FAA-74-6

REPORT NO. FAA-RD-74-82

HUMAN FACTORS EXPERIMENTS FOR DATA LINK
Extended Summary of Interim Reports 1 Through 4

Edwin H. Hilborn

APRIL 1974

INTERIM REPORT

DOCUMENT IS AVAILABLE TO THE PUBLIC
THROUGH THE NATIONAL TECHNICAL
INFORMATION SERVICE, SPRINGFIELD,
VIRGINIA 22151.

Prepared for
DEPARTMENT OF TRANSPORTATION
FEDERAL AVIATION ADMINISTRATION
SYSTEMS RESEARCH AND DEVELOPMENT SERVICE
Washington DC 20591

CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. HUMAN FACTORS EXPERIMENTS FOR DATA LINK EXTENDED SUMMARY OF INTERIM REPORTS 1 THROUGH 4

1.1 INTRODUCTION

Over the past 18 months, four Interim Reports, all bearing the generic title "Human Factors Experiments for Data Link" have originated at the Department of Transportation/Transportation Systems Center, and are identified respectively as FAA-RD-72-150, FAA-RD-73-55, RAA-RD-73-69 and FAA-RD-74-81. The objectives of these experiments have been two-fold; to determine how meaningful short ATC messages are when coded in various ways and presented in different formats, and to evaluate candidate display devices and technologies for the presentation of such short ATC messages, so as permit a reduction in the number of such devices requiring further simulator evaluation and eventual flight testing.

The four previously published reports document the findings of eleven discrete experiments. The sheer volume of data collected and reported (approximately 18,000 data points, organized and presented in some 125 tables and graphs) has resulted in reports much too massive to permit easy access to the more important findings by any but the most devoted reader. The present extended summary attempts to put these major findings into perspective and into a form which can be more easily assimilated by a reader interested only in the conclusions which have been drawn thus far, and not in the techniques employed for obtaining the data, Section 6 of the present document lists these findings.

While the emphasis throughout this series has been on the use of simulators for the evaluation of concepts and prototype equipment for Data Link, this work has been supplemented by laboratory studies in those areas where useful information can be collected more efficiently and rapidly than can be done under the constraints of operations in a flight simulator. The material in this summary is organized according to the types of studies run, and thus does not necessarily follow the historical sequence of the studies. For purposes of organization, the material covered includes

(1) simulator studies of visual short message ATC (SMATC) displays, (2) laboratory tests of message format and coding schemes, and (3) preliminary evaluation of synthetic speech as a means for providing ATC information.

Throughout the remainder of this extended summary, the original source of any reported experimental procedures or findings is identified with parentheses by a two-digit number. Thus, (3-2), identifies material presented in more detail in the second section of Interim Report #3.

2. SIMULATOR STUDIES OF SHORT MESSAGE ATC VISUAL DISPLAYS

With the exception of clearances, weather reports and ATIS, nearly all other ATC commands and advisories can be presented on a visual display limited to some reasonably small number of characters. Since such short messages constitute a large percentage of ATC transactions, and since these messages place more demands for short reaction times by flight crews than do longer messages such as clearances or ATIS, the emphasis during the experiments to date has been on the evaluation of such short messages.

In a preliminary experiment, run on the Transportation Systems Center's GAT-1 simulator* (1-1), an attempt was made to validate the concept that Data Link could substitute for a voice channel in presenting a limited repetoire of ATC commands. A simple display capable of presenting heading, altitude and speed commands was fabricated, and was evaluated by eight TSC pilots on simulated flight paths around the Boston area, leading to an eventual landing at Boston's Logan Airport. It was found that the commands could be generated manually with sufficient precision to permit accurate control of the flight path. All of the pilots completed all of their simulated flights satisfactorily. Pilot opinion of the display was universally favorable. The display used is depicted in Figure 2-1. It was identified as the WIDCOM.

The success of this very preliminary experiment justified the fabrication and testing of additional displays having capabilities for the presentation of a somewhat greater variety of short ATC messages (3-1). Four such additional displays were fabricated and are depicted in Figures 2-2 through 2-5.

The 7-Window display, depicted in Figure 2-2, used 16-segment alphanumeric readouts from Master Specialties Company as the means for character generation. Characters were generated using incandescent lamps and fiber optics, had a height of 0.42 inches, and a

*The GAT-1 simulates a light, single-engine aircraft with a one man crew.

Figure 2-1. The WIDCOM Display

Figure 2-2. The 7-Window Display

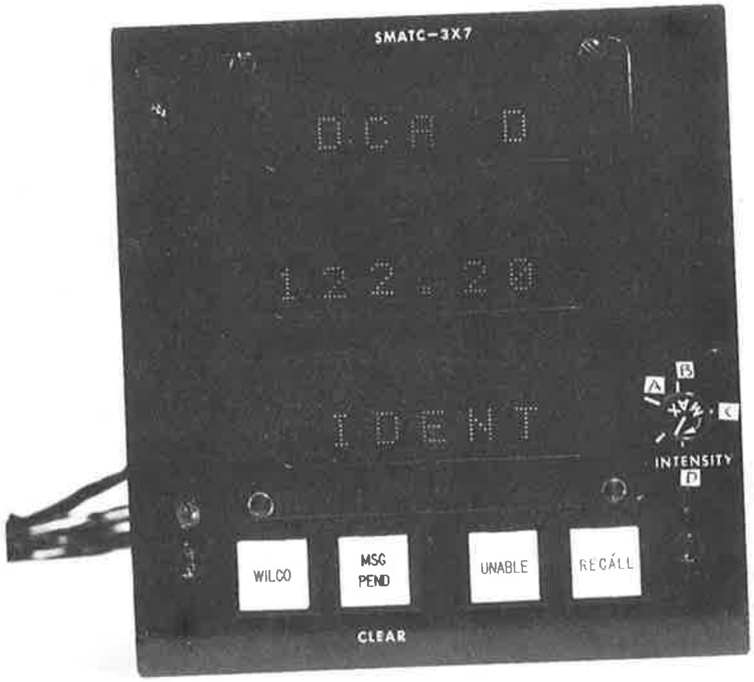

Figure 2-3. The 3x7-Window Display

Figure 2-4. The 32-Window Display

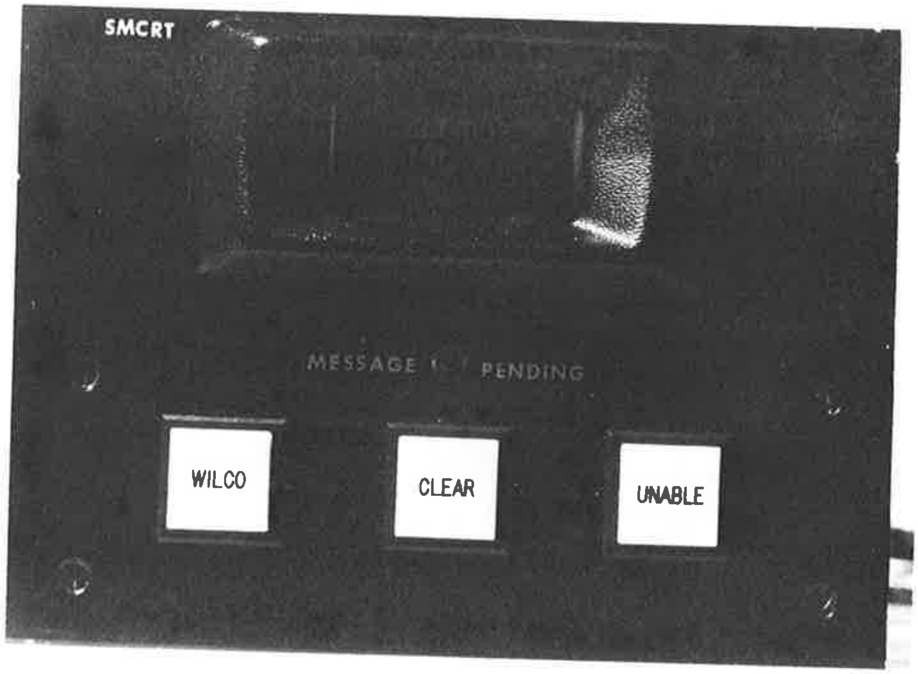

Figure 2-5. The NIMO Display

typical brightness of 400 foot-Lamberts. Storage registers were provided for three 7-character messages, and these could be acessed sequentially by depression of a "Message Pending" pushbutton located below the readout.

The second prototype display, depicted in Figure 2-3 presented three lines of seven characters each, using Monsanto red light emitting diodes. Characters were based on a 5 x 7 dot matrix, character height was 0.35", and brightness typically 300 foot-lamberts.

Figure 2-4 depicts a plasma display utilizing a Burroughs Self-Scan Panel 8.50 inches wide by 2.25 inches high. On this panel, it was possible to present a linear array of 32 characters, each in a 5 x 7 dot matrix format with characters 0.20" high. Nominal light output per dot was 25 foot-Lamberts.

The final prototype, depicted in Figure 2-5, utilized a special NIMO tube, a variety of miniature Charactron CRT produced by Industrial Electronic Engineers, Inc. The tube contained an array of 64 cathodes and 64 areas of metal stencil, so arranged that any cathode and its associated stencil mask area could produce a character or a message on a 3/4 inch square area on the end of the tube. No deflection circuitry was employed; the position at which information appeared on the tube face was entirely a function of the geometry of the cathodes and their associated stencil mask areas. For the TSC application, a special mask was employed which made it possible to present messages to a maximum of six characters on each of three lines. Additionally, certain mask positions were reserved for individual digits at specific locations, so that by time-sharing cathodes and mask positions at a flicker-free rate, it was possible to display messages along with any required numerical values.

In addition to the display proper, each of the above units provided pushbuttons to permit "Wilco" and "Unable" responses, to the previous message by the pilot. On each display, the appearance of a new message was accompanied by an audio alert, consisting of a "beep" repeated three times per second at 50% duty cycle until a response was made.

Each of the four displays along with its required drive circuitry was packaged in the same standard-size chassis to permit the displays to be installed and interchanged in a common location on the panel of the GAT-1 simulator. Panel installation was at pilot eye level and approximately one foot to the left of his centerline. Figure 2-6 is a block diagram of the experimental setup. Indicated here is the teletype/paper tape input for messages, and the readouts to permit the experimenters to monitor flight parameters and pilot performance.

Eight FAA NAFEC test pilots each made a total of eight simulated flights to evaluate each of the displays under simulated daylight and night conditions. Four different scenarios, each involving

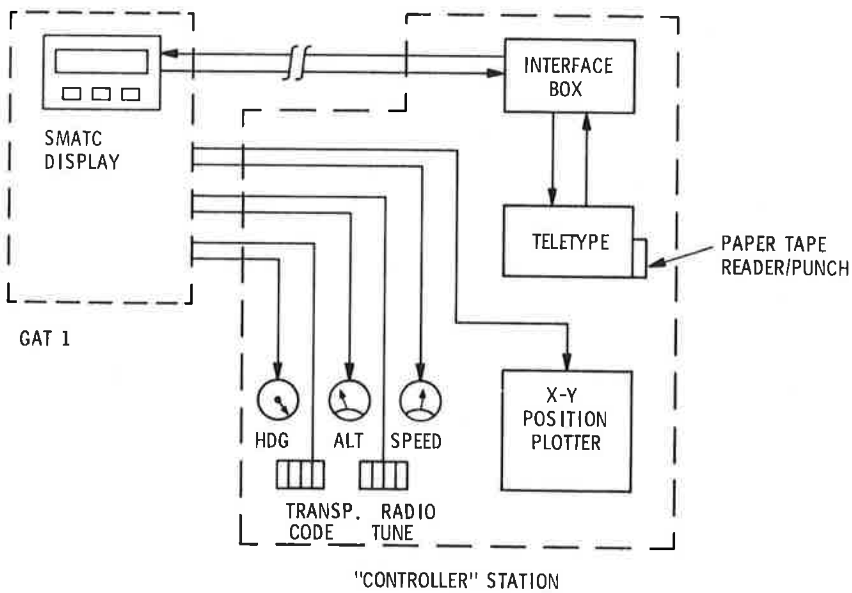

Figure 2-6. Block Diagram of Experimental Setup

simulated flights around the Boston area, were employed in a counter-balanced experimental design.

Data as to pilot response time to each message were recorded during the runs.* Table 2-1 shows the mean response time to each each of the displays under simulated day and night conditions.

Here, it should be noted (1) that response time to the 7-Window display was slowest because of the requirement for scrolling this display to view the separate portions of those messages which required more than seven characters, (2) that response time to the 32-Window display was appreciably slower under simulated daylight conditions than under night conditions, confirming that the readability of this display in the daylight is marginal, and (3) that

*While response time is not the only indicator of the "success" of a display, it is easily measured, and certainly gives some indication as to how readable and understandable a message may be. It also provides useful data for the design of an associated ground system.

TABLE 2-1. MEAN REACTION TIMES (IN SECONDS) TO THE DISPLAYS

DISPLAY	DAY	NIGHT	MEAN
NIMO	3.31	3.36	3.33
7-Window	6.04	6.10	6.07
3x7-Window	4.71	4.66	4.69
32-Window	4.67	4.12	4.39
Mean	4.68	4.56	4.62

response times to the NIMO were the fastest of the four displays. Part of this difference is attributable to the fact that only a single message could be presented at any one time on the NIMO, whereas multiple commands such as heading, altitude and speed could be presented on the other displays. When reaction time obtained with each of the displays was plotted as a function of the number of "information units" present (an "information unit" in this case being defined as a discrete message), it was first found that reaction time to two-unit messages was somewhat greater than for three-unit messages for the 3x7-Window and 32-Window displays. However, it was noted that this apparent discrepancy resulted from the extremely long reaction times to two-unit messages involving radio frequency plus transponder code settings. Here, because of the number of digits involved, pilots were reluctant to "Wilco" a message prior to making the appropriate settings because of the possibility of losing this information should a new message arrive. With these two-unit messages eliminated, there was a linear and only very modest increase in reaction time as the number of information units increased with 3x7-Window and 32-Window displays, as indicated in Figure 2-7. However, it should also be noted that with the requirement for manual scrolling of the 7-Window display, response times increased markedly for two-unit messages and would probably not be acceptable for three-unit messages.

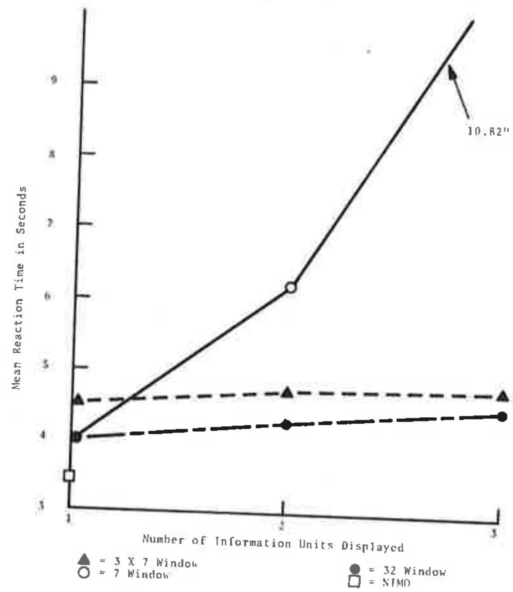

Figure 2-7. Reaction Times as a Function of Number of Information Units in Messages. 2-Unit Messages of Radio Frequency and Transponder Code Settings Are Omitted.

Pilots were informed prior to the start of the experimental runs that occasional impossible commands would be issued so that they would not routinely press the "Wilco" buttons without first interpreting the meaning of a message. Typical impossible messages asked for a speed of 900 knots or a heading of 540 degrees. A total of 44 such messages were introduced into the 64 experimental runs. Pilots responded with an immediate "Unable" on thirty of these, responded with "Wilco" immediately followed by "Unable" on five, and failed to detect erroneous "Wilco" responses to nine of the impossible messages.

Only two of the eight pilots failed to make any errors. Mean response time for the messages correctly detected as erroneous was 5.71 seconds as compared with a mean value of 4.62 seconds for all messages.

At the completion of the experimental runs, a 28-item questionnaire was administered to each of the pilots in an effort to elicit additional information. There was general agreement on only a few of the items. Pilot opinion of Data Link as a concept was generally favorable and a majority felt that it would reduce pilot workload. Pilot comments included: "It's great to get rid of that incessant chatter." The scratchpad capability of the Data Link also drew favorable comments. There was, on the other hand, complete lack of agreement as to the preferred color for the Data Link displays, with comments ranging from "red" to "anything but red." When asked to rank-order the four displays, and with weighting of these rank orders, the 32-Window display proved an overwhelming favorite, the 3x7-Window and NIMO displays appeared usable and acceptable, and the 7-Window display (in its current version) appeared unacceptable.

The encouraging results from this series of simulated flights prompted further exploration of Data Link using two-man crews in the GAT-2* simulator at FAA/NAFEC (4-1; 4-2; 4-3). Planning for these newer experiments included certain modifications of the displays used in the GAT-1 tests. An automatic scrolling feature was incorporated into the 7-Window display so that up to three lines of seven characters each could be presented sequentially without a requirement for pilot intervention. A storage register was built into the 3x7-Window display to provide scratchpad callup of heading, altitude and speed commands on demand. Dimming controls were added to all of the displays. Twelve character spaces on the right hand side of the 32-Window display were reserved for the continuous display of heading, altitude and speed commands, thereby limiting other messages to a maximum of 20 characters. Also, a "Wilco" button was provided on each simulator control column as well as on the display itself.

As in the previous experiment, messages were generated from punched tape through a TSC-constructed interface box which provided the required decoding, storage and control functions.

*The GAT-2 simulates a light, twin-engine aircraft with a two man crew and space for an on-board observer.

Eight 2-man crews of FAA/NAFEC test pilots each made four simulated flights, two on each of two scenarios, one involving a flight from JFK to Atlantic City, and the other from Philadelphia to LaGuardia (4-1). The experimental design was replicated for eight additional crews consisting of airline and ALPA Pilots (4-2). A single crew of AOPA pilots also participated (4-3). The greater complexity of the experiment over that of the earlier experiment made it possible to make a much greater number of comparistons between and among data points than had been possible previously. In addition to differences among crews, performance differences within crews were also measurable. Performance on each of the displays could be measured, differences between "Wilco" responses made on the control column versus those made directly on the display panel, day versus night differences and of first versus second runs on a scenario. Response times as a function of message type and length were, of course, measured, as well as responses to "Unable" messages. Comparisons among groups (test pilots versus airline pilots versus AOPA pilots), "old" (test pilots who had previously participated in the tests on the GAT-1) versus "new" (test pilots with no previous Data Link experience), and between the results of the tests conducted on the GAT-1 and GAT-2 were also possible.

Some explanation as to the reasons for the separation of the data from the pilot groups may be in order. The FAA/NAFEC pilots represented a more homogeneous population than the Airline/ALPA pilots. The age range for the NAFEC pilots was 45 to 53 as compared with an age range of 30 to 56 for the airline pilots. Flying experience of the NAFEC pilots varied from 4,200 to 20,000 hours, while the experience of the airline pilots ranged from 800 to 25,000 hours. On the average, the NAFEC pilots had some 2,200 extra hours of experience over that of the airline pilots. The fact that the NAFEC pilots were test pilots by profession and thus experienced in the evaluation of new equipment in a variety of aircraft provided further justification for the separation of the data from the two groups.

Due to an oversight in data collection, no record was kept as to which crew member made the responses on the display panels; only those responses made on the control column "Wilco" buttons can be attributed to a specific crew member. This defect will be rectified in future experiments.

For those data in which the crew member making the response can be identified, it can be seen that different crews handled Data Link in different manners. Under usual commercial procedures, the co-pilot is responsible for the majority of communications transactions, while the pilot handles control of the aircraft. During the experiment, the crews were given no instructions as to which crew member had responsibility for Data Link, and different crews interpreted Data Link on their own as either a control or communication function, even though they alternated as pilot and co-pilot on successive runs. A majority favored the co-pilot communication function. Despite the alternation of cockpit seating, crew members were remarkably consistent in having shorter response times while serving as co-pilot. For the NAFEC crews, mean response time for pilots was 7.2" as compared with 5.5" for co-pilots. For the airline crews, the times were 6.4" and 5.5" respectively. For both groups, responses made on the display panel were measurably longer, with means of 7.8" and 6.7" respectively. Some definition of Data Link function is indicated in the instructions for crews in future experiments.

Mean reaction times of the NAFEC and airline crews for the four displays are presented in Table 2.2. For both groups, mean response times were fastest for the 3x7-Window and NIMO displays, somewhat slower with the 32-Window display and slowest for the 7-Window display. However, with each of these pilot groups, there was wide variability in the response of individual crews to the four displays. The ratios of the reaction time of each crew on each display to the mean reaction times of all crews to that particular display are presented in Tables 2.3 and 2.4 The more disparate performances are underlined. Among the test pilots, at least one crew performed poorly on each of the displays, and

TABLE 2-2. MEAN REACTION TIMES OF TEST PILOTS AND AIRLINE CREWS TO THE FOUR DISPLAYS (IN SECONDS)

DISPLAY	TEST PILOTS	AIRLINE CREWS
NIMO	5.9	5.4
7-W	8.5	6.9
3x7-W	5.9	5.1
32-W	6.4	6.8
Mean	6.7	6.1

TABLE 2-3. RATIO OF THE MEAN RESPONSE TIME OF EACH NAFEC CREW ON EACH DISPLAY TO THE MEAN RESPONSE TIME OF ALL NAFEC CREWS ON THAT PARTICULAR DISPLAY

DISPLAY	CREW								
	A	B	C	D	E	F	G	H	RANGE
NIMO	.71	1.08	1.00	1.02	.73	.81	1.46	1.14	0.75
7-W	.98	1.46	1.26	.69	.94	.84	.69	1.18	.71
3x7-W	1.07	.86	1.17	.63	.76	.75	1.58	1.14	.95
32-W	.94	1.40	.84	.75	1.06	.97	1.11	.91	.65

TABLE 2-4. RATIO OF THE MEAN RESPONSE TIME OF EACH AIRLINE CREW ON EACH DISPLAY TO THE MEAN RESPONSE TIME OF ALL AIRLINE CREWS ON THAT PARTICULAR DISPLAY

DISPLAY	CREW								
	I	J	L	L	M	N	O	P	RANGE
NIMO	1.07	1.00	.89	.96	1.15	1.06	.96	.89	0.26
7-W	1.13	1.14	.81	.68	1.65	1.14	.75	.71	.97
3.7-W	.90	1.06	1.16	.80	1.12	.94	1.12	.92	.36
32-W	.91	1.04	1.03	.50	2.00	1.18	.59	.72	1.50

two crews (B and G) performed poorly on two of the displays (but different displays). Only one crew performed better than average on all four of the displays. Among the airline pilots, there was much greater consistency in performance on the NIMO and 3x7-Window displays. One crew performed poorly on the 7-Window display, and that same crew (M) performed poorly on the 32-Window display, with a response time twice as great as the mean. On the other hand, crew "L" had a response time on the 32-Window display which was only half of the mean value for all crews. The above results indicate the need to use several crews in the evaluation of any given display if meaningful data are-to-be obtained. The use of fewer than eight crews does not seem to be advisable in future experiments, since some of the conditions tested proved to be excessively difficult or easy to only one out of the eight crews who participated.

If variability among crews on their performance with a given display is to be minimized, it is important that the scenarios be equated for difficulty insofar as possible, since with an incomplete block experimental design as in the present experiments, all crews did not evaluate each display with each scenario. Mean reaction times for the NAFEC crews on Scenarios "A" and "B" were 6.8" and 6.4" respectively; for the airline crews 6.1" and 6.1", indicating that there was little difference in difficulty between the two scenarios. There was, however, a noticeable practice effect between first and second runs on a scenario as indicated in Table 2.5

TABLE 2-5. MEAN REACTION TIME ON CREWS ON 1ST AND 2ND RUNS ON EACH SCENARIO (IN SECONDS)

CREWS	SCENARIO "A"		SCENARIO "B"	
	1st Run	2nd Run	1st Run	2nd Run
NAFEC	6.9	6.6	7.3	5.4
Airline	6.9	5.1	7.0	5.2

These data indicate that a given scenario should definitely not be run more than twice by any given crew, and that preferably it should be used only once.

For the "daylight" condition in the tests run on the GAT-1, a high intensity light source was used which produced an uncomfortable glare. Limitations of the layout in the GAT-2 area did not permit similar high intensity lighting, and the "daylight" condition instead approximated that of a well lit office. While differences in performance were found between the "night" and "day" conditions on the tests on the GAT-2, these differences were inconsistent, as indicated in Table 2-6.

TABLE 2-6. REACTION TIME DIFFERENCES BETWEEN DAY AND NIGHT CONDITIONS

DISPLAY	NAFEC CREWS			AIRLINE CREWS		
	DAY	NIGHT	MEAN	DAY	NIGHT	MEAN
NIMO	5.9	5.7	5.8	5.3	5.6	5.4
7-W	8.2	9.2	8.7	7.8	6.4	7.0
3x7-W	4.9	6.8	5.8	4.7	5.5	5.1
32-W	6.4	6.9	6.7	7.8	5.7	6.7
Mean	6.4	6.9	6.7	6.4	5.8	6.1

(IN SECONDS)

As with the GAT-1 tests, certain of the messages presented on each of the displays, excepting the NIMO, contained more than one unit of information. The use of such multiple units of information for Data Link could increase its efficiency since the long string of characters required to establish sync., aircraft ID, parity, etc. would require transmission only once while the message *per se* could provide multiple units of information. Such an increase in transmission efficiency must, however, be predicated on the ability of crews to assimilate such multiple messages readily and without large increases in response times. Table 2.7 indicates that response time is increased only modestly when two units of information are presented in a single message.

TABLE 2-7. RESPONSE TIMES AS A FUNCTION OF NUMBER OF INFORMATION UNITS IN MESSAGES (IN SECONDS)

DISPLAY	NAFEC CREWS		AIRLINE CREWS	
	# OF INFO. UNITS 1	2	# OF INFO UNITS 1	2
NIMO	5.2	---	5.0	---
7-W	9.1	11.8	6.3	5.9
3x7-W	5.5	5.5	4.9	5.2
32-W	5.9	7.4	4.8	6.3
Mean	6.3	8.3	5.3	5.9

For the tests using the GAT-2, the differences in response times to two-unit messages containing combined radio frequency and transponder code settings and other two-unit messages were trivial, probably because the two crew members could share the load of tuning the radio and transponder.

Again, as in the tests on the GAT-1, responses to "Unable" messages were longer than for "Wilco." For the NAFEC crews, mean "Unable" response time was 7.7 seconds as compared with 6.3 seconds for "Wilco." For the airline crews, these figures were 10.0 seconds and 6.1 seconds respectively.

The two pilot groups handled Data Link responses in a somewhat different manner. With the test pilot group, a majority of the responses were made by the co-pilot using his control column response button; with the airline pilots, the largest number of responses were made directly on the display panel. Despite this, and the different age and experience makeup of the two groups, they were remarkably consistent in their relative ranking under the several conditions studied. Table 2-8 compares the two groups for a number of the parameters which were studied.

A single pilot was used for the tests run on the GAT-1, and somewhat different results might therefore be anticipated from those obtained on the GAT-2 (4-1). Responses were appreciably faster on the GAT-1 tests, probably because the single crew member felt no requirement for waiting until he was certain that the other crew

TABLE 2-8. SUMMARY OF RESPONSE DIFFERENCES BETWEEN TEST PILOTS AND AIRLINE PILOTS

(IN SECONDS)

		Test Pilots	Airline Pilots
% Responses:			
By Pilot		15.4	15.7
By Co-pilot		51.1	31.6
On Display		33.5	52.7
Response Times:			
Mean		6.7	6.1
NIMO		5.9	5.4
3x7-W		5.9	5.1
32-W		6.4	6.8
7-W		8.5	6.9
By Pilot		7.2	6.4
By Co-pilot		5.5	5.5
On Display Panel		7.8	6.7
Scenario A:	1st Run	6.9	6.4
	2nd Run	6.6	5.1
Scenario B:	1st Run	7.3	7.0
	2nd Run	5.4	5.2
Daylight		6.4	6.4
Night		6.9	5.8
1 Info. Unit Messages		6.3	5.3
2 Info. Unit Messages		8.3	5.9

member has also absorbed the information before acknowledging and thereby running the possibility of losing that information. The differences in relative response times to the four displays on the two series of tests probably resulted from (1) the newly installed dimming capability on the 3x7-Window display, which made it possible to avoid some of the glare present during the tests run on the GAT-1, (2) because 12 characters on the right hand side of the 32-Window display were reserved for the continuous display of heading, altitude and speed, there may have been a tendency for certain crews to scan the entire length of the display before making a response, even though the heading, altitude and speed information on the right hand side of the display frequently duplicated the information presented on the left, and (3) the location of the displays midway between the crew members in the GAT-2 made it difficult to read the small characters on the NIMO, whereas no similar difficulty was present with the larger characters of the 7-Window and 3x7-Window displays. The data from the tests on the GAT-1 and GAT-2 are compared in Table 2-9.

TABLE 2-9. MEAN RESPONSE TIMES (IN SECONDS) FOR TESTS ON THE GAT-1 AND GAT-2

Response Times	GAT-1	GAT-2
Mean	4.6	6.4
NIMO	3.3	5.7
3x7-Window	4.7	5.5
32-Window	4.4	6.6
7-Window	6.1	7.7
Day	4.7	6.4
Night	4.6	6.4

Eight of the test pilots who served as subjects in the GAT-2 experiment had previous familiarity with the Data Link equipment. This did not, however, result in performance faster than that of the "new" test pilot group. Mean reaction time for the "old" pilots was 6.8 seconds as compared with 6.6 seconds for the "new" pilots.

The data obtained from the single crew of AOPA pilots do not permit broad generalization since no counterbalancing of the sequence in which displays were evaluated, scenarios, and day versus night conditions was possible. As might be expected from pilots used to flying their own aircraft and frequently alone, the co-pilot functioned mainly in the role of observer on all four simulated flights made by this crew. The crew member serving as co-pilot made no responses using his control column. Two responses were made directly on the display panel and the remaining 157 responses by the pilot on his control column.

Mean response time for the two AOPA pilots was 6.1 seconds, comparable with that of the airline pilots. The rank ordering of the response times to the several displays is comparable with that of the other pilot groups; these data however could be subject to modification if a larger number of AOPA crews were used in a counterbalanced experimental design such as that employed for the NAFEC and Airline Pilots.

The larger number of pilots available to answer a questionnaire after the tests on the GAT-2 as compared with those on the GAT-1 made possible an even greater diversity of opinions. Both groups were highly but not universally favorable to Data Link, with the test pilots appearing to be slightly more enthusiastic than the airline pilots. Thirteen out of the 16 test pilots and only 9 out of the 16 airline pilots thought that Data Link would reduce pilot work load. Airline pilots expressed concern that a visual communications system would interfere with other visual tasks, and that information presently heard during ATC communications transactions with other aircraft would be lost.

Scratchpad was a highly popular feature, although the airline pilots preferred the recall feature of the 3x7-Window display to the continuous scratchpad capability of the 32-Window display by 9 to 6 whereas the test pilots preferred the 32-Window scratchpad by 9 to 4.

White was the favorite display color, preferred by 7 test pilots and 5 airline pilots. Red, though it had some adherents, aroused the most opposition.

The limited data from the two AOPA pilots did not produce opinions which differed from those expressed by one or more members of the test pilot and airline pilot groups.

The rank order of preferences for the several displays by the several pilot groups are presented in Table 2-10.

TABLE 2-10. NUMBER OF PILOTS EXPRESSING THEIR RANK ORDER OF PREFERENCE FOR EACH DISPLAY

Display & Pilot Group	# of Pilots ranking display as:			
	Best	2nd	3rd	Worst
Test Pilots:				
NIMO	0	2	3 1/2	10 1/2
7-Window	3	3	6 1/2	3 1/2
3x7-Window	2	9	5	0
32-Window	11	2	1	2
Airline Pilots:				
NIMO	0	0	2	14
7-Window	0	3	11	2
3x7-Window	6	7	3	0
32-Window	10	6	0	0
AOPA Pilots:				
NIMO	0	0	1	1
7-Window	0	0	1	1
3x7-Window	1	1	0	0
32-Window	1	1	0	0

3. LABORATORY TESTS OF MESSAGE CODING AND FORMATS

The discussion of Section 2. of this report has concentrated on how to display Data Link information. Of equal importance to the Data Link project is the determination of what information requires display and what its format should be to make it as meaningful as possible. While study of this aspect of the problem is possible in a flight simulator, this is inefficient since a realistic simulation can provide data points only every 60-90 seconds. In the laboratory, on the other hand, it is possible to accumulate data points every few seconds if proper procedures are employed, and investigation of the "what" aspects of Data Link has accordingly concentrated in this direction.

Slides flashed onto a projection screen, with a requirement for the viewer to make an appropriate response, have provided a means for TSC Human Factors Laboratory personnel to investigate schemes for message coding and formatting. The shortage of prime panel space on the flight deck of commercial airliners makes it imperative that messages be as brief as possible while still remaining meaningful. Brevity may be accomplished in two ways; either by using short abbreviations or by avoiding spaces between words or abbreviations.

In one experiment to evaluate these variables, slides were prepared containing 25 typical short ATC messages in four different forms: (1) short abbreviations without spaces, (2) short abbreviations with spaces, (3) longer abbreviations without spaces, and (4) longer abbreviations with spaces. Thus, a message such as "Resume Speed" could be presented as "RESS", "RES S", "RESSPD" and "RES SPD". The slide sequence was randomized and presented individually to 12 TSC engineers, all of whom had some knowledge of air traffic control technology and had been given the opportunity to memorize the abbreviations in their various forms (2-1). Subjects were told to depress a response button as soon as they recognized the meaning of a message. Depression of the response button blanked the screen to prevent further reference to the stimulus material and the subjects were then asked to verbalize the meaning of the message. The same slides were later presented to the eight FAA/NAFEC Test Pilots who participated in the GAT-1 simulator tests previously reported (3-3).

Table 3-1 presents the mean response time in seconds and the total errors for the two subject groups for the several conditions. An error was recorded either when the subject failed to respond or failed to respond correctly.

TABLE 3-1. REACTION TIMES (IN SECONDS) AND ERRORS OF TWO SUBJECT GROUPS TO FOUR DIFFERENT MESSAGE FORMATS

Stimulus Material	TSC Engineers		Test Pilots	
	Time	Errors	Time	Errors
Short abbrev. No spaces	2.08	19	2.68	16
Short abbrev. With spaces	1.61	9	1.98	6
Long abbrev. No spaces	2.20	22	2.33	5
Long abbrev. With spaces	1.58	13	1.82	2

With both subject groups, reaction time was reduced by approximately one-half second when spaces were used, and the error rate was approximately one-half. The use of longer and supposedly more meaningful abbreviations, on the other hand, made little difference. Statistically, the differences resulting from the use of spaces were significant respectively at the .999 and .95 levels for the engineers and test pilots, while differences as a function of abbreviation length were non-significant.

Figure 3-1 represents the mean reaction times for the four message types as a function of message length, "message length" in this case being defined as the number of characters in the shortest version of a particular message. Here, it should be noted that with short abbreviations with spaces, reaction time remains relatively constant regardless of message length. On the other hand, long messages, particularly those using long abbreviations without spaces, can be comprehended only slowly and with difficulty. The

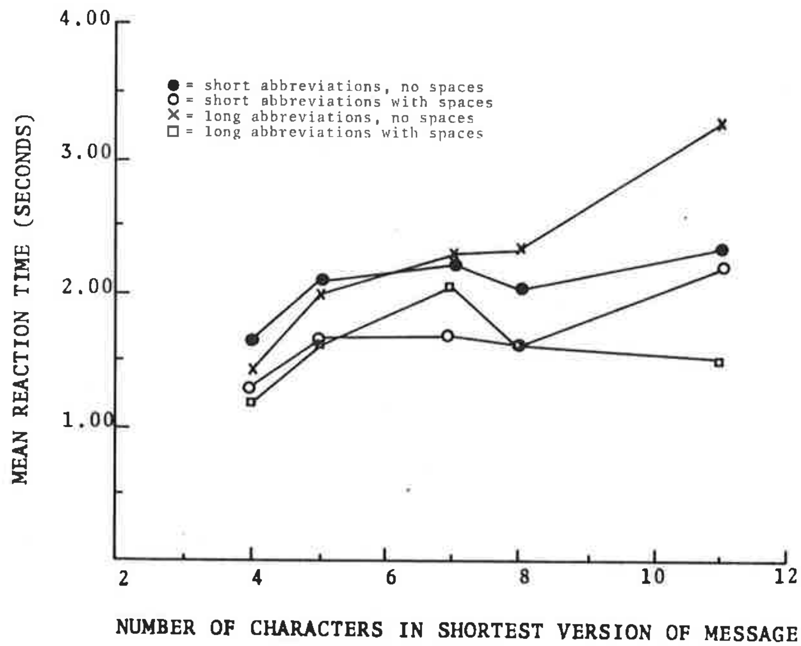

Figure 3-1. Mean Reaction Time as a Function of Message Length

message appears as so much "alphabet soup." The data depicted are for the TSC engineers; data for the test pilots were comparable.

Since all short or long abbreviations were not equally short or long, reaction times were plotted as a function of the number of "information units" in a message ("information units" in this context meaning an abbreviation for a word, or a group of digits), and are presented in Figure 3-2 for the TSC engineers. Again, it should be noted that increases in message length produce only a modest increase in reaction time when spaces are used.

The relatively low error rate achieved by both subject groups in these experiments makes it impossible to plot smooth curves for error rate versus message length for the separate experimental conditions. Figure 3-3 depicts error rate as a function of message information units for the lumped experimental conditions. Since error rate increases sharply as the amount of information increases, messages should be as short as possible if correct interpretation is to be achieved with a high degree of regularity.

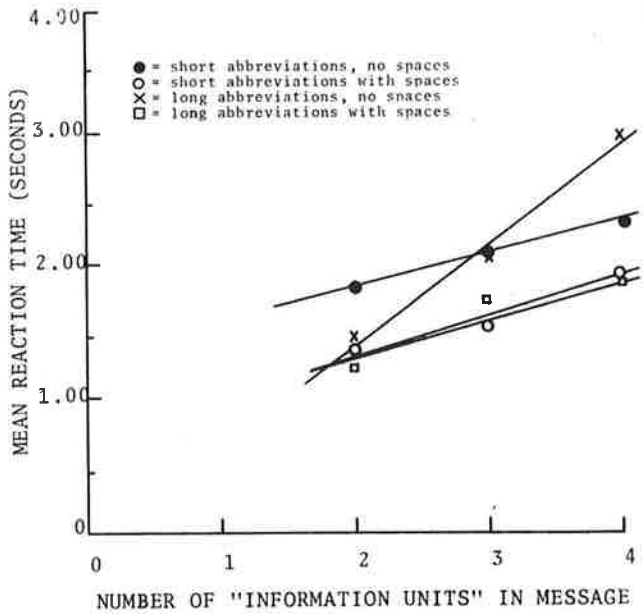

Figure 3-2. Mean Reaction Time as a Function of Number of Information Units in a Message

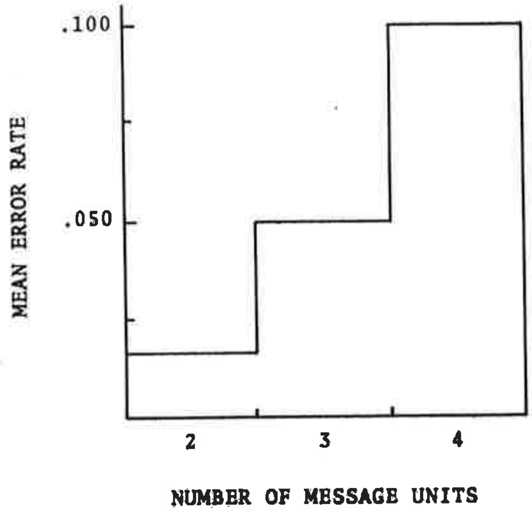

Figure 3-3. Mean Error Rate as a Function of Number of Message Units

For a second and somewhat more ambitious series of laboratory experiments, 144 slides were prepared containing ATC commands and advisories of several types using three different type fonts and a variety of coding and formatting schemes. As the slides were presented in random sequence, experimental subjects were required to select the appropriate responses from among multiple choices according to the type of information being presented. For this, they used the response box depicted in Figure 3-4. The four buttons in the center, arranged in a diamond-shaped pattern represented their "controls" for "fly left", "fly up","fly down", and "fly right". The button on the extreme left represented the "control" for radio frequency settings, and that on the extreme right their acknowledgement of receipt of a message which did not fit any of the other categories.

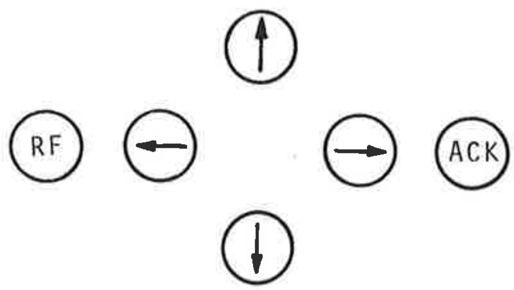

Figure 3-4. Response Box Layout

The legends on the response buttons of Figure 3-4 are to aid the reader in understanding these functions. Blank buttons were used during the experimental runs, and the subjects were given a cardboard facsimile indicating every possible symbol or abbreviation that might be used with each button.

Each message was reproduced with three different type fonts, simulating a 16-segment array, a 5x7 dot matrix array, and stencil-type characters similar to those produced by a Charactron CRT.

During a first running of the experiment, data were obtained from ten TSC engineers (1-3). A second run was made later using ten TSC engineers, with the procedural variation that when the subject responded, the information was automatically removed from the screen and he was then required to verbalize the meaning of the message (2-2). In a third replication, the eight NAFEC test pilots were run using the same latter procedure (3-2).

The results of the three experiments are summarized in Table 3-2. In general, the three type fonts were found to be equally readable. Three-line messages were easier to read than extended single-line messages. New information should preferably be presented at the top or left of a display. For an emergency situation, a single word or arrow should first be presented, such as "CLIMB" or "↑", followed later by the desired altitude. In the majority of cases, arrows and words were found to be equivalent, except in messages such as "HDGR160", where the "R" was difficult to detect. Here, the use of arrows drastically reduced reaction time. The general agreement of the results from engineers and test pilots validates the use of engineers for an experiment of this type.

TABLE 3-2. COMPARISON OF EXPERIMENTAL CONDITIONS. REACTION TIMES IN SECONDS. n.s. = a non-significant difference.

Variable	Typical Message	First Run Reaction Time	First Run Signif. of Dif.	Second Run Reaction Time	Second Run Signif. of Dif.	Third Run Reaction Time	Third Run Signif. of Dif.
Type Font Differences: Dot Matrix Segmented Stencil		1.24 1.31 1.21	n.s	1.33 1.38 1.27	n.s.	2.09 2.10 1.89	n.s.
Type Font Differences with Buried Arrows: Dot Matrix Segmented Stencil	A↓120	1.24 1.47 1.31	.001	1.56 1.55 1.41	.02	2.39 2.57 2.41	n.s.
Single Words Arrows	LEFT ←←	0.93 0.79	.02	0.96 0.84	.05	1.16 1.11	n.s.
Words + Numbers Arrows + Numbers	DOWN120 A↓120	1.06 1.00	n.s.	1.19 1.16	n.s.	2.05 2.19	n.s.
3-Line Messages using words 3-Line Messages using words	TURN RIGHT 180 vs TURN →→ 090	1.02 0.96	n.s.	1.15 1.07	n.s.	1.56 1.58	n.s.
Arrows Alone Arrows with Numbers	→→ →120	0.79 0.92	.02	0.84 1.11	.001	1.11 1.57	.001
Text with Numbers Text without Numbers	LEFT220 RIGHT	1.04 0.93	.05	1.17 0.96	.01	1.81 1.16	.001
Linear Messages 3-Line Messages	*HDG 070* ALT 190 SPD 210 (HDG 160 (*ALT 110* (SPD 165	1.64 1.33	.001	1.75 1.58	.01	2.73 2.54	n.s.
New Info. on Left New Info. "Buried"	*HDG 070* ALT 190 SPD 210 HDG 260 *ALT 140* SPD 220	1.62 1.73	.05	1.76 1.81	n.s.	2.54 2.88	.001
New Info.on top in 3-line format New Info. "buried" 3-line format	*HDG 180* ALT 160 SPD 200 vs HDG 090 *ALT 080* SPD 165	1.29 1.40	.05	1.45	n.s.	2.42	.001
"Buried" Arrows "Buried" "L" or "R"	HDG←080 HDGR080	1.02 1.67	.001	1.12 1.59	.001	1.72 2.77	.001

4. SYNTHETIC SPEECH

Synthetic speech generated within the cockpit in response to digital signals transmitted from the ground can provide transmission efficiencies comparable with that possible with digitally transmitted and generated visual displays. While the emphasis thus far on the Data Link program has been on the evaluation of visual displays, the possibility for a requirement for the use of synthetic speech has not been overlooked, since such synthetic speech would not add to the pilot's already busy visual scan and search workload. Voice synthesizers based upon both word and phoneme storage have been investigated.

At the beginning of the Data Link program, a voice synthesizer based upon word storage was procured from McDonnell-Douglas Electronics. Its original vocabulary of 75 words was shortly thereafter expanded to 128. During the first simulator runs (1-1), it was tested alone and in parallel with the WIDCOM for the generation of simple heading, altitude and speed commands. Four of the eight pilots tested felt that the intelligibility of the speech synthesizer was marginal, although all eight pilots also stated that the intelligibility improved with practice. Pilot response times on the trials in which the synthesizer was used in parallel with the WIDCOM visual display were appreciably longer than with the visual display alone, indicating that when information was supplied to both sensory modalities, pilots tended to confirm the information on one against the other before making a response.

Because of the marginal intelligibility of the voice synthesizer as received, intelligibility testing of the individual words was done in a speech laboratory. (1-2) Those words found to be deficient were reprogrammed, and later testing confirmed that improvements had been made.

The limitations inherent in a limited size, preprogrammed vocabulary have, however, limited further simulator testing of synthetic speech. In particular, the absence of a reasonable number of airport and navaid identifiers has made it impossible to duplicate with synthetic speech the variety of commands which can be generated very specifically on a visual display. Pilots on subsequent simulator runs (3-1), (4-1), (4-2), (4-3) have been given limited opportunities to listen to synthetic speech and questioned concerning its possible merits or deficiencies, but actual ATC commands via synthetic speech have not been employed during these simulator runs.

More recently, a new speech synthesizer based upon phoneme storage and thus not vocabulary-limited has been procured from the Vocal Interface Division of Federal Screw Works, and is scheduled for evaluation during the next series of simulator tests. However, it should be pointed out that, while intelligibility testing can be accomplished in a simple simulator such as the GAT-1 or GAT-2, the validation of pilot preference for a visual or an auditory display for Data Link can be accomplished only in a simulator having a sophisticated out-the-window display, or in the real world where the competition for the use of the crew's visual capacity is more realistic.

5. FUTURE PLANS

The simulator studies thus far have concentrated on short message ATC displays and limited downlink capability because (1) this represents a large proportion of ATC transactions, and (2) represents the messages requiring rapid crew responses. For the next series of tests on the GAT-2, a newly designed display capable of presenting two lines of eight LED characters each, and having storage capability for the recall of the latest heading, altitude and speed commands, and packaged in a 3-AT1 case, will be the primary source for visual short message ATC information. This will be supplemented by synthetic speech and by printers for longer ATC messages such as clearances and ATIS. Input devices having two different levels of complexity and capability will also be evaluated. A total of seven such hardware combinations will be compared with a control condition wherein all communication is verbal and follows present ATC procedures. A second similar experiment is scheduled later using airline simulators.

The experiments to date have met our objective of reducing the number of SMATC displays requiring future evaluation. The 7-window display has been abandoned because of the excessively long response times which occurred during its use, the NIMO because of poor character alignment, and the 32-window because of difficulty in reading the display under high ambient illumination levels. The deficiencies of the existing 3x7-window display are less apparent. However, message analysis* has indicated that a reduction from the 21 windows of the 3x7-window display to two lines of eight characters each yields no significant reduction in the number of message types which can be presented. Limiting the SMATC to two lines of eight characters, on the other hand, permats packaging in a 3AT-1 case and thus permits greater flexibility in the choice of location of installation on the panel of most simulators and aircraft. Such a prototype will be evaluated during the next series of simulator tests.

*Mpontsekaris, M. "Correlation of Short Message Display Capabilities with ATC Message Lengths," Interim Report #1, Report KHL-TSC-73-1064, Prepared by Kentron Hawaii, Ltd. for Data Services Division of the Transportation Systems Center.

6. SUMMARY OF RESULTS TO DATE

The tabulation below summarizes the results from the various experiments previously reported. The two-digit numbers within parentheses refer respectively to the number and subsection of the four Interim Reports wherein the data are presented to substantiate these findings.

(1-1) Evaluation of WIDCOM and Voice Synthesizer on GAT-1:

a) Synthetic speech could be generated via keyboard at sufficient speed to permit accurate control of a flight patch.

b) Intelligibility of the synthetic speech was marginal.

c) The WIDCOM visual display was highly acceptable to pilots.

d) Responses may be delayed when information is presented both visually and auditorially.

(1-2) Intelligibility Testing of Voice Synthesizer:

Of 128 words tested, 9 were completely unacceptable and 12 marginal

(1-3) Laboratory Experiment comparing Message Formats:

1) There were no differences in reaction time resulting from the use of different type fonts.

2) Arrows were better than words or abbreviations for simple IPC Commands.

3) For an emergency situation, arrows only should be presented, followed later by a numerical value if required.

4) Multiple information units are preferably presented on multiple short lines rather than one extended line.

5) New information should be located at the top or left of the display.

6) A command such as HDGLXXX (where X represents a digit) should be avoided and arrows substituted for the "R" or "L".

(2-1) Laboratory Experiment Involving Coding Schemes:

a) The shortest possible abbreviations are meaningful after only brief training.

b) Spaces between abbreviations are required for rapid and error-free interpretation.

c) Response times increase linearly as the number of information units in a message increases, but error rate increases exponentially.

(2-2) Laboratory Experiment involving further Comparison of Message Formats:

The data confirmed all findings of (1-3) on previous page.

(3-1) Evaluation of Four Prototype Visual Displays on GAT-1.

1) Response time was fastest to NIMO, followed by 32-W, 3x7-W and 7-W.

2) Response time to 32-W increased under daylight conditions.

3) There appeared to be a negative correlation between reaction times and flying experience.

4) Multiple unit messages (e.g., heading and altitude change) cause only a modest increase in reaction time except with 7-W display.

5) Reaction times were slower to "Unable" messages.

6) Pilots most preferred the 32-W, rated the 3x7-W and NIMO intermediate and approximately equal, and 7-W last.

7) All pilots favored the inclusion of some form of "scratchpad" capability on future experiments.

8) Pilots generally felt that Data Link would reduce pilot workload.

9) There was complete lack of agreement as to what color should be used for the display.

10) An audio alert is necessary with visual displays.

11) Pilots liked to have the new information set off within asterisks.

12) The quickness of response to a display does not always correlate with pilot opinion as to how good the display is.

13) Test pilots were generally favorable to the Data Link concept.

(3-2) Laboratory Experiments involving further Comparison of Message Formats:

a) The findings of (1-3), a, b, c, d, e and f were all confirmed.

b) The data obtained here from test pilots validates the data obtained from TSC engineers on previous experiments using this same stimulus material.

(3-3) Laboratory Experiments involving Further Tests of Coding Schemes:

1) All findings of (2-1) were confirmed.

2) The data from test pilots confirmed the data obtained from TSC engineers.

(4-1), (4-2), (4-3) GAT-2 Tests:

a) Most crews handled Data Link as a communication function, with the identifiable responses made mainly by the co-pilot (4-1), (4-2).

b) Responses by co-pilot were faster than those by pilot (4-1), (4-2).

c) Responses to 7-W display were slowest. (4-1), (4-2), (4-3)

d) Response times to NIMO and 3x7-W were fastest. (4-1), (4-2), (4-3)

e) Relative response times for different crews on different displays varied widely. (4-1), (4-2) If statistically significant data are to be obtained on future experiments, at least eight crews should be run on every experimental condition.

f) Response times were much faster with "Wilco" on control column than on the display panel. (4-1), (4-2)

g) Response times were approximately equal on the two scenarios. (4-1), (4-2).

h) There were definite practice effects on the second run of a scenario. (4-1) (4-2) On future experiments, no scenario should be used more than twice by a crew, and preferably, it should be used only once.

i) Response time differences to the different displays under daylight and night conditions were inconsistent. (4-1), (4-2) Some of this may have been due to inability to simulate true daylight conditions.

j) Increasing the number of information units in a message causes only a modest increase in response times. (4-1), (4-2)

k) "Unable" responses were slower than "Wilco". (4-2), (4-2)

l) Airline pilots responded slightly faster than test pilots. (4-1), (4-2)

m) Responses were faster on the GAT-1 tests than on the GAT-2 Tests. (3-1), (4-1), (4-2).

n) AOPA Pilots reacted similarly to airline pilots. (4-2), (4-3)

o) AOPA Pilots response times to the four displays were similar to the differences found with test pilots and airline pilots. (4-1), (4-2), (4-3). Generalization should be limited, however, because of the small number of pilots tested.

p) Both test pilots and airline pilots were generally (but not universally) favorable to Data Link. (4-1), (4-2).

q) A higher percentage of test pilots than airline pilots thought the Data Link would reduce workload. (4-1), (4-2)

r) The NIMO display was rated worst by both test pilots and airline pilots. (4-1), (4-2)

s) Scratchpad was highly popular with both test pilots and airline pilots. (4-1), (4-2), but they preferred it on different displays.

t) Test pilots were in favor of the concept of having Data Link set heading bugs and altitude alert, while airline pilots were against this. (4-1), (4-2).

u) There were strong differences in display color preference for both the test pilot and airline pilot groups. (4-1), (4-2)

v) The suspected negative correlation between reaction time and flying experience of (3-1) was not confirmed with the larger population of the later experiment. (4-1), (4-2).

7. REFERENCES

1. Hilborn, Edwin H., "Human Factors Experiments for Data Link: Interim Report #1," Report #FAA-RD-72-150. November, 1972.
2. Hilborn, Edwin H., and Devanna, Leonard R. "Human Factors Experiments for Data Link. Interim Report #2," Report #FAA-RD-73-55. April, 1973.
3. Hilborn, Edwin H., and Wisleder, Robert W. "Human Factors Experiments for Data Link. Interim Report #3," Report #FAA-RD-73-69. August, 1973.
4. Hilborn, Edwin H., "Human Factors Experiments for Data Link. Interim Report #4," Report #FAA-RD-74- 81. April, 1974.
5. Mpontsekaris, P., "Correlation of Short Message Display Capabilities with ATC Message Lengths. Interim Report #1," Report #KHL-TSC-73-1064. Prepared by Kentron Hawaii, Ltd. for Data Services Division of Transportation Systems Center.