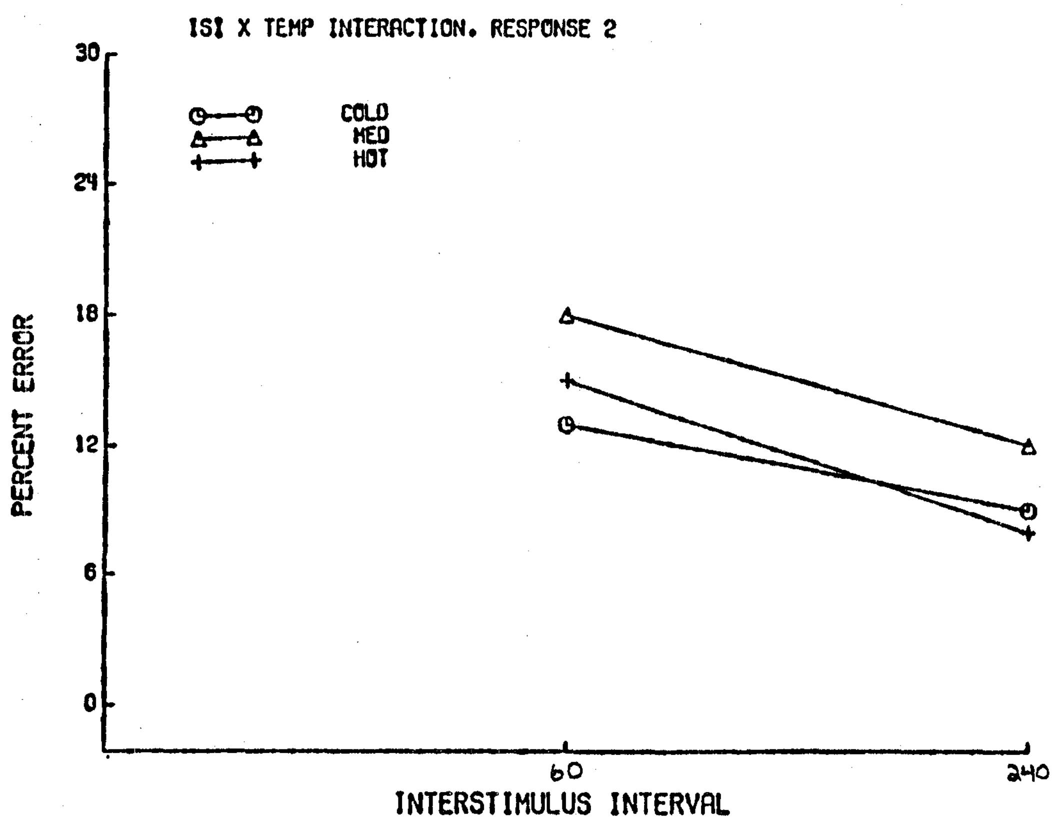
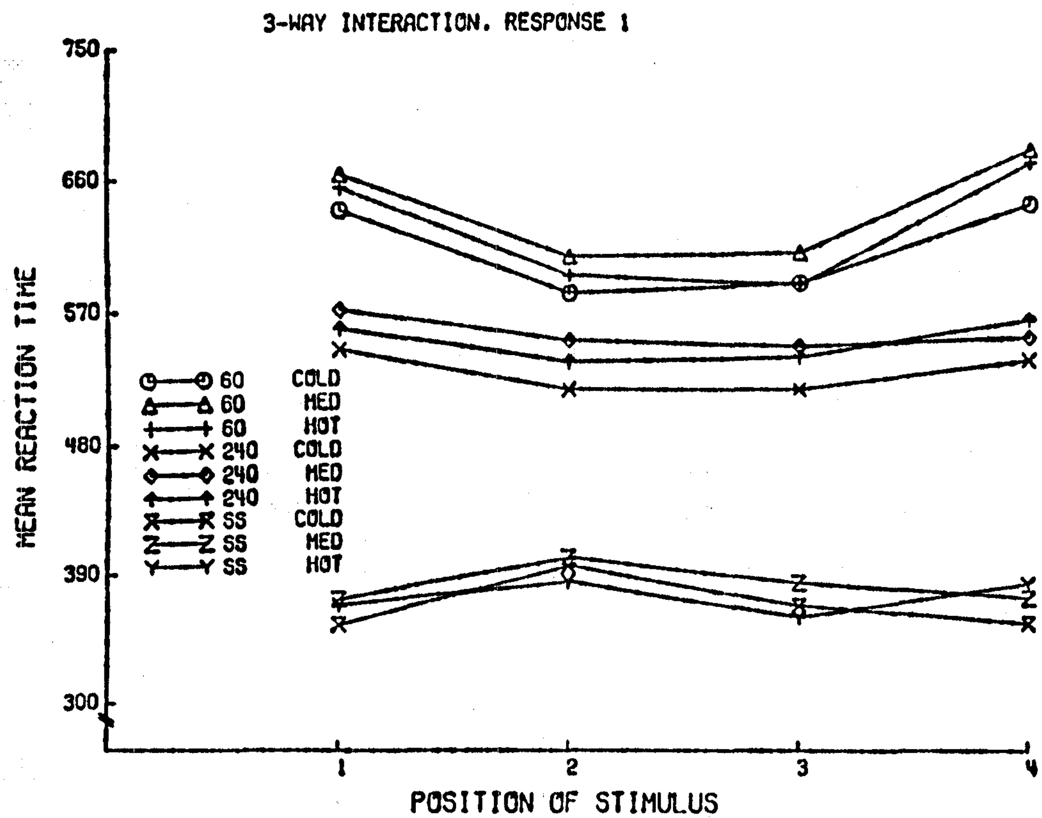

ISI X TEMP INTERACTION. RESPONSE 2
COLD
MED
HOT
PERCENT ERROR
30
24
18
12
6
0
60
240
INTERSTIMULUS INTERVAL

Figure 8


3-WAY INTERACTION, RESPONSE 1
750
660
570
480
390
300
MEAN REACTION TIME
60 COLD
60 MED
60 HOT
240 COLD
240 MED
240 HOT
SS COLD
SS MED
SS HOT
1
2
3
4
POSITION OF STIMULUS

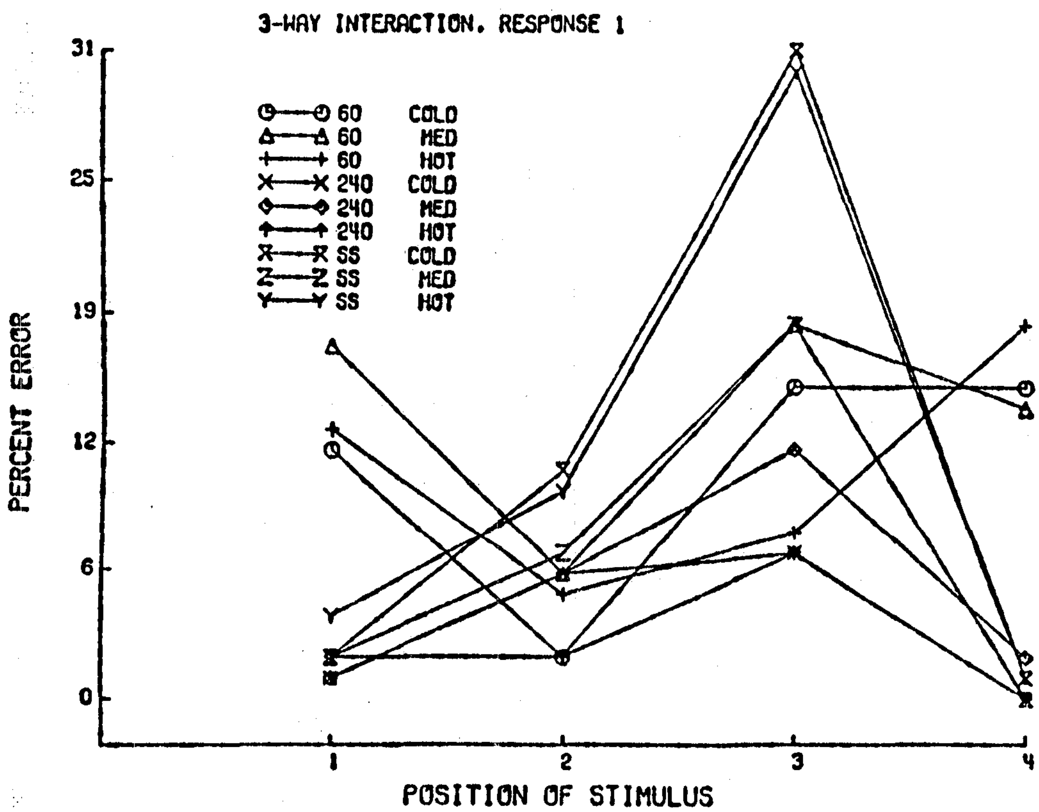
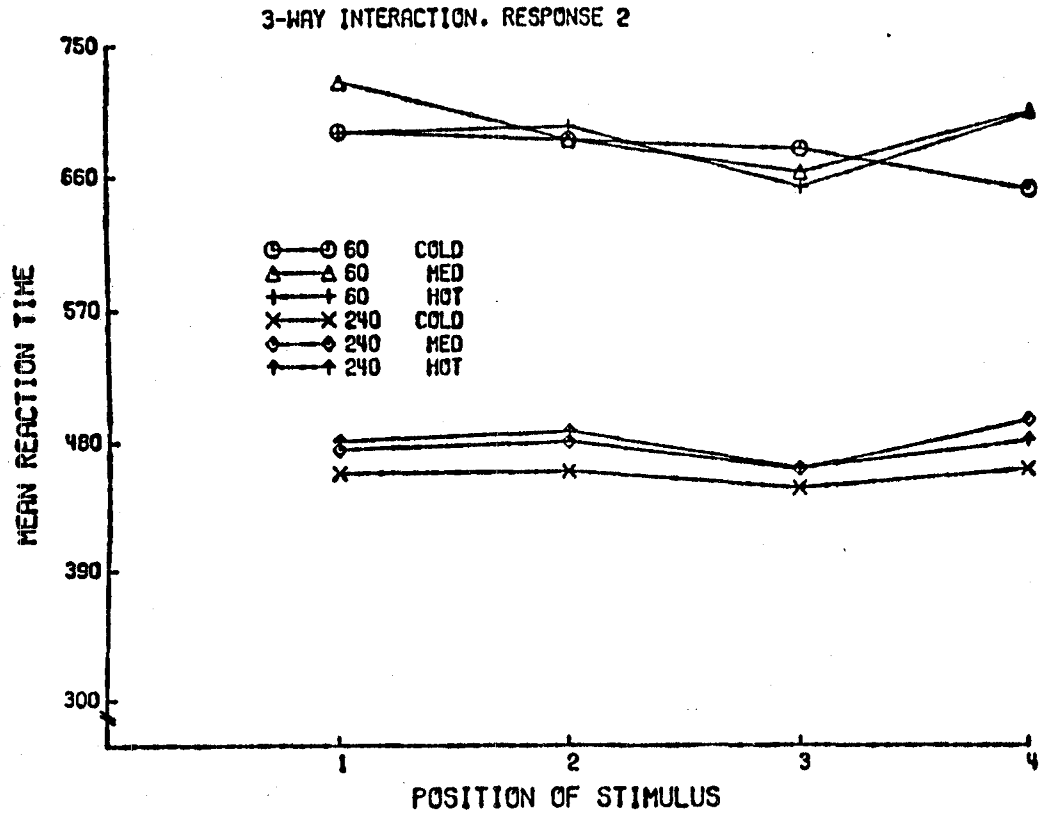

3-WAY INTERACTION, RESPONSE 1
31
25
19
12
6
0
PERCENT ERROR
60 COLD
60 MED
60 HOT
240 COLD
240 MED
240 HOT
SS COLD
SS MED
SS HOT
1
2
3
4
POSITION OF STIMULUS

Figure 9


3-WAY INTERACTION. RESPONSE 2
750
660
570
480
390
300
MEAN REACTION TIME
60 COLD
60 MED
60 HOT
240 COLD
240 MED
240 HOT
1
2
3
4
POSITION OF STIMULUS

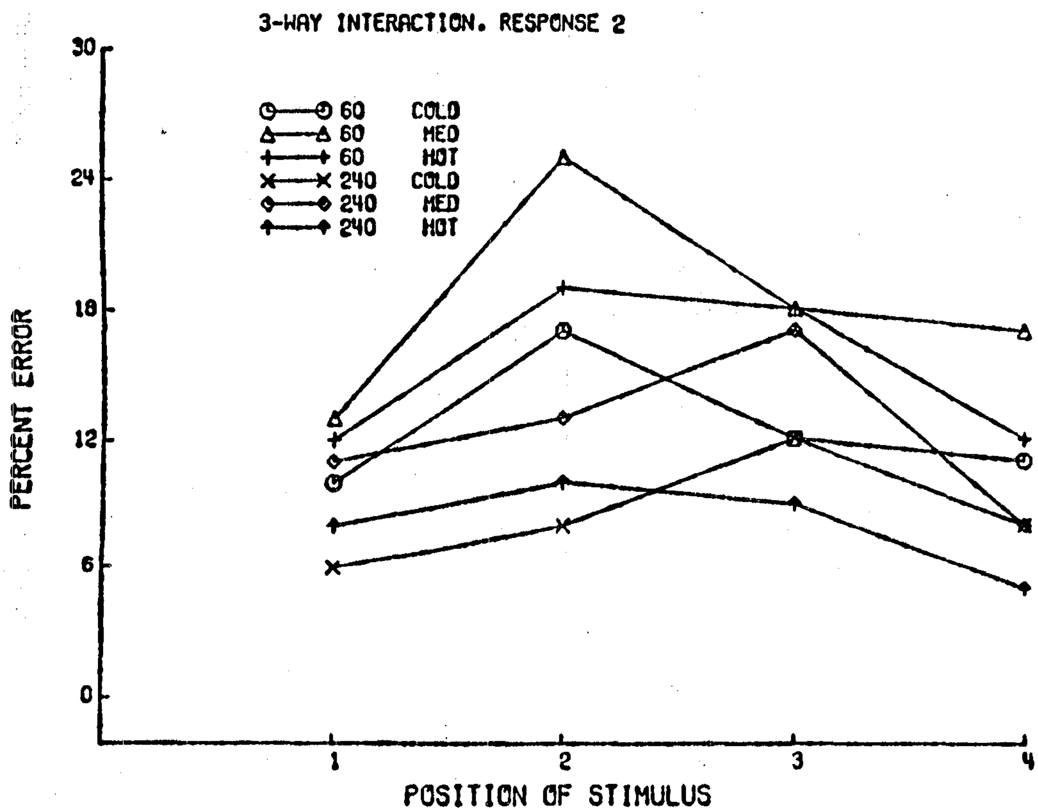

3-WAY INTERACTION. RESPONSE 2
30
24
18
12
6
0
PERCENT ERROR
60 COLD
60 MED
60 HOT
240 COLD
240 MED
240 HOT
1
2
3
4
POSITION OF STIMULUS

TABLE 1A.

A. Analysis of Variance Summary Table for Response 1 reaction time (RT1), Response 1 accuracy (ACC1), Response 2 reaction time (RT2), and Response 2 accuracy (ACC2)

		RT1		ACC1	
SOURCE	DF	MS	F	MS	F
STIM	3	230478.	8.99***	4.31	54.74***
ISI	2	$.264 \times 10^8$	1030.92***	3.24	44.93***
TEMP	2	212413.	8.29***	.06	.76
STIM × ISI	6	224744.	8.77***	2.67	37.03***
STIM × TEMP	6	13365.	.52	.06	.84
ISI × TEMP	4	9497.	.37	.14	1.98
STIM × ISI × TEMP	12	4970.	.19	.08	1.05
RESIDUAL	4847	25629.			
	5307			.07	

* SIG. AT .05
** SIG AT .01
*** SIG AT .001

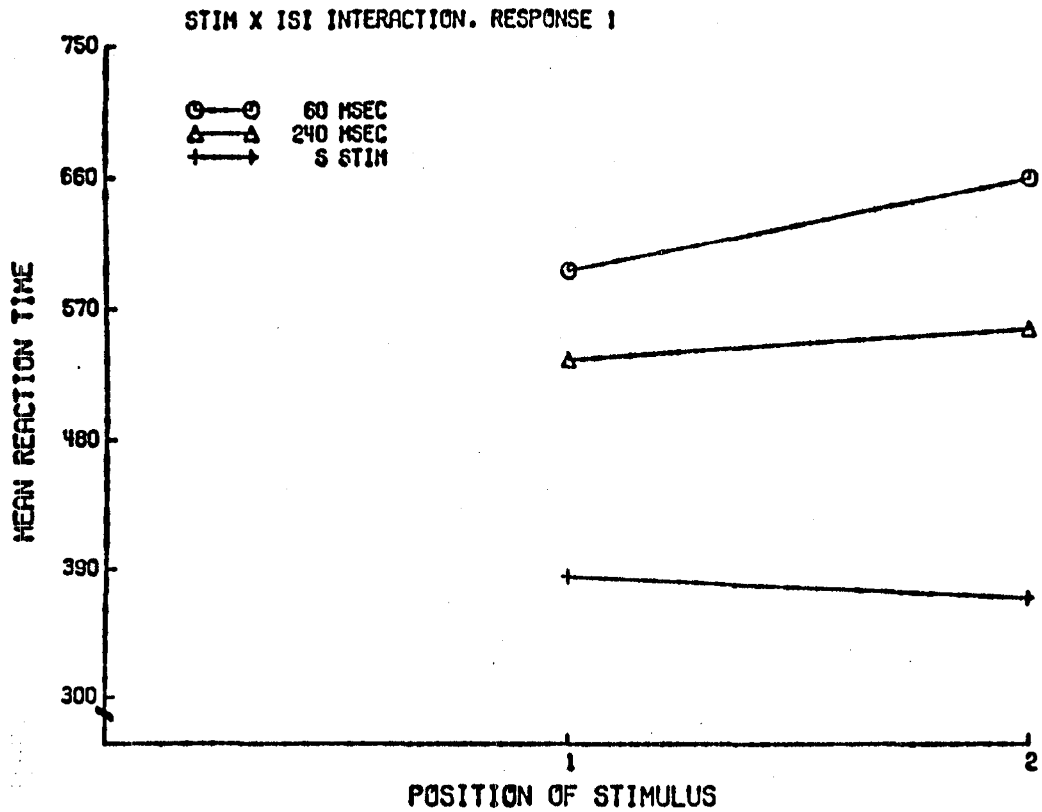
TABLE 1B.

B. Analysis of Variance Summary Table for Response 1 reaction time (RT1), Response 1 accuracy (ACC1), Response 2 reaction time (RT2), and Response 2 accuracy (ACC2).

		RT2		ACC2	
SOURCE	DF	MS	F	MS	F
STIM	3	101817.0	3.09*	.71	6.67***
ISI	1	$.18 \times 10^8$	548.97***	1.49	14.10***
TEMP	2	113945.2	3.46*	.72	6.81***
STIM X ISI	3	22803.4	.69	.21	1.99
STIM X TEMP	6	31890.4	.97	.03	.29
ISI X TEMP	2	11646.7	.35	.08	.74
STIM X ISI X TEMP	6	23255.5	.71	.06	.57
RESIDUAL	3101	32945.6			
	3536			.11	

* SIG. AT .05
** SIG. AT .01
*** SIG. AT .001

Figure 10


STIM X ISI INTERACTION. RESPONSE 1
750
660
570
480
390
300
MEAN REACTION TIME
60 MSEC
240 MSEC
S STIM
1
2
POSITION OF STIMULUS

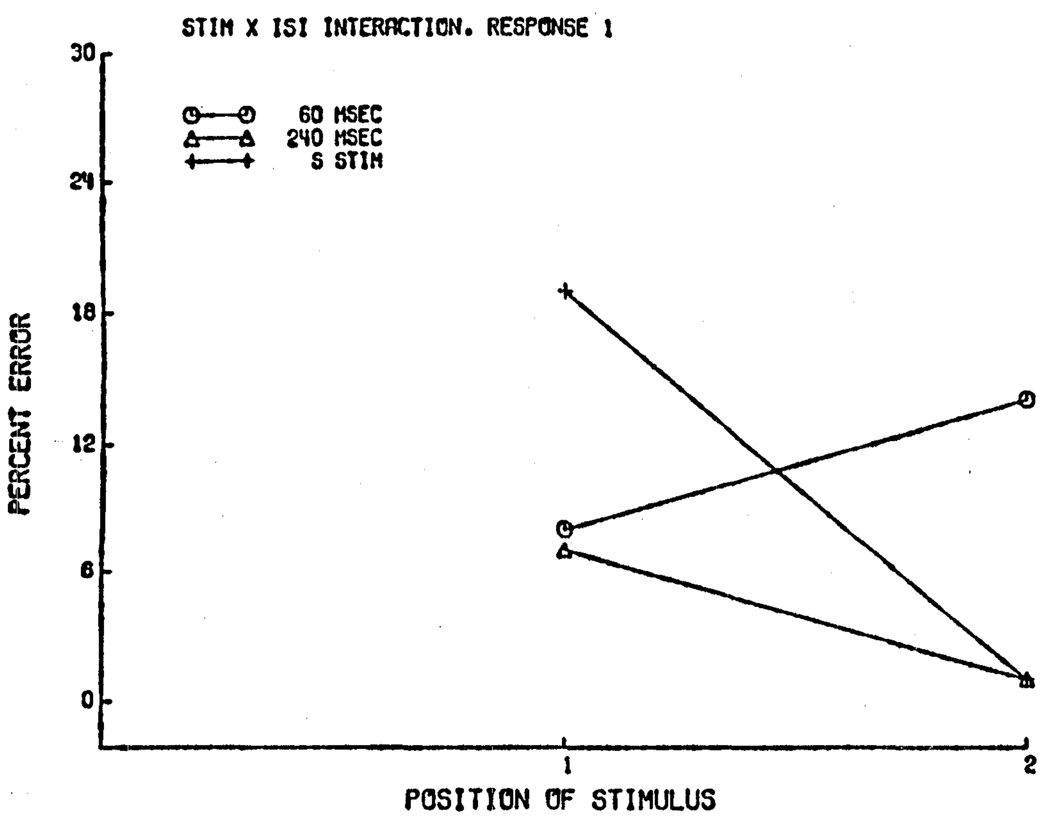
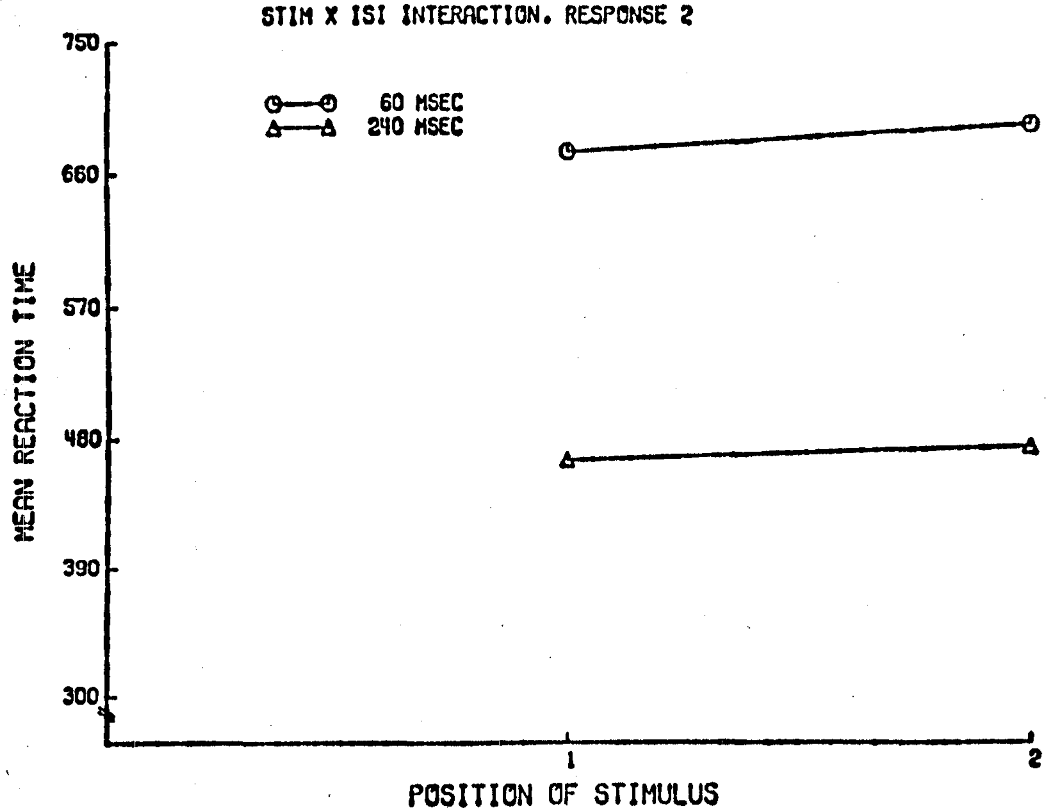

STIM X ISI INTERACTION. RESPONSE 1
30
24
18
12
6
0
PERCENT ERROR
60 MSEC
240 MSEC
S STIM
1
2
POSITION OF STIMULUS

Figure 11


STIM X ISI INTERACTION. RESPONSE 2
60 MSEC
240 MSEC
MEAN REACTION TIME
750
660
570
480
390
300
1
2
POSITION OF STIMULUS

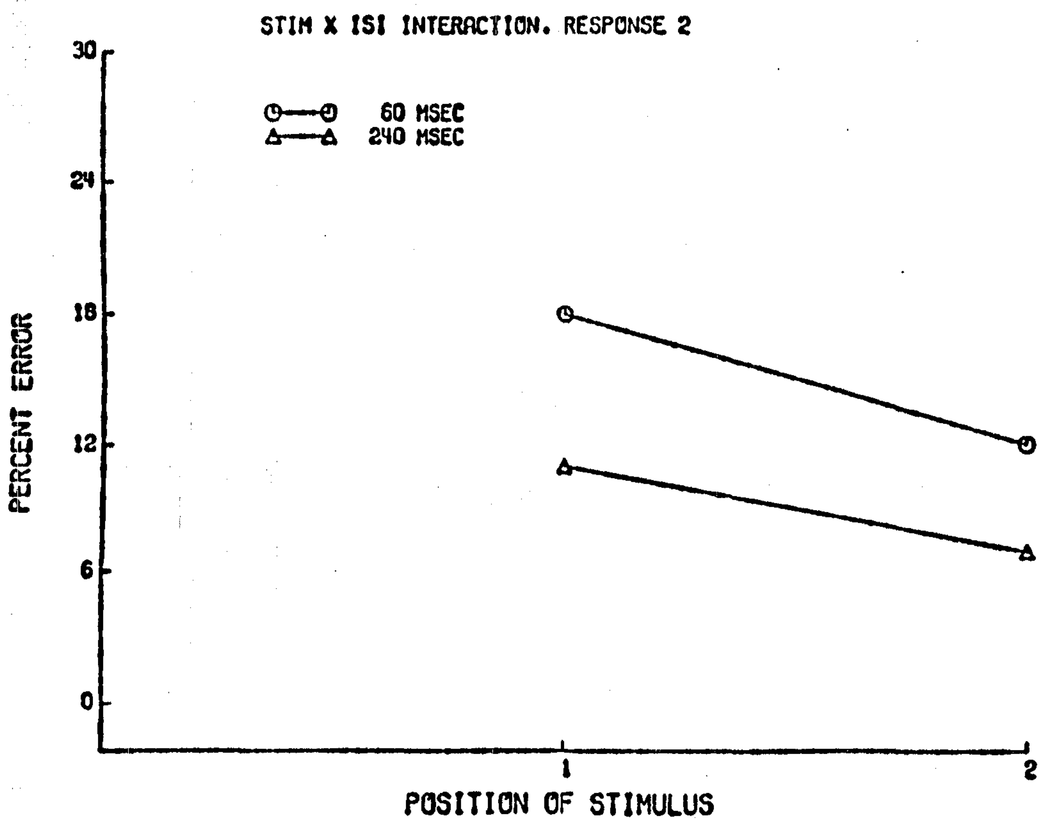
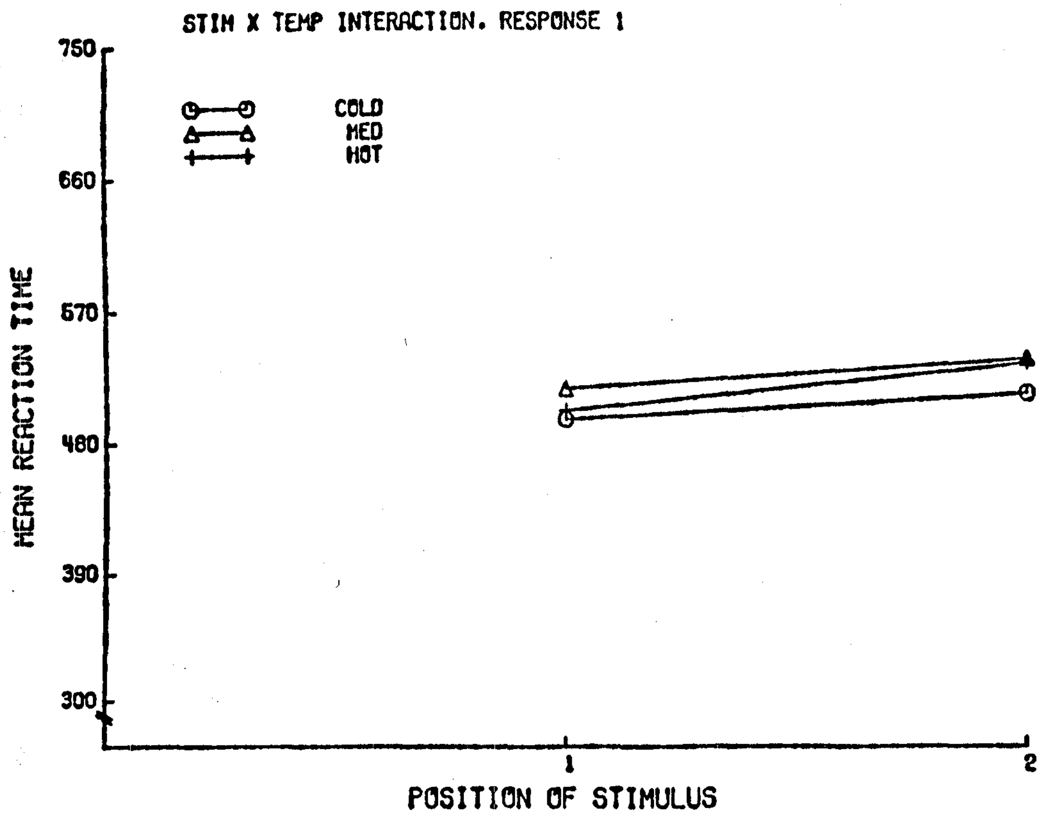

STIM X ISI INTERACTION. RESPONSE 2
60 MSEC
240 MSEC
PERCENT ERROR
30
24
18
12
6
0
1
2
POSITION OF STIMULUS

Figure 12


STIM X TEMP INTERACTION. RESPONSE 1
COLD
MED
HOT
MEAN REACTION TIME
750
660
570
480
390
300
1
2
POSITION OF STIMULUS

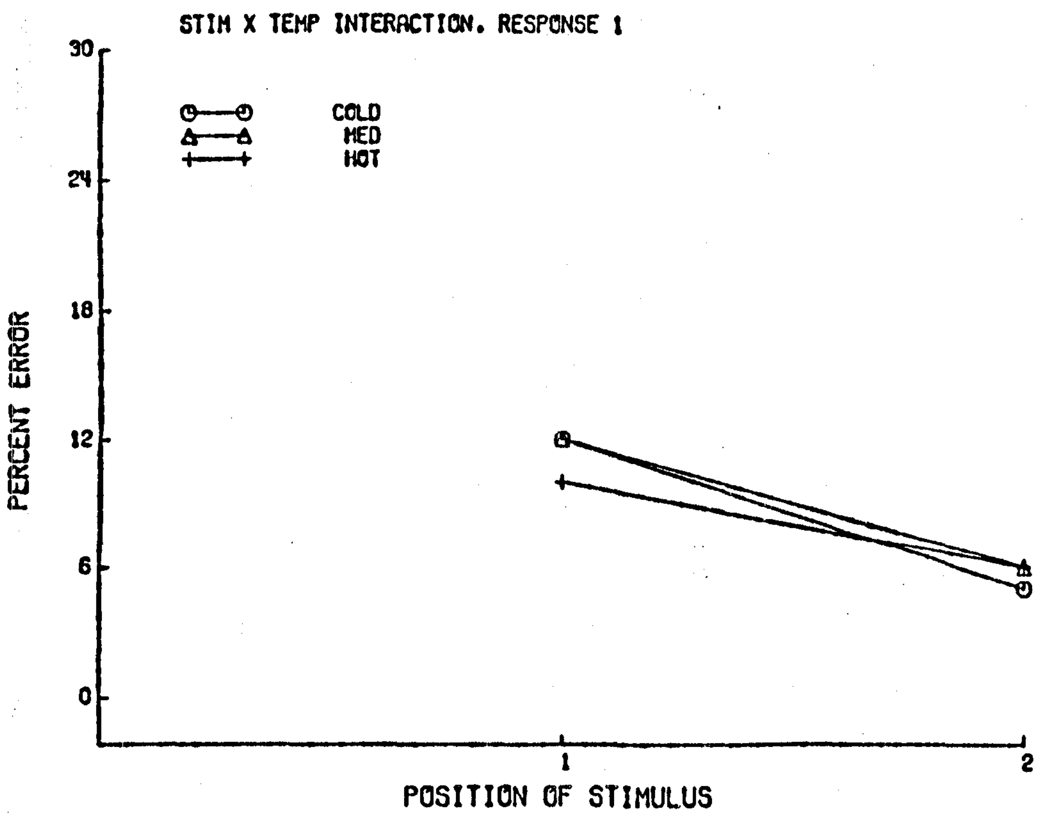
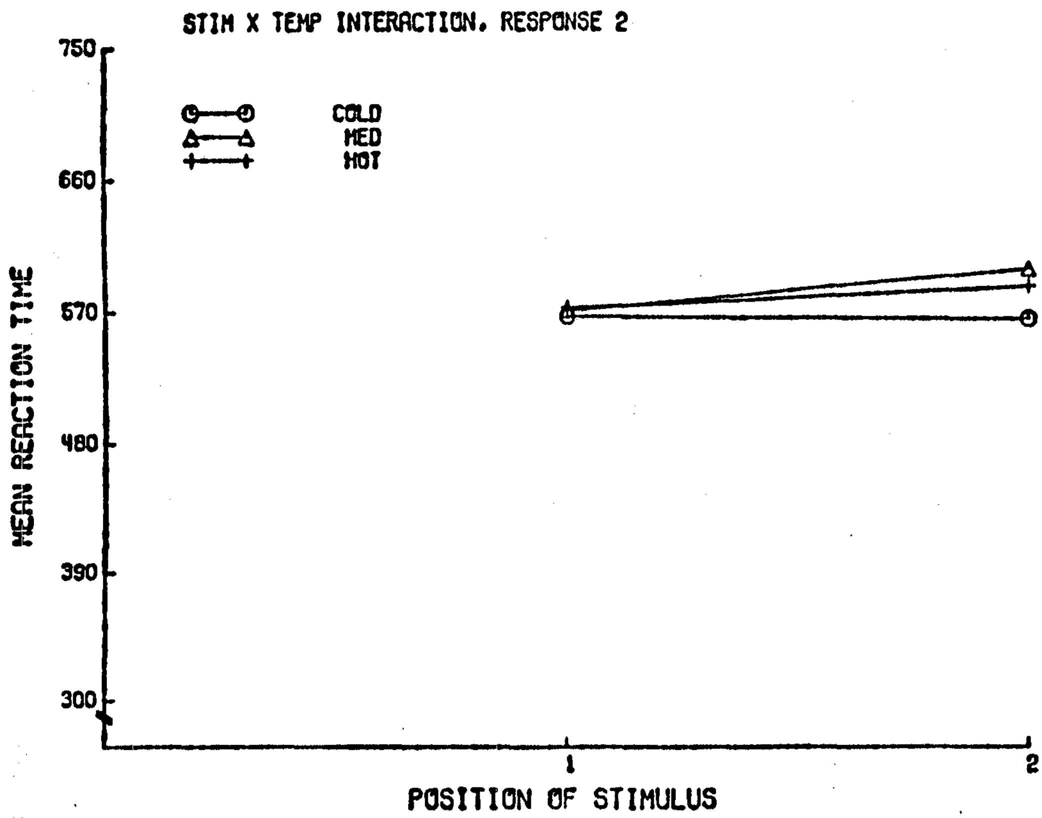

STIM X TEMP INTERACTION. RESPONSE 1
COLD
MED
HOT
PERCENT ERROR
30
24
18
12
6
0
1
2
POSITION OF STIMULUS

Figure 13


STIM X TEMP INTERACTION, RESPONSE 2
750
660
570
480
390
300
MEAN REACTION TIME
COLD
MED
HOT
1
2
POSITION OF STIMULUS

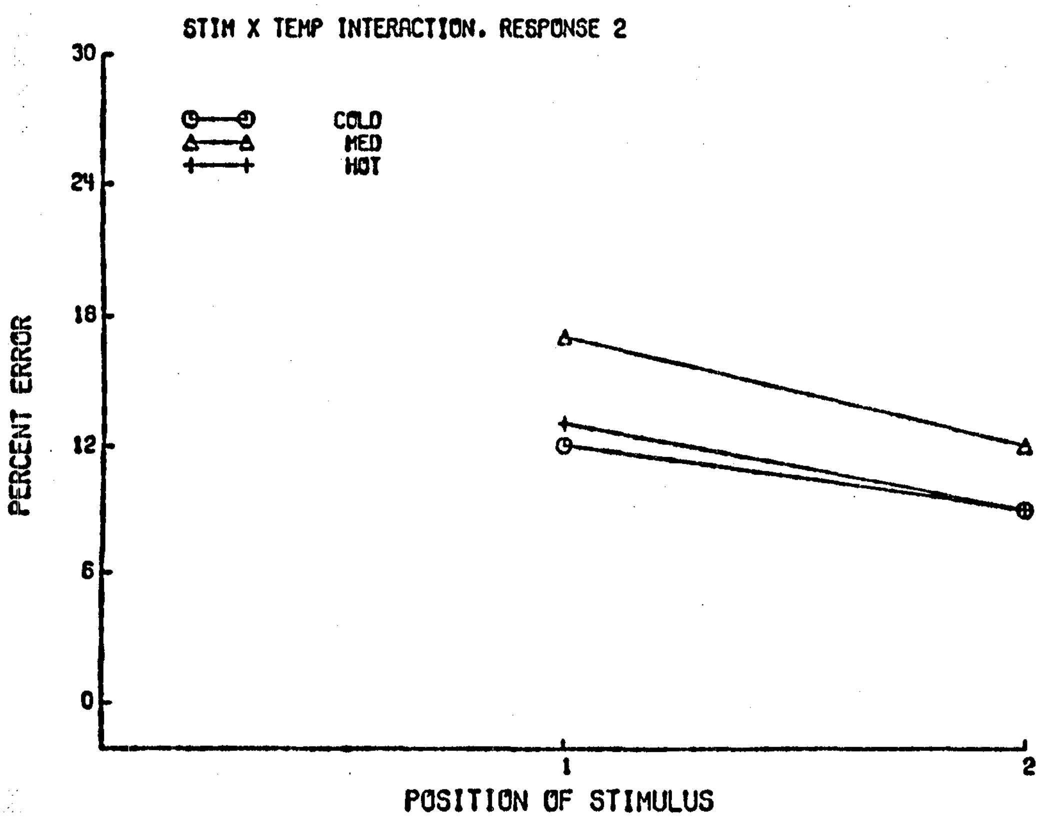
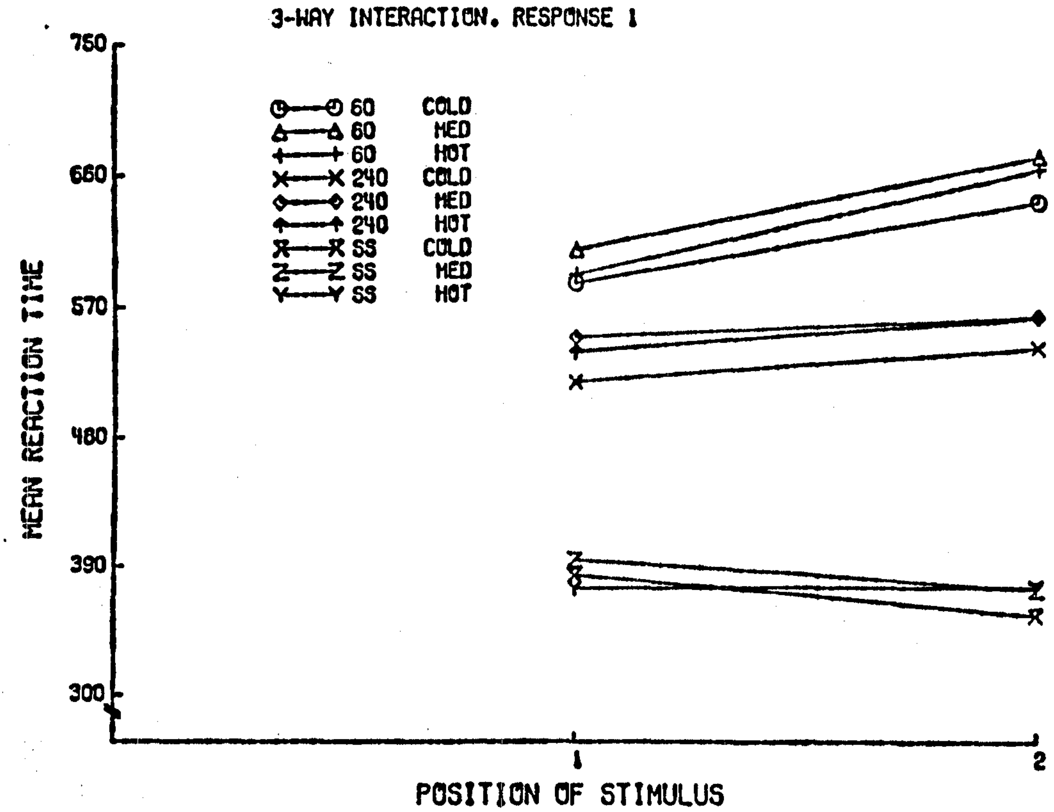

STIM X TEMP INTERACTION, RESPONSE 2
30
24
18
12
6
0
PERCENT ERROR
COLD
MED
HOT
1
2
POSITION OF STIMULUS

Figure 14


3-WAY INTERACTION. RESPONSE 1
MEAN REACTION TIME
750
660
570
480
390
300
60 COLD
60 MED
60 HOT
240 COLD
240 MED
240 HOT
SS COLD
SS MED
SS HOT
1
2
POSITION OF STIMULUS

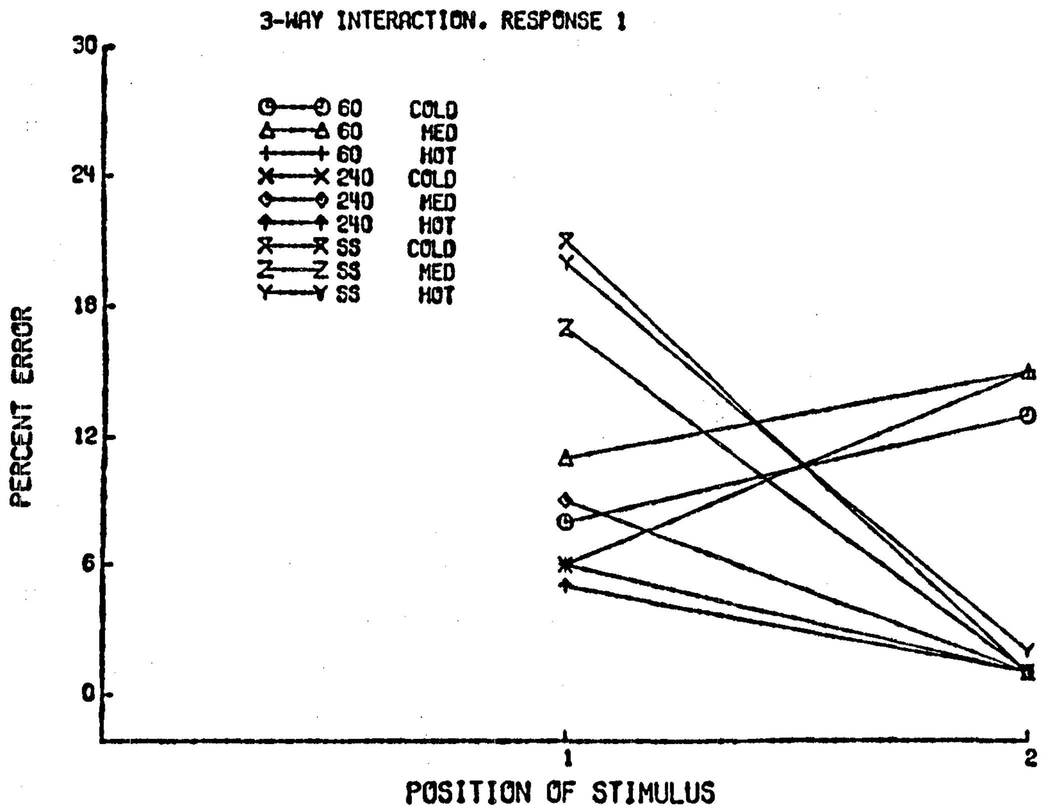
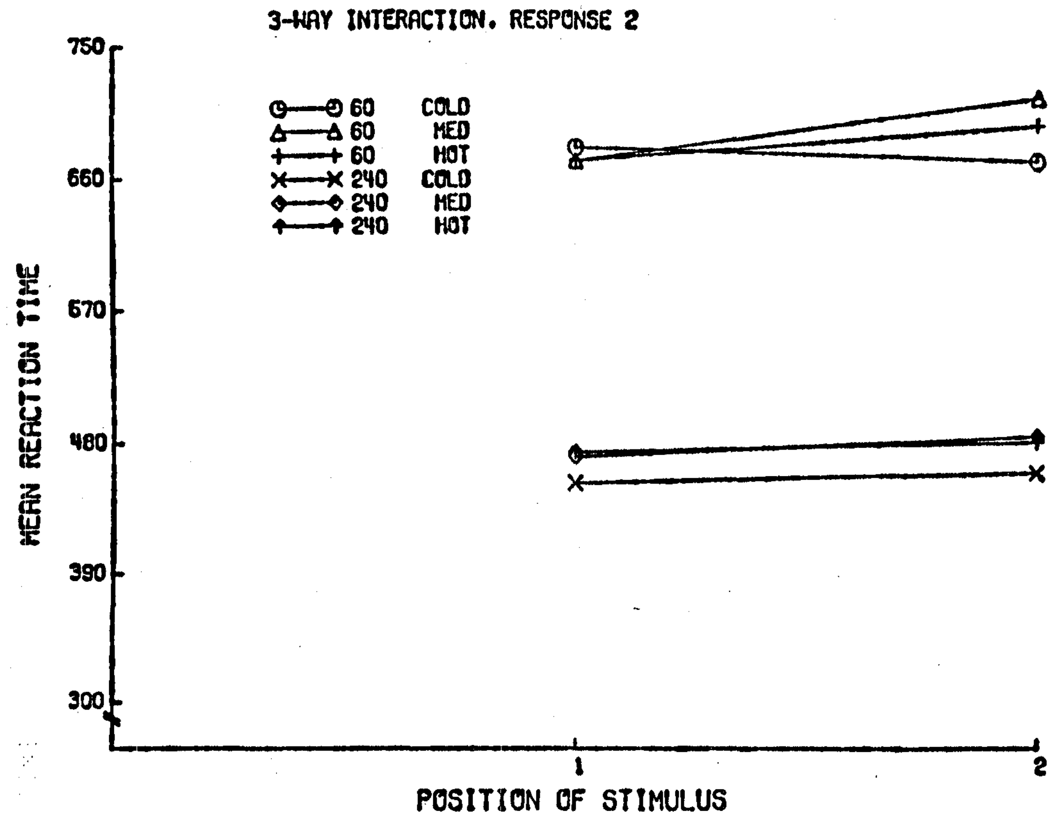
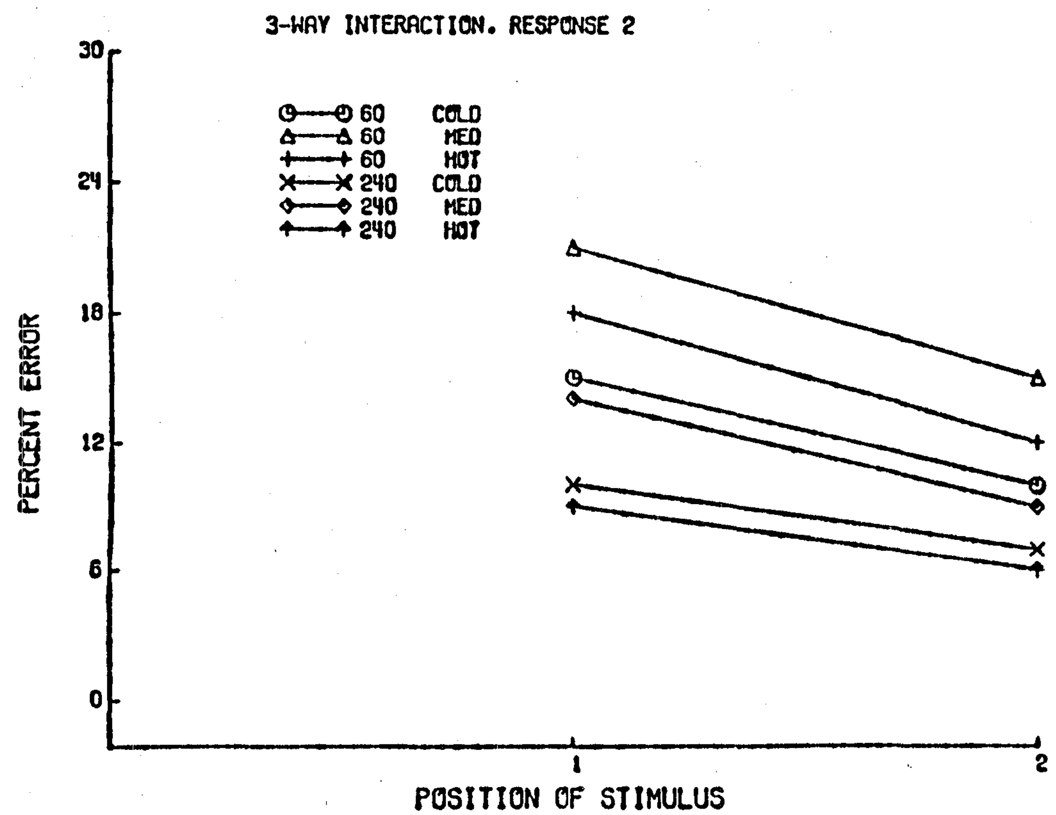

3-WAY INTERACTION. RESPONSE 1
PERCENT ERROR
30
24
18
12
6
0
60 COLD
60 MED
60 HOT
240 COLD
240 MED
240 HOT
SS COLD
SS MED
SS HOT
1
2
POSITION OF STIMULUS

Figure 15


3-WAY INTERACTION. RESPONSE 2
750
660
570
480
390
300
60 COLD
60 MED
60 HOT
240 COLD
240 MED
240 HOT
MEAN REACTION TIME
1
2
POSITION OF STIMULUS


3-WAY INTERACTION. RESPONSE 2
30
24
18
12
6
0
60 COLD
60 MED
60 HOT
240 COLD
240 MED
240 HOT
PERCENT ERROR
1
2
POSITION OF STIMULUS

End of Document