N82-13697

2-D RESULTS ON HUMAN OPERATOR PERCEPTION

Alex A. Siapkara
Control Division
Hellenic Technology Center
402 Mesogeion Av.
Athens, Greece

Thomas B. Sheridan
Man-Machine Laboratory
Massachusetts Institute of Technology
Cambridge, Mass. 02139

ABSTRACT

This paper presents some preliminary results of the application of multi-dimensional scaling methodology in human factors engineering. The non-orthogonality of internally perceived task variables is exhibited for first and second order plants with both dependent and independent task variables. Directions of operator preference are shown for actual performance, pilot opinion rating, and subjective measures of fatigue, adaptability and system recognition. Improvement of performance in second order systems is, in addition, exhibited by the use of bang-bang feedback information. New dissimilarity measures for system comparison are suggested in order to account for human operator rotations and subjective sense of time.

1. INTRODUCTION

In comparing the objective performance of a human operator (H.O) with his/her subjective evaluations it was found helpful in reference [1] to use the methodology provided by multi-dimensional scaling (MDS). According to this methodology, first proposed in detail by Torgerson in 1952 [2] in the area of mathematical psychology, unidimensional pilot opinion rating scales (POR) are only to be thought of as the vecotrs in a multi-dimensional space which represents the "perceptual model" the H.O. has "internally constructed" on the nature and performance of his/her own task.

This internal task space (ITS) is conceptually different from the ones arrived at by deviate internal models of the H.O., as for example done in reference [3], in that the MDS formulation does not rely on the LQG models of the H.O. and is in essence completely model-free. What this ITS really depends on is the metric assumed to apply in constructing it from comparison measurements provided subjectively by the H.O. and interpreted mathematically via the MDS formulation. These comparisons fix the distances between the compared objects within the multi-dimensional space and correlate with the unidimensional subjective (SE) and objective evaluations (OE) of various system parameters via the use of a vector-fitting algorithm.

A complete analysis of the mathematical model employed including the réal-time computational aspects involved in updating the ITS can be found in reference [4a]. The dimensionality of the matrix POR depends on the total number of tasks, compared, and more precisely

$$R^{n} \ni \quad n \geqslant N \;\; \& \;\; \log_{N} (n_{max} + 1) = \bar{n}$$

where n the dimensionality of the ITS
N the number of task variables
n_i the # of selected values for the i^{th} variable, and
$\bar{n}$ the geometric average of n_i's.

This logarithmic complexity of MDS experiments makes it imperative to:

i. partition the original set of tasks into partially overlapping smaller groups which will still retain the property of clustering them back together "under the same roof" of a single ITS;

ii. design separate experiments with the smaller groups employing a small number of selected values for each variable;

iii. reduce mathematically via a statistical hyperplane fit the total number of dimensions towards the number of actual task variables (if these are known beforehand and the experiments well controlled).

The prescribed technique has been successfully utilized by Siapkara [4b] in performing a series of three sets of experiments with three subjects and two task variables. The resulting two-dimensional ITS, though certainly of limited validity, does effectively portray many of the associated MDS issues. Section 2 of this paper describes the experiments and section 3 deals with the experimental results. All accompanying figures can be found at the end of the paper.

2. THE SCOPE OF THE EXPERIMENT

Figure 1 shows in a self-explanatory way the experimental set-up used. The signal source is a smoothed slowly-varying output from a random number generator (RNG), so that the configuration of the systems to be controlled [Figure 2] corresponds to both a first and second-order system. Which of the systems is actually controlled by the H.O. depends on which of the two displays is accessible for consultation, and not on any difference in plant dynamics. In order to keep a small number of variables at hand, the parameters of both systems do not vary independently [Figure 3]; they are all instead in terms of a common parameter λ.

The second task variable K comes from the input. An effort was made to use a random-number generator which experiences minimal statistical variations over time and different starting values, so that only its mean strength (amplitude) will count. The generation method was based on a combination of techniques proposed by references [5] and [6]. Consistency of the smoothed random input was checked for two completely different sets of task variables and measurement policies. [Figure 4]. It also checked comparatively well with the sinusoid method used in reference [7], with its statistics attaining an asymptotic stability much sooner, though the variations of the sinusoidal RNG are less frequent.

The use of fixed analog displays did not allow for the appropriate scaling of feedback indication, as would be the case with a flexible CRT screen. So, instead of performing the image range calibrations as outlined by equation (3) of reference [1], it was necessary to keep the number of display overshoots as an objective performance index in addition to tracking error scores. The spring characteristics of the vertical level knob used in the experiment was what Dommasch [8] would call a "bungee" (or down spring) control element, with a built-in center-wards pull which requires a constant off-center push in part of the subject in the case of proper control-force static stability. Transformation of the results to a situation which uses different control-element characteristics can be done via Rothbauer [9] [Figure 5]. Performance curves of these types of controlled elements is given in reference [10] [Figure 6].

Finally, provision was made in the experimental set-up to include a white-red light depending upon whether the tracking error was positive or negative. Experiments based solely on this type of feedback information will here-in be referred to as sign experiments. Justification for the inclusion of such an experiment is based on previously acquired insight in the field of experimental psychology. Johnson [11] mentions a 28-person 1968 experiment where multi-dimensional judgements correlated well with uni-dimensional equivalents when simple combinatory transformations of the various variables were performed. It was found that 42% judged on linear scales, 10.5% on quadratic, 45.5% on signed cues, and the rest in other configural modes. The signed cues were indicators with either +1 or -1 values, and they contained the sign information of the judgement only. For example, in the MDS context of paired comparisons the subjects would actually judge as if the stimuli were near the vicinity of just-discriminable differences: instead of rendering refined estimates of similarity, they would rather ask themselves some more foundamental questions:

> "Are the two situations compared different enough so as to bother giving out discrete and even more so continous estimates on scales beyond binary? And, if I admit there is a perceptible difference between them, will I be able later on on subsequent pair comparisons to maintain some credible consistency of how I rate these minute differences? Or, is it possible that I am going to develop the tendency of accentuating the dissimilarity scale near the similarity end of it, and compress thus my judgements on the truly dissimilar cases?"

3. EXPERIMENTAL RESULTS

Nine first and nine second order systems in all were evaluated. They were gotten by combining three values from each of the two variables as shown in Figure 7. The same figure shows also the uncertainties and just-discriminable differences related with each of these task variables. Various types of subjective judgements rated on a [illegible] 10 scale were collected in the experiments for these systems. The verbal characterization of these scales was only fixed at the ends, and is presented in Figure 8. For each category of systems there were essentially three kinds of runs performed: combined familiarization-evaluation runs, dissimilarity runs, and finally identification runs. For each run, in addition to the particular subjective judgement aimed at, objective performance records were kept, as well as subjective judgements on the level of fatigue experienced with the experiment and the level of difficulty (effort) in pronouncing the subjective evaluations themselves.

Familiarization - evaluation runs lasted 2 mins for each system. In the first 20 secs of familiarization SE on the success of familiarization was taken verbally around every 5 secs [Figure 9, located under Figure 6]. For 1-st order systems as $\lambda\uparrow$ so in general did the difficulty for familiarization (ex. VIII), but in no circumstance did the degree of familiarization decrease as time progressed. On the contrary 2-nd order systems with $K\downarrow$ experienced such a drop (ex. VII). The next 10 secs were considered a break between the familiarization portion of the run and the evaluation portion. During this interval a combined SE was provoked that indicated the efficacy of this relaxation period and the degree of comfort the subject experienced [Figure 10(a) ——]. The direction of relaxation increase is definitely different for 1-st and 2-nd order systems. The middle 60 secs of the runs are devoted to system evaluation. The subject with uninterrupted concentration on the task was previously instructed to perform his best. At the end of this period he/she gives an overall POR on the task [Figure 10(b) ---] , while OE are recorded for future comparison [Figure 10(b)——]. While OE increases with $K\downarrow,\lambda\downarrow$ for both 1-st and 2-nd order systems, SE's in general do not. Separated areas indicate entries which did not comform with the general direction of the property vectors and differ by more than one point in the psychological scale from what would have been considered as a value in acceptable deviation from the rule.

Along with the POR, the subject indicates his/her own mental and dexterity fatigue status, and the effort expended by him/her in rendering these SE's [Figure 10(c)]. A general kind of agreement can be seen for both SE measures used, while the scales utilized by the subject differ by one and two points, the subject being more harsh on rating the fatigue factor. The last 30 secs of these runs are used for deadaptation purposes till the start of the subsequent run. Figure 10a (vectors in segmented line) shows the degree of comfort felt by the H.O. It can be seen that inter-run deadaptation does not relate to intra-run relaxation, though both are comfort accomodating. This is so because relaxation on the same task is viewed by the H.O. simply as a means to reduce his/her fatigue, whilst deadaptation seems to depend more upon the anxiety of what comes next.

Dissimilarity runs lasted 70 secs for each pair of systems compared. Each member of the pair was controlled for 30 secs at the end of which OE was recorded. Figure 11 shows the variations in performance relative to the OE of evaluation runs as a standard of reference for both types of systems. 10 secs in between the individual presentations served as a relaxation during which the degree of ease for remembering the behaviour of the system presented first was recorded [Figure 10(d)]. The incomplete nature of the memorization space results from the factorial design of the dissimilarity runs which forms a minimum number of combinations to be compared. Despite the lack of additional information the essential character of the memorization vectors is evident, and suggests that the impression of remembering a system remains in direction the same for 1-st and 2-nd order systems. However, this does not necessarily mean that actual ability for such memorization is so. This point is discussed in more detail later on. At the end of the dissimilarity runs the H.O. judges the similarity of the systems compared, as well as his/her own degree of comfort in pronouncing this judgement. An off-line procedure for computing two-dimensional ITS's is invoked at the end of all the dissimilarity experiments.

The three sets of experiments involved tasks V_1-VI_1-IX_1, I_1-II_1--III_1-IV_1-VII_1-IX_1 and I_2-II_2-III_2-IV_2. After finding the task vectors based on these partitioned experiments, the ITS's are brought together by modifying the point dispersions so that the task vectors coicide [Figure 13(a)]. The rest of the systems $VIII_1$ and V_2-IX_2 are then placed within this combined ITS by a straight-forward two-dimensional interpolation procedure. The same is done with objective performance and subjective evaluation vectors for both 1-st and 2-nd order systems [Figures 13(b)+(c)].

4. DISCUSSION

Many could be the implications of these configurations, if it was not for the limited evidence for these internal space constructions. What is for sure is that the evidence collected on the few subjects of the experiments exhibits on the average tendencies which were sort of anticipated, and which motivated this study in the first place, anyway. These tendencies are:

i. the task vectors are not in general perceived independently,
ii. objective and subjective ratings of manual tracking tasks do not necessarily coincide.

The large inconsistencies in comparison distances (offsets that are 30% off from the Euclidean point of view), and the quite considerable variations in performance (uncertainties of the order of 36.4%) raised a number of questions on the validity/utility of MDS in the multi-dimensional assessment of POR's.

This motivated a third stage in the experiments beyond familiarization-evaluation and dissimilarity runs. More specifically, identification runs were specially conducted to test the hypothesis inherent in these experiments, that the H.O. could identify successfully the differentiating character of the systems he/she is confronted with. Figure 10d (vectors in segmented line) shows the directions of maximum increase in actual memorization, assuming that identifiability is strictly speaking a measure of memorization. In certain cases an almost complete lack of ability to identify a specific task is evident, as for example: III_1,VII_2,IV_2 and III_2. [Figure 12]. The short duration of the tasks can be cited as a main cause, because identifiability is a cumulative property which combines together the temporal elements of a task; and if the task has not fully developed its essential idiosyncracies this plays a negative role on the H.O. perceiving its global nature. Figure 14 shows, for example, the immense variability in the character of an almost marginal 2-nd order task in 10 secs intervals. This particular task reveals its true collapsing character after 50 secs, and certainly not within the first 30 secs from its activation. On the other hand, difficulties in memorization are unrelated to the factorial design of the similarity experiments. Figure 15 clearly shows that fatigue accumulated on entire groups of comparisons (based on the same first member of the pairs) does not seem to relate directly to the difficulty associated with identifying those systems.

Finally, the improved performance shown with sign rather than complete information on tracking error [Figure 16(a)] motivated a correlation between the various systems and the number of bang-bang pieces of feedback information. Figure 16(b) suggests that the implicit strategy used by the H.O. in optimizing manual tracking performance is to try to reach a uniform level of acceptance in the number of tracking error crossovers.

5. FUTURE DIRECTIONS

Despite some of the disheartening aspects of applying MDS on POR pronouncement, this has much more to do with H.O. inconsistencies than with a methodological difficulty and/or inability inherent in MDS itself. To the contrary, reference [4c] suggests that an MDS approach to the problems of mental workload in a multi-task environment and of multi-operator judgement and control (collective task-attending) would simplify their study by avoiding the use of the law for comparative judgement [12] and of group probability partitioning [13] , respectively.

Relating individual ITS's of various tasks or various operators, —firstly between themselves and secondly with the more complex ITS's resulting from multi-task or multi-operator situations—, would in the opinion of the authors provide us with usuful POR matrix transformations; for example:

i. between groups of people with different levels of aptitude in performing manual tracking tasks,

ii. for increasing the reliability of the operation by providing feedback information to the H.O. about discrepancies in his/her ITS between objective performance and its subjective evaluation,

iii. for the design of flexible control/display configurations which will automatically adapt their dynamics according to the ITS peculiarities of the H.O. involved in the operation so as to improve performance in a way transparent to the H O., etc.

Finally, it has to be noted that Euclidean or even Minkowski spaces provide metrics for ITS that are not suitable to express the contribution of terms that correspond to situations where H.O. space rotations and his/her subjective sense of time (the time thought of as having been ellapsed in the controlling action) play an important role. This is so because the base vectors representing a space rotation combine in a multiplication group [14] and those representing time belong to the spinor class [15]. This can be illustrated as follows:

Consider $\underline{y} = x_1\hat{e}^1 + x_2\hat{e}^2 + \ldots\ldots + x_n\hat{e}^n$

and $\|y\| = [s_1x_1^p + s^2x_2^p + \ldots\ldots + s^nx_n^p]^{1/p}$ (1)

where only superscript p is a power. Then for variables x_i that constitute

i. an inner product group, $s^i = +1$

ii. elements of a space rotation,

$$\hat{e}^{k_1} \odot \hat{e}^{k_2} \odot \ldots\ldots \hat{e}^{k_{p-1}} = \hat{e}^{k_p}$$

where $k_1, k_2, \ldots\ldots k_p$ in circular order

iii. a designant for the H.O. sense of ellapsed time (when experiments seem to depend on it), $s^i = -1$.

It can be seen that (1) applies only in the first and third cases, whereas the metric corresponding to the second case is given by the more general form

$$\|y\| = [\hat{e}^{k_1} \odot \hat{e}^{k_2} \odot \ldots\ldots \hat{e}^{k_p} \cdot x_{k_1} x_{k_2} \ldots\ldots x_{k_p}]^{1/p} \qquad (2)$$

following reference [14], the summation convention and the definition

$$\|y\| = [\underline{y} \odot \underline{y} \odot \ldots\ldots \odot \underline{y}]^{1/p}$$

The double-circle operator might be any legitimate operator, such as an inner or outer product, an integration of base functions, or even a convolution integral in the case of cascaded moving vectors, or a meaningful mixture of the above.

Whereas solving MDS with metrics of the form (1) seems to be a trivial extension of the case where $s^i = 1 \ \forall\, i$, this is certainly not the case with the much more general form (2), where coupled terms do in general appear.

ACKNOWLEDGEMENT

The authors would like to pay tribute to Dr. R.E. Curry of NASA Ames for shaping while at MIT in 1975 the original idea of using MDS in multidimensional POR, and suspecting the non-ortogonality of internally perveived task variables.
A. Siapkaras would, in addition, like to thank Dr. J.A. Argyris, Director of the ISD Aerospace Institute of Stuttgart University, FRG for partly releiving him from his contractual arrangements for preparation of the manuscript.

REFERENCES

1. Siapkara, A.A. "Multi-Attribute Subjective Evaluations of Manual Tracking Tasks vs. Objective Performance of the Human Operator" 13-th Annual Conference on Manual Control, June 1977, 118-125

2. Torgerson, W.S. "MDS Theory and Methods" Psychometrika 17 (1952), 401-419

3. Baron, S. and Berliner, J.E. "The Effects of Deviate Internal Representations in the Optimal Model of the Human Operator" Procs. of the 1976 IEEE Conf. on Decision & Control, 1055-1057

4. Siapkara, A.A. "Multi-Attribute Evaluations of Piloting Tasks" M.I.T. Eng.D Thesis, December 1978; (partially referred to as a = ps. 180-232, b = ps. 233-261, c = 262-287)

5. Zelen, Marvin "Random Number Generation" Handbook of Mathematical Functions, US/NBS 1970, 949-953

6. Hull + Dobell "Random Number Generators" SIAM 4 (1962), 230-254

7. Shirley, R.S. "A Comparison of Techniques for Measuring Human Operator Frequency Response" 6-th Annual Conference on Manual Control, 1970, 803-825

8. Dommasch "Airplane Aerodynamics" Ed. Pitman, GB 1961, 456

9. Rothbauer, Gunter "Influences of Joy-Stick Spring Resistance on the Execution of Simple and Complex Positioning Movements" 13-th Annual Conference on Manual Control, June 1977, 447-451

10. Frost, G. "Man-Machine Dynamics" The Van Cott & Kinkade U.S. Armed Forces Guide to Human Engineering Design, Ed. 1972, 236-241

11. Johnson, D.M. "Multi-Dimensional Judgements on Continuous and Discrete Scales" Systematic Introduction to the Psychology of Thinking Ch. 8, p. 379, Harper & Row 1972

12. Daryanian, B. "Subjective Scaling of Mental Workload in a Multi-Task Environment" 16-th Annual Conference on Manual Control, May 1980, 172-188

13. Ferrell, W.R. "A Model of Subjective Probabilities from Small Groups" 16-th Annual Conference on Manual Control, May 1980, 271-284

14. Glazman, I.M.; Ljubic, J.I. "Finite-Dimensional Linear Analysis", The M.I.T. Press 1974, (§5.6, §6.3)

15. Cartan, Elie "The Theory of Spinors" The M.I.T. Press, Mono. 37, 1966 (§5.6, §6.6)

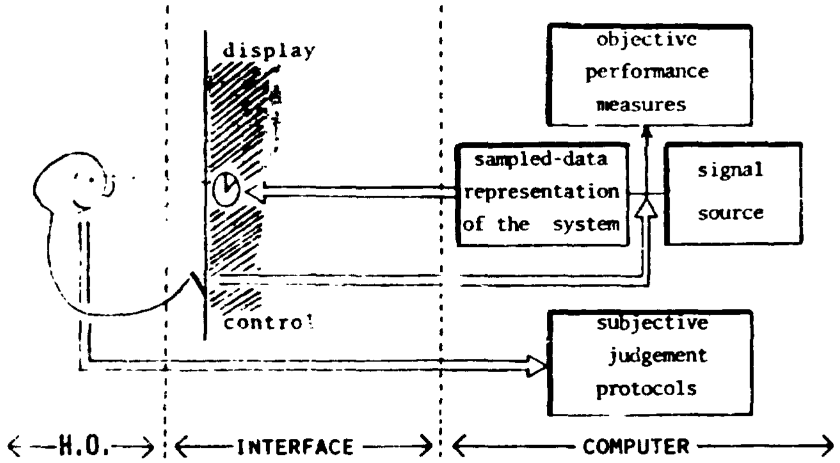


FIGURE 1

THE EXPERIMENTAL SET-UP

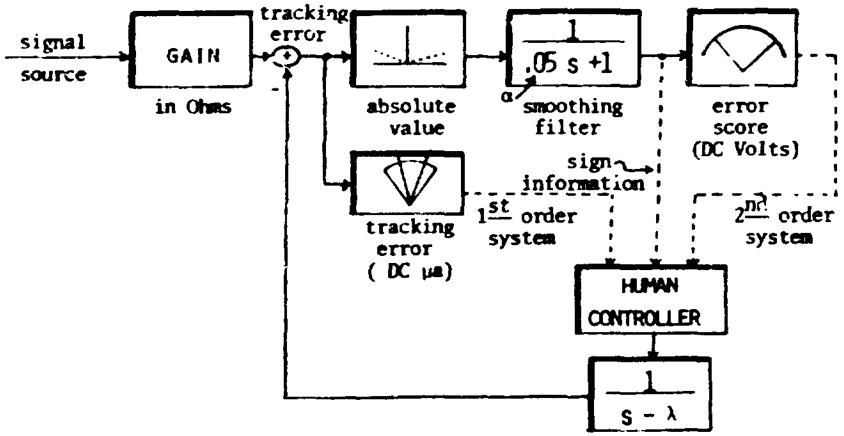


FIGURE 2

THE SYSTEM TO BE CONTROLLED

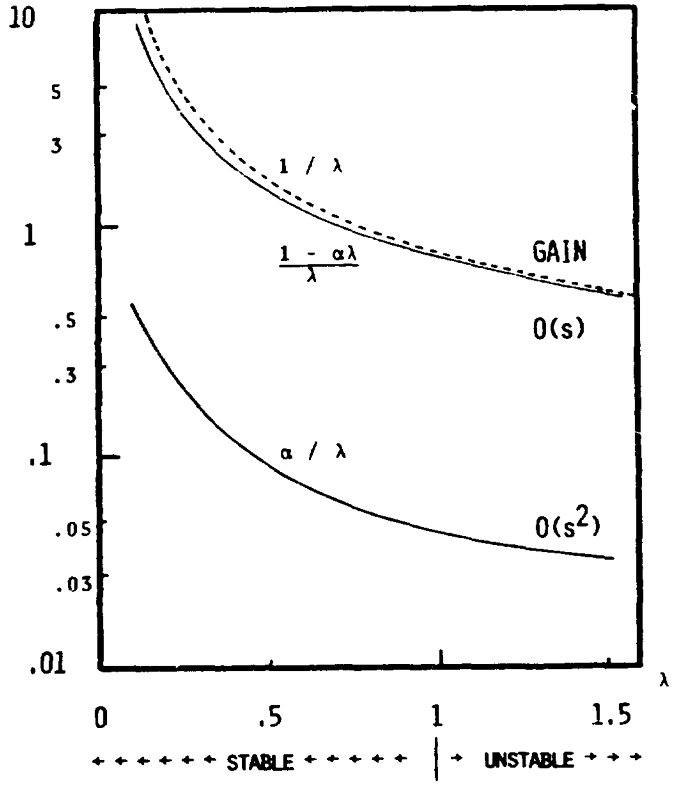
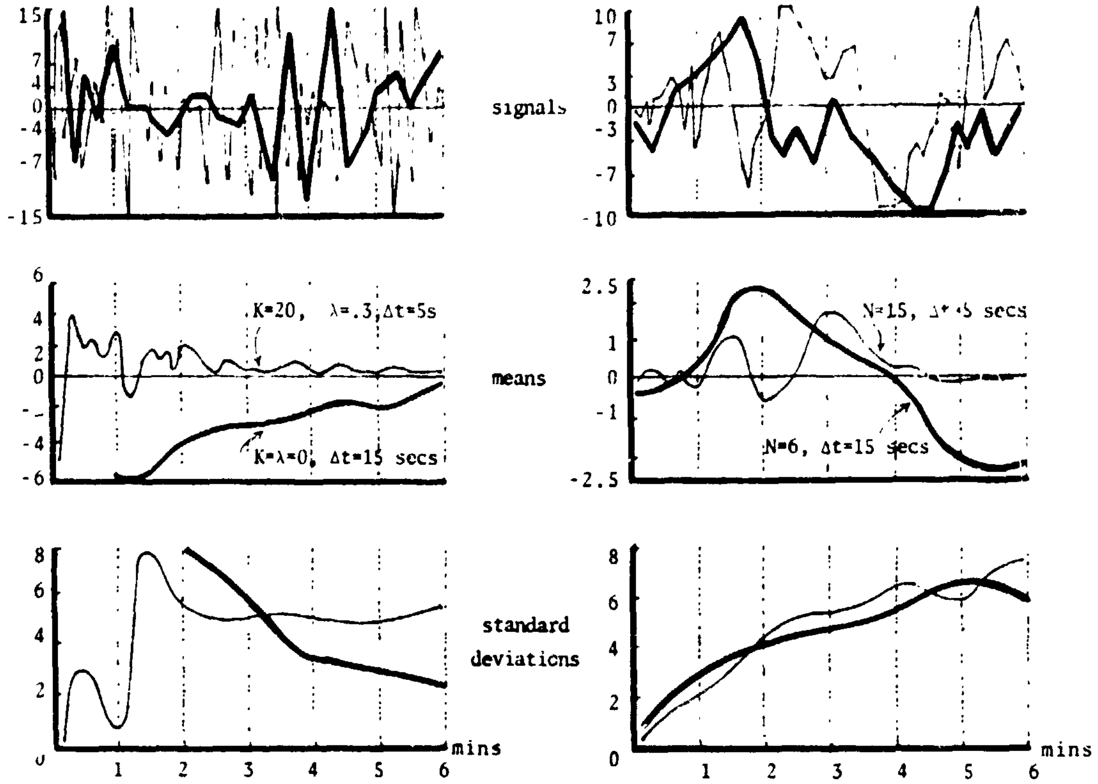


FIGURE 3

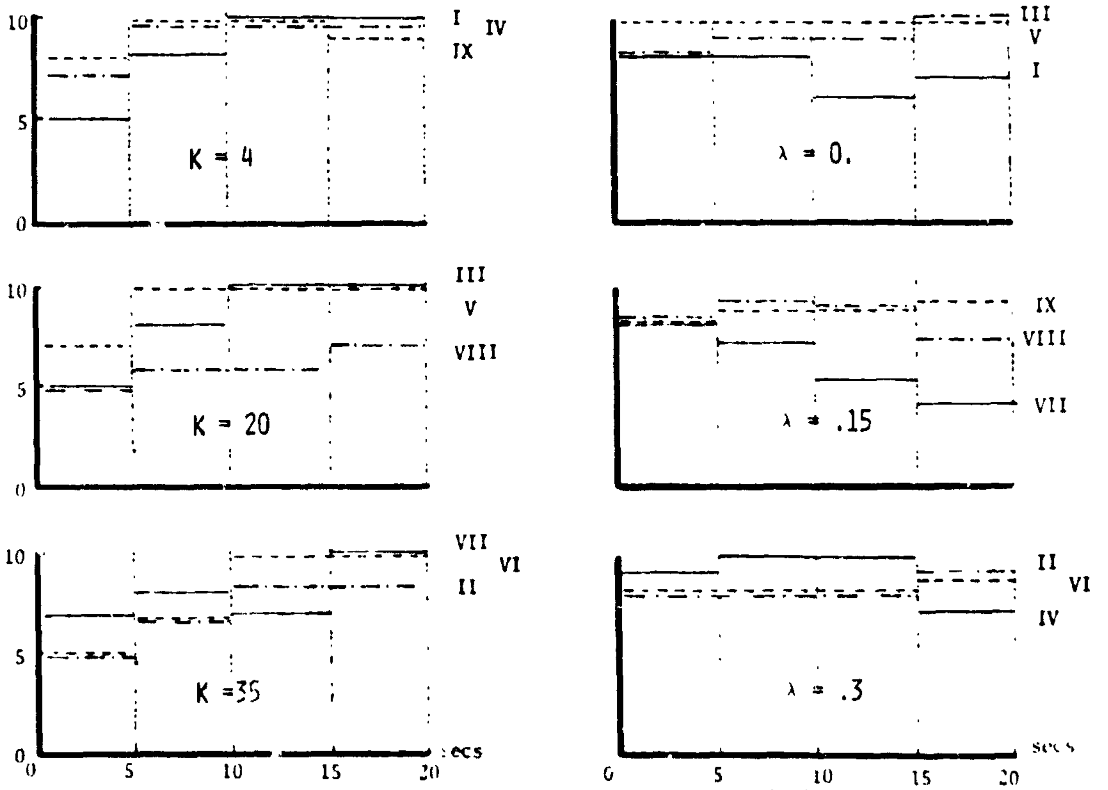
THE VARIABILITY OF SYSTEM PARAMETERS



the one actually used in the experiment

example of the sinusoidal RNG

FIGURE 4
CHARACTERISTICS OF THE SMOOTHED RNG



1-st order

2-nd order

FIGURE 9
HISTORY OF SUBJECTIVE EVALUATIONS ON FAMILIARIZATION

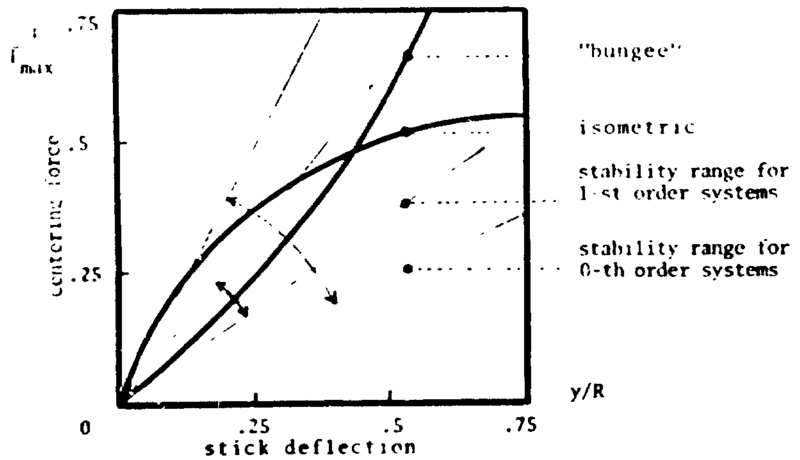
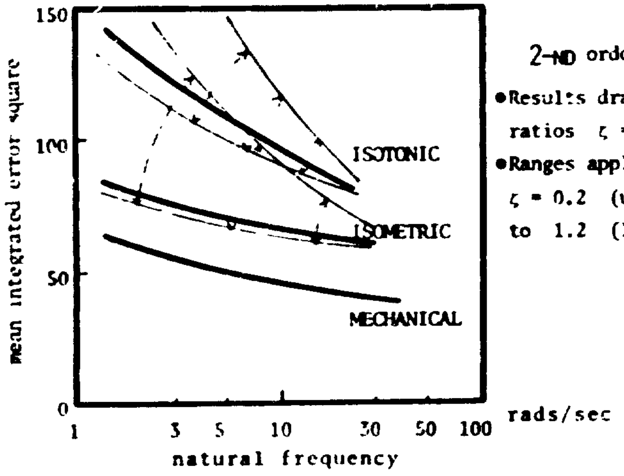


FIGURE 5

CONTROLLED ELEMENT TRANSFORMATIONS
(adopted from ROTHBAUER)



2-ND order systems
- Results drawn for damping ratios $\zeta = .8$
- Ranges apply for the region $\zeta = 0.2$ (upper margin) to 1.2 (lower ")

FIGURE 6
PERFORMANCE CURVES FOR VARIOUS CONTROLLED ELEMENTS

1st order

λ \ K	4	20	35
.1	I	III	VII
.3	IX	V	VI
.5	IV	VIII	II

2nd order

λ \ K	0	10	20
0.	I	V	III
.15	VII	IX	VIII
.3	IV	VI	II

thresholds

K	2	5
λ	.025	.1
	±2secs ∀ t	5 for t=30

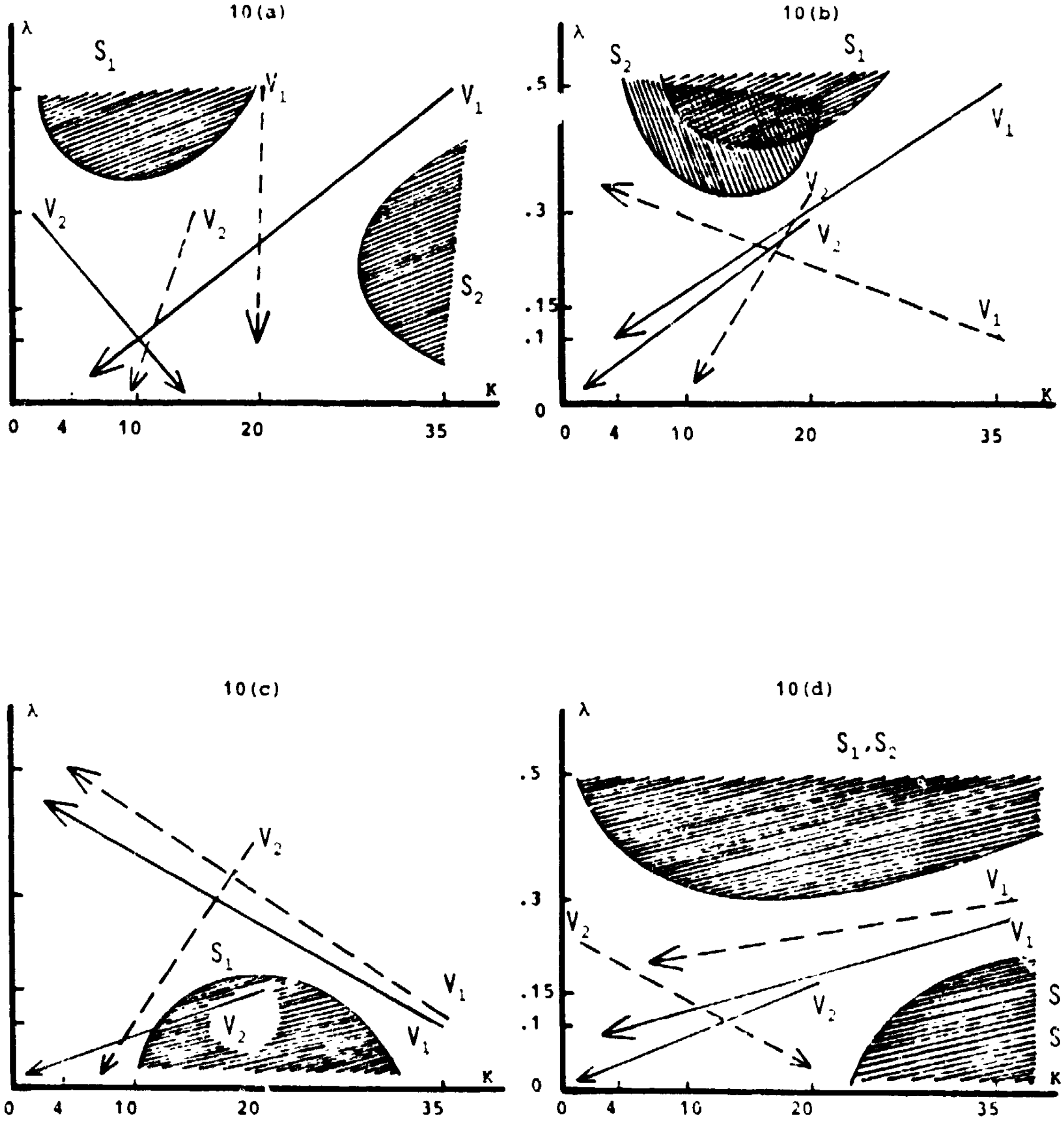
time intervals
jdd in perceiving
uncertainty in the settings

FIGURE 7
SYSTEM PARAMETERS

	1	5	10
FAMILIARIZATION :	uncomfortable		comfortable
DEADAPTATION :	unsatisfactory		satisfactory
RELAXATION :	low		high
SE / POR :	bad		good
FATIGUE :	fair		no
EFFORT :	fair		no
SIMILARITY :	low		high
COMFORT :	low		high

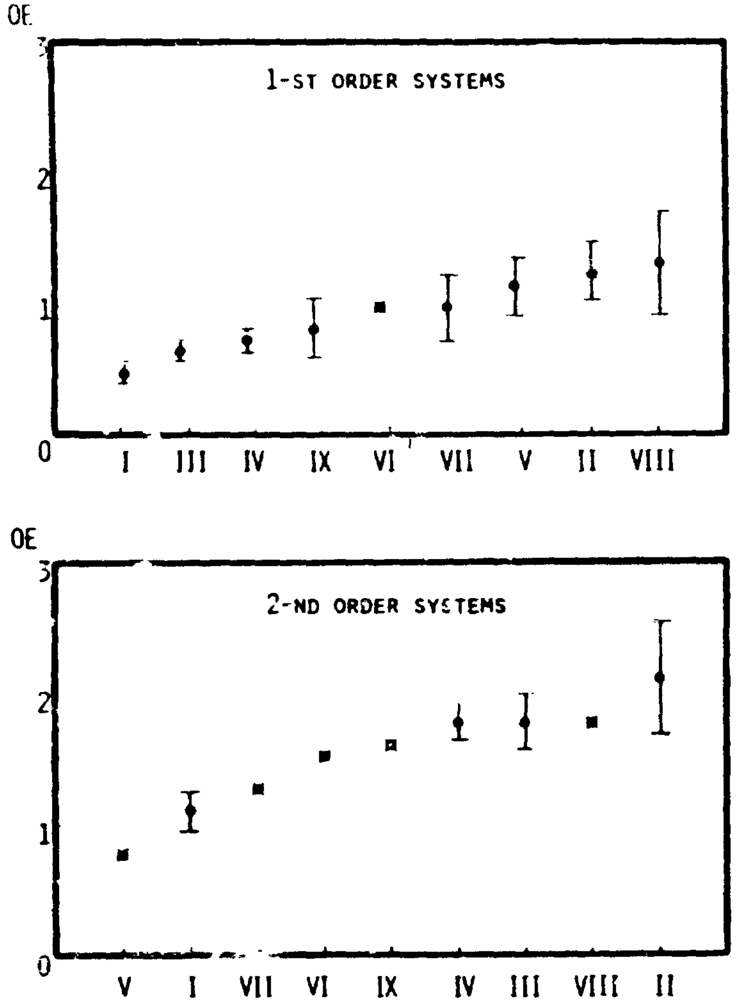
FIGURE 8

VERBAL CHARACTERIZATION
(in order of appearance
within an experimental session)



FIGURES 10

(a) INTRA- AND INTER-RUN RELAXATION VECTORS
(relaxation ——, deadaptation - - - -)
(b) PERFORMANCE AND JUDGEMENT VECTORS
(objective performance ——,
subjective evaluation - - - -)
(c) SUBJECTIVE STRESS STATUS VECTORS
(fatigue ——, effort - - - -)
(d) MEMORIZATION VECTORS
(subjective ——, objective - - - -)

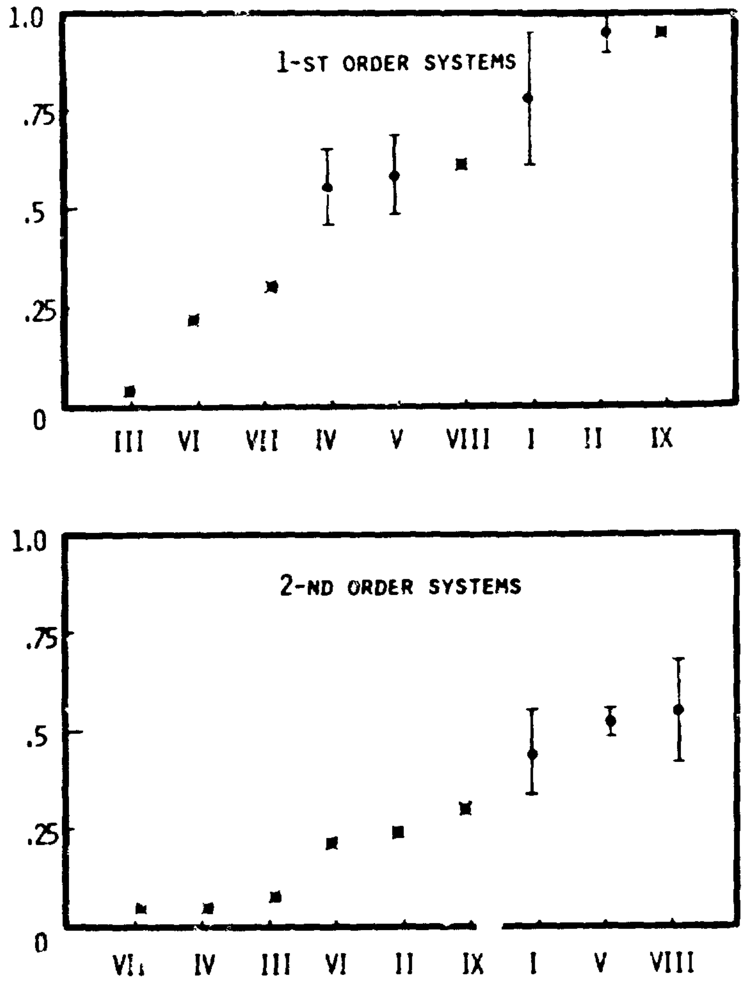


FIGURES 11

PERFORMANCE VARIATIONS

(systems are rank-ordered)

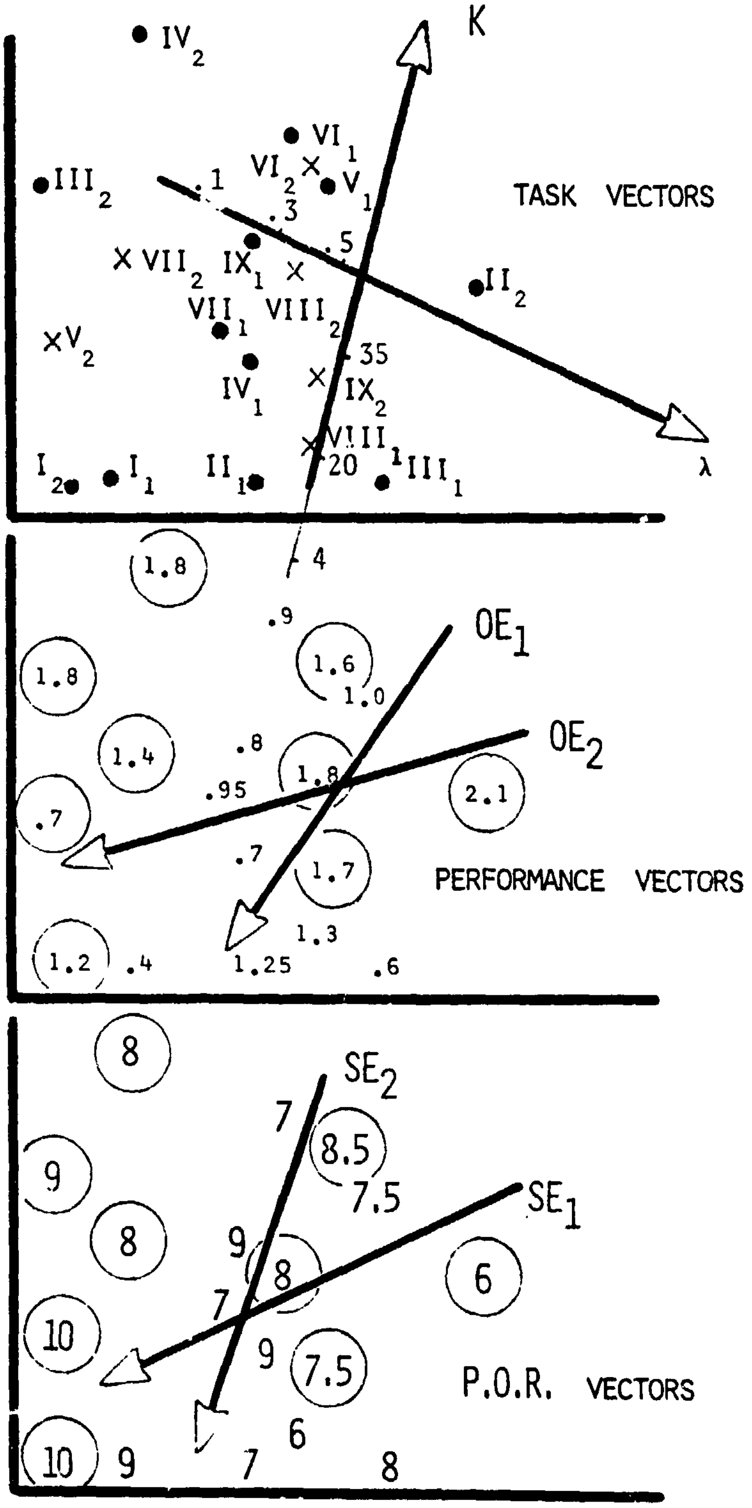
[■ there is no sufficient sampling to derive valid σ measurements]



FIGURES 12

SYSTEM IDENTIFIABILITY

(by the H.O.)



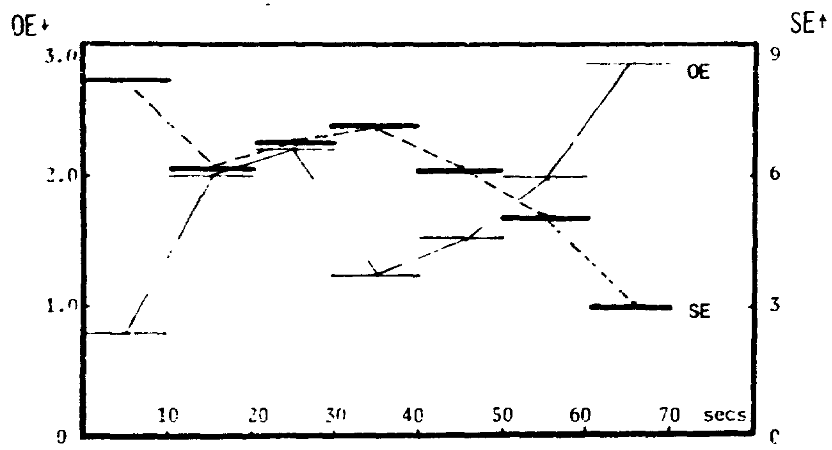
FIGURES 13

PROPERTY VECTORS

(IN THE DISSIMILARITY ITS)

[subscripts signify order of the system; x signify predicted locations of the remaining systems; circled numbers signify 2-nd order systems]

C-5



FIGURES 14

VARIABILITY OF TASK II_2

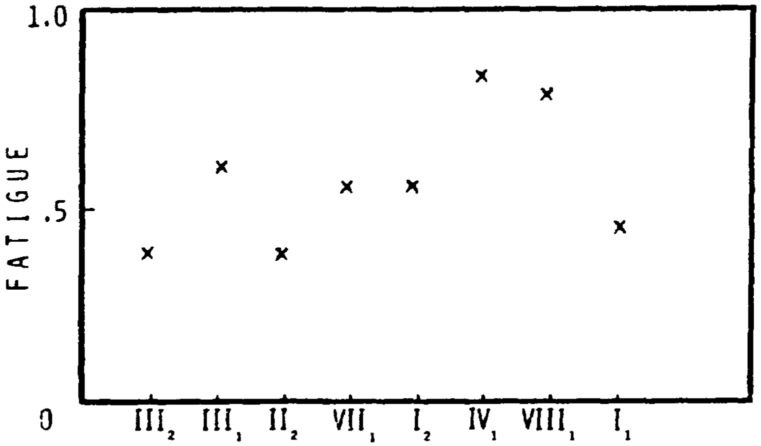
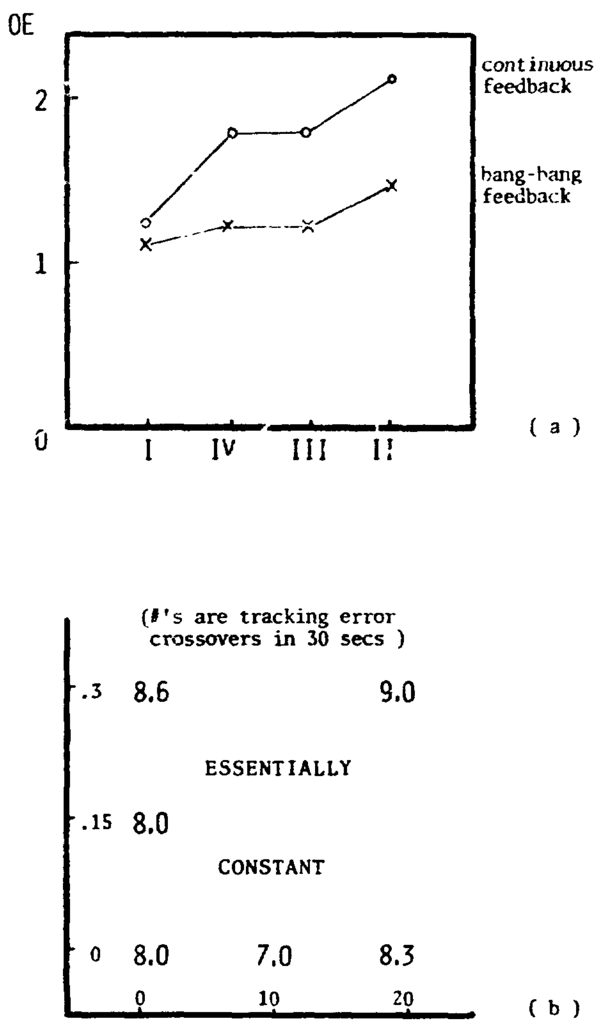


FIGURE 15

IDENTIFIABILITY VS. FATIGUE

(tasks are rank-ordered according to identifiability)



FIGURES 16

BANG-BANG FEEDBACK INFO

(2-nd order systems)