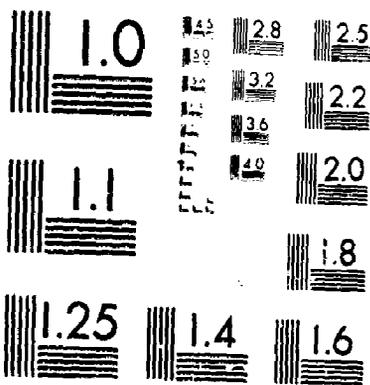


OFF

5821

UNCLAS



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

NASA Reference Publication 1055

Some Human Factors Issues in the
Development and Evaluation of
Cockpit Alerting and Warning Systems

Robert J. Randle, Jr., William E. Larsen,
and Douglas H. Williams
*Ames Research Center
Moffett Field, California*

NASA

National Aeronautics
and Space Administration

**Scientific and Technical
Information Office**

1980

TABLE OF CONTENTS

| | Page |
|--|------|
| <u>SUMMARY</u> | 1 |
| <u>INTRODUCTION</u> | 1 |
| GENERAL BACKGROUND | 1 |
| FORMAL DEFINITION OF THE PROBLEM | 4 |
| OBJECTIVES OF THIS DOCUMENT | 5 |
| Goals | 5 |
| Guidelines | 6 |
| <u>HUMAN FACTORS ISSUES IN COCKPIT ALERTING AND WARNING SYSTEMS</u> | 7 |
| <u>CATEGORIES OF WARNING SYSTEMS</u> | 7 |
| Performance Deviations | 7 |
| Configuration Deviations | 8 |
| System Deviations | 9 |
| <u>SINGLE ALERT LOGIC</u> | 9 |
| Alerting the Crew | 13 |
| Nuisance | 14 |
| Attention and vigilance | 14 |
| Response conditioning | 15 |
| Informing the Crew | 16 |
| Crew Option and Control | 19 |
| <u>ALERTING SYSTEMS IN THE AGGREGATE</u> | 20 |
| CAWS as a Subsystem | 20 |
| Master Warning Systems: Groupings | 21 |
| Integration Concepts | 21 |
| <u>METHODOLOGICAL CONSIDERATIONS IN MAN-MACHINE SYSTEMS EVALUATION</u> | 24 |
| SYSTEMS ENGINEERING: APPLICABILITY | 24 |
| TESTING FOR DESIGN DECISIONS | 27 |
| The Task Fidelity Continuum | 27 |
| Classical vs Engineering Studies | 28 |
| Criteria | 31 |
| KINDS OF TESTS IN SYSTEM DEVELOPMENT | 32 |
| Exploratory | 33 |
| Resolution | 33 |
| Verification | 33 |
| THE TOTAL SYSTEM | 35 |
| An Old Problem, a New Scenario | 35 |
| System Induced Errors | 36 |
| A System Information Processing Model | 37 |

| | Page |
|--|------|
| <u>PERFORMANCE EVALUATION IN CAWS</u> | 37 |
| WHERE TO TEST (CONTEXT)..... | 38 |
| In-Flight Testing | 38 |
| Flight Simulation | 39 |
| Part Simulation | 40 |
| WHAT TO TEST (CONTENT)..... | 40 |
| System Analysis | 41 |
| The operational requirement | 41 |
| Requirements from operational experience | 42 |
| System Synthesis | 44 |
| A sample alerting system | 44 |
| The informing function | 45 |
| Crew control | 47 |
| System Evaluation | 48 |
| Breadboard testing: system exploration | 49 |
| Component testing: design resolution | 51 |
| Operational testing: system verification | 53 |
| <u>REFERENCES</u> | 57 |

SOME HUMAN FACTORS ISSUES IN THE DEVELOPMENT AND EVALUATION OF COCKPIT ALERTING AND WARNING SYSTEMS

Robert J. Randle, Jr.
Ames Research Center

William E. Larsen
Federal Aviation Administration
Ames Research Center

and

Douglas H. Williams*
Ames Research Center

SUMMARY

The purpose of this report is to provide system development personnel with a set of general guidelines for evaluating a newly developed cockpit alerting and warning system in terms of human factors issues. Although the discussion centers around a general methodology, it has been made specific to the issues involved in alerting systems. The approach has been to look to the future in preparation for next generation commercial aircraft and the application of a more mature technology of automation. An overall statement of the current operational problem is presented, with an attempt to describe the more salient human factors problems with reference to existing alerting and warning systems. Next, the methodology for proceeding through system development to system test is discussed, with special emphasis on the differences between traditional human factors laboratory evaluations and those required for evaluation of complex man-machine systems under development. The last section deals more explicitly with performance evaluation in the alerting and warning subsystem using a hypothetical sample system. A further implicit purpose of this report is to engender an industry consensus as to a logical, efficient, and economical way to proceed to a new generation solution of the alerting system problem.

INTRODUCTION

NASA and the FAA have undertaken, at Ames Research Center, a review of the human factors associated with cockpit alerting and warning systems (CAWS). The purpose was to study – and where possible to outline – a method for assessing these systems in terms of human performance, acceptance, and general operability.

GENERAL BACKGROUND

There is a growing awareness that the principal cause of commercial aircraft accidents is human error. However, because the error is often embedded in a series of events, all of which contribute in

*NRC Associate.

a very complex way to a given mishap, it is usually difficult to identify a specific probable cause. Nevertheless, it is usually compellingly obvious that some kind of human error occurred and it is equally obvious that aircraft systems are seldom culpable.

Although it is difficult to identify, and almost impossible to classify, specific instances of human error in accident causation, the descriptions of the sources of human error — for example, high workload, fatigue, division of attention, cognitive and judgmental load, and crew coordination — are unanimous. One author (ref. 1) states:

Complaints about our lack of understanding of the "why" of human error are actually expressions of frustration about the inability to counteract every human shortcoming with technology. Although accident rates in all forms of flying made a nosedive during the last 25 years, one elusive factor has remained relatively stable: the percentage of accidents attributed to the pilot.

There is little doubt that any further significant advances in commercial aviation safety will be brought about by simplifying the considerable resources-management task of the flight crew. And this must be accomplished without reducing crew vigilance or resourcefulness, and without limiting the opportunities to exercise those aspects of the human subsystem which are superior to technological systems, such as adaptability, response to unprogrammed contingencies, flexibility, task performance, memory and discrete and analog intelligence.

One solution to some of the human weaknesses in monitoring complex systems was the provision of a sensory indication to the pilot whenever an out-of-tolerance condition occurred or was imminent. Cockpit alerting and warning systems were installed as a "reasonable" means of assisting the aircrew to maintain safe, reliable, economical system operation in the face of high workloads. However, these systems, intended to reduce hazard, are themselves becoming hazards. They introduce new problems related to an almost uncontrolled proliferation of CAWS on the newer more complex wide-body aircraft. In going from the B-707 to the B-747, the number of warning and caution alerts increased from 188 to 455 or 142%. The increase from DC-8 to DC-10 was from 172 to 418 or 143% (ref. 2, p. 48).

These problems associated with the various CAWS are summarized as follows:

1. The proliferation of CAWS has caused aircrews to frequently view the systems as a nuisance rather than a help.
2. More alerts require more memorization of meanings, higher workloads, and greater probability of error.
3. The credibility of alarms decreases as the number of false alarms, due to equipment failure and incorrect setting of sensor thresholds, increases.
4. Due to "response extinction" some frequently heard alerts actually go unheard or unheeded.
5. Aural alerts may function in competition with each other and their design may be tacitly governed by that consideration rather than by that of simply attracting attention.

6. Because cockpit alerting and warning systems have not been treated as true subsystems which have a common purpose, cohesiveness, and functional interrelationship, their proliferation has not been controlled.

7. An alerting signal - aural, visual, or tactile - increases the workload even if the alert is false.

8. Warning and alerting devices are sometimes used as shortcut solutions for problems that should be alleviated by better overall system design, for example, automation, operational procedures, etc.

9. The alerting value of any signal decreases dramatically as workload and attentional demands increase. Channellized attention

10. The absence of systems integration in the cockpit as a whole tends to retard the treatment of CAWS as a true subsystem thereof.

Some examples of problems associated with cockpit alerting and warning systems are paraphrased here to illustrate some of the problems of human response to them:

Case A (ref. 3): As a DC-9 passed through rotational speed - the point just prior to lifting off the runway - a false stall warning occurred. The stick shaker activated and the aural warning sounded. The pilot, reacting to these warnings, attempted to stop the takeoff but the jetliner overran the runway, struck several approach light stanchions, caught fire and burned. The pilot was convinced by the steady and persistent nature of the warning that it was valid and that the aircraft would not fly.

Any one of the numerous possible malfunctions could have activated the stall warning at rotation, the NTSB examining board said, but an examination of the stall warning system components produced no evidence of malfunction.

With a stall warning, there is little time to accomplish a validity check. And, particularly on takeoff, reaction must be immediate. The effects of this accident on the credibility of subsequent similar alarms is difficult to assess but will probably be significant.

Case B (ref. 4): A captain told of making an appropriate decision but of being placed in a situation in which only fortunate circumstances prevented a possible catastrophe. The aircraft flaps stuck in position 5 on approach and would extend no farther. The captain elected to transition to a 2° glide slope in order to be ready for faster spool-up should a go-around be necessary. While affecting the approach descent angle change, the sink rate increased to 1500 ft/min and glide slope was departed. The ground proximity warning system (GPWS) was triggered and, since there is no punch-out capability and the audio level of the voice annunciation is extremely high: "... all cockpit and tower communications were blocked at this time. Speed callouts, sink rate callouts, height above the runway callouts were not possible. This, in effect, denied a coordinated crew function." Under these conditions a go-around, which was possible given the reported environmental conditions, seemed to be hazardous and a very unattractive alternative. This probably would not

have been the case had the crew had the capability temporarily to deactivate the offending audio.

The GPWS has some special problems related to synthetic voice annunciation. The captain who related *Case B* also had some comments regarding the lack of identification of the reason for the "pullup" annunciation for four of the five GPWS modes.

Case C (ref. 5): This example is referenced as a pilot incident report but it describes a general response to a particular warning system -- the altitude capture and deviation alert. After describing the many alerting audios attributable to this system during a single flight, the author states, "this incident illustrates poor system design in which a 'warning sound' is heard repeatedly during normal operation. The 'warning' sound becomes a normal sound and its warning value is negated. In order to be effective, a warning sound should only be heard when there is a discrepancy. In operations each pilot hears this 'warning' sound approximately 360 times per month. Now, if once every 6 months a pilot makes an altitude error, he is faced with hearing a 'warning' sound to which he has been conditioned 2000 times in normal operation to ignore. This is an FAA requirement and should be changed. The light should be retained as at present, but the sound should only be heard during an abnormal operation such as an 'altitude error'."

The altitude alerting system, due to the operational difficulties associated with it, was recently the subject of an FAA rule change (September 21, 1977). Part 91.51 "Altitude alerting system or device; turbojet powered civil airplanes," now offers the option of having the aural alert not operating in the altitude *capture* mode; but it still must be active for deviations above and below the captured altitude. For a full discussion of this issue see reference 6.

The *Case C* report is illustrative of the response to very high false alarm rates. It is necessary to differentiate between false alarms occasioned by an actual sensor-alerting system malfunction and those due to noncritical or "routine" deviations or those due to sensor thresholds that are set too low. Though due to different sources the net result is decreased credibility and increased annoyances. These interfere with confident flight deck resource management by increasing perceived (perhaps actual) system indeterminacy. Indeterminacy in a system is roughly equivalent to uncertainty in an information processing sense. It refers to lack of clarity as to the number and kind of system inputs, the number and kind of options in an output set, and their interrelationship in a decision situation (refs. 7 and 8). Note that a determinate system may appear indeterminate to a novice or to one who has been improperly or inadequately trained on the operation of the systems.

FORMAL DEFINITION OF THE PROBLEM

In order to provide a conceptual model for use in thinking through problems associated with CAWS, it is necessary to identify basic functions that do not change as a function of individual warning systems. A successful technological approach to the solution of CAWS-related problems requires a formal model that is consistent and unchanging. The development of such a model is attempted here through a system orientation and by an adherence to functional categories identified as basic to all CAWS.

These functional categories provide a framework through which the several human factors areas may be identified and discussed. The approach begins to suggest the broader categories of behavior which may or may not be amenable to performance measurement. These behavior categories are identified as the alerting function, the informing function, and crew option and control.

The purpose of this document is not to design an ideal alerting and warning system but to suggest what is important to consider in the design and evaluation of such systems. However, in taking a systems view and in dealing in fundamentals, it is soon seen that optimism is justified for the next generation systems which combine the many seemingly disparate systems into a truly integrated subsystem, particularly if a microcomputer is used as a unifying information processor.

What also begins to be apparent is that *observable* human behavior vis-a-vis warning systems, is discrete, of low frequency, and dubiously quantifiable except in limited circumstances, for example, part-task studies of the alerting power of an aural alert. The human factors problems are seen, rather, as a particular subset of the more general problems of flight management, which is, in turn, a subset of the flight crew's overall resources-management task. Factors that influence that performance are: crew workload; crew physical and mental state; crew coordination; flight phase; aircraft type; crew training and backgrounds; emergency and abnormal procedures; company policy; navigation and Air Traffic Control (ATC); and weather.

The problems associated with the measurement of human performance with respect to CAWS are similar to problems associated with other flight deck activities with special emphasis on attention, division of attention, vigilance, and monitoring. A single statement of the human factors problems in crew response to CAWS is not possible; however, significant elements of the problems will be identified.

OBJECTIVES OF THIS DOCUMENT

Goals

The objectives of this document are defined as follows:

1. To provide a general statement that is representative of and in harmony with the desires of concerned individuals in the industrial, user, and government technological community in their pursuit of a significantly improved cockpit alerting and warning subsystem;
2. To provide a cohesive and useful description of the burgeoning CAWS problems;
3. To suggest a point of view from which CAWS may be regarded as an aircraft subsystem rather than as a collection of independent entities;
4. To describe the detailed task elements as a function of operational task requirements;
5. To describe human factors problems and to discuss the extent to which these can be examined independently of all other flight deck activities; and

6. To discuss human performance assessment methods and the extent to which they are relevant in terms of operational criteria such as system output.

Guidelines

The following guidelines define the internal constraints on the purpose, scope, detail, and, especially, the approach of this document:

1. A time frame is assumed that has constraints comparable to those existing in the development of a real system: for example, contractual, monetary, and end-item delivery.

2. An attempt is made to retain a system-orientation throughout; crew members are thus treated as human subsystems in a system.

3. Human factors evaluation – performance assessment – refers to the relevance and efficiency of the output of the human subsystem with regard to system purposes and goals.

4. The ground rules, strategies, procedures, and issues in behavioral, part-task laboratory studies are obtainable elsewhere. As topics in this report they qualify only to the extent that they contribute to and support the overall goal – human subsystem performance evaluation.

5. The context of system development and end-item delivery is not an appropriate context for the development of human factors research technology, explicit human performance hypothesis testing, or the pursuit of infinitely branching research questions regarding fundamental human capabilities.

6. The authors of this report are unaware of techniques for assessing complex performance unless such techniques are available to the community at large.

7. Experimental investigations in full-task situations contain all the problems involved in part-task, laboratory situations; the converse is not true. To be inclusive, this report is concerned with the former.

8. The human factors effort refers to assisting in determining design requirements and identifying criteria. Research requirements will proceed from the subsystem mission objectives and constraints, functional analysis, and activity analysis for candidate integrated systems. Research that may be required is motivated by specific questions arising in the system engineering design process.

9. A general human factors approach is outlined, illustrated by its application to a specific subsystem – the cockpit alerting and warning subsystem.

10. The evaluation strategy discussed is with reference to a CAWS that has been designed with knowledge of the larger system context and of explicit operational requirements. This means that system optimization rather than system improvement is the guiding philosophy.

11. Only the more salient points in a wide body of related literature are included in this report.

HUMAN FACTORS ISSUES IN COCKPIT ALERTING AND WARNING SYSTEMS

CATEGORIES OF WARNING SYSTEMS

There are several ways that the many kinds of alerting, warning, caution, and advisory displays might be categorized. To some extent the method chosen depends on the purposes to which the categorization scheme is to be put. Since the purpose here is to clarify a seemingly disparate collection of independent entities, a starting point is selected based on the kinds of deviations that make the alerting signals necessary. Three general operational sources of aircraft deviations are: (1) performance deviations; (2) configuration deviations; and (3) system deviations. Within each of these classifications it is possible to further categorize the deviations in terms of the urgency with which the deviation must be corrected.

Each of these categories will be discussed below. In each case examples will be cited using the three major wide-body aircraft: the Boeing 747, the Lockheed L-1011, and the McDonnell-Douglas DC-10. A word of caution is in order. It is understood that many variations in the specific embodiment of even "standardized" alerting and warning systems may be found due to aircraft evolution, user practices and procedures, and changing regulatory requirements. The sources of the examples discussed below are references 9-11. Although the validity of the examples is limited to the scope of the referenced documents it is still sufficient for illustrative purposes. The B-747 operating manual is that of Pan American Airways. The other two manuals were obtained from the airframe manufacturer and reflect operating procedures used by several airlines.

Performance Deviations

Performance deviations refer to aircraft departures from safe flight profiles. They have a high level of urgency and frequently require immediate action to correct a potentially hazardous condition. They are usually signaled by an audio alert and, in conjunction with configuration deviations (discussed below), are of prime importance in the increasing problems associated with cockpit alerting and warning systems generally.

Examples of this class of CAWS are: ground proximity warning system (GPWS), altitude alert, and excessive Mach airspeed alerts. When the aircraft enters one of the five GPWS warning envelopes (excessive sink rate, terrain closure, descent during takeoff, not in landing configuration below 500 ft above ground level, low on glide slope) a "Whoop, whoop, pullup!" message is activated and a warning light is flashed for the first four conditions. The "whoop" is a swept tone, the "pullup" is a synthetic voice. For the fifth mode, the synthetic voice iterates "glide slope." The GPWS is standard on air transport aircraft.

The following altitude-alert modes are in conformance with FAR 91.51, dated August 31, 1971. This regulation was rewritten to eliminate the aural warning on approaching altitude. It was incorporated on September 21, 1977.

1. B-747: In the capture mode, a 2-sec chord tone and a steady amber light are activated ±900 ft from the selected altitude. The amber light extinguishes at ±300 ft from the selected

altitude. In the deviation mode a 2-sec aural sounds and the amber light flashes when ± 300 ft is exceeded. It continues flashing until ± 900 ft is exceeded at which time the system automatically resets.

2. L-1011: The L-1011 is similar to the B-747 using a C-chord tone and amber light for the functions described.

3. DC-10: The DC-10 uses a 2-sec dual air horn and an amber light for these functions.

The setting of the two bands bracketing the selected altitude is a function of the airline choice so will vary from carrier to carrier.

The overspeed alert warnings are as follows:

1. B-747: The 747 uses a "clacker" for overspeed warning.

2. L-1011: The L-1011 uses a clacker sound.

3. DC-10: The DC-10 uses a "clucking sound" for the alerting function. This sound is also used as the slats extended warning. Some aural warnings have a dual meaning and are assigned to functions in separated flight phases.

Configuration Deviations

These deviations differ from performance deviations in that some positive action on the part of the crew has been omitted preceding a transition from one flight phase to another. The action omitted is one that preconfigures the aerodynamic profile of the aircraft for controlled, safe flight in the intended flight regime. As with performance deviations, immediate remedial action is usually required when configuration alerts activate. Examples include unsafe takeoff and landing configuration and open doors.

Alerts for unsafe landing configuration are as follows:-

1. B-747: A steady horn is sounded when the gear is not down and locked and any thrust lever is retarded to idle with flaps at 1° , 5° , 10° , or 20° or the gear is not down and locked with flaps at 25° or 30° , thrust levers in any position. In the first case, the horn can be silenced by activating the warning horn cutout switch on the aft section of the center console. In the second, it can only be silenced by pulling the aural warning circuit breaker. In both cases the horn is silenced by selecting the correct configuration.

2. L-1011: A steady horn is sounded if the gear is not locked, the flaps are not extended more than 30° , airspeed is less than 180 knots and any throttle is retarded more than 57° . The horn cannot be "punched out" when the first two conditions exist.

3. DC-10: A continuous "car" horn sounds when thrust is retarded to idle and the gear is not down and locked and airspeed is less than 215 knots. The horn may be silenced by the "horn off"

button if the flaps are in an approach configuration ($<28.5^\circ$). If the flaps are extended beyond 28.5° (say, 35° landing configuration) the horn can be silenced only by extending the gear.

Prior to flight, all aircraft doors must be closed for aerodynamic, environmental, engineering, and safety reasons. An open door or a faulty door condition is usually indicated by amber lights. Reference 2 gave the following numbers of unsafe door alert lights for the three wide-body aircraft: B-747, 16 lights; L-1011, 12 steady and one flashing amber light; DC-10, 20 lights.

System Deviations

Aircraft systems and subsystems are numerous, complex, and frequently redundant. Malfunctions and faults in these many systems are usually signaled by needle positions, instrument safe operating bands, flags, and lights. It should be noted that the deviation of an instrument from a nominal value or band is of itself a form of alarm but one that is not explicit as is the case with flags and lights. As with other alerting signals there has been a steady proliferation of system deviation warnings. The increase seems to be in response to and parallel with the concurrent growth in the demands made on the crew by dividing their attention among the many functions involved in the complex flight-management task.

Reference 2 shows that the B-747 has 37 bands, 65 flags, and 655 lights; the L-1011 has 7, 152, and 639; the DC-10 has 110, 53, and 463, respectively. Most of these do not have a high urgency level and frequently require only that an alternate subsystem be selected. However, if engine fire may be included in this category of deviations it is a notable exception, one in which the urgency level is high and one that requires immediate action. Because of the seriousness of on-board fires, the fire warning (engine, wheel-well) is always accompanied by an aural alert. A stereotyped sound, a bell, is used whose specific characteristics may vary somewhat (e.g., intermittent vs steady ringing).

Finally, deviations can be indicated by controls and displays *not* being activated; for example, gear, flap, throttle, speedbrake, out of position, and normal operation lights being extinguished (blue, green, white lights usually). And, to complete the set of deviation indicators, all cockpit displays provide deviation signals. Reference 2 provides a full listing of the many CAWS in current aircraft and is an excellent source of data that have been categorized in several ways. The three categories chosen above are for convenience only.

SINGLE ALERT LOGIC

For the purpose of human factors analysis, CAWS can be further compartmentalized as *single* alerting systems versus alerting systems in the aggregate. A discussion of the single alert directs one's attention to the operational and functional requirements that underlie the instrumented system. A schematic diagram of the simplified alert response logic is shown in figure 1.

The primary purpose of an alerting and warning system is to direct the attention of the crew to an impending or current system or subsystem deviation. This kind of signal is necessary for two main reasons: (1) it is not always possible to fully monitor aircraft system status due to the many

facets of flight management demanding attention, decisions, and cognitive activities; and (2) many subsystem faults are not readily apparent by reference to the displays normally associated with them. Implicit in the alerting function are requirements to *alert*, or capture the attention of the crew, and to *inform* the crew as to the source of the deviation.

What is not so apparent is the possible requirement that the crew be given some *control* of the alerting signal *itself* other than by correcting the signaled deviation. Such is already the case with some alerts. For example, the landing gear warning horn can be "punched out" under some conditions and master caution lights can be reset, some while an amber, labeled light remains on until the deviation is corrected (the causes of system deviations are not actually corrected; rather, an alternate system or system configuration is selected).

In figure 1, the significant events related to the onset of a single deviation and alert and the associated response logic are shown in schematic form. Events beyond junction (3) in figure 1 are human responses related to the alerting function. The success or failure of the sensor (2) and/or associated circuitry is of interest here only to the extent that it may influence subsequent human performance (for instance, junction (5)). Once it is assumed that the alerting signal (the alarm) has been activated, the input boundary of the CAWS task can be discerned. In order to establish criteria for human performance evaluations it is necessary to make explicit the task boundaries by reference to system function *and* task demands both by abstraction, as here, and, ultimately, in context, as embedded in flight-management activities.

In junction (4) an alerting signal may go unperceived for a variety of reasons, not the least of which is a conditioned response to ignore an alert that is activated frequently and is only conditionally informative (altitude alert). Workload and division of attention are strong influences at all junctions. If the alert is not perceived, a missed alert (outcome (A) in fig. 1) ensues. If it is perceived, a validity check (junction (5)) may be in order. If the alert is not valid there may or may not be a response to it (junction (9)). If there is a response, it is a false response and outcome (C) occurs. Both the missed alert (A) and the false response (C) can lead to system state contingencies. If the alert is not valid the false alert conclusion (B) is a correct response.

There are several ways in which false alarms may be triggered. One is simply failure of one or more elements in the sensing-alerting loop. Another is the triggering of a momentary alert in making external or internal configuration changes. A third is due to the lack of mission phase-adaptive cockpit alerting and warning systems; for instance, a gear warning horn during other than the approach phase. Another source of "false" alerting signals is that due to setting the sensitivity threshold of the sensing system too low. As is well known, the detection probability in a signal detection system determines how large the ratio of real to false signals will be and this can be predetermined. Some such process is involved in the altitude alerting system in presetting the band above and below selected altitude within which (e.g., ± 250 ft) deviations will not be signaled. Where the band is to be preset depends on the values deemed to be significant in terms of user system performance criteria. Wider limits (± 500 ft) would increase the number of missed alarms; narrower limits (± 50 ft) would increase the number of false alarms (note that, in the latter case, taken to the extreme, the altitude alerting system could be used as a precision altitude tracking system with auditory feedback). These remarks refer to all deviation sensors that monitor continuous variables but not to binary variables; that is, to such variables as temperature, altitude, sink rate, pressure, but not to switch closures, valves, equipment go-no-go states, signal loss, etc.

If the signal is valid, a response occurs or it does not (junction (6), fig. 1). If there is no response it may have been ignored (junction (10)) in which case event (E), missed response, occurs, which is the identical outcome to an unperceived alert, event (A). However, it may not have been ignored but rather postponed to be given attention later (junction (11)). Some alerts signal continuously as long as the deviation persists – GPWS, stall warning, gear horn, alert lights. Others are momentary – for example, altitude alert. Some, thus, have an innate “storage” function, even if it is only a panel light remaining on, but have varying degrees of attention-demand both initially and subsequently.

If the alert is forgotten (junction (12)) beyond a critical delay interval, event (E), missed response, occurs and, again, the reason cannot be determined by objective observation. If the postponed response is activated then its timeliness is a consideration. It may be delayed too long because of activities in the “no immediate response” loop ((10), (11), (12)) or it may be accomplished too quickly as in too hasty engine shutdown. Even if the response is not postponed the timeliness of the response is a major consideration. Both the “too soon” and “too late” responses can result in overall system state changes.

If the correct response has been selected in a timely manner (junction (8)) event (F), a true response occurs and the system continues toward its goal. The system state may be unaltered or altered. **Whether a crew member responds to an alert by correction of the deviation depends on several things:**

- 1. The physical characteristics of the signal -- its alerting power**
- 2. The urgency level of the deviation**
- 3. The results of the validity check**
- 4. Phase of flight**
- 5. Workload level**
- 6. Crew option**

Figure 1 indicates, simplistically, continuation of the deviation for an incorrect response, but an incorrect response can also lead to new deviations and other than normal operations. This is also true of a correct response. The non-normal operations shown in the “A/C System” box is arbitrary, meant only to refer to configurations of varying criticality.

Figure 1 is meant to illustrate single-alert logic. However, the several single alerts in the case of multiple alerts do not take on a different logic. What happens then is that the rationale of postponement and the timeliness of the responses become more critical because of urgency levels and optimum sequencing of deviation correction activities. A newly conceived CAWS might allocate those functions to the machine.

It should be recognized that the single-alert logic has been stripped to its bare essentials in terms of human response. The development and evaluation of CAWS demands a detailed task analysis describing the actions and interactions of the crew over time in response to various and combined alerting signals in concert with on-going flight deck activities.

Alerting the Crew

Channelized attention is an example of perceptual narrowing

The prime functions of current alerting signals are to capture the attention of the crew and to direct that attention to a source of deviation so that appropriate action can be taken. To add to an alerting signal coded information pertaining to the nature of the deviation is a design economy that can be effective if there are only a few alerting systems. However, there are so many, unstandardized alerts in modern cockpits that the several meanings of coded aural cannot be retained with confidence, particularly by transitioning pilots or under conditions of high workload or stress when perceptual narrowing can occur. Perceptual narrowing is associated with human performance under high stress. Related phenomena under such conditions are attention fixation, reversion to well-learned automatized responses, and reduced ability to accept new inputs. Perceptual narrowing may be manifested in a sensory narrowing such as tunnel vision – a decreasing sensitivity to peripheral visual stimuli and events. In the extreme, stress may produce a “white-knuckle freeze.”

A great deal of attention has been given to determining how well an alerting signal can attract the attention of an observer. The kind of evaluation employed, however, has been of single alerts in experimental settings where other flight activities and busyness were usually only analogically or inferentially represented. It is an assumption of this report that attempts to make current, single-alerting systems better is not profitable. This is partly because to do so fosters a competition between alerting systems with the most recent being the most attention-getting; moreover, the emphasis may be tacitly placed on competition between alerting systems themselves rather than between an alert and ongoing flight-management activities. The capture of the crew's attention seems much more a matter of vying for a limited attention capacity and how well the alert does this varies not so much as a function of alert stimulus parameters but as a function of the constantly shifting human threshold of attention and arousal to the reception of new inputs, that is, new load. Pilots state that some aural alerts are not *heard* even though the sound levels (stimulus intensity) are usually set above 90 dB (90 dB is about equal to the sound of a subway train at a distance of 20 ft for the 75 to 1200 Hz frequency band (see ref. 12)). Reference 13 discusses the current alerting and warning system problem and, at one point, states:

It is clear that our present-day warning systems are inadequate and unsatisfactory. They may be further sophisticated by the addition of colors, horns, bells, and buzzers, but in their present form they will never really become fully compatible with the human being in the cockpit. The time has come for a careful reevaluation of our design philosophy.

There is a relatively large literature reporting the results of human performance in response to warning signals where the independent variables were stimulus dimensions. For vision, stimulus parameters are size, brightness, contrast, location, format, color, workload, vigilance, coding, and subject age. For auditory signals they are frequency (bandwidth), intensity, location, background noise, signal number and rate, and vigilance. Tactile signals may be of interest also but they would seem to have limited application except for the universally used stick shaker stall warning.

Part-task simulation such as these laboratory evaluations of single alerting systems where the system context (flight deck) is represented by simplistic analogs are not the major concern of this report. Techniques for conducting these studies are well known and available to the human factors practitioner. Their use in system development is most beneficial if the configuration of the

developing system can be more faithfully represented and the task more clearly particularized (for a full discussion of this point and related methodological considerations see ref. 14).

Specific problems related to the human response to the CAWS alerting function are the following:

Nuisance- Alerting signals are increasing in number as systems increase in complexity. These are not all aural signals but the annoyance expressed by aircrew members is usually in response to the intensity, inadvertent activation, frequency of activation, false alarm rate, and intrusiveness of the aural (17 in the B-747, 15 in the L-1011, 15 in the DC-10). The consequences of this kind of response may be nil in terms of criteria for the successful completion of flight phases and ultimate system output. However, reduction and simplification of CAWS for reasons more relevant to system performance criteria would produce benefits even in this area.

Attention and vigilance- The alerting signal is supposed to be an attention-getting device. Under quiet cockpit conditions a single aural will probably accomplish that and the crew would not be hard-pressed to attend to it. With flags and lights the state of vigilance of the crew would be an important factor due to the need to visually scan the panels or to be actively engaged in using a display upon which a flag appeared or next to which a light came on. Thus, the state of arousal or vigilance level of the crew is important for signal detection where flight deck activity requirements approach quiescence. System and flight profile monitoring can become overly relaxed and subtle deviations missed, especially non-aural signals. On the other hand, the intrusion of a loud aural alert during, say, the performance of a precision task requiring concentrated attention can have a "startle" effect; its effect on task performance is not known but could be considerable.

At the other end of the task activity scale (high workload) arousal may be high, but human channel capacity becomes a significant factor. It is not necessary to posit the human as a *single* channel processor of information to give this fact credibility. Many behavioral studies have shown the decrement in performance related to task demand, particularly in the reception and processing of information (see ref. 14). The findings are that humans are limited in input load processing. The view taken here is that of Chiles (ref. 15): "... the human operator is influenced by too great a variety of factors to try to permanently settle the single channel hypothesis at this time."

At this end of the scale, high workload, the alerting signal competes for the attention of the crew. A signal not attended to can mean that a decision has been made to postpone it or temporarily ignore it for reasons of priority of other ongoing activities. It could also mean that it was missed - that is, not perceived. The evaluation of human performance in terms of whether or how quickly the alerting signal was responded to cannot have relevance unless the conditions under which the signal occurred are taken into account. Criterion performance in responding to an engine fire warning bell would include a period of assessment of varying duration dependent on the aircraft, flight phase, and, perhaps, immediately antecedent events. Also, a decision (covert) to postpone action on an alert cannot be observed and a missed alarm (bad performance) looks exactly like a postponed alarm (good performance)

The binary nature of the effect of workload on human performance is reflected in the often-referred to, hypothetical relationship between workload level and performance. The relationship is illustrated by an idealized, inverted U-shaped curve, as in figure 2. The important functional

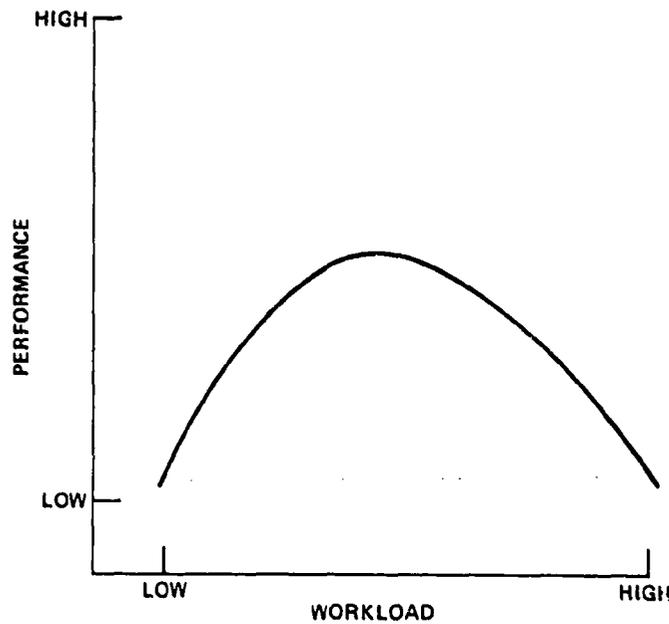


Figure 2. - Idealized hypothesized relationship between workload level and performance.

relationship shown is the reduced performance for both low and high workloads, even though the internal behavioral content is quite different under the two conditions.

Whatever the response outcome, the response to the alerting signal is like many other flight deck tasks: it is discrete and it is of low recurrence rate. Unlike other flight deck tasks, it has an unpredictable probability of occurrence. In any full-mission simulation - flight simulator or actual aircraft - the programmed occurrence of this independent variable for reasons of experimental strategy has an innate weakness. The alerting signals rise to the attentional foreground and become the focus of primary interest with high repetition rates, uncharacteristic of contingency events. This may defeat the purpose of the evaluation. This is also a weakness of laboratory studies of alerting signals in which the full-task context is not represented except by analogy. The alerting signals occur by design rather than exigency and are unrelated to crucial life circumstances. Their results, then, may be more germane to the revelation of performance under *low* states of vigilance and arousal. The soporific behavior of many laboratory experimental subjects is all too frequently observed.

Response conditioning- In December of 1972, an Eastern Airlines L-1011 flew into the ground 18 miles northwest of Miami International Airport. The National Transportation Safety Board report of the accident (ref. 16) contains, in a list of 17 Board findings, the statement that: "The flight crew did not hear the aural altitude alert which sounded as the aircraft descended through 1,750 feet m.s.l."

It is not known, of course, whether the crew did not hear it, ignored it, or postponed action on it. A real possibility, however, is that the psychological phenomenon of "response extinction" may have been responsible for the "missed alert." A graphic description of the role of extinction stated

in operational terms is provided by the communique to the NASA Aviation Safety Reporting System quoted in *Case C* in the introduction above. The altitude alerting system is a prime example of over-alerting and the production of alarms that, while not strictly false, have only occasional value. It is a commonly observed fact in learning and in behavior modification theory that unreinforced behavior will extinguish. Responding to the altitude alert frequently does not have a reinforcing value because its utility to the maintenance of a desired altitude is absent.

Although the crew has been trained to respond to the altitude alert, operational experience provides for a deconditioning or extinguishing of this response. This process is retarded only by the crew's knowledge that what they are being warned of may be catastrophic if allowed to proceed. It is also complicated by the ever-present doubt about the truth (a critical altitude deviation has occurred or a critical altitude deviation has not occurred) of the alert.

Credibility is thus severely compromised under these circumstances. There may be several consequences of this.

1. The extinction process is related to innate human neural and cognitive processes and cannot always be successfully mediated by rational processes.

2. The tendency to extinction and lessened credibility may *generalize* to other alerting systems and alerting philosophy, particularly where deviation sensor thresholds are so low as to foster high false alarm rates.

Our system is "coded" Informing the Crew

A second function of the alerting system is to indicate to the crew the source of the deviation and, to some extent, the urgency of the required corrective response. This triple coding (alerting, informing, assessing) has both advantages and disadvantages. The advantages are compactness, economy, and some possible standardization. The disadvantage is that with an increasing number of systems the requirement to remember several different codes results in an increase in workload and a further source of uncertainty, particularly for crew members in cockpits with which they are not completely familiar. A quote from the experienced test pilot-engineer quoted above (ref. 13) underscores this disadvantage:

In fact, we almost expect a pilot to demonstrate "perfect pitch" hearing during his medical. Even if this were the case, it is a very poor way out to rely on this memory to *identify* a failure, especially when the flight is in a critical phase. In addition, an especially loud, penetrating noise may cause overreaction and most certainly will interfere with normal mental processes. Also, there is a potential risk of confusion caused by past experience on previous aircraft.

The most ubiquitous form of coding is that of position. Amber or red warning lights can usually be found positioned somewhere close to or within the cluster of displays and controls for the many aircraft subsystems. Flags, of course, are integral with the instrument for which they serve as an alert that the instrument itself or a supporting subsystem has malfunctioned. Many instruments also have small amber or red lights that activate when an operating range is exceeded or not

attained. Their meaning is clear with regard to the result of a malfunction but not always to its source (the specific subsystem).

Labels on these warning lights also describe the problem; for instance, "HYD.SYS. PRESS.1," "LOW QTY," "OVERHEAT," and "RESERVE VALVE." The urgency level is indicated by universally accepted color codes: red for the most critical warnings, yellow or amber for caution, green for critical functions whose status is correct, and blue and white for advisory and status information (see FAR 25.1322).

A very natural coding scheme is that represented by the stick shaker stall warning employed on all jet air carriers in compliance with FAR 25.207. The stick shaker, a tactile stimulus, is the only alerting system that not only alerts the pilot to an impending hazard — the stall — but does so in a way that simulates the actual reaction of an aircraft as it approaches the stall condition. All aspects of a complete alerting and warning system (as herein conceptualized) are included: the pilot is alerted, he is informed of the problem, and the way to alleviate the problem is even suggested in a single nonverbal stimulus. Some systems go further and activate a "stick pusher" after a period of nonresponsiveness by the pilot (e.g., Trident I aircraft).

One aspect of the coding of information is the evolution of certain stereotyped aural signals such as the gear warning "horn," the fire "bell," and some alerts that are achieving a level of standardization through common usage, such as the overspeed clacker, the intermittent horn for unsafe takeoff configuration, and the altitude alerting "tone." These are similar on the three wide-body aircraft used above as illustration. Where stereotypy has occurred, a strong connection has been made between the aural signal and the deviation source so that interpretation is immediate. The dichotomy between these stereotyped aural alerts and recent or future added aural signals is particularly significant when some future integrated system is contemplated.

The use of synthetic voice annunciation for the informing function has already become a reality with the inclusion of the GPWS into the alerting and warning system aggregate. Voice annunciation has brought with it certain special problems in human factors. Also, some early operational shortcomings were recognized and resolved by modifications to the technical specifications.

The GPWS was required in all air carriers by December 1, 1975 in accordance with FAR 121.360. However, a petition was made to provide relief until September 1, 1976, due to the large number of nuisance (false) alerts being generated in operational use. Appropriate modifications were made to "Envelopes of Conditions" for warning in Radio Technical Commission for Aeronautics (RTCA) document DO-161 and a new document issued, DO-161A, May 27, 1976 (see ref. 17). These documents are referenced in FAR 37.201, Technical Standard Order — C92b as providing minimum performance, environmental, and test procedures requirements for ground proximity warning — glide-slope deviation alerting equipment.

The functional requirements for the GPWS specify the following modes for alert activation:

1. Excessive descent rates
2. Excessive terrain closure rate
3. Excessive sink after takeoff (up to 700 ft)

4. Terrain closure (not in landing configuration)
5. Excessive downward deviation from glide slope

Both an aural and visual signal are required. For modes (1) through (4), the aural warning is to be swept tone (400 Hz to 800 Hz) repeated once -- "whoop, whoop" -- followed by the synthetically annunciated word "pullup." This is iterated until the deviation is corrected. A simultaneous visual warning, a red light clearly labeled GPWS, is to be activated. For mode (5), the aural annunciation is to be "glide slope." As terrain clearance decreases and/or as the glide-slope deviation increases, the repetition rate and the loudness of the aural, or both, are to be increased.

The requirements for crew deactivation of the GPWS aural make it clear that it is not desired to facilitate this procedure. Modes (1) through (5) may be deactivated by a circuit breaker or by a guarded safety-wired switch. Independent deactivation of mode (5) is to be possible with a crew control separate from the above switch. This latter has been implemented in the B-747, by pressing the red pullup light on each pilot's panel. Automatic reactivation is required for a subsequent approach. This automatically resets in the B-747 GPWS when the airplane is again flown above 1000 ft above ground level (AGL).

There appears to be little question concerning the advisability of the GPWS for the prevention of controlled-flight-into-terrain (CFIT) accidents. In fact, one vendor has shown data indicating a decrease in CFIT accidents for carrier, Part 121/123, operation for the period following the mandatory GPWS requirement in the United States (ref. 18).

However, the GPWS is a first instance of the use of synthetic voice annunciation for both alerting and informing the crew about deviations, in this case performance deviations from safe flight profiles. It thus introduces a host of problems in psycholinguistics -- for example, semantic and syntactic structure, signal-to-noise ratios, operational or pragmatic context, and redundancy -- all of which require serious consideration in the design of voice annunciation systems. If synthetic voice is used, the message set that can be drawn from becomes virtually infinite and the informing function in CAWS can be extremely articulate with very high information transfer rates.

A fundamental question, or perhaps operational quandary, is whether the message is to be presented as an imperative or an advisory. When presented as a command "pullup" rather than as an advisory -- "terrain" -- the tacit assumption is made that there exists a single action solution. But a high sink rate can result from either diving or sinking, each of which requires different remedial techniques. Four of the five GPWS modes are signaled by a "pullup" command without identification as to the source of the problem.

There is a new generation GPWS in which the modes are identified by more appropriate annunciations and which indicate to the crew why the altitude gain has been commanded (Sundstrand Data Control, Inc., MARK II GPWS). The new GPWS, based on operational experience, has been modified to provide fewer nuisance warnings, longer warning times, more warnings for excessive descent just short of the runway, and a warning for descent below minimum descent altitude (MDA). Also, the synthetic voice now enunciates seven warnings, six of which refer to the reason for pulling up, for instance, "terrain," "too low gear," and "sink rate." The command to "pullup" is used only for deeper penetrations of the warning envelopes of the several GPWS modes. This second-generation system is already in operational use.

Reference 19 discusses the linguistic variables in synthetic voice annunciation and the problem of advisory versus command information; it also lists the independent variables and parameters that must be controlled in any human factors research in electronic-voice-warning-system design. The reader is urged to read that report as an adjunct to the material in this section.

Crew Option and Control

A third function related to cockpit alerting and warning systems is that of crew control once a deviation has occurred. Similarly with the other two functions, the provision of crew control seems to have been a matter of evolution and inadvertency rather than purposeful design. This is a natural outcome given the already high workload environment of the flight deck and the desire not to add to it. However, there are instances in which it may be desirable to have the opportunity to delay the response to an alert for reasons that may not be determinate beforehand. The case cited above, where the inability to silence the GPWS could have had disastrous results, is an example, although the probability of it occurring is admittedly low.

The kind of option referred to here is that of control over the alerting system itself rather than that of deviation correction; that is, the inner loop in figure 3. This is not to say that the two loops shown are independent of each other, only that the concern herein is not with deviation correction, per se.

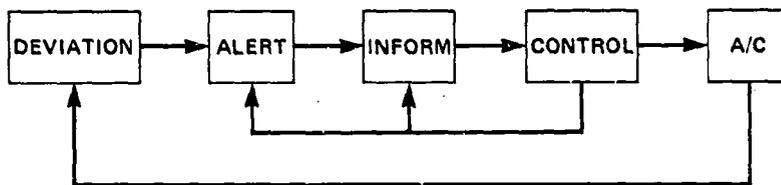


Figure 3.— Single alert control loops.

One form of control is to let the crew deliberately postpone action on a deviation until a more desirable time. Recognition of the disruptive effect of inappropriate alerts has led to the concept of the "phase adaptive" CAWS in which the various alerts would be armed only for flight phases in which they had meaning (see ref. 13). For instance, the cabin altitude alert would be inhibited in approach and landing as would all alerts unimportant for that phase. Punching out an alerting aural is a form of postponed action, for example, the gear horn, under certain circumstances.

A major problem with storing an alert for future action is the loss of the imperative of the original alerting signal. Also, the need to remember is an added load on the crew if no other display of the deviation is available. Proposals to code the alerts according to an urgency level scheme seem to have this problem in mind. For instance, a single-alerting aural could be used and the urgency level could be related to frequency, loudness, or repetition rate. In this way the crew would have some discretion in reacting to the deviation, and possible disruption of important ongoing tasks could be avoided.

The subject of crew control of the alerting and warning subsystem may be premature in that CAW's are not treated as, nor are they in fact, true subsystems. If alerting systems were regarded as

subsystems, it might be possible to provide for such system capabilities as coherent and logical prioritization, control functions, and alarm inhibits, through the use of a microprocessor. Moreover, this may be achieved through the more general use of a microprocessor for the larger complex of flight-management tasks.

ALERTING SYSTEMS IN THE AGGREGATE

To this point, the discussion has been mainly concerned with single alerting systems. However, the major problems emerge with the proliferation of these systems, in varying forms, on the flight deck. With each deviation identified as being worthy of a special alerting signal, a new flag or new aural appears in the cockpit. In addition to gear horns and fire bells there are now buzzers, tones, clackers, wailers, chimes, beepers, klaxons, gongs, crickets, and, in the Concorde, something called a "cavalry charge" to signal autopilot disconnect. This listing itself is offensive to the ear and perhaps conveys more ominous implications than is intended.

However, the collection of warning systems is a set of independent entities whose purposes are identical but which have no unifying mechanism. That it may be desirable to provide some unity and to conceive of the warning system array as an integral whole is becoming apparent within the industry. Approaches to the problem are evolving that point to this eventuality.

CAWS as a Subsystem

Despite the many problems associated with current cockpit alerting and warning systems, there is reason for optimism. This optimism is based on the emergence within the industry and user organizations of unifying and simplifying design approaches and the readiness of the state-of-the-art to accommodate integration of CAWS. This is in the direction of the attainment of true subsystem status for what is now merely an aggregate of similar entities. (For an interesting viewpoint on the tractability of the problem see ref. 20.)

A system, subsystem, or suprasystem has certain characteristics that distinguish it from a collection of similar but unrelated components. It has a purpose to which all components are subservient. Elements and components of a system interact with each other and with system output in support of the purpose of the system. Systems interact with other systems and the interfaces with other systems are environmental boundaries across which inputs and outputs are processed. Systems may be modeled and conceptualized as information processing and signal generating devices. Cockpit alerting and warning systems are, preeminently, information processors; as currently designed they have, in the aggregate, none of the other characteristics of systems listed above. The integration of CAWS thus means the transformation of the aggregate into an ensemble — into a true subsystem.

Indications in the trade and technical journals, in the general aviation community, and in hardware design are that in concept, at least, integration (and simplification) is near at hand. There are, however, many forms that an integrated subsystem could take, so the problem becomes one of the design and evaluation of candidate ensembles.

Master Warning Systems: Groupings

The three wide-body aircraft cited in this report - the B-747, DC-10, and L-1011 - have cockpit displays called master warning or master warning/caution annunciator panels. They give so-called "gross" indications of system deviations which are signaled in more detail in locations not within the pilots' immediate field of view; for instance, on the flight engineer's panel. The flight engineer may also have a similar summary panel but displaying different subsystems. In addition, an annunciator panel may be included at the latter station to provide a grouped status indication of the many doors on these wide-body aircraft. Aural warnings are not used with these summary panels but the warning light may be flashed to increase its attention-getting value.

Grouping of several critical subsystem warning indications in this fashion is a step in the direction of alleviation of the general proliferation problem. In addition to proximal grouping, there has been an assessment and inclusion of those subsystem deviations of higher urgency in terms of system safe performance. These are steps toward fuller integration for advanced cockpits.

Integration Concepts

The widely recognized problems associated with CAWS proliferation have spawned some functional design solutions that incorporate some of the principles of integrated subsystem design. These proposals arise from an appreciation of the problem and, more importantly, from a recognition of the plethora of solutions available, given that a CAWS may be completely revamped and reinstrumented using, for example, visual or voice annunciation, priority coding, mission-phase-adapted inhibits, and pilot option and control. This also assumes that a significant step forward will be taken in flight-deck avionics, notably automatic data processing in the manipulation and display of information.

Reference 2 is a complete survey and analysis of cockpit alerting and warning systems. It includes not only a compendium of current systems but, in a second volume (ref. 21) provides an extensive compilation and review of relevant human factors studies and guidelines. Also included is a set of system design guidelines. Some of the major recommendations are:

1. **Prioritization:** Alerts would be categorized as a function of criticality and flight phase. A unique audio, visual, or combination audio-visual alerting method would be associated with each priority level. The priority system has four levels: (1) emergency - requires immediate crew action; (2) abnormal (caution) - requires immediate crew awareness and corrective action; (3) advisory - requires crew awareness and *may* require action; and (4) information - indicates system condition but not necessarily as part of the integrated warning system.
2. **Inhibits:** The number and type of alerts active during critical phases of flight would be limited.
3. **Annunciation:** An alphanumeric display would be placed in front of each pilot to identify warning/caution type alerts. Aural alerts would be kept to a minimum (less than four). Voice annunciation would be preceded by an "attention getting" identifier.

4. Coding: Recommendations are provided for position, color, size, and brightness coding of visual alerts in terms of the priority scheme. **Recommendations are also provided for the coding of aural alerts in terms of number, intensity, and signal frequency.**

The Society of Automotive Engineers S-7 committee for Flight Deck Handling Qualities Standards for Transport Category Aircraft has been concerned with the problem of simplification of CAWS. A set of design objectives is given in their proposed ARD-450D "Integrated Flight Deck Alerting System." This proposal is in the form of a functional specification and only its major features are repeated here.

1. Attention: A *single aural alert* is recommended for the three most critical urgency levels. Their four urgency levels are those given in reference 21, above. The attention is to be modulated to correspond to a given urgency level (attention: an attention-getting sound).

2. Discrete aural alerts: It is recognized that some aural alerts are deeply stereotyped and may be included in the integrated system as "discretes" in addition to the attention. It is recommended they be restricted to the highest urgency level and be limited to four in number.

3. Annunciation: A centrally located alphanumeric display is recommended for the visual display of the three highest urgency level deviations. Color coding may also be used to enhance urgency recognition. Voice annunciation is recommended to supplement the attention and visual display for the most urgent deviations. Some deviations of lesser urgency may also be selectively signaled by voice annunciation.

4. Inhibits: Although the integrated system is to have provision for alert inhibits this requirement is not given in detail except that the urgency level of a given alert is to be varied as a function of flight phase.

5. Pilot option and control: There are several statements in the SAE specification that indicate a recognition of the need for pilot interaction with the integrated subsystem:

a. "Warnings, cautions, or advisories shall be automatically cleared from the system when fault conditions no longer exist."

b. "Capability to cancel signals for uncorrected faults and to recall such canceled signals shall be provided."

c. "System test capability shall be provided that will give the crew maximum confidence with minimum activity and complexity."

d. "The attention should be self-canceling for some alerts and manually cancellable for others."

e. "Alphanumeric readouts shall be self-canceling for corrected faults."

f. "Capability to cancel signals (on the alphanumeric display) for uncorrected faults shall be provided."

g. "Capability to recall manually cancelled readouts shall be provided."

Since this specification details functional requirements, specific human engineering requirements are not included. The design of visual and aural displays in terms of such factors as size, brightness, and intensity, are presumably specifiable from data in reference 2 and in other human engineering guideline documents.

The subject of selective or flight-phase-adapted-inhibits is not fully developed in the above guidelines. However, reference 13 treats this as a distinct issue and provides a rationale for a Phase Adaptive Warning System - PAWS. This proposal centers on a switching logic module that determines, from inputs from selected sensors (airspeed, altitude), what warnings to present to the crew and what to hold for a more propitious time. An example from reference 13:

A generator failure at 80 knots at takeoff will cause a red light signifying "abort." The principle is, of course, that below 100 knots you can stop at leisure for every failure, take your time to evaluate the urgency, and decide on either returning to the ramp or taking off again. But the same generator failure above 100 knots will be "held" by the switching logic until passing 1500 ft at which point it causes an amber light.

It was learned in a personal communication with the author that PAWS is the subject of a proposal for installation in an aircraft like the Fokker F-28. The decision is not yet final because an attempt is being made to cooperate with other airframe manufacturers (other than Fokker-VFW B.V.) in order to arrive at a standardized alerting system for the next generation of aircraft. The name has been changed to Phase Adapted Alerting System - PAAS.

Reference 22 describes the cockpit alerting and warning system objectives for the production Concorde aircraft. Several of its features coincide with design guidelines given above, such as alert prioritization, the use of functional grouping in a master warning system, visual and aural coding, and a design specification for the CAWS signal to: (1) alert; (2) inform as to the deviation source; and (3) direct the crew to appropriate corrective (control) action.

Reference 23 describes the viewpoints of the Royal Dutch Airlines on CAWS design. In reviewing that document one is impressed by the recurring themes of functional grouping, prioritization, flight-phase-adapted-inhibits, the separation of the alerting function from the informing function through voice and/or alphanumeric display, independent CAWS validity verification, and "dedicated" aural for those already accepted as a stereotype (e.g., selcal and altitude alert).

The above is only a sample of possible integration ideas. Each of the above, however, is distinguished by positive viewpoints and related to what an integrated alerting and warning subsystem ought to be and do. That these suggested solutions are in consonance with a much wider aviation community is indicated by a recent report accomplished under NASA funding. Reference 24 is an industry survey of CAWS. It clearly reveals the universality of the opinion regarding the seriousness of the problem, and the need for integration, simplification, and system-oriented solutions along functional lines. The report is too extensive to paraphrase here but is recommended reading for anyone interested in the prevailing opinions related to the material in this section.

METHODOLOGICAL CONSIDERATIONS IN MAN-MACHINE SYSTEMS EVALUATION

The many difficulties involved in the evaluation of performance in complex task situations are common to all flight-deck task elements. The present focus of interest is on CAWS because there is widespread agreement that it is a distinct problem complex. The CAWS problem may have unique aspects, but in terms of methods of system development and test, the procedures do not differ.

One example of the way CAWS differs from other subsystems is that the display has (usually) no clearly related control. To close a loop, correct the deviation, and neutralize the display may require the completion of many emergency operating procedures. The performance associated with the CAWS display is not the completion of the procedures, per se. As illustrated in figure 1, the important performance is related to decision processes and to the initiation of corrective action in a timely and accurate manner. Timely does not mean "right now"; timely only has meaning in terms of system requirements and, definitely, under specified conditions.

This leads to a discussion of performance criteria. The purpose of this section is to deal with the criterion problem from the systems engineering point of view. To that end a very brief discussion of the system development process will be presented followed by some of the salient aspects of system test and evaluation procedures.

SYSTEMS ENGINEERING: APPLICABILITY

In the development of a complex man-machine system there is a definite starting point provided by mission and operational requirements which a customer (e.g., military, industrial, commercial airline) has deemed necessary to one or more operations he wants to accomplish. These operational requirements are the beginning and the end for the design team – and a team is indeed required. The requirements are functional statements specifying what the system must do and they reflect requirements related to criteria of success; for example, economy, efficiency, military deterrence, profit, and safety, of the proposed system. A system cannot be designed unless the purpose to which it is to be put is known at the outset. Later design effectiveness can only be appraised in terms of these system requirements, not in terms of general design recommendations. What is true for the total system is also true for subsystems. The latter will have operational requirements emanating from the parent system with which they interface and whose purpose they must support.

All the steps taken in the system engineering design and development paradigm are done in the interest of objectivity, completeness, design relevancy and practicability, and, very importantly, management planning and control. This paradigm is thus a grand scheme for establishing ground rules and objectives for the design team in its progress toward implementing the stated mission objectives. Design experience alone does not automatically provide this progress nor ensure its success.

Given the operational requirements, the next general step is to transform them into *functional* requirements to meet the objectives. Functional requirements are conceived as jobs or tasks that are critical to system objectives; this must be a complete catalog of functional elements. These

ultimately must be stated in terms of performance limits and design constraints for use in making decisions at a later stage. This is also a stage in which performance criteria begin to be enunciated, both for man and machine and, more importantly their combination and contribution to system output performance. The function analysis moves, in increasing detail, from macrolevels to microlevels of analysis. The larger requirements are figuratively torn apart into elementary functional particles but without their relatedness to each other and to the system and mission objectives being ignored.

With the functions (jobs and tasks) identified, performance standards established, and design criteria proposed, the synthesis process begins with the initial allocation of functions to either man or machine. This is the point at which the human factors practitioner can make a most important contribution to system design and operability. This is not to say that this is where he first appears; he should have been involved from the start.

The next step in system synthesis is that of deciding on design solutions. These are selected on the basis of *relevancy* and *practicality*. The first involves performance, the second involves such factors as cost, weight, size, and availability.

These steps are not accomplished in a serial order but interact with each other throughout the whole design process. There is an ongoing iterative process which is required not so much for expediency but because the requirements and product are integrally related. This is graphically illustrated in figure 4 (fig. 4 and much of this discussion are taken from ref. 25). Even so, system development and integration *starts* with functional description and *ends* with design.

To be most effective, human factors personnel should be included on the design team and involved throughout the design process. If instead they are asked to help solve a specific human factors issue and then excused, the solution may only confound the issue. This is too frequently the manner in which system-related human factors problems are dealt with. There are several good publications on the methods and principles involved in the application of psychology in system design and development (refs. 8, 14, 26-32). These should be consulted for detailed discussion of those procedures.

There are several reasons for providing the above brief description of the elemental aspects of system design. First, it is a process with a great deal of intuitive appeal because of a technological utility in addition to its very obvious fiscal and management utility. Note that the schemes discussed at the end of the last section relating to CAWS integration are sets of functional requirements. The system requirements that spawned them are implicit in the experience and expertise of their authors. Though not formally delineated, as is recommended in system development procedures, many of the operational requirements have become well known to the community of users, at least in general.

A second reason for outlining the system development process is to emphasize that the orientation of this report and subsequent research and technology efforts related to CAWS is seen as fully dependent on that model. This endorsement is in the interest of and is vital to a closer working relationship between the private sector and NASA/FAA human factors activities. That it is vital ensues from the need for agreed upon goals, means, and ground rules.

AS YOU PROGRESS THROUGH THE SYSTEM EVOLUTION 4 STEPS, NOTE THAT ALL STEPS ARE ITERATED - CHECKED WITH MISSION OBJECTIVES AND EACH PRECEDING STEP.

ANALYSIS/DEFINITION PHASE
(FUNCTIONS TO BE PERFORMED AND PERFORMANCE, OPERATIONAL, AND DESIGN CONSTRAINTS)

SYNTHESIS SOLUTION PHASE
(DESIGN COMPROMISE BETWEEN FUNCTIONAL EFFICIENCY AND PRACTICALITY)

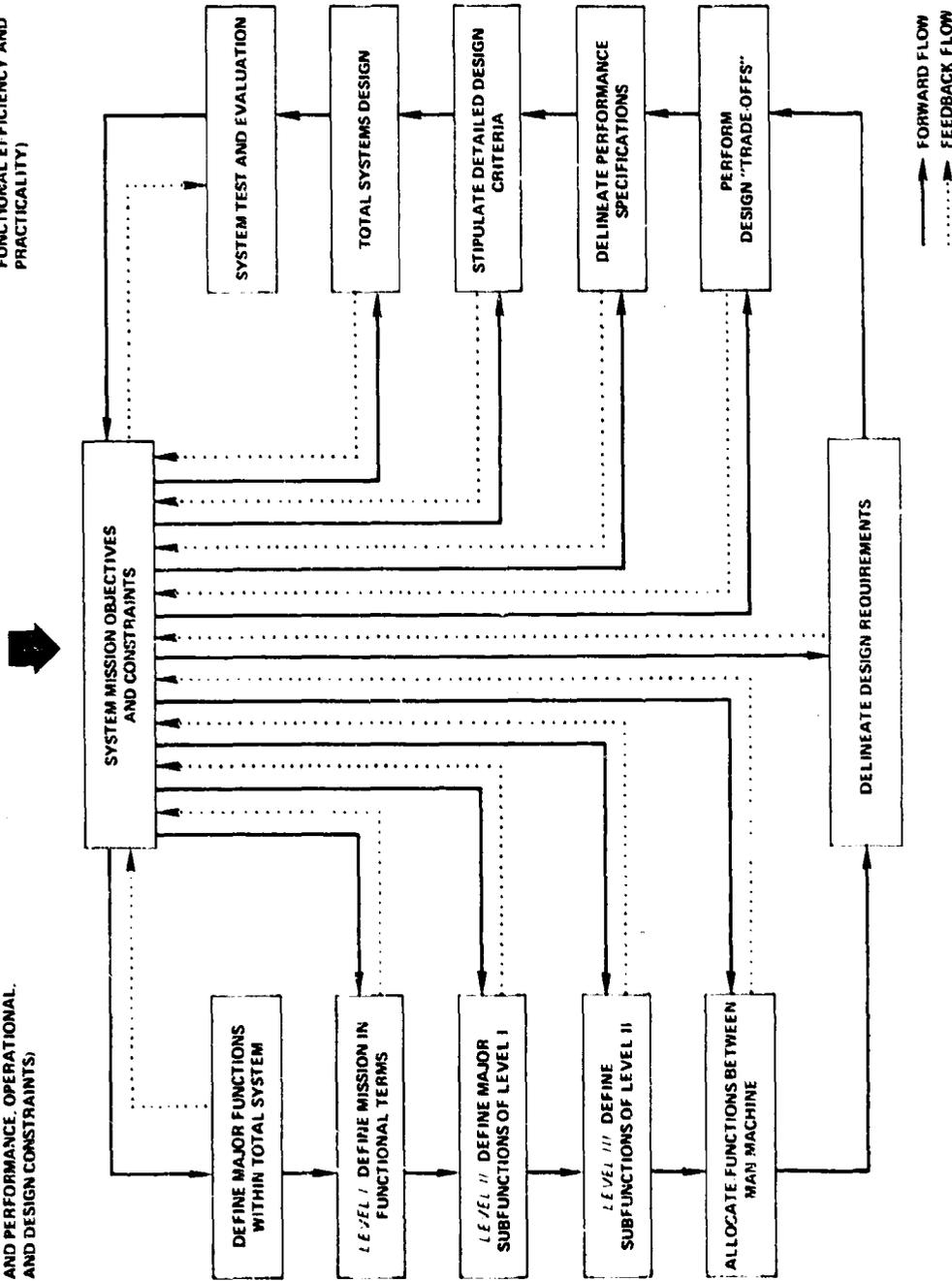


Figure 4. System engineering functions involved in producing a total system. (Reproduced courtesy of Fearon Pitman Publishers, Inc., Belmont, California, 94002.)

A third consideration is that there are crucial and fundamental differences between the problem-solving research accomplished in the development of a system and the general hypothesis-testing laboratory research in human factors. The remainder of this section addresses this topic in more detail for the purpose of providing some guidelines for evaluating the relevance of proposed studies for system test or system development support.

Finally, it should be made clear that the human factors effort is integral with but subordinate to the larger effort in the development of a large technological system. A human factors effort undertaken expressly to anticipate the configuration of a new subsystem and one that might be completed outside of or precedent to an actual development cycle would still best be governed by the above considerations.

TESTING FOR DESIGN DECISIONS

The function of human factors studies accomplished in the course of system development is solely that of providing a basis for design decisions. They are only justified when appropriate data do not exist or exist in forms unsuitable for the purpose. There is a vast human factors literature pertaining to human performance capabilities in man-machine related problems. However, these are seldom of help: they are frequently too general, too system specific, include too few independent (real-world) variables, and too often focus on individual performance variables rather than the combined man-machine system variables.

In an examination of two volumes of the Human Factors Journal, 1965-1967, Van Cott estimated that about 25% of the reported research fell into a category called "information or principles directly applicable to one or a family of systems." Of these only about one-half or 12% of the total would "... see the light of day in terms of actual implemented applications" (quoted in ref. 30, pp. 367-368). However, what is not possible to estimate are those human factors contributions that are made and not formally reported. They do contribute to system development but are of insufficient generality to warrant a reporting of results. Sometimes the human factors in the development of a real system is masked by high-level security classifications in military or other sensitive projects -- as are the efforts of other technologists. Or, perhaps even more to the point is Burrows comment (ref. 33):

Our effective human factors practitioner, who is politically astute enough to arrange for himself to be early enough in the programme and influential enough to steer it well, tends to disappear (except in the eyes of perceptive management) the better he is at the job.

The Task Fidelity Continuum

One of the continuing problems in human factors research is caused by the lack of differentiation between the pure research of the laboratory and that accomplished for purposes of decision-making in work situations -- engineering test and evaluation. The natural dichotomy of these aspects of technological investigations is also roughly reflected in the activities and character of government research laboratories versus human factors practitioners in the world of system design and fabrication. The goals of the workers on the two sides of the dichotomy are quite different. For

the laboratory researcher the goal of the investigation is frequently a contribution to a body of knowledge. To the extent that that is accomplished the results have an application, somewhere, sometime. The results of the engineering study must contribute *immediately* to system design for a given set of operational requirements.

The dichotomy is frequently pictured as a continuum with classical scientific investigations at one extreme and the engineering studies at the other. Knowles (ref. 34) shows a "continuum of research studies" with "models" at one end and "demonstrations" at the other. Obermayer (ref. 35) shows a "levels of abstraction" continuum with "math models" at one extreme and "real world" at the other. Chapanis and Van Cott (ref. 31) picture a similarly bounded continuum for the "test fidelity" dimension.

Similarly, figure 5 is intended to convey the characteristics associated with levels of completeness of the task structure. Moving the arena of research in the directions indicated will enhance the presence and visibility of the associated, listed attributes. At the top are listed units of behavior in ascending order (from left to right) of inclusiveness, complexity, and operational relevance. A task taxonomy would be the result of an effort to relate manageable, measurable component behaviors at the left of the figure to their complex combinations at the right. Measurement opportunities and techniques become scarce to the right as the researcher includes more and more of the total task context in the design decisionmaking process. Note that the categories delineated are reasonable but arbitrary, an inherent feature of classification ("your system is my subsystem").

The reason for including figure 5 is to provide a graphic model for assisting in classifying proposed research in the course of and for the purpose of system development. There seems to be a considerable fuzziness in the human factors community – not about *how* to proceed but *where* to proceed. This is evident at all management levels and is manifested in the individual researcher's tendency toward activities to the left end of the continuum, when he has a choice. It seems that he gravitates to that which most affirms the ideals and sources of reward inculcated in him by his formal classical training. But that is conjecture. What is important is that the usual continuum is more than just that; the extremes represent wholly different sets of purposes, criteria, methods, and products. They are both necessary, are of high purpose and value, and, at worst, can only be charged with inappropriateness. The major difference is that the goals of the practitioner in the left part of the arena are conceptualization and understanding; those of the one in the right are empiricism and control.

Classical vs Engineering Studies

The fundamental differences between classical and engineering studies were explored by Finan (ref. 36). The differences are stated as design options in formulating and conducting a research study. They are quoted here in order to preserve the elegance of the original statements.

An initial option is taken at the stage of selecting and formulating the research problem. In theoretical research, the problem is transposed into a more controllable context, and the variables involved are translated into conceptualized dimensions. In contrast, the prime requirement for the results for engineering research to be relevant to practical goals restricts this latter type of study to situations and variables that closely simulate the complex of operational conditions.

| <u>TASK ELEMENTS</u> | <u>SUB-TASKS</u> | <u>TASKS</u> | <u>FUNCTIONS</u> | <u>JOB ELEMENTS</u> | <u>JOB</u> |
|---|---|--|--|---|--|
| <ul style="list-style-type: none"> ● STIMULUS TONE LIGHT STICKSHAKER ● INTERMEDIATION COGNITION MAKING MEMORY ● RESPONSE SWITCH BUTTON | <ul style="list-style-type: none"> ● SENSE ALERT ● DETERMINE SOURCE ● CHECK VALIDITY ● DECIDE ACTION ● CORRECT PROBLEM ● EVALUATE RESULT ● CONTINUE, HOLD, CYCLE | <ul style="list-style-type: none"> ● MAINTAIN CORRECT ALTITUDE THRU AWARENESS OF: <ul style="list-style-type: none"> ● ALTITUDE ALERT ● GPWS ● ALTIMETER ● SPATIAL LOCATION HISTORY AND PROGNOSIS ● ATC | <ul style="list-style-type: none"> ● COMMUNICATE ● CONFIGURE FOR FLIGHT SEGMENT ● MANAGE POWER ● NAVIGATE ● COORDINATE WITH CREW ● PLAN APPROACH | <ul style="list-style-type: none"> ● TERMINAL OPERATION AND POSITIONING ● APPROACH TO LAND ● LAND, TAXI, PARK, DEPLANE | <ul style="list-style-type: none"> ● COMMERCIAL AIRCRAFT CAPTAIN ● FIRST OFFICER |

BEHAVIOR FUNDAMENTALS

APPLICATIONS

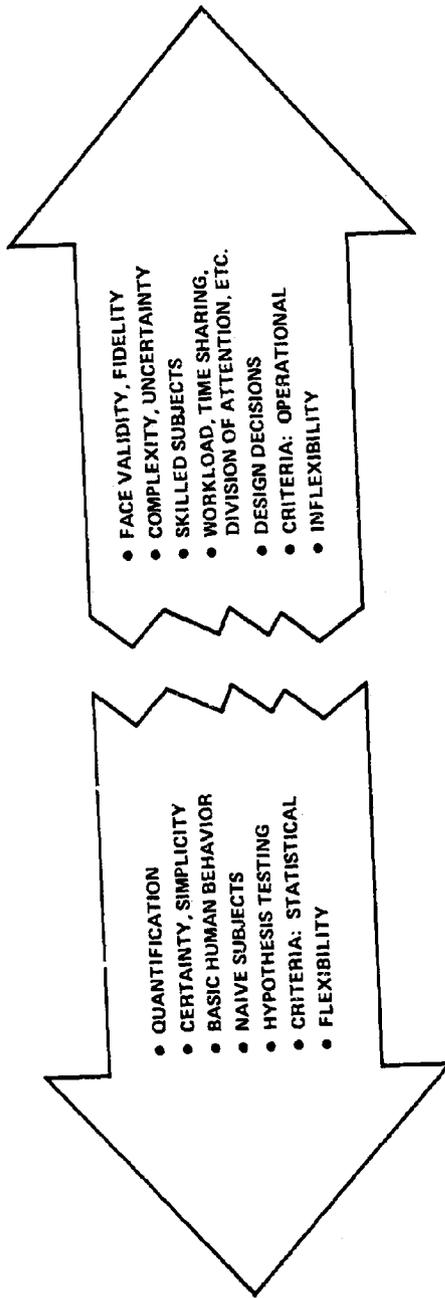


Figure 5.- Task fidelity continuum.

A second option concerns the use of analogies in research. The explanatory model of theoretical inquiry is essentially a symbolic idealization of observations, while the forecast formula of engineering psychology can be considered an empirical summary of results. A related distinction is made between *validity* of theoretical research – the correspondence between a concept and germane phenomena – and *fidelity* of engineering research – the degree of relationship between forecaster and criterial terms required for a specified practical purpose.

A third option is taken with respect to the differential role of hypothesis. In theoretical research, hypotheses are explicitly linked to a model, and if experimentally corroborated, make possible attribution of effect to cause. When employed in engineering psychology, hypotheses may serve to suggest the content of forecaster and criterial terms, and may, *post hoc*, be used to interpret the observed relationship between them. The hypothesis-testing and demonstration types of experiment exemplify this contrast.

A fourth option is taken in dealing with the problem of the variability of observations. Causal attribution depends, in the ideal, on the possibility of rigorously controlling all relevant experimental conditions other than the one manipulated. Forecasting, however, depends on representing with maximum fidelity *whatever* sources of variability may operate within the criterial situation. This is the important difference between systematic and representative design.

A fifth option is taken in order to define units of analysis. In theoretical research, units are selected for the purpose of demonstrating behavioral uniformities; in engineering research, the requirement is to define a unit that proves manageable for producing or forecasting a particular operational system.

A sixth option is taken with reference to the criteria of acceptable inference. Statistical hypothesis-testing is considered more appropriate to the demonstration of model relationships, while statistical estimation techniques are deemed more suitable for forecasting to a criterion. The appropriateness for engineering studies of the conventional 0.05 probability level for acceptance of findings will subsequently be questioned.

A seventh option is taken when the conclusions of the research study are extended to new situations. In theoretical research, generalization proceeds by demonstrating the extensibility of the model dimensions to the new conditions. Limited engineering generalizations can be based on inference from populations and guessed inter-class relationships.

A final option is taken in connection with utilization of research outcomes in practical situations. Results obtained under the pure conditions of the laboratory yield *abstract predictions* which may have implications that are adaptable to operational systems. Engineering studies yield *forecasts* which constitute a direct and immediate basis for action.

Criteria

Websters Seventh New Collegiate Dictionary defines a criterion as "a standard on which a judgment may be made." A standard is defined as "something established by authority, custom, or general consent as a model or example: criterion."

The criterion is essential to measurement that is done for a purpose. And measurement *can* take place without purpose in the absence of clear and relevant criteria. The definition of criteria turns on the word "standard" which in turn is seen to be arbitrary in the sense that consensus is involved in its definition. At the extreme of ultimate criteria, from which system operational requirements ensue, a value system will be discernible that is related to such larger system considerations as political, social, military, humanistic, or economic. However, criteria following lower in the system hierarchy are generally not arbitrary but are constrained by the need to be relevant to the judged system goals. Thus, criteria will range from valuelike statements ("an alerting and warning system should be unobtrusive until really needed") through relevant functional statements ("the pilot response to the aural should be timely and accurate without degradation of ongoing tasks") to measurable attributes in the criterial chain ("pilot response time to the GPWS under x conditions should be y seconds; to an hydraulic failure under x conditions, it should be y seconds").

Operational criteria are already available by taking note of the generally agreed upon complaints with current CAWS: there are too many; they are disruptive; there are too many false and/or needless alarms; it is difficult to remember deviations associated with aural; they are not standardized (even functionally) within or between aircraft; they are not phase adaptive; they are too loud; they can be missed (lights); there are only gross indications of priority and urgency. Criteria for the design of a new CAWS subsystem would be the obvious inverse of these. The current CAWS may be considered to have been exposed to "operational test and evaluation" and the results – user judgment – are being "tabulated" in the literature and in anecdotes. This appears to be one area of cockpit design that has been, and is being, given considerable attention by users and has yielded not only a wealth of criticism but a wealth of design criteria. The unanimity of opinion as expressed in emerging functional requirements is unlike other problem areas in cockpit design (e.g., display integration, HUD).

It might be supposed that the crew reports require validation in terms of *behavioral* criteria. That may sometimes be so, but it should not stand in the way of identifying and utilizing *psychological* criteria that refer to crew satisfaction with cockpit equipment and procedures. These latter criteria cannot be ignored in design by a slavish adherence to the principle that man-machine system (MMS) output (measurable) is all that counts. Dissatisfaction with (even approaching hostility toward) job design, and the tools to do it, when interfacing with desynchronization, fatigue, high workload, and keenly felt responsibility can lead to errors in judgment or to actions that terminate in accidents for which there is usually only a tentatively identified cause.

There is such a paucity of measurable or clearly defined and relevant criteria for the evaluation of MMS performance generally that the subject requires much broader and incisive treatment than can be given here. In CAWS the problem is made even more difficult by the lack of regularly recurring behavior and its predominantly covert nature (see fig. 1, decisionmaking, alert validity checking, etc.). However, the selection of criteria; or the perception that none exist, for a given purpose requires a firm grounding in the meaning and function of criteria in MMS development and

evaluation. The reader is referred to references 26, 27, 31, 34 -40 for advice and is offered the admonition that criteria must be selected based on relevancy, efficiency, definitiveness, and measurability. These standards for criteria will hold for any systems, but the actual criteria selected for use will, in general, be different for each system.

If it is agreed that the outlines of the system development process sketched above describe the "ball park," then criteria selection is the "ball game." The selection of suitable quantitative and qualitative criteria is crucial. Unfortunately, in view of the large element of indeterminacy, it appears to be largely a creative process without formal guidelines.

KINDS OF TESTS IN SYSTEM DEVELOPMENT

One of the fundamental differences between classical and system engineering studies not made explicit by the eight options given by Finan (ref. 36) above, is the etiology of the problem that requires evaluation or controlled study. This has to do with where and when to revert to studies to solve design problems. The system development process cannot be delayed by these activities so constraints can become severe. The remarks in this section refer to that context and not to investigative activities carried out by human factors personnel employed by the system developer in *anticipation* of the *next* system to be processed.

Most human factors research reports are addressed to systems that will be developed later. Generally, the classical researcher "comes out of" a body of knowledge (behavioral) to an application in broad terms (MMS) and wishes to make a contribution to the technology for that knowledge. A problem is defined that places a man, machine, and task in controllable proximity and a study is accomplished. The goal is to report findings that will have sufficient generality so as to be broadly applicable. In this sense, "basic" research can promise more "applicability" than "applied" research. This is a worthwhile endeavor and is necessary to the growth and sustenance of the human factors technological base. It is usually not of much use to the system designer, however, in that it frequently is not in a form he can use (continuous or extrapolable). His design problem is similar to yet different from the study-task configuration that was severely truncated in the interest, necessarily, of experimental control. This research is generally characterized as a solution looking for a problem.

On the other hand, studies accomplished in support of system development are, or should be, exactly that. Anticipation is, of course, a desired attribute of this support and should be exercised. In fact, anticipation is already apparent with respect to CAWS, as witness the development of functional requirements alluded to above. Problems which arise in the development of a system come from the need for the resolution of design decisions. Specific solutions are sought to specific problems right away. The problems are not hypothesized but surface as nodes of reality in the system development process.

Recognition of the chronology and source of human factors issues requiring attention has led to suggested testing paradigms for the practitioner which are tailored to the larger system development testing paradigm. One such testing paradigm is that offered by Shapero and Erickson (ref. 41) and included in references 29 and 42. These authors divided system testing and evaluation into exploratory, resolution, and verification.

Exploratory

This level and class of test would normally occur early in system design and would be motivated by problems arising in the course of "roughing" in the broad outlines of the hardware in response to functional requirements. Also, these tests provide information where none exist in the records of previous research. Because of the tentativeness of hardware configurations the most opportunity exists for manipulation and control of variables so this level of testing resembles traditional laboratory research. This is generally the stage in which decisions about allocation of functions can be empirically validated. For instance, does it appear to have been a good decision to have the avionics logic determine priorities for queuing multiple alarms or should they be displayed as they occur chronologically and let the crew make the decisions? Questions of this sort are typical of the "newer" generation of human factors issues occasioned by the trend toward digital avionics and automation and the apparently changing role of crew members into one of information processors, resource managers, and decisionmakers as against controllers (ref. 43). These changes are related to changes in the evolving national aviation system and it would appear that important decisions concerning the allocation of man-machine functions are in the offing. This will probably require extensive use of the exploratory technique involving, as it does, probing of limits and ranges and the clarification of roles and performance requirements for both man and system. Of importance here is the consideration of the integrating of a newly conceived CAWS subsystem into a higher level subsystem supporting all flight management tasks.

Resolution

While exploratory testing is not tied specifically to the system under development, resolution testing is. The purpose of resolution testing is choosing from among two or more candidate configurations. Although the level of the statistical significance of the difference between mean performance values may be the most appealing criterion on which to base a decision, it should be regarded in view of the qualification, "other things being equal." Other things being equal, the choice can be based on statistical criteria; but such factors as costs, space requirements, reliabilities, weight, pilot acceptance, and development time are seldom equal. This is true because this kind of evaluation simply seeks to predict which configuration will result in a more effective system when measured against selected mission criteria variables. Resolution testing begins to look like system testing because it may involve operational configurations. This class of testing also resembles conventional research in its control of variables for comparability of testing for the several systems under consideration. Also, the side-by-side comparison affords the opportunity for performance contrasts.

Verification

Verification tests usually occur later in system development and are related to the question of whether the fully composed system satisfies functional performance criteria, that is, operational criteria. However, the fitting of performance to operational requirements is the prime goal from the beginning in system development so the underlying rationale for verification testing is also present throughout. Because a standard or "absolute criteria" is used, all results have an either-or cast and require critical analyses. Changes to a completely configured system can be costly and can seriously delay product delivery. Decision criteria must be clear and compelling. The extreme of this level of

testing in terms of the development cycle would be, in the case of an aircraft, flight testing including subsystem verification.

These three categories of tests – exploratory, resolution, and verification – are not mutually exclusive in terms of either chronology or tactics. They may be used and reused in the general process of iteration. Table 1 from reference 41 lists the major differences between the test types.

TABLE 1.- MAJOR DIFFERENCES AMONG TEST TYPES

| Characteristics | Types of tests | | |
|---------------------------------------|--|---------------------------------------|--|
| | Exploratory | Resolution | Verification |
| Typically performed in | Predesign and early development | Early development | Throughout development |
| Control of independent variables | High | Moderate | Low |
| Number of measures recorded | Few | Few to many | Many |
| Repeatability of test conditions | High | Intermediate | Low |
| Number of conditions compared | Any reasonable number (e.g., factorial design) | Few (gross configuration differences) | One (comparison with performance standard) |
| Control over test environment | High | Moderate | Low |
| Number of dependent variables | Few | Few to many | Many |
| Factors initiating test | Ambiguity of system | Need for design decision | Need to verify system adequacy |
| Part/system testing | Part | Part to subsystem | Subsystem to system |
| Resemblance to operational conditions | Low | Moderate to high | High |

(Reproduced courtesy of John Wiley & Sons, New York.)

THE TOTAL SYSTEM

What has been said so far in support of an orderly (system engineering) approach to design, synthesis, and evaluation needs to be extended to the bounds defined by the National Aviation System (NAS). The system called CAWS is a subsystem of the aircraft system which is a subsystem of the commercial system which is a subsystem of other subsystems. Previously, human factors work has tended to emphasize the more tightly coupled man-machine activities such as tracking performance, control/display relationship, workplace design, accuracy and time in task performance. This, with increased engineering excellence, has led to better cockpit design and to a general reduction of interest in problems associated with performance at the level of "inner loop" control.

However, the emphasis is now shifting to "outer loop" activities variously called flight management, resource management, monitoring activity, etc. This relates to the crew members as decisionmakers, in an enormously complex information processing, utility assessment, and even, social system. But the major underlying human factors issue has not really changed from the time the first instrument was installed in an aircraft to complement the pilot's sensory and control capabilities.

An Old Problem, a New Scenario

The problem is one of transforming the real-world spatial orientation of the aircraft – position and rate of change of position – into a veridical human *percept*, where such a percept cannot be accomplished by "natural" means, that is, directly or visually. This is traditionally accomplished by fragmentating the information into numerical indices of the various flight parameters and displaying them to the pilot; no one of these is sufficient by itself. The task of the pilot-controller is to integrate these fragments into an operationally useful percept of his constantly changing spatial orientation. With appropriate training, humans are adequate for this task. However, this performance is extremely susceptible to degradation under conditions of high workload, divided attention, and low levels of arousal for monitoring system states. This refers not just to misreading instruments – the documentation relating to misread altimeters is voluminous – but to combining that information with all other system information for the formation of a valid and continuously updated veridical percept. Visual illusions which occur during the transition from head-down (instrument) to head-up (visual) flight reference would seem to be at least partly due to the discrepancy between the *synthesized* cognitive percept and the revealed *visual* percept, particularly since a percept is unitary, will not be formed unless it is believed, and tends to persist.

In attacks on this problem there have been attempts to present information relating to spatial orientation in "contact analog" form. The attitude indicator is a mechanical analog for displaying inertial space and the aircraft's rotational orientation therein. Many schemes have been advanced using cathode-ray tubes (CRT) and transparent, flat plate CRT's in the wind screen position for displaying information in analog form – "highway in the sky" – which would also conform to the real visual world at breakout. Conformal type information (see refs. 44 and 45) is also being included in proposed formats for head-up displays (HUD). A recent development for study is that synthesized by Baty (ref. 46) for a head-down display. It is called a coordinated cockpit display (CCD) and it was designed by reference to perceptual and human factors research over the last 25 years related to spatial orientation and localization. Also, the design was guided by the

operational requirements of an air traffic control system. It is generic in nature and uses three CRT's for the presentation of three orthogonal projections of the aircraft situation: perpendicular to the forward line of sight, parallel to the ground, and perpendicular to those two. The latter is thus a side view. This provides a pictorial altitude profile, heretofore not available in the cockpit. (Would a GPWS be required if such a vivid altitude record were available?). This effort is oriented in a philosophy of integrating the information in inner loop control (aircraft subsystem) with that of outer loops, the air transportation system context.

System Induced Errors

Underscoring the necessity to consider human factors issues in the broader system context is a paper by Wiener, titled "Controlled Flight into Terrain Accidents: System Induced Errors" (ref. 47). It is noteworthy that the "system" referred to is not the aircraft or its subsystems, but rather, "... a complex air traffic control system with ample opportunities for system-induced errors." The author further states:

The human factors profession has long recognized the concept of design-induced errors. This paper simply extends the concept to a large-scale system, whose principal components are vehicles, traffic control, and terminals. These three components are embedded in two other components: regulations and weather.

The contradiction suggested by the title is not really that, but is more indicative of a failure of the integration of the two levels of system management being discussed: aircraft control (inner loop) and aircraft management (outer loop). The author cites several examples of system-induced errors. Two of these (ref. 16 cited previously and ref. 48) are of particular interest because they involve the altitude alert. In one of these controlled-flight-into-terrain accidents the altitude alert sounded 1 min 34.5 sec before impact; in the second, 1 min and 17 sec before impact. In the first case it was ignored; in the second it was silenced by one of the crew.

The impotency of the alerting system is thus reflected in two ways neither of which response is seen to have any relationship to the physical attributes of the stimulus except, perhaps, its ability to annoy. Wiener (ref. 47) has this to say:

An appealing though perhaps somewhat simplistic answer to vigilance and attention problems is to install warning devices which are *assumed* to be attention demanding. One must recognize these from the outset for what they are: one form of sensory stimulation replacing or *simplifying* another, and the extent to which they are useful is determined primarily by their novelty. *There is never a guarantee that any stimulus, regardless of its psychological dimensions will be correctly attended to, "correctly" because one of the usual ways of dealing with warning devices, particularly in the auditory mode, is to shut them off* (emphasis ours).

Obviously, for an alert to be effective, credibility must be high. If it is not, then even when it is valid but does not agree with the crew's incorrect perception of the situation it will be deemed invalid by deduction from false premises. It is thus, tragically, too late in terms of contributing to the formation of a veridical percept.

At the national level, a recognition and concern with system-induced errors, as defined by Wiener, is reflected in reference 49, which is a congressional subcommittee report on the needs and opportunities of the air traffic control system. It is recommended reading for gaining an appreciation of the total system concept.

A System Information Processing Model

A fruitful model of the total operational system would be an information processing one. This would afford a common and germane set of standards within which to establish system performance criteria. It would provide the needed integrity for bringing together subsystems for the support of the larger system purposes and goals. In talking about workload we are not usually referring to physical work but to cognitive effort in the acquisition and processing of information. Diversion of attention usually refers to attending to several sources of information in close temporal sequence. Monitoring has to do with clear and ongoing attention – right now and into a reasonably extended future – to information available in the system concerning the aircraft subsystem. Decisionmaking is a process of weighing information and deciding on a course of action tempered by utility and cost. All of these human performance aspects of information processing can be critically influenced by stress, high workload, fatigue, desynchronization, psychological and physiological states, etc. However, their degree of influence could be assessed in terms of criteria related to information processing activities and ultimately to suprasystem processes.

A good example of a descriptive information processing model is given in reference 50. In that report a method is given for the study of human factors in aircraft operations. It consists of describing categories of information sources in the aviation system and how this is utilized by *people* in the system according to a brief taxonomy of behavioral functions – decisionmaking and decision implementation. The model and method are for the purpose of analyzing system activities for the identification of “human errors.” The authors recognize the limitations of that term and state: “. . . the investigator might well reason: that what appeared to be an obvious ‘pilot error’ was, in fact, a *system problem* which led the pilot to an *incorrect decision*, and therefore, an incorrect course of action” (emphasis ours).

The approach and method proposed by these authors would appear to start analysis at the total system level – from the top down, as it were. This provides a governing framework that permeates all task structures in the system. In this model, cockpit alerting and warning systems would appear as remedial displays for information sources not monitored, for information not adequately synthesized, or for information not made available. In this light, they also seem to signal not only that some remedial action is required, but also that a system operational failure has occurred.

PERFORMANCE EVALUATION IN CAWS

It should be made clear that there are no “recipe” or “checklist” type sets of guidelines available for use in human factors evaluations of MMS's. There are only sets of procedures and methods that seem to produce cost-effective activities in the pursuit of rational technical design. This is probably a technically healthy circumstance given the fuzzy state of our art in the

extremities of its applications. The predetermination of test details – criteria, variables, controls – and their application by fiat rather than by rational analysis is not possible and could result in stifling creativity in solving problems in the system development process. The material in this section will therefore be general and advisory rather than specific.

WHERE TO TEST (CONTEXT)

The question of where, when, and how to test during the development process is difficult to answer beforehand because the need to test and its identity come out of that process. Pointing out the relative merits of part versus full simulations appears an endorsement of the latter – and it was. That was on rational, technical grounds. However, availability and cost problems could prohibit the use of full-mission simulation. That this might be obviated by a pooling of resources in a community facility seems an increasingly feasible option to the extent that “commonality” can be realized. However, departing from the fuller *task* context increases the burden of proof of validity and relevancy. The discipline has not furnished tools and techniques for systematically estimating the validity of varying degrees of task abbreviation so this may be difficult.

Before discussing traditional test contexts it should be mentioned that the design process itself is an ongoing evaluation. Every design decision implies value judgments and thus the solution has gone through an evaluative process in qualifying for implementation. The more conscious and explicit this process is made, generally, the more relevant is the solution. However, the following remarks refer to testing in which, for example, objectivity, repeatability, and quantification are desired in support of design decisionmaking and end-product assessment.

In-Flight Testing

Testing in the actual aircraft, whether of a newly fabricated CAWS subsystem or a totally new aircraft, is difficult and expensive in relation to other methods. If special instrumentation is required one can quickly become enmeshed in regulatory and airworthiness considerations. The small subject population sample (test pilots) that would be used with a new aircraft would limit the usefulness of the results. Moreover, one is faced with the question of what to measure. In-flight performance is unique in that acceptable performance deviations are variable. During cruise, thousands-of-feet off-altitude and miles-off-course can be acceptable data points in an experiment; during an approach and landing, the acceptable deviations, that is, those tolerances which if exceeded would cause a safety pilot to take over, are reduced to a few feet. Add this to a strong desire not to interfere with the ongoing behavior, and an understandable subject resistance to being instrumented or wired to equipment, and the range of acceptable performance measures becomes fairly restricted.

The list of practical difficulties with in-flight evaluation could be further extended – lack of repeatability, lack of experimental control, etc., but a more interesting consideration in the case of CAWS is that all *test* stimuli are, a priori, “false” alarms. That is so because the subsystem under test cannot be fully conceived as integral with the aircraft system and its in-flight scenario. So much effort and attention would have to be expended in the separation of test and real alerts that the test alerts would become unrealistically prominent. This would obviate a valid estimate of their alerting

power. The other two CAWS functions -- informing and pilot control -- may not similarly be compromised, but the mix of simulation with flight may be an overly hazardous stratagem. Flight tests of new aircraft would place the test in the category of verification testing of the aircraft and its subsystems. Flight tests of new subsystems for current aircraft would be *exploratory*, *resolution*, or *verification* tests, or all three, at appropriate points.

Flight Simulation

A full-mission simulator has many obvious advantages over an actual aircraft as a test bed. The major advantage is, of course, the freedom to manipulate operational and experimental variables. A general rule is that the more the full flight-deck tasks are included and are made realistic, the more valid will be the conclusions drawn from investigations in its context. The representation of tasks and information management activities rather than fidelity of aircraft configuration would seem to be a reasonable priority. These are not mutually exclusive options, however, and in the absence of knowledge of how to sacrifice physical fidelity while retaining psychological fidelity it is usually considered better to opt for full physical fidelity. The full-mission simulators used in military and commercial airline training programs represent the extreme in sophistication and functional completeness, but they are not usually intended to serve as research devices. However, they are extremely cost effective in playing the role for which they were designed.

In the best of all possible worlds it might be recommended that a full-mission training simulator be especially configured for research and development purposes, but such would come too late in the development cycle and be too expensive. A limitation of the high-fidelity simulator is that it is limited to the specific aircraft it represents and may be of limited value for developing new aircraft except, perhaps, for retrofit of new or modified subsystems. However, it can also be reasonably argued that CAWS' are not aircraft specific and could be assessed in any similar aircraft cockpit simulation of like complexity.

A good example of both resolution and verification testing of a newly conceived subsystem in a full-mission training context is reported in a paper by Carol A. Simpson ("A Synthesized Voice Approach Callout System for Air Traffic Transport Operation," unpublished data) in which a synthetic voice subsystem for making deviation callouts in the final approach to landing (SYNCALL) was tested. Criteria for the evaluation were derived from operational performance requirements (aircraft position and flightpath deviations) and expert opinion (pilot judgments).

When a simulator is to be used in the design of a new configuration it cannot, of course, yet have that configuration, except in a general way. The simulator would evolve only as does the article it simulates and would then be available for testing at the level of verification. That is a reality that forces early design simulations to be tentative and functional in nature and where exploratory and resolution testing would appear to be most apropos, particularly in the case of a totally new cockpit configuration. However, in the case of CAWS it appears feasible to install, say, a brassboard model of a proposed subsystem into a currently configured simulator cockpit for all three levels of testing -- exploratory, resolution, and verification. Reaching a decision in the selection of one from two or more proposed designs could, then, also be accomplished in this context.

The conflict between "test" alerts and "real" alerts mentioned above for flight testing disappears in the simulator context. This gives the simulator a decided advantage over in-flight testing, tempered only by the extent to which test alerts are considered real by test subjects in their simulated flights.

Part Simulation

Most human factors studies and evaluations are accomplished in so-called part-task simulations and, as discussed previously, are thus more frequently used for research purposes rather than for the *direct* support of the developing system. As used in system support they can be a relatively low-cost way of making design decisions by side-by-side comparisons of proposed solutions. This can be done in a test site removed from the main stream developmental activities. However, the problem being studied cannot; it is always integral with and motivated by the design process. Part-task simulations refer to the completeness of the *task* under investigation, not to the *physical features* of the apparatus employed (see fig. 5). A small task element can be investigated in a full-flight simulator and, in some cases, profitably studied in a simple mockup. As the task for study assumes more and more the dimensions of the full task it becomes less easy to isolate it for study. The part-task can almost always be placed in the full-task simulator for study; the reverse is not true. In addition, in part-simulations there is also the option of the part-mission, full-task option. In this method only selected segments of the full mission are used in the study. Those that are used use a full-task context as in full-mission simulations. This is appropriate for CAWS evaluations.

However, the size and the physical configuration of the space used in system-related tests are not the determining dimensions. What is important is that the task dimensions be veridical with respect to such aspects as information input and output, procedural activities, functional relevancy and situational factors (workload, crew interaction, and operational contingencies). The fidelity continuum discussed before is cast in terms of task complexity not apparatus complexity.

WHAT TO TEST (CONTENT)

The technology of flight management in the National Airspace System is on the verge of revolutionary changes. The reasons for this will not be pursued here other than to say that automation of functions previously requiring pilot decision and action appears to be reliable and economically feasible using digital computer techniques. Many new technologies are available for new solutions to complex problems. Automation will, of course, bring new problems but they will relate to the appropriate mating of human intelligence with machine logic for optimum system management. The position taken here is that redesign of current CAWS is retrogressive and would result in only superficial changes. Table 2 lists the key differences between methods for change based on "systems improvement" and "systems change." The table is taken from reference 51 which was "... written to emphasize the difference in intent, scope, methodology ... and results between improvement and design." And further: "The treatment of system problems by improving the operation of existing systems is bound to fail. Systems improvement can work only in the limited context of small systems with negligible interdependencies with other systems - a condition that does not occur very often." Finally, "... the Systems Design Approach is basically a

TABLE 2. COMPARISON OF TWO METHODOLOGIES OF CHANGE: SYSTEMS IMPROVEMENT AND SYSTEMS DESIGN

| | Systems improvement | Systems design |
|-------------------------|--|--|
| Condition of the system | Design is set | Design is in question |
| Concern | Structure and operation | Purpose and function |
| Paradigm | Analysis of systems and component subsystems (the Analytical Method) | Design of the whole system (the Systems Approach) |
| Thinking processes | Deduction and reduction | Induction and synthesis |
| Output | Improvement of the existing system | Optimization of the whole system |
| Method | Determination of causes of deviations between intended and actual operation (direct costs) | Determination of difference between actual design and optimum design (opportunity costs) |
| Emphasis | Explanation of past deviations | Predictions of future results |
| Outlook | Introspective: from system inward | Extrospective: from system outward |
| Planner's role | Follower: satisfies trends | Leader: influences trends |

(Reproduced courtesy of John P. Van Gigh, "Applied General Systems Theory," 2nd Ed., Harper & Row Publishers, Inc., New York, Copyright 1978.)

methodology of design, and as such it questions the very nature of the system and its role in the context of the larger system."

System Analysis

To get at the "very nature of the system in the context of the larger system" means that the source of the requirement must be made explicit. This is accomplished by rational analysis of the purpose of the system of interest in light of the purpose of the parent system. In so doing, operational performance criteria emerge and the question of what to test has its best chance of being answered. The following discussion follows the general guidelines presented in "A Systems Approach to the Semiautomation of the National Airspace System" (ref. 52). That approach proceeds from system analysis to system synthesis to system evaluation.

The operational requirement—The need for a mechanism to alert aircrew members to deviations in aircraft and subsystem performance arises out of limits in the ability of humans to continuously monitor the total system operation. For instance, the pilot cannot directly observe internal subsystem states and must otherwise be advised when malfunctions occur. High task complexity, high workload, low workload, crew interaction, fatigue, personal problems, contingency events, all contribute to the diversion of attention from the aircraft system monitoring.

task. In order for the aircraft system to carry out its operational function, deviations must be made known to – transmitted to – the management and control subsystem – the crew.

The operational purpose and function of CAWS, based on these requirements may be stated as follows: The purpose of the cockpit alerting and warning system is to assist the aircrew in their monitoring, control and management activities by sensing critical aircraft performance, configuration, and subsystem deviations, calling them to the attention of the crew, and informing the crew of the problem source through aural, visual, and tactile displays; the essential functions are thus: sensing, alerting, and informing.

The sensing system is the domain of credibility (false and missed alerts due to malfunction or the setting of deviation detection thresholds). The alerting function is the domain of monitoring and vigilance with its concomitants of disruption and intrusion. The informing function is the domain of information transfer with its concomitant of interpretation. Except for aircraft systems malfunctions, it is notable that the CAWS display system, amplifies or transforms or dichotomizes continuous information normally available from other cockpit displays. These other displays have a control associated with them (lever, switch, button, knob, pedal, handle, etc.) and the crew member closes a loop by activating the control that changes the displayed value or state. Cockpit alerting and warning systems are different from these other displays: they are only displays. So, while for most cockpit man-machine functions a clear task requirement can be stated in terms of observable activities, the same is not true of CAWS. It is awkward to state the crew's tasks in relation to these displays: "to be alerted" or "to be informed." These are cognitive states, not activity states. They are not trainable in the sense that conventional tasks are routinely trainable. The response logic diagram in figure 1 indicates that their adequacy can only be evaluated by inferences from activities "downstream," that is, response outcomes A, C, E, and F.

Thus, the operational requirement that underlies CAWS is fairly straightforward and comes from a real need. However, it is extremely difficult to obtain objective evidence that the need is being satisfied. Consequently, one is heavily dependent upon subjective indices of performance adequacy.

Requirements from operational experience— In the past, the operational requirement stated above has been instrumented somewhat piecemeal by appending to each new or redundant subsystem its own deviation detection and alerting system. This has resulted in disorderly growth because of the lack of a unifying system approach. This is partly to blame for the general dissatisfaction with current configurations. But as experience is gained with more and more complex systems it is becoming apparent that instrumentation of the simply stated operational requirements leads to some paradoxes which are due to a fundamental incompatibility of discrete machine logic and human intelligence. Thus, in addition to alerting and informing, CAWS should:

1. Alert but not "alarm"
2. Intervene but not disrupt
3. Inform but not surprise
4. Aid monitoring but not replace it
5. Decrease false alarms without increasing missed alarms
6. Initiate timely but not too hasty action
7. Provide information but not increase workload

These almost mutually exclusive goals suggest an underlying antithesis. The machine is activated by a fixed, predetermined risk assessment paradigm; the man, as perceiver, brings to the interface an adaptive, judgmental, anticipatory paradigm. This basic incompatibility is a core problem and is the design challenge in CAWS.

This incompatibility leads to two extremes – one in which the alert may be perceived as a redundant nuisance, the other in which it comes as a surprise with a high probability of being ignored. For example, if a pilot is fully aware of the aircraft system's position and progress in four dimensions and changes (deviates) the system as a result of judgment and by choice (contingency) the alert will activate in its automatic fashion. This is at best confirmatory, at least redundant, and at worst a nuisance, possibly disruptive of other ongoing tasks (see cases B and C in the introduction). In a recent scan of ASRS data bank covering the period August 1976 to November 1977, six GPWS activations were reported. By pilot report of these incidents four of the six were considered unnecessary and disruptive. (See also ref. 53 for an interesting discussion of similar problems with the altitude alerting system.)

On the other hand, the second kind of extreme is encountered when the pilot does not have a veridical percept of the aircraft system's position and progress in four dimensions and does not know that it is nonveridical. The activation of a deviation alert can be perceived as a complete surprise and perhaps deemed incredible (false, unbelievable) given the shaky state of alert credibility, generally. This may have been the case in the two controlled-flight-into-terrain accidents cited in the previous section in which the altitude aural alert presented crucial information.

The identification of the information required for control has a special meaning with regard to CAWS. In order to "control" the CAWS, the aircraft and other systems are controlled, that is, the deviation is remedied. Control requirements over the CAWS itself are not clear. Loop closure is mainly considered to be through the pilot to the system monitored by CAWS and then the CAWS display (see the functional schematic in fig. 3). Ideally, this might suffice if the alerting system itself did not introduce problems, such as disruption, startle, and other problems of "bad timing." Since some or all of these may be obviated in a new system design, the question of the control of the alerting system as a system should be deferred until a configuration has evolved.

The process of analyzing the system and identifying the functions and tasks to be performed leads to some conclusions about the basic nature of CAWS:

1. CAWS do not serve the system as do other subsystems; they serve the *monitor* of all those subsystems. They can, thus, only be evaluated at the flight-management level of human response.
2. The application of the usual task analytic paradigm, display-control-response, is seen not to apply. Only the *display* is present. The relevant *control* is elsewhere. The *response* may be delayed, complex (several controls), and is also elsewhere. Therefore, given that the appropriateness of a response to an alert has been assessed, further evaluation is irrelevant and is in the domain of standard and contingency operating procedures.
3. An appropriate paradigm for CAWS might be display-response-activity, where the response is cognitive and covert, thus not directly observable. The responses are the decision options shown in figure 1.

4. Given these departures from the usual discrete task structure, a "reaction-time" criterion for performance evaluation does not appear to be a valid discriminator in full-flight management context. However, an activity delay interval following an alert may be a relevant dependent variable. Such a measure would have to be evaluated against operational criteria, such as allowable safe delay, flight phase, concurrent workload, and level of urgency.

System Synthesis

It is well, at this point, to note that the purpose of providing a discussion of system analysis and system synthesis is to arrive logically and naturally at the topic of system evaluation. Although the contents of the discussion may only be illustrative and sometimes even incorrect, the process is not. The proof of the process is in the doing rather than in its discussion; however, in the doing, frequently the overall goals are lost sight of and one finds that short-term goals obscure long-term goals. It is not productive to synthesize a system without the knowledge and insights provided by the system analysis exercise. It is also not productive to begin synthesis without use of the deductive processes. In the deductive process the system operational requirements are developed into functional requirements of CAWS which appear as: (1) an explicit system logical structure and function, and (2) a hardware subsystem totally in support of that structure.

The selection of CAWS displays before the logical structure is set will result in an implicit structure determined by lower order elements and inadvertent rather than deliberate choices. That approach is inductive and more characteristic of system improvement than system design (see table 2). With a deductive approach, system evaluation will be seen to be a verification of the system's adequacy in terms of the operational requirements that guided its design.

A sample alerting system— As discussed above, several CAWS integration schemes have been proposed by interested groups in the technical community. These proposals are in the form of functional requirements. They reflect the deductive process in which operational and system requirements have been made explicit. Taking a cue from these efforts, a sample alerting system will be described with which to anchor the discussion to follow.

TABLE 3. URGENCY MATRIX USING DUAL CATEGORIZATION

| | Performance | Configuration | Systems |
|-----------|-------------|---------------|---------|
| Emergency | S1-1 | S2-1 | S3-1 |
| Warning | S1-2 | S2-2 | S3-2 |
| Caution | S1-3 | S2-3 | S3-3 |

The proposed system logic is tabulated in table 3. It will be seen that the first step in system synthesis — functions allocation between man and machine — has appeared. In keeping with other proposed schemes, urgency level coding and prioritization have been assigned to the machine but these assignments are modifiable by software manipulations and are adaptable if and when required. The sample logic system (SL) shown in table 3, categorizes the alert two ways: by the three deviation

categories, performance, configuration, system, and by the level of immediate response required as indicated in the urgency level column.

Definitions of the column headings in table 3 are adapted from the section on "Categories of Warning Systems," above. They were:

1. Performance deviations: departures from safe flight profiles
2. Configuration deviations: departures from aircraft aerodynamic profile for a given regime
3. System deviations: malfunctions, faults, inoperability of aircraft and subsystems

These sources of alerts can interact with each other so they are not totally independent. There is a general hierarchy in that configurations subserve performance and systems subserve both configurations and performance.

The titles for the rows on the left of the matrix refer to the urgency level of the given deviation: (1) "Emergency" is to indicate that immediate action is required; (2) "Warning" is to mean that action may be delayed but if not taken an "Emergency" may result; (3) "Caution" means that the crew needs to program remedial activities into its workload and that the aircraft may proceed in an abnormal mode of operation -- a longer delay is allowed but if left untended the situation could degrade to "Warning" and then to "Emergency" levels.

Some logic schemes propose a fourth category to display information on system states such as "system armed -- unarmed," "system on -- off," and "go -- no -- go." However, in this sample system these are arbitrarily excluded for both simplicity and to enhance the alerting system purpose.

In each cell, a signal, S1-1, S1-2, . . . , S3-3, is required. The coding possibilities for this nine-cell matrix could range from one aural with nine levels of dimensionality coding (frequency, repetition rate, intensity, etc.) to nine distinct aurals. An option at midrange is to use three aurals each with three coding levels. The scheme selected should retain simplicity so that meaning is immediately known and training beyond familiarization is not required. However, at this stage in development of the system logic, a choice is not required. This is not to say, however, that the exploration of the set of choices should be delayed.

The sample system will also include an alert prioritization scheme so that multiple deviations will be presented in order of criticality without overlap or confusion of signals. Another feature that would appear to be needed is the inhibition of certain alarms as a function of flight phase. Those deviations that have no effect on the immediate performance of the system would be postponed. Workloads are particularly high during takeoff and landing and these flight phases would be the most critical with regard to attentional and monitoring activities by the crew.

A problem with urgency-level coding, prioritization, and alert inhibits is how to assign deviations to the several categories. This requires a set of nonarbitrary rules that involve a host of relevant parameters. This is, itself, a major effort in the design process and, obviously, cannot form a part of this document. Reference 2 recognizes this need and suggests possible methods for satisfying it. The relative importance or criticality of various faults, failures, and malfunctions may also be derived through the use of other guidance documents. For instance, the SAE APP 926A (proposed) "Design Analysis Procedure for Failure Mode, Effects, and Criticality Analysis" and MIL-STD-2070 "Procedures for Performing a Failure Mode Effects and Criticality Analysis." Failure mode effects analysis provides for a description of the crew response required for various failures.

The informing function -- Having considered a way to structure an alerting logic, it is next necessary to establish functions and rules for informing the crew. For instance, a "system/warning"

level alert has activated. The crew now needs to know: Which system? What component or function? What to do?

The informing function could have three purposes. One is the initial message to the crew concerning the particular component of performance, configuration, or system that has deviated. Another is the presentation of the details of the deviation. A third is the presentation of remedial measures to correct the deviation. If such remedial text is to be included in a CAWS memory system it might be better treated separately. The issue of the interface of CAWS with other flight-management systems is a real one and ultimately must be considered. For instance, if one or more cathode-ray tubes (CRT) are used for area navigation (RNAV) and cockpit display of traffic information (CDTI) it may be possible to use these displays for CAWS annunciation and information retention. That kind of holistic approach is being pursued in an interesting series of studies by British Aerospace Enterprises using nine CRT's in a simulator cockpit. Such considerations are not basically different from the present effort but are beyond its scope.

The informing function is the process of elaborating the initial information in the coded alert. There is a finite number of ways in which humans acquire information from the environment. In the order of their ability to provide information they are visual, aural, tactile, kinesthetic, olfactory, and gustatory. The last three can quickly be disqualified for informing the crew on the basis of articulation and convenience. (Tactile stimulations can be strong alerts for human beings – for example, electric shock and seat prods – but their acceptability is questionable. The stick-shaker stall warning is an exemplary application.)

There are only, then, two ways to carry through the informing function that was initiated by the coding scheme in the system logic discussed above. These are voice (synthetic) and visual annunciation (alphanumeric or CRT) display.

Voice has the advantage that it is attention-getting and does not require shifting of the direction of gaze. It is also direct and can be simple. Also, given that the voice is distinctly different from all other voice messages, it can be used also for the alerting function. It is recommended in the SAA, S-7 committee functional scheme that the single aural "attenson" be supplemented by synthetic voice for the most urgent deviations. This tends to agree with the results of a recent questionnaire (ref. 54) administered to 50 line pilots. Its purpose was to elicit pilot opinion on several aspects of alerting and warning system design. The pilots generally were in favor of the use of synthetic voice but only for very urgent deviations. A distinct disadvantage is that the message is not storable except by further crew action. That is, it could be held in silent storage by a "hold" button and recalled for later processing by a recall button or automatic timing circuitry. An alternative is reiteration of the voice message but that could become obnoxious for deviations that are not or cannot be tended to immediately.

Visual displays in support of the informing function would probably be best selected from those with a broad message content capability – alphanumeric or CRT displays driven by preprogrammed digital logic. User operating procedures would be strong determinants of the logic structure, information content, format, etc., of the displayed information. But the selection of options is based on basic capability provided by the design engineer.

The visual display has the advantage over voice that a much greater amount of information can be presented. Also, the information can be retained on the display for later use. The visual display

does require one's attention in that it must be looked at to be used. That disadvantage, which is true for the initial informing function stage where task disruption can occur, is offset by its power in deviation alleviation and the establishment of contingency operating procedures. This provides a natural interface for further subsystem integration into the full flight-management task.

Three display types have been suggested in the course of this hypothetical system synthesis:

1. An alerting tone or tones coded for category and urgency level
2. Synthetic voice for categorical deviation identification and as a possible alerting aural
3. A visual-textual display for full deviation identification and remedial information

No one of these can fill all of the identified requirements. It can be seen, however, that the required flow from attention-getting to explanation is well supported by the combination. All have at least rudimentary application in current CAWS. The sample system being synthesized would encompass the three levels of information articulation with specific displays to be determined.

Crew control— The last element in the CAWS functional structure is crew option and control over the CAWS itself. This is a difficult problem because such control could, in certain instances, defeat the purpose of the alerting system. In allowing the crew to cancel, postpone, or inhibit control modes there is the possibility of making the alert ineffective unless a "remembering" and reactivation capability is included. The GPWS is an example of the "double bind" that control or lack of control implies. The system is certainly of value as is, but as one pilot reporter stated in an incident from the ASRS survey reported above: "... it (GPWS) can endanger more lives than it saves." But if it can be turned off (postponed), when is it appropriate to turn it on again? Just as some deviations can be inhibited for flight phase, so can these be postponed — stored — for later manipulation. Almost all system deviations are of this kind but usually performance and configuration deviations are not.

Again, an explicit need is seen to allocate functions to man or machine. The phase-adaptive alerting system concept is proposed as a machine function that "unloads" the human component. The sample system being discussed will be hypothesized to have such a capability. This might obviate the requirement for an individual alert postponement command. That is to the good because a major design goal is to provide a CAWS that does not increase workload.

Urgency level coding is also a means for softening the requirement for crew alert cancellation or postponement. It does so by providing an instantaneous risk-assessment decision: "respond right now," "be ready to respond," or "schedule a response" (emergency, warning, caution). Given these rules, the reiteration of aurals would be limited to only the most critical deviations. An option would be visual holding of the alert with display flashing and color, intensity, or frequency coding for criticality levels as appropriate.

The need to place alerts on hold or to "punch out" aurals comes not so much from operational requirements but seems to derive from CAWS as currently embodied. It is fairly apparent that any "punch-out" function allocated to crew option should allow cancellation or postponement of the *alerting* function but not the *informing* function. As suggested for phase-adaptive inhibits, this should be an alert inhibit but not an information-display inhibit. In the sample system being synthesized all aural alerts except those for the most urgent performance deviations will be transferable to visual display with information content intact, using a single control for all

deviations. (The use of a voice recognition system for the control function is to be considered.) Reiterated aural for critical performance deviations would only be cancellable by appropriate and immediate control action.

To this point an information processing system structure has been synthesized as a sample. Many functions previously allocated to the pilot are now allocated to the machine. Its purpose is illustrative; it is not a *proposed* system. The system is still "soft," and changes to the functional and logical structure are easy to make. This will be less true as implementation proceeds to a hardware configuration. Even then, however, since displays will be versatile and articulable, significant changes can be effected in the character of the system by software changes rather than hardware retrofit. This is the conceptual first step in the design process and it is not costly. Several CAWS functional structures can be designed as candidates for implementation and checked against operational requirements in the exploratory phase of system evaluation. Multiple conceptual designs are to be encouraged because there is no single design solution. On the contrary, there are likely several concepts that would satisfy operational requirements.

The process of actually putting the system together with hardware components is logically, but not necessarily chronologically, the next step. System test and evaluation is integral with that process in making design decisions and in assessing the appropriation of the total system for operational use.

System Evaluation

It has been stated that system evaluation is an on-going process, one that starts with the enunciation of system functional requirements. From that point onward each design decision is itself an evaluative step. Sources of criteria for these decisions range from engineering experience, intuition, and expertise, through user requirements and existing relevant data, to the more formal system-keyed studies and tests.

The application of system-keyed studies is decided upon during the development process. Whether the test is exploratory, resolution, or verification depends on the state of evolution of the system. However, the thinking, planning, and design process should run well ahead of the development activities because the operational requirements are known in advance. Thus, exploratory testing is as much governed by those requirements as is verification testing but these events, as activities, require hardware. Table 4 shows the gross steps in system evolution and, in general, where testing activities would occur. The framework for the evaluation of the sample system is now to be referred to the emergence of system hardware.

TABLE 4.- SYSTEM DEVELOPMENT CYCLE WITH TEST PHASING

| Development phase | Exploratory | Resolution | Verification |
|-------------------|-------------|------------|--------------|
| Concept | X | | |
| Breadboard | X | X | |
| Brassboard | | X | X |
| System | | | X |
| Use | | | X |

Note that exploratory testing can begin with the "concept" using paper and pencil techniques for eliciting pilot (user) critique. The term "breadboard" is understood to mean a shop or laboratory integration of equipment selected for the assessment of functional concepts. The detail of such an ensemble could range from a simple tabletop model to a cockpit mockup. The "brassboard" is understood to mean a system that embodies the functions and standards developed with the breadboard and used in a particular application, for a specific aircraft. The utilization of the brassboard would be from cockpit mockup to full-mission cockpit and the aircraft. The "system" is that which appears in the production aircraft, a fully developed and integrated brassboard.

Breadboard testing: system exploration— For the exploration of the system in terms of its logical functions, its dynamic operation, and the integration of system components, a digital computer or a microprocessor could be used as a central test apparatus. Such a scheme could profitably be used throughout all phases of test and, indeed, may itself be central to a newly conceived CAWS subsystem. This option is conceived as a functions and concepts evaluation technique and will henceforth be called "*facet*" to facilitate exposition.

The microprocessor, if such is used, could be programmed and easily reprogrammed to provide logical functions related to alert prioritization, inhibits, urgency levels, holding, and pilot inputs ("store," "give me some information," "recall"). All aircraft deviation sensors would be assumed to input to the microprocessor. It may also be utilized to store checklists, emergency operating procedures, and other textual material. The microprocessor would thus be the means for information reception, ordering, processing, routing, and transmittal to displays, whatever they might be.

Peripherals to this central facility would be all those control and display schemes that are considered to be candidates for the ultimate design:

1. Visual presentation of the centrally manipulated information is provided by an alphanumeric display, a cathode-ray tube, or it may be displayed by such means as a central annunciator matrix panel or even in the visual field of a head-up display.
2. Aural presentation of information is by speaker or headset and can include alerting stimuli and synthetic voice, each provided by appropriate electronic modules.
3. A means would also be provided for crew information input to the central microprocessor. As with peripheral displays, the peripheral control unit can take as many forms as it is desired to explore: keyboard, touch panel, and even a microphone and a voice recognition module for possible verbal control.
4. It must be understood that, given the central information processing capability, one has a choice of peripherals, limited only by practicality and relevancy. The purpose of *facet* is not initially to recommend particular display/control schemes but to furnish a test context for their evaluation, one that is subordinate to the exercise and evaluation of the information acquisition, processing, and transmission logic.

A *facet* may be used at many levels of simulation, fidelity and complexity. These range from tabletop demonstrations and game playing through mockups and part-task simulations to research flight simulators and full-mission training simulators. At the simple level, the forcing function is

derived from scenarios, scripts, and role playing; when a simulated aircraft is the test bed, the aircraft simulated systems and aircraft responses generate *facet* operation. The game playing becomes less a matter of dissimulation and more a matter of realistic simulation as task fidelity increases.

A *facet* would be "loaded" with the logic proposed in the sample system outlined above. The system would be "exercised" through appropriate inputs and the critique would begin. Questions that would be asked would be those leading to "proof of concept" evaluations. The source of these are the operational requirements as revealed by processes illustrated above including the core task structure shown in figure 1. For example:

1. Does the dual categorization scheme in the sample system provide for immediate crew understanding without complexity?
2. Does the reduction of aural alerts to one or, even three, provide the needed information? Is the reduction acceptable?
3. Does this scheme seem to be attention-getting but not disruptive; that is, is disruption minimized by inhibits and prioritization?
4. Can the system include the stereotyped alerts without loss of their current urgency level coding?
5. Which deviations should be announced by (1) synthetic voice, (2) by coded aural?
6. Should information for validity checking be included? How?
7. Should an ignored signal be reiterated? How long before such is activated? Repetition rate?
8. Should prioritization be fixed in the microprocessor? Modifiable? By whom?
9. The system was designed to give a first-level indication of required response time. Do observers or experimental subjects behave accordingly?
10. Does the control for transferring auditory information to visual information appear helpful? Should the transfer be automatic?

The foregoing questions are a small sample of the perhaps infinite number of questions that might be posed during system exploratory activities. The better, more pointed ones will only arise in the actual doing.

Since it is fairly obvious from the single-alert logic shown in figure 1 that there is little behavior to observe or score, early dimensionalization of responses may not be appropriate. However, system logic can be checked because *facet* would make it explicit and visible. *Facet* would also be flexible and could be easily changed to accommodate other logic structures. For instance, the sample system might be changed by rejecting the present column categories ("Performance," "Configuration," and "System") and categorizing in terms of "urgency." This, then, would be

simply a finer urgency level indicator than is provided already by the required level of response in table 3.

During the breadboard, exploratory phase of evaluation and, in fact, throughout the whole development testing the major method for relating the system to criteria will be expert judgment. This is because measurable phenomena are rare and because pilot acceptance is a prime (psychological) criterion. This means that user response is to be solicited using objective techniques. There is a body of human factors technology that deals with means for objectifying subjective impressions — rating scales, questionnaires, interviews, polls, utility assessment schemes, etc. (ref. 55). It should be made clear that what is not proposed is system design and evaluation by committee. The emotional arguments of a forceful, but biased, individual or group is a poor source of criteria.

In summary, exploratory testing in early system development is for the purpose of functions allocation and checkout of those choices. It is best accomplished with a wide-ranging opportunity to "cut-and-try."

The *facet* approach could be used throughout system development and system test. It should include the capability for simulating all the options that any candidate system would draw from. A candidate CAWS is one that is proposed for a given system; it is developed into a brassboard or prototype. *Facet* thus can provide for fast reconfiguration and the checkout of system changes as soon as they are contemplated.

Component testing: design resolution— Much of the previous human factors studies of CAWS have been at the component level. Also, they have been in the main, attempts to evaluate, in isolation, the alerting power of a proposed or existing alerting system. Component testing in isolation from the "present" CAWS system is not appropriate in the current development scheme. The system components are no longer separate entities. Thus, they are not in competition with each other for the attention of the crew as much as they are in support of the system logic which, in turn, is supporting flight and resource management activities.

Components of CAWS are those "peripheral" devices used in a *facet* to interface the system with the crew to provide information transfer. Alerting tones, synthetic speech, and visual displays have been mentioned as transfer media.

It is assumed for the sample system that the alerting function will be accomplished by auditory means. The sample system, as conceived, constrains this to a 3 X 3 code. A signal or signals are to be selected which best satisfy the dual function of alerting and providing initial information with regard to the nature and urgency of the deviation. Three basic tones might be selected that can be interrupted at rates corresponding to the three urgency levels. Another scheme might be to consider a chime, a tone, and, say, a clacker, each of which is to be repeated three times for highest urgency, two times for the next level and once for the least. There are a great many possible combinations of alerting sounds and coding dimensions. A *facet* that has the characteristics suggested could provide for early trial and selection from a very large set.

The selection of a few combinations for resolution testing can be accomplished by exercising the breadboard system with experienced line pilots acting as in-flight users. This requires the development of at least simple scenarios, but the effort is not wasted because scenarios will be required for later full system evaluation.

The set of candidate aural coding schemes can be evaluated by techniques appropriate to the elicitation of preverbal pilot judgment. That is, an opinion is not solicited. What is sought is a judgment in which each scheme is forced into a position on a utility rating or ranking scale such as through the use of a card-sort process (see ref. 55). Utility assessment methods are developing that lead to objectification and quantification of subjectively evaluated multifaceted phenomena. This kind of judgmental response can, of course, be acquired from paper and pencil techniques but *facet* provides an opportunity to see and use the set of options to be evaluated. Unlike many other decisionmaking tasks, the set of outcomes is visible and even modifiable.

If the options for aural alert coding can be reduced to a single scheme, then no further testing is required. This will generally not be the case because there probably will be several that can satisfy the operational requirements. Other criteria that will enter into component selection and which may be brought to bear at this point include cost, size, weight, availability, and power consumption. If these, and even such things as product aesthetics, do not provide discriminators, then a formal side-by-side test would be in order. This would involve a conventional experimental paradigm, control of dependent and independent variables, and the selection of an appropriate criterial performance score. The latter might be the number of correct responses to the alerting code in unit time. The test stimuli would be presented in a "fast-time" scenario with subject workload to include at least one other attention-diverting task. Fast-time means that many events are compacted into a shorter interval than they would normally occupy. Obviously, a test of this kind requires subjects who are willing to "play the game." It is suggested that naive subjects not be used anywhere in the system development and evaluation process. The use of such subjects assumes that *fundamental* processes are being evaluated. They are not. The population sampled must be representative of experienced, highly trained, and equipment-sophisticated line pilots.

In a manner similar to the above for the coded alerting signal, other candidate components are processed. Another example is the need to determine when voice annunciation is to be used either in addition to or in lieu of the primary audio alert. Starting with the notion that synthetic voice ought to be used only with the most urgent deviations several combinations are possible. Should voice be used only for the "emergency" row in the dual category alerting scheme in table 3? Should it be used only for the "performance" deviations column in that matrix? And what about the first two columns and the first two rows?

These options can be "brought up" on *facet* and evaluated by the judgment of experienced subjects. Again, obviously inadequate schemes will become apparent, particularly if there has been no attempt to limit options. And, again, selection between two or more adequate candidates is accomplished by resolution testing, other things being equal.

The fitting of the display interface to the information processing scheme also has a set of relevant questions that characterize the nature of the decisions required:

1. What is the nature of the auditory alerts which best mate with human capabilities? Intensity? Frequency? Timbre? Presentation rate? Acceptability? Signal/noise ratio (see ref. 21 and related human factors data)?
2. Where is the voice to be used? When used, does the alerting aural precede it? Or is it inhibited?

3. The GPWS would fall in the "performance-emergency" cell in table 3. Can the GPWS be integrated into the sample matrix and use the same synthetic voice as all the other cells?

4. Is it desirable to integrate all other discrete or stereotyped alerts into this basic scheme? Should the fire bell be voice annunciated at a "system-emergency" level or be kept separate?

5. Is the flow of information from alerting and initial informing to full informing a meaningful progression that has optimized the contribution of the alerting aural, voice annunciation, and visual annunciation?

6. Is there a need for an alphanumeric, single line display for visual display of short messages? How could this be made parallel to coding logic for auditory displays? Would changes in flash rate, color coding, or brightness for "emergency," "warning," "caution" levels of urgency be appropriate to evaluate?

7. Can the "one liner" alphanumeric display be accomplished through the use of a CRT that can also present textual material for problem analysis and remedy?

8. Or, from a human engineering viewpoint, since there is as yet no central display for the sample CAWS, might the alphanumeric stand as such? It could be mounted in the same central position as current master warning panels. It would add to the information contained in the coded alerting aural but full information would be presented automatically and simultaneously on a CRT.

9. If the CRT is used also as an alerting element what coding is to be used such that visual coding is parallel to auditory coding?

10. How are inhibited alerts to be displayed and activated when the inhibit regime is departed?

11. How are multiple alerts to be displayed assuming predetermined prioritization rules?

Again, this sample set of questions is far from complete but illustrates the kinds of concerns that govern the component (display) selection process and that require evaluation activities.

In the normal flow of logic that takes place in moving from operational requirements to hardware it is now reasonable to apply human engineering design principles, for example, display location, brightness, legibility of visual material, formats, control location, and size and shape (see refs. 2 and 21). An interesting human engineering question is whether the many amber and red caution and warning lights which currently clutter the cockpit can be eliminated.

Operational testing: system verification— As has been alluded to several times there is no single, perfect CAWS that is the only answer to the operational requirements. Many candidate systems could be designed that would be satisfactory in that respect. Thus, the testing is of a system against those requirements rather than a system against a system, as is appropriate in resolution testing.

Obviously, it is no longer possible to refer to the sample system except in basic function. That is because of activities carried out as outlined in the previous two sections. They would have resulted in a configuration that cannot be visualized without the test results. Also, examples, like

analogies, should not be pushed too far. However, what is still perceptible are the operational requirements.

Performance criteria: Given that the operational requirements are reasonable the criteria for operational test are:

1. The system is in consonance with the stated operational requirement. This is supported by clear system logic, functional integration, and integration into the larger monitoring task.
2. The seven principal paradoxes of CAWS, as currently embodied, have been minimized (p. 42, above).
3. Nuisance, disruption, startle, and pilot dissatisfaction (psychological criteria) have been minimized or eliminated.
4. The response alternatives shown by the junctions in the single alert logic diagram in figure 1 are clear and executable under all flight deck task loads.

Items (1) through (3) are criteria to which system characteristics can be compared mainly, perhaps only, by expert judgment. Item (4) includes elements that allow counting: missed alerts, false responses, incorrect responses, true responses, and response timeliness (see fig. 1). But the comparison of each of these to a criterion number is not possible because such numbers do not exist. (They would be ideal in a resolution test between two systems, though the "winner" would still require operational verification.) These too, then, would be subject to evaluation by expert judgment, except for the last one, response timeliness. The same parameters and decision rules that were used to develop urgency levels and alert prioritization may be used to establish critical response times.

Thus, in addition to the use of expert judgment, which has been used extensively throughout development testing, there appear to be the following performance characteristics that are relatable to objective indices: (1) Response time with respect to time required/time available; (2) Response correctness and scheduling with respect to deviation urgency and priority; and (3) Timeliness and correctness of validity checking where possible.

The test context is to be designed to "tease out" these performance characteristics. It is important to distinguish between the task-oriented time measure (response time) and reaction time. Although reaction time has frequently been used as a performance measure in laboratory evaluations of alerting systems, there are difficulties associated with its use. Almost anything that affects human behavior affects reaction time. Forbes (ref. 56) found a full stomach slowed a reaction time to sound, but not to light, for example. It reveals something about a *neural* response to given stimuli but nothing about response appropriateness in real world tasks. As a matter of fact, Fitts's classic analysis of 480 pilot-error accidents does not include any that are due to slowness in responding (ref. 57).

Contextual considerations: The context for verification testing must include all the tasks, conditions, contingencies, and workloads that are to be expected in actual operation of the system. The testing arena is characterized by the attributes shown at the far right of figure 3, at least in terms of task-fidelity.

Even in this full-task context a problem arises immediately. The problem is peculiar to any flight-deck oriented research. That is the problem of having a sufficient number of events occur in a necessarily limited time period for observation and counting. This problem is particularly acute in the case of CAWS because of the low probability of occurrence of deviations.

A way to provide enough events for evaluation purposes is to use the previously mentioned "fast-time" simulation in which events are compacted into a shorter time period than would be true in operation. With appropriate planning and scenario building these can be made realistic in all aspects except their probability of occurrence. This is known to be acceptable to pilots for training purposes. Although not realistic, valid results can be obtained if the test subjects are willing to "play the game," and are prebriefed to accept the probabilities as being temporarily real. Note that the events and related performance are in real time; what is time-shortened are the intervals between events and their frequency of occurrence is increased.

A related problem is that of the psychological "set" of test subjects who are always fully conscious that CAWS is the reason for the test. This is a state of sensitization totally different from real-world monitoring readiness levels. Even if it is not made explicit, the purpose of the test will be transparent to experienced pilots. As mentioned previously, however, this may compromise the evaluation of *only* the *alerting* function. The measurement of response time and deviation dispensation is not necessarily similarly compromised.

Another difficulty is one that is peculiar to the nature of the task. That is the element of covert behavior which makes ignored, missed, and postponed alarms (fig. 1) all look alike, at least initially. If, however, a postponement action is required of a crew member the behavior becomes observable. That may be a pilot control option in the CAWS itself or it may be injected into the full-task simulation as an element of the experimental apparatus.

It has been stated that the important consideration in the design of the test context is that it be totally representative of the flight-deck task structure. This requires sample scenarios based on real world operations and operational contingencies, opportunities to exercise resource management and flight management activities, and representative workloads. Workload is an area of technological applications that is ill-defined in theoretical construct and difficult to quantify. There is no generally accepted definition of workload at a formal and comprehensive level (see the survey of concepts in ref. 58). However, this lack does not prohibit the inclusion of workload in many applied studies. It would, however, be appropriate to have at least an operational definition in hand. Chiles (ref. 5) offers such a descriptive definition for the case of pilot workload. The Chiles report is recommended for guidelines in conceptualizing and planning for the inclusion of appropriate workload levels for the CAWS test context. Artificial and/or secondary tasks unrelated to flight deck activities are not herein recommended. What is recommended is the design of scenarios that include high *operational* workloads.

Finally, it will be necessary to observe performance in the test simulation in order to collect data on those behaviors that can be scored. Also, observation of the total task activities will yield a great amount of information with regard to general CAWS operability, the appropriateness of the test design, and other information that cannot be anticipated. It will be necessary to train observers to be knowledgeable in the operational tasks and to be sensitive to the occurrence of test related events.

The character of the questions to be asked during verification testing may be indicated by the following sample:

1. Does the CAWS man-machine task appear consonant with other ongoing tasks?
2. Is the crew able to schedule CAWS events without undue disruption?
3. Has the ideal of a "quiet cockpit" (both auditory and visual) been realized?
4. Are all deviations displayed? Have some arisen which were not anticipated but will be of help if included now?
5. How easy is the system to modify, given the availability of sensing methods?
6. What is the general reception of experienced crews to the integrated system? Enthusiastic? Approval? Conditional approval? Disapproval? Enthusiastic disapproval?
7. The operational test may be construed as a precursor to certification. Does the system comply with all FAR's? Any conflicts?
8. Is the operability clear and simple requiring less training than current systems?
9. CAWS, now seen as an integrated subsystem, can be included in the aircraft operating manual, as a subsection. Has such been written and used in the operational test?

Ames Research Center
National Aeronautics and Space Administration
Moffett Field, Calif. 94035, August 6, 1979

REFERENCES

1. Bruggink, Gerard M.: Human Error Accidents and Character Assurance. Flight Safety Foundation, Incorporated, Human Factors Bulletin, Nov./Dec. 1976.
2. Veitengruber, J. E.; Boucek, G. P., Jr.; and Smith, W. D.: Aircraft Alerting Systems Criteria Study, Vol. I: Collation and Analysis of Aircraft Alerting System Data. Report No. FAA-RD-76-222, I, May 1977.
3. Texas International Airlines, Inc., Douglas DC-9-14, N9104, Stapleton International Airport, Denver, Colorado, November 16, 1976. National Transportation Safety Board, Aircraft Accident Report, Report No. NTSB-AAR-77-10, Oct. 1977.
4. Billings, C. E.; Lauber, J. K.; Funkhouser, H.; Lyman, E. G.; and Huff, E. M.: NASA Aviation Safety Reporting System Quarterly Report, April 15, 1976–July 14, 1976. NASA TM X-3445, 1976.
5. NASA Aviation Safety Reporting System: Second Quarterly Report, July 15–October 14, 1976. NASA TM X-3494, 1976.
6. NASA Aviation Safety Reporting System: Sixth Quarterly Report, July 1, 1977–September 30, 1977. NASA TM-78511, 1978.
7. Katz, Fred E.: Indeterminacy in the Structure of Systems. Beh. Sci., vol. 19, no. 6, Nov. 1974, pp. 394–403.
8. Meister, D.: Heresies, Brief Essays on Human Factors. Navy Personnel Research and Development Center, San Diego, California, Mar. 1977.
9. Aircraft Operating Manual, Boeing 747 Aircraft. Pan American World Airways, Inc., undated.
10. L-1011 Tri Star Operating Manual, Vol. 2. System Description. Lockheed Aircraft Corporation, 1971.
11. DC-10 Flight Crew Operating Manual. McDonnell Douglas, 1977.
12. Webb, Paul, M.D., ed.: Bioastronautics Data Book. NASA SP-3006, 1964.
13. Vanderschraaf, Abe: Problem Area: Warning Systems. Fokker, VFW Aircraft, B.V., paper presented at the 29th International Air Safety seminar of the Flight Safety Foundation, Anaheim, Calif., Oct. 25–29, 1976.
14. Meister, David: Behavioral Foundations of System Development. John Wiley and Sons, New York, 1976.
15. Chiles, Dean W.: Objective Methods for Developing Indices of Pilot Workload. Civil Aeromedical Institute. FAA, Oklahoma City, Oklahoma, FAA-AM-77-15, July 1977.
16. Eastern Airlines, Inc., L-1011, N310EA, Miami, Florida, Dec. 29, 1972. National Transportation Safety Board Accident Report, Report Number, NTSB-AAR-73-14, June 14, 1973.
17. Minimum Performance Standards – Airborne Ground Proximity Warning Equipment. Radio Technical Commission for Aeronautics, RTCA Document No. D-161A, Revised May 27, 1976.
18. Bateman, C. D.: Introduction of the Ground Proximity Warning System (GPWS) into Airline Service. Prepared for 29th International Air Safety Seminar, Flight Safety Foundation, Inc., Anaheim, Calif., Oct. 25–29, 1976.

19. Simpson, C. A.; and Williams, D. H.: Human Factors Research Problems in Electronic Voice Warning System Design. Eleventh Annual Conference on Manual Control. NASA TM X-62,464, 1975, p. 3081.
20. Green, R.: The Psychology of Auditory Warnings. Pilot, Spring 1977.
21. Baucek, J. P., Jr.; Veitengruber, J. E.; Smith, W. D.; and Osgood, A. G.: Aircraft Alerting System Criteria Study, Vol. II, Human Factors Guidelines for Aircraft Alerting Systems. FAA-RD-76-222 II, May 1977.
22. Systems Engineering, Production Concorde Flight Deck, Warning Philosophy. British Aircraft Corporation Limited, Ref. ss/B67-0/2027, Issue 2, June 1975.
23. Viewpoints on Design of Warning Systems. KLM Royal Dutch Airlines, Ref. AMS/05.230/Pe/lb, Nov. 16, 1976.
24. Cooper, George E.: A Survey of the Status and Philosophies Relating to Cockpit Warning Systems. NASA CR-152071, June 1977.
25. Corrigan, Robert E.; and Kaufman, Roger A.: Why System Engineering. Fearon Publishers, Palo Alto, Calif. 1965.
26. Chiles, W. D., ed.: Methodology in the Assessment of Complex Performance. Conference proceedings. Wright-Patterson AFB, Ohio, March 29-30, 1966, Human Factors, vol. 9, no. 4, Aug. 1967, pp. 325-394.
27. Singleton, W. T.; Fox, J. G.; and Whitfield, D., eds.: Measurement of Man at Work. An Appraisal of Physiological and Psychological Criteria in Man-Machine Systems. Taylor and Francis Ltd., London, 1971.
28. Gagne, Robert: Psychological Principles in System Development. Holt, Rinehart, and Winston, New York, June 1962.
29. Meister, David; and Rabideau, Gerald F.: Human Factors Evaluation in System Development. John Wiley and Sons, New York, London, Sydney, 1965.
30. Meister, David: Human Factors: Theory and Practice. Wiley-Interscience, a division of John Wiley and Sons, Inc., New York, London, Sydney, Toronto, 1971.
31. Chapanis, A., and Van Cott, H. P.: Human Engineering Tests and Evaluations in Human Engineering Guide to Equipment Design. Van Cott, H. P. and Kinkada, R. G., eds., U.S. Government Printing Office, 1972.
32. Kidd, J. S.: A New Look at Systems Research and Analysis. Human Factors, vol. 4, no. 4, Aug. 1962, pp. 209-216.
33. Burrous, A. A.: The Structure of an Effective Human Factors Effort. In Measurement of Man at Work, An Appraisal of Physiological and Psychological Criteria in Man-Machine Systems. Singleton, W. T., Fox, J. G., and Whitfield, D., eds., Taylor and Frances, Ltd. London, 1971.
34. Knowles, W. B.: Aerospace Simulation and Human Performance Research. Human Factors, vol. 9, no. 2, Apr. 1967, pp. 149-159.
35. Obermayer, R. W.: Simulation, Models, and Games: Sources of Measurement. Human Factors, vol. 6, no. 6, Dec. 1964, pp. 607-619.
36. Finan, J. L.: The System Concept as a Principle of Methodological Decision. Psychological Principles in System Development. Gagné, R. M., ed., Holt, Rinehart, and Winston, New York, 1963.

37. Chiles, D. W.: Complex Performance: The Development of Research Criteria Applicable in the Real World. Measurement of Man at Work, An Appraisal of Physiological and Psychological Criteria in Man-Machine Systems. Singleton, W. T., Fox, J. G., and Whitfield, D., eds., Taylor and Frances Ltd., London, 1971.
38. Shackel, B.: Criteria in Relation to Large Scale Systems and Design. Measurement of Man at Work, An Appraisal of Physiological and Psychological Criteria in Man-Machine Systems. Singleton, W. T., Fox, J. G., and Whitfield, D., eds., Taylor and Frances Ltd., London, 1971.
39. Chapanis, A.: The Relevance of Laboratory Studies to Practical Situations. Ergonomics, vol. 10, no. 5, Sept. 1967, pp. 557-577.
40. Chapanis, A.: The Search for Relevance in Applied Research. Measurement of Man at Work, An Appraisal of Physiological and Psychological Criteria in Man-Machine Systems. Singleton, W. T., Fox, J. G., and Whitfield, D., eds., Taylor and Frances Ltd., London, 1971.
41. Shapero, A.; and Erickson, C. J.: Human Factors Testing in Weapon System Development. Paper presented at the Missile and Space Vehicle Testing Conference, Preprint 1644-61, American Rocket Society, Los Angeles, Calif., Mar. 13-16, 1961.
42. Rabideau, G. F.: Observational Techniques in Systems Research and Development. Measurement of Man at Work, An Appraisal of Physiological and Psychological Criteria in Man-Machine Systems. Singleton, W. T., Fox, J. S., and Whitfield, D., eds., Taylor and Frances Ltd., London, 1971.
43. Wempe, T. E.: Flight Management - Pilot Procedures and System Interfaces for the 1980-1990's. Paper presented at the AIAA Sciences and Systems Conference, Arlington, Texas, Nov. 6-8, 1974.
44. Naish, J. M.: A Conformal Head-Up Display for the Visual Approach. Paper presented to the Ninth Annual Conference on Manual Control, M.I.T., Cambridge, Mass., May 23-25, 1973 (Douglas Aircraft Co. paper 6126).
45. Chopping, D.: A Revolutionary All-Weather HUD: Interavia Previews the Thomson - CSF TC 125. Interavia, vol. 31, 1976, pp. 548-549.
46. Baty, D. L.: Rationale and Description of a Coordinated Cockpit Display for Aircraft Flight Management. NASA TM X-3457, Nov. 1976.
47. Weiner, E. L.: Controlled Flight into Terrain Accidents: System Induced Errors. Human Factors, vol. 19, no. 2, Apr. 1977, pp. 171-181.
48. Eastern Airlines, Inc., Douglas DC-9-31, N8984E, Charlotte, N.C., Sept. 11, 1974. National Transportation Safety Board Report Number NTSB-AAR-75-9, May 23, 1975.
49. Future Needs and Opportunities in the Air Traffic Control System. U.S. Congress, House Subcommittee, on Transportation, Aviation, and Weather, of the Committee on Science and Technology, 95th Congress, 1st session, Nov. 1977.
50. Barnhart, W.; Billings, C.; et al.: A Method for the Study of Human Factors in Aircraft Operations. NASA TM X-62,472, Sept. 1975.
51. Van Gigch, John P.: Applied General Systems Theory. Harper and Row, N.Y., Evanston, San Francisco, London, 1974.

52. Roscoe, S. N.; Beringer, Dennis B.; and Maillard, Morris H.: A System s Approach to the Semiautomation of the National Airspace System. Concept Paper (for Office of Systems Engineering Management, FAA). Illiana Aviation Sciences Limited, Champaign, Ill., Mar. 1978.
53. NASA Aviation Safety Reporting System: Sixth Quarterly Report, July 1-Sept. 30, 1977. NASA TM-78511. July 1978.
54. Williams, D. H.; and Simpson, C. A.: A Systematic Approach to Advanced Cockpit Warning Systems for Air Transport Operation: Line Pilot Preferences. NASA Aircraft Safety and Operating Problems Conference. NASA Langley Research Center, Oct. 18-20, 1976 (In NASA SP-416).
55. I.E.E.E. Transactions on Systems, Man, and Cybernetics. Special Issue on Behavioral Decision Making. vol. SMC-7, no. 5, May 1977.
56. Forbes, G.: The Effect of Certain Variables on Visual and Auditory Reaction Times. J. Exp. Psycho., vol. 35, 1945, pp. 153-162.
57. Sinaiko, H. W.: Selected Papers on Human Factors in the Design and Use of Control Systems. Dover Publications, N.Y., 1961.
58. Gartner, W. B.; and Murphy, M. R.: Pilot Workload and Fatigue: A Critical Survey of Concepts and Assessment Techniques. NASA TN D-8365, Nov. 1976.