

A SYSTEMATIC APPROACH TO ADVANCED COCKPIT WARNING SYSTEMS FOR
AIR TRANSPORT OPERATIONS: LINE PILOT PREFERENCES*

Douglas H. Williams and Carol A. Simpson
NASA Ames Research Center

SUMMARY

Fifty line pilots (captains, first officers, and flight engineers) from 8 different airlines were administered a structured questionnaire relating to future warning system design and solutions to current warning system problems. This was followed by a semantic differential to obtain a factor analysis of 18 different cockpit warning signals on scales such as informative/distracting, annoying/soothing. Half the pilots received a demonstration of the experimental text and voice synthesizer warning systems before answering the questionnaire and the semantic differential. A control group answered the questionnaire and the semantic differential first, thus providing a check for the stability of pilot preferences with and without actual exposure to experimental systems. It was hypothesized that preferences for warning method and cancellation method would vary as a function of warning urgency or priority and as a function of expected false-alarm rate. It was also thought that age and position flown might influence pilot preferences. There were no significant differences between the two groups for overall preferences for text and voice warnings compared to other warning methods, suggesting a high degree of stability and reliability of pilot preferences for warning methods. Warning urgency and expected false-alarm rate did produce significant differences in pilot preferences for some, but not all, warning methods. Warning urgency also produced significant differences in preferred cancellation methods for some warning methods. Generally, the preference data obtained revealed much consistency and strong agreement among line pilots¹ concerning advance cockpit warning system design.

INTRODUCTION

There seems to be substantial agreement among members of the aviation community that current cockpit warning systems for commercial jet transports suffer from a wide range of human factors design problems. In a recent study by the Boeing Company (ref. 1), funded by the Federal Aviation Administration, Vietengruber documented the warning systems in aircraft now current in the fleet. He found insufficient standardization of warning signals between

*This research was supported by NASA Ames Research Center Grant NGL-05-046-002, San Jose State University Foundation Account 02-01-3414.

¹The term "line pilots" refers to pilots who regularly fly commercial transport aircraft. It does not include airline check pilots, chief pilots, or instructor pilots.

aircraft types and even for the same aircraft type between airlines. He also noted a trend toward increasing numbers of different warning signals in both the visual and auditory modes. George Cooper (personal communication), under contract to the National Aeronautics and Space Administration, conducted structured interviews with foreign and domestic aircraft and avionics manufacturers in order to document current philosophies and to identify specific guidelines which might assist in improving warning system design. He found general agreement that cockpit warning systems are inadequate and may be adding to cockpit workload at times when this is already heavy due to additional demands on crew attention. While general agreement was found for many guidelines, the Cooper study also revealed some major points of disagreement among the parties interviewed regarding preferred methods of alerting, that is, tones, bells, voice, alphanumerics, labeled lights, and tactile warnings such as the stick shaker. However, a need for improved standards or guidelines was recognized.

There are a number of proposals that define general approaches to cockpit warnings and assign particular alerting methods to particular hazardous conditions. Aeronautical Radio, Inc. (ARINC) has a project paper (ref. 2) that outlines specifications for airborne audible warning generators. By assigning specific aural alerts to specific hazardous conditions, they have attempted to standardize the "meanings" that pilots would have to learn to associate with each of the different types of sounds. They also recommend a visual annunciator that would remain on until an existing fault is corrected. They provide for the possibility of voice warnings in place of or in addition to the nonspeech aural warnings. The Society of Automotive Engineers (SAE-7) Steering Committee on cockpit design is currently working on design standards for future warning systems. British Airways, in a paper for the International Air Transport Association (ref. 3), presented strong arguments against the use of nonverbal aural alerts, pointing out that such alerts are limited in the amount of information they can transmit and often are startling and distracting. They suggest audio alerts, preferably voice, for high priority, quick-action dangers and visual warnings, color coded for priority, for all priorities of warnings. They recommend that high priority voice warnings be noncancellable while providing for a cancel button for lesser priority voice warnings. Clearly, there is no industry-wide consensus regarding the types of alerting signals or the system logic that should be used for cockpit warning systems.

One type of data frequently overlooked is objective measurements of user preferences. All too often, experimental systems are designed and tested in the simulator first, with pilot debriefings afterwards. Perhaps this approach is popular because of a belief that there is little agreement among pilots concerning new cockpit displays and therefore little to be gained by asking them what they want in advance. This deprives an investigation of the vast resource of flying experience of the pilots who are destined to use and depend on the new system until *after* major commitments have been made to particular design elements or types of systems. This study had the dual purpose of sampling line pilot preferences for cockpit warning system design and also providing data that would be useful in guiding subsequent flight simulation research aimed at the determination of design principles for warning systems for air transport aircraft.

For this investigation, only U.S. Customary Units were used in the test booklets; therefore, the equivalent values in SI Units are given as follows:

3/4 in. = 1.9 cm

3/8 in. = 0.95 cm

HYPOTHESES

Some general hypotheses regarding pilot preferences for warning systems were derived from pilot debriefings in connection with previous research on Area Navigation (RNAV) CRT displays (ref. 4) and on voice warning system design (ref. 5). Additional input was derived from flight deck observations made during a series of flights as part of an ongoing collaborative research project with American Airlines Flight Training Academy on synthesized speech displays.

During discussions with pilots on the Ground Proximity Warning System (GPWS), it was observed that they seemed to have a generally negative attitude toward voice warnings and used the GPWS to illustrate their opinion. This observation was supported as well by various articles appearing in publications written for and read by airline pilots Connes, 1975, and Rawlings, 1976 (refs. 6 and 7). However, when pilots were asked if voice warnings with extremely low false-alarm rates would be acceptable, many responded that voices could be very useful for high priority warnings if one could depend on them to be accurate. This suggested that pilots would not be pro- or anti-voice warning per se, but would instead want voice only under conditions of high priority and low false-alarm rate. To test this, it was hypothesized that pilot preferences for warning method would depend both on the urgency or priority of the problem signaled by the warning and on the expected false-alarm rate.

Another point that pilots emphasized in the earlier discussions was the difficulty created by loud sounds and voices that continued during decision-making and intracrew communication. The engine fire bell was frequently given as an example of a signal that prevented or disrupted attention to the decision-making process and masked crew checklist callouts and other important auditory events. On the other hand, pilots gave two types of comments about cancellation of visual warnings. Some wanted to cancel all lights and other visual warnings as soon as they occurred to prevent distraction from other visual displays. Others wanted visual signals to remain as long as the hazardous conditions remained. From these observations came the hypothesis that preferences for warning cancellation would depend on both warning urgency and on the warning method used (i.e., auditory, visual, tactile). In addition, it was hypothesized that a limited priority assignment scheme for visual signals would satisfy the majority of pilots.

It was also expected that age and position flown might have an effect on pilot preferences for warning system design. However, no specific predictions concerning these possible effects were formulated. To summarize, then, the hypotheses were that

1. Pilot preferences for warning method will depend on the urgency of the problem signaled and on the expected false-alarm rate.
2. Preferences for warning cancellation will depend on warning urgency and the warning method.
3. A limited priority assignment scheme for visual signals would satisfy most pilots.
4. Age of pilots and crew position flown might produce differences in preferences.

METHOD

Often, when a pilot preference survey is proposed, resistance is encountered because of supposed characteristics of pilot preferences. In fact, these characteristics, if they do exist, can be allowed for in the construction of the measuring instrument.

First, it is often stated that pilots have too many opinions to be adequately measured: "There are as many opinions as there are pilots." To solve this, subjects were offered reasonable alternatives to rate or rank or were offered forced choices among alternatives. In most cases, they were also offered spaces for free responses where they could write in their own opinions or suggest their own system, if they thought the ones offered were totally inadequate. If this alternative was used by a significant proportion of respondents, it would be evidence of an unmanageable diversity of opinion among pilots, inadequate test items, or both.

Further, it is often suggested that pilot opinions are too changeable because they consider whatever new system they saw last to be best. This problem was handled by splitting the sample group into two subgroups and then by making a vigorous effort to convince one of the subgroups of the usefulness and potential of two types of systems, CRT or voice, which would then be represented in the test measures. Any significant differences between the responses of the subgroups on these two systems would be evidence for changeability of the pilot preferences obtained.

Measuring Instrument

The measuring instrument was a 32-page booklet to be filled out by each test subject. It consisted of 2 pages of biographical information and 30 pages of free response, rating scale, preference grid, and ranking items. There was a second, optional test, the semantic differential, in a separate booklet. It consisted of judgments of 18 concepts on 17 polar opposite scales. It was only administered to subjects who finished the first booklet within the 3 hr allowed for the complete session. If administered, it took approximately 15 min to complete.

Subjects

Subjects were line captains (13), first officers (20), or flight engineers (14) currently flying transport aircraft or recently furloughed (3). The oldest was 60, the youngest 27, with a mean age of 41.3 years. Total time averaged 9300 hr, and 8 airlines were represented. (See table 1.) This sample was not randomly chosen, and so may not represent a true cross section.

Procedure

Subjects were obtained in sets of 1 to 9 persons. They were drawn from a pool of airline pilots based in the San Francisco Bay Area who had expressed interest in participating in research at Ames Research Center. They were paid for their participation. Each set was randomly assigned to treatments, except that the last set was picked to exactly complete the group sample size of 25.

Demonstration-first group — Sets of pilots who were assigned to the "demonstration-first" group were offered coffee and a short, purposely vague introduction² to the purpose of the study, and assurance that their name and airline could not be connected to their individual responses on the test items. They then participated in a 20-min experiment that exposed them to synthesized speech warning messages. Following that, they were shown a video tape of various possible CRT display warnings, and then color slides of this type of CRT system and another type of alphanumeric warning system in several simulators, aircraft, and artists' conceptions. Finally, they were given the two test booklets and allowed to work on them at their own pace, for a maximum of 3 hr. Most completed them in less time.

Questionnaire-first group — The "questionnaire-first" pilots were given the short introduction, coffee, anonymity assurances, and told they would have 2 1/2 hr to complete the test booklets, after which they were given the synthesized speech experiment and the video and slide demonstrations.

The experimenters were always present during administration of the test booklets to answer questions. Discussions were not allowed to become established. If opinions were offered spontaneously, subjects were politely encouraged to write them in the appropriate spaces on the booklets. Motivation was good, and subjects willingly worked on the booklets without complaint. Spontaneous comments offered at the end of the session were encouraging. All subjects were offered an opportunity to fill out a name and address sheet to receive a copy of the report on the study. More than two-thirds of them chose to do so; 41 of the 50 also chose to do the optional semantic differential. These are rough indications of good motivation and interest.

²The introduction was left vague to prevent biasing subjects. If any particular system was even mentioned in a positive or negative statement, it could have influenced their responses.

RESULTS AND DISCUSSION

In the following, whenever a result is presented or a conclusion offered, assume that the "demonstration-first" and the "questionnaire-first" groups were NOT significantly different, and the data were combined after the significance test had been run. The few instances when there was a significant difference are indicated. The term "pilots" is used to describe any of the subjects, whether they were captains, first officers, or flight engineers.

Throughout this paper, there are repeated references to priority (= urgency) levels, 1 through 5. These were adapted from a priority assignment scheme developed by the Boeing Company and were presented to subjects in the form shown in table II. Warning methods are repeatedly referred to, and these were initially presented to the subjects as shown in table III. Subsequently, they were referred to in an abbreviated form.

Warning Method Preferences

The preference grid shown in table IV was repeated four times, once for each of four false-alarm rates: 50 false:1 real alarm, 1 false:50 real alarms, 1 false:1000 real alarms, and 1 false:1,000,000 real alarms. Pilots were asked, for each of these rates, to place an X under the system or systems they would want for warnings of that urgency, given that the system false-alarm rate could be no better than stated. Results are shown in figure 1. There is much information in this figure, but it will reward close study. First consider one cell concerned with voice warnings for priority 1 problems (fig. 2). Note that with the high false-alarm rate, 50:1, few respondents will accept the voice warning, but as the false-alarm rate improves, more and more pilots are willing to accept this method of warning for priority 1 situations. Returning to figure 1, the larger histogram, note, in the row for voice warnings, that for the lower priority situations, the number of pilots desiring voice warnings declines, regardless of the false-alarm rate. In short, pilots do not want voice for "information only" or low priority warnings, and they do not want them if the false-alarm rate is high. But they are willing to accept voice warnings for very important, high priority warnings if the false-alarm rate is low.

Now consider the row "text message." Note that it is generally acceptable no matter what the false-alarm rate is, as shown by the evenness of the histograms within cells. But also note the slightly greater concentration of responses within the cell for urgency level 3. This would tend to indicate that text messages are seen as more valuable for moderate priority items.

Two further observations can be made from this figure. For the auditory warnings (top three rows), responses are concentrated in the higher priority columns, indicating that pilots want sounds only for important problems. A χ^2 test for the effect of false-alarm rate is also consistent ($\chi^2 = 22.6$, $df = 12$, $p < 0.05$).

Now consider visual displays. A relative insensitivity to false alarm rates and urgency level is shown. These responses suggest that visual warnings can be more easily ignored if they are a false alarm, and that they can be tailored to suit the urgency of the situation. χ^2 tests were performed on the data for each false-alarm rate to test the relation of warning method to urgency level. For all false-alarm rates, significance levels of $p < 0.05$ or better were obtained. Inspection of the tabled data showed that the auditory methods, particularly "other sound" and "voice," were preferred for the more urgent warnings provided the false-alarm rate was low. Visual methods, particularly "labeled light" were preferred for moderate and less urgent warnings for all false-alarm rates and were also preferred for level 1 and 2 urgency, "action now" warnings when the false-alarm rate was high.

When differences between the responses of the "questionnaire-first" and "demonstration-first" groups were tested on this item, none of the expected ones were found, despite this being one of the most likely sections of the questionnaire to show such differences.

Warning Cancellation

Part of matching a warning to a situation is providing a way to cancel the warning when it is no longer wanted. Responses were collected by means of a preference grid with cancellation options and warning methods, given different priorities. The results are shown in figure 3. For clarity, one cell is shown in figure 4, voice warnings for priorities 1 and 2. (For brevity on the questionnaire, priorities 1 and 2 were combined as were priorities 3 and 4.) For these, "cancel button" is the method most preferred, and this is a general finding.

Note the very small number of "noncancellable" responses, which indicates that the respondents *do* want to be able to cancel a voice warning. This is true for nearly all auditory warnings. Also, few subjects checked "don't use" for this warning method, for this priority, indicating that they *do* find voice warnings acceptable. Figure 3 also shows that "noncancellable" is very frequently checked for lights, text messages, and flags; therefore they should stay on until the problem is solved. When χ^2 tests were applied to test for a statistically significant relation between urgency and cancellation preferences, they were significantly related ($p < 0.001$) for auditory warning methods but not for visual methods ($p > 0.20$).

Finally, figure 3 shows that as priority decreases, there are many more "don't use" responses for audio displays, the trend being reversed for unlabeled lights and flags.

System Logic

Another question at issue in the design of warning systems is the assignment of priorities for the warnings and filtering or inhibiting them. This

problem was assessed in the context of a single-channel warning system that could not present multiple simultaneous warnings. Pilots were asked to rate the following systems:

A. An onboard computer decides priority, presents the most urgent warning until the condition is removed, then presents the next most urgent, etc.

B. A primary display presents the most urgent warnings; a subsidiary display presents any others which occur simultaneously.

C. Priorities are not assigned--warnings are presented on the primary display as they occur.

D. A warning is presented for 5 seconds on the primary display then replaced with another warning, until all warnings in the stack are exhausted; then the entire thing repeats until all conditions requiring warnings are removed. All warnings are also displayed on a subsidiary display.

E. An onboard computer analyzes the pattern of warnings, then presents the crew with the best course of action in command format.

Two primary display systems were to be considered -- an "alphanumeric display block of 3/4-in.-high letters"³ or a "synthesized voice display in earphones and speakers." Results are presented in figure 5. In all cases, the visual warning was slightly more desirable. System B is the clear favorite. The other systems have mean ratings of 3, "no preference" or worse, so the pilots seem to indicate that they do not like any of these other systems very much. Clearly, more thought must be given to priority assignment schemes; perhaps system B could be used as a starting point.

Text Displays With and Without Alerting Tone

The next item concerned the use of a flashing versus a nonflashing display and the use of an auditory alerting tone. A grid was presented showing different systems and urgency levels (table V). Subjects went through the grid twice, first making an X for any systems they would want for warnings of a given urgency level, and the second time making an A for any systems they would want for a given urgency level if a single audio alerting tone were presented at the same time. The results for a warning on the bottom line of a CRT are shown in figure 6. The "not flashing" version is most desired for low-urgency warnings, while the audio tone is not considered particularly helpful for any priority. When the same warning is flashing, however, the preferences move to priority 2 with audio alerting and to priority 3 without.

Next consider figure 7, "warning on whole CRT screen." When this is not flashing, the preferences center around priority 2; but when it is flashing, the preferences move to priority 1, the most urgent flight safety items. For these, the audio alerting tone also becomes more desirable.

³All material enclosed in quotation marks is a direct quotation from the questionnaire.

Figure 8 shows the results for a "single line of twelve 3/4-in.-high alphanumeric." Again, when not flashing it is viewed as a device for presenting normal information or less vital warnings, priority 3 or below; in this case, no audio alert is wanted. When flashing, these alphanumerics can be used for the highest priority items. In this case, the audio tone becomes acceptable as an alerting device. It should be noted that some of the demonstration given to the subjects consisted of illustrations of exactly these three systems. They were shown by video tape on a 12.7-cm (5-in.) by 14.0-cm (5.5-in.) CRT monitor and by color slides of the 12-alphanumeric block used in flight on the CV-990 at Ames Research Center. Surprisingly, no significant differences in responses were noted between the "demonstration-first" and "questionnaire-first" groups for these items.

Voice Warnings

Another group of questions concerned voice warning systems. A technologically possible future system was described in which each of the present aural warnings would be replaced by a voice warning with a different voice message for each different malfunction. The description is reproduced below.

It would be possible to replace all of the current aural warnings with voice warnings. Such a system would be able to have different voice warnings for unsafe conditions which are now signaled by the same aural warning, e.g. TAKE-OFF warning and CABIN PRESSURE warning. Such a system would also include a volume control to adjust to different listening conditions. The warnings would be presented to your headset at the same volume as your own adjusted volume level for ATC communications. As a back-up, the warnings would also be presented over a speaker with their volume automatically adjusted to be just sufficiently above the volume of the ambient cockpit noise so you would hear them clearly -- as the noise level changed, the volume of the warnings would be automatically adjusted up or down. A visual status display would also be included to display all unsafe conditions as long as they continued to exist. You would have the option of leaving the visual display on continuously or turning it on only when you wanted to look at it. There would be a cancel button for voice warnings.

In response to the question, "Would you want such a system in your cockpit?", far more pilots said they would want the proposed voice warning system. Eighty-two percent of the 50 pilots gave a "yes" or a "qualified yes" response while only 18 percent responded "no." This difference was highly significant as tested by the 50-Percent Probability test ($x = 9$, $n = 50$, $p < 0.002$).

The pilot responses to the proposed voice warning system were potentially among the most susceptible to possible influence from the demonstration of experimental synthesized speech and text displays. In Figure 9, the responses of the questionnaire-first group are shown on the left, and those of the demonstration-first group are shown on the right. The proportions of "yes" (including "qualified yes") to "no" responses were nearly identical for the two

groups. The only possible difference between the two groups might be in the proportion of "qualified yes" responses given. A smaller percentage of "yes" responses for the demonstration-first group were "qualified yes" responses. The pilots in the questionnaire-first group, however, were responding on the basis of no prior experience with the capabilities of electronic voice warning systems. In this context, it would be reasonable to find a higher degree of uncertainty concerning the characteristics of a proposed voice warning system. This, in turn, could have caused more of these pilots to give a "qualified yes" response compared to the demonstration-first group. However, this apparent difference between the two groups was not significant as shown by Fisher's test ($a = 8$, $b = 3$, $c = 12$, $d = 18$, $p > 0.05$).

The next question was designed to determine which features or components of the proposed voice warning system were responsible for producing "yes" or "qualified yes" responses to the system. Figure 10 shows the percentage of "essential" judgments received by each voice warning system component. These responses were given by the 41 pilots who had responded affirmatively to the proposed system. Each pilot placed a check beside each component he thought was essential to make the voice warning system acceptable. The differences in numbers of "essential" judgments for the different components were highly significant ($\chi^2 = 54$, $df = 6$, $p < 0.005$). Clearly, the two most essential components are the voice cancel button (65 percent) and the visual status display (63 percent). We interpret this to mean that pilots want voice warnings only if they can cancel them and only if they have a visual status display that will continue to make the warning information available. Each of the remaining components received some "essential" judgments, but in each case, from less than 50 percent of the pilots. These other components should certainly be regarded as desirable. In contrast, the voice cancel button and the visual status display have to be included in any voice warning system.

Figure 11 also supports the finding that pilots want to be able to cancel voice warnings. In this question, they were asked to choose among several types of voice warning repetition; 77 percent of the pilots wanted the warnings to repeat until they pressed a cancel button or the problem was corrected, whichever happened first. This compares to only 19 percent who wanted the warnings repeated a fixed number of times — once, twice, or three times — and a mere 4 percent who thought voice warnings should be noncancellable. These differences were highly significant by χ^2 tests ($\chi^2 = 44.8$, $df = 2$, $p < 0.001$).

The pilots as a group expressed no strong preference on the question concerning the effect of warning urgency on the type of voice warning repetition. Figure 12 shows that 39 percent of the pilots thought the type of repetition should depend on urgency level, while 61 percent thought urgency should not affect the repetition of voice warnings. This difference was not significant as tested by the 50-Percent Probability test ($x = 18$, $n = 46$, $p > 0.10$). (Four pilots did not respond to this question.) Those pilots who felt voice warning repetition should depend on warning urgency wanted more repetitions and/or more stringent cancellation conditions for high priority warnings than for lower priority warnings.

The pilots also gave their preferences for the uses of voice warnings (fig. 13). First, the number of pilots who responded affirmatively to one or more of the proposed uses for voice -- alerting, specific problem, immediate action, or "other" -- was compared to the number of pilots who responded "don't use" voice. This gave a ratio of 43 pilots responding affirmatively to only 3 pilots responding "don't use." This difference was highly significant as tested by the 50-Percent Probability test ($x = 3$, $n = 46$, $p < 0.002$).⁴ Note that the 6-percent "don't use" responses compare roughly with the responses to the proposed voice warning system where 18 percent of the 50 pilots responded "no." This can be taken as an internal crosscheck of the earlier finding that pilots generally are in favor of the concept of voice warnings.

A three-way comparison of pilot affirmative responses to the three proposed functions of voices -- alerting, tell specific problem, and tell immediate actions -- also resulted in significant differences in preferences ($\chi^2 = 6.75$, $df = 2$, $p < 0.05$). Mostly, the pilots wanted voice warnings to tell them the specific problem (78 percent). In addition, 64 percent wanted an alerting word such as "warning." Only 36 percent wanted to be told immediate action items. Under "other" uses (12 percent), suggestions were made by a few pilots that checklist items or immediate action items should be available on demand by voice or CRT display. These responses imply that a voice warning format consisting of an alerting word followed by a statement of the specific problem would be acceptable to most pilots.

Age and Position Flown

Neither age nor position flown resulted in significant differences for acceptability of the proposed voice warning system nor in the pilot ratings of the proposed text warning systems. Fisher's test for differences between younger (21 to 40 yr) and older (41 to 60 yr) pilots for the number of "yes" and "no" responses to the voice warning system yielded $a = 5$, $b = 3$, $c = 19$, $d = 23$, and $p > 0.05$. Similarly, a χ^2 test for position flown (captain, first officer, or flight engineer) by number of "yes" and "no" responses yielded $\chi^2 = 1.09$, $df = 2$, $p > 0.10$.

The sums of individual pilot ratings for the five proposed visual text systems were also compared for the same younger and older pilot groups using Wilcoxon's sum of ranks test ($n_A = 16$, $n_R = 16$, $n_B = 24$, $R = 305$, $z = 0.63$, $p > 0.10$). And, finally, a 3×2 comparison of position flown by low (7 to 13) versus high (14 to 19) sums of ratings for the proposed visual text systems resulted in no significant effect for position flown ($\chi^2 = 4.45$, $df = 2$, $p > 0.10$).

⁴Four of the 50 pilots did not respond to this item. Assuming they had checked "don't use" the ratio of 43:7 would still have been significant at the 0.002 level.

Semantic Differential

The analysis of pilot responses to the semantic differential also revealed strong agreement among pilots regarding the features of the different warning concepts. For those readers unfamiliar with this instrument, the semantic differential is a technique pioneered by Osgood in 1957 (ref. 8). It is generally useful for finding related concepts in a diverse collection. In use, each concept or item to be judged is placed at the top of a page which has a number of polar opposite scales. The concepts used here were all warning-related items (table VI). The scales are shown in table VII. Subjects were given a 19-page booklet, 1 page of instructions and 18 pages, each with one concept. They placed an X on each 7-point scale, closer to one of the polar opposite adjectives or the other, depending on the one they felt the concept was most closely related to. If it was unrelated, or related to both adjectives by the same amount, they placed an X in the middle space on the scale.

Data from the semantic differential are usually analyzed several ways. The analyses presented here involve mean responses of all pilots on each scale for each concept. Two-way comparisons between pairs of warning concepts are shown in figures 14(a) through (g). Figure 14(g), for example, shows that when "VASI lights" is compared to "Whoop, whoop, pull up, pull up," the lights are less startling, more informative, far more beautiful, more valuable, more passive, far more quiet, and far more soothing. The useful conclusion, then, is that to startle and annoy one would use "whoop, whoop, pull up." To present informative unobtrusively, one would use VASI lights.

Further use of the semantic differential for evaluation of experimental cockpit warning systems seems warranted. Factor analysis techniques are expected to extract groups of warnings that have similar values on the 17 polar opposite scales and to determine how factors such as evaluation, utility, and intensity characterize the different types of warnings. The aim is to standardize a set of semantic differential scales which could be used to characterize a new warning system in relation to existing systems merely by having pilots fly a simulation of the new system and then fill out a semantic differential booklet.

CONCLUDING REMARKS

Several conclusions can be drawn from the data analyzed and presented. It was shown that preferred warning methods depend on urgency or priority of the warned condition, and that false-alarm rate has a major impact on the preferred presentation mode. For example, these data would indicate that, if a system has an inherent high false-alarm rate, a visual warning method such as labeled lights or a CRT is preferable to any audio system. If a low false-alarm rate can be achieved, the audio systems, particularly voice, are preferable to visual systems for high priority warnings.

It can be inferred from these data that pilots would like a limited priority assignment scheme, that all warnings which are current should be displayed somewhere, and that the pilot should decide the course of action rather than

being told what to do. However, within these broad guidelines, much more work must be done to define the priority assignment schemes and to optimize the warning displays. Candidate schemes and displays must be thoroughly tested in simulations and in flight before they are recommended for airline use.

It has also been shown that preferred cancellation options depend on whether the warning is auditory, visual, or tactile, as well as on the priority or urgency of the warning. The data in figure 3 will allow a choice of the preferred cancellation option, given the priority and warning mode.

Finally, the results of this study show that a systematic, objective measurement of pilot preferences for warning system design reveals consistency and strong agreement among this sample of line pilots. While the user cannot entirely dictate the system design, especially in airline cockpits where regulatory and cost considerations are so important, it would seem useful to include input from experienced line pilots in the development of aircraft warning systems for civil transport aircraft. The subjects whose collective opinion is represented by the data presented here have much and varied experience flying in different environments and aircraft types, and this should be given due regard in the design of future warning systems.

REFERENCES

1. Vietengruber, James E.: Collation and Analysis of Alerting Systems Data, Midterm Review, FAA Contract DOT-FA73WA-3233 (MOD II), The Boeing Co., Seattle, Washington, June 1976.
2. Audible Warning System, ARINC Characteristic 577-1. Prepared by the Airlines Electronic Engineering Committee, Aeronautical Radio, Inc., 2551 Riva Road, Annapolis, Maryland, first published March 15, 1975, revised March 31, 1975.
3. Warning Systems. Part I, Voice Warnings; Part II, A General Consideration of Audible Warnings. Paper presented by British Airways at the International Air Transport Association, 20th Technical Conference, Istanbul, Turkey, Nov. 10-15, 1975.
4. Williams, Douglas H.: Comparison of Numeric Keyboard and CRT Line-Labeled Buttons for Information Access. NAECON '76 Record, May 1976, pp. 419-423.
5. Simpson, Carol A.: Effects of Linguistic Redundancy on Pilots' Comprehension of Synthesized Speech. 12th Annual Conference on Manual Control, NASA TM X-73,170, 1976.
6. Connes, Keith: The Coming of TWIT. Air Progress, Aug. 1975, p. 18.
7. Rawlings, Barney: SQUAWKS: Super-Quick Universal Aural-Warning Kibitz System. Airline Pilot, Aug. 1976, p. 28.
8. Osgood, Charles E.; Suci, George J.; and Tannenbaum, Percy H.: The Measurement of Meaning. Univ. of Illinois, Urbana, 1957.

TABLE I.- BACKGROUND AND EXPERIENCE OF 50 PILOTS

<u>POSITIONS FLOWN</u>		<u>AIRLINES</u>		<u>ROUTES FLOWN</u>	
CAPTAIN	13	AMERICAN	20	INTERNATIONAL	
FIRST OFFICER	20	PAN AM	8	DOMESTIC	
FLIGHT ENGINEER	14	UNITED	7	SHORT HAUL	
OTHER	3	WESTERN	5	CHARTER	
		FLYING TIGERS	4	FREIGHT	
		TWA	2	FERRY	
		HUGHES AIR	1		
		BIRDAIR	1		
		OTHER/FURLOUGH	2		

<u>AGE (yr)</u>			<u>TOTAL TIME (hr)</u>		
YOUNGEST	MEAN	OLDEST	LEAST	MEAN	MOST
27	41.3	60	500	9,300	30,000

TABLE II.- URGENCY SCALE USED IN QUESTIONNAIRE

LEVEL	TYPE OF PROBLEM
1	IMMEDIATE ACTION REQUIRED BY CREW TO SAVE AIRCRAFT
2	IMMEDIATE ACTION REQUIRED BY CREW AS SOON AS AIRCRAFT IS STABLE
3	ACTION REQUIRED AS SOON AS TIME AVAILABLE
4	ACTION REQUIRED LATER IN THE FLIGHT - FLIGHT PLANNING MAY BE AFFECTED
5	ABNORMAL EVENTS SIGNALLED FOR INFO ONLY; NO ACTION REQUIRED - MAY AFFECT FLIGHT PLANNING

TABLE III.- WARNING METHODS AS DEFINED IN QUESTIONNAIRE

TYPE	EXAMPLES
MUSICAL SOUNDS	TONES, CHIMES, CHORDS
OTHER NON-SPEECH SOUNDS	BELLS, CLACKERS, HORNS, BUZZERS
VOICE	ELECTRONIC SPEECH (LIKE HAL IN 2001 OR LIKE GPWS VOICE)
LIGHTS WITH LABELS	LIGHTS WITH PRINTED LABELS, ALL COLORS, STEADY OR FLASHING
LIGHTS WITH NO LABELS	PLAIN, UNLABELED LIGHTS, ALL COLORS, STEADY OR FLASHING
TEXT MESSAGES	ALPHANUMERIC WARNINGS ON A TV-LIKE SCREEN OR AN ALPHANUMERIC DISPLAY BLOCK
FLAGS	MECHANICAL FLAGS IN FLIGHT INSTRUMENTS, DISPLAY COVERS, DOLL'S EYES
TACTILE	STICK SHAKERS, RUDDER SHAKERS, SEAT BOUNCERS

TABLE IV.- TYPICAL PREFERENCE GRID FOR WARNING METHODS

50 FALSE ALARMS PER 1 REAL ALARM

URGENCY LEVEL	MUSICAL SOUND	OTHER SOUND	VOICE	LABELED LIGHT	UNLABELED LIGHT	TEXT MESSAGE	FLAG	TACTILE
	1	2	3	4	5	6	7	8
1 IMMEDIATE ACTION TO SAVE AIRCRAFT								
2 IMMEDIATE ACTION AFTER AIRCRAFT STABLE								
3 ACTION WHEN POSSIBLE								
4 ACTION LATER								
5 NO ACTION/ INFORMATION ONLY								

TABLE V.- PREFERENCE GRID USED FOR FLASHING/NOT FLASHING
AND AUDIO ALERT QUESTION

WARNING SYSTEM	1 IMMEDIATE ACTION	2 ACTION WHEN STABLE	3 ACTION WHEN STABLE	4 ACTION LATER	5 INFO ONLY
WARNING ON BOTTOM LINE OF CRT- -NOT FLASHING					
WARNING ON BOTTOM LINE OF CRT- -FLASHING 3/sec					
WARNING ON WHOLE CRT SCREEN- -NOT FLASHING, BUT WHATEVER WAS DISPLAYED BEFORE IS REMOVED					
WARNING ON WHOLE CRT SCREEN- -FLASHING, AND WHATEVER WAS DISPLAYED BEFORE IS REMOVED					
SINGLE LINE OF 12 3/4" HIGH ALPHANUMERICS- -NOT FLASHING					
SINGLE LINE OF 12 3/4" HIGH ALPHANUMERICS- -FLASHING 3 TIMES/sec					

TABLE VI.- SEMANTIC DIFFERENTIAL
CONCEPTS USED

Altitude alert tone
3/4-in.-square yellow light, flashing
3/4-in.-high alphanumeric display, not flashing
3/4-in.-diameter red light, flashing
3/4-in.-square yellow light, not flashing
Synthesized speech
3/4-in.-diameter red light, not flashing
3/8-in.-high lettering on a CRT
Whoop, whoop, pull up, pull up
ATC controller
Mechanical flag in glideslope indicator
Engine fire bell
3/8-in. blue light, not flashing
Stick shaker
SEL CAL tone
VASI lights
REIL lights
Gear horn

TABLE VII.- SEMANTIC DIFFERENTIAL
SCALES USED

Startling	---	Tranquilizing
Informative	---	Distracting
Good	---	Bad
Ugly	---	Beautiful
Soft	---	Hard
Strong	---	Weak
Worthless	---	Valuable
Loud	---	Soft
Unpleasant	---	Pleasant
Hot	---	Cold
Nice	---	Awful
Dark	---	Bright
Active	---	Passive
Noisy	---	Quiet
Safe	---	Dangerous
Alerting	---	Imperceptible
Annoying	---	Soothing

WARNING METHOD PREFERENCES

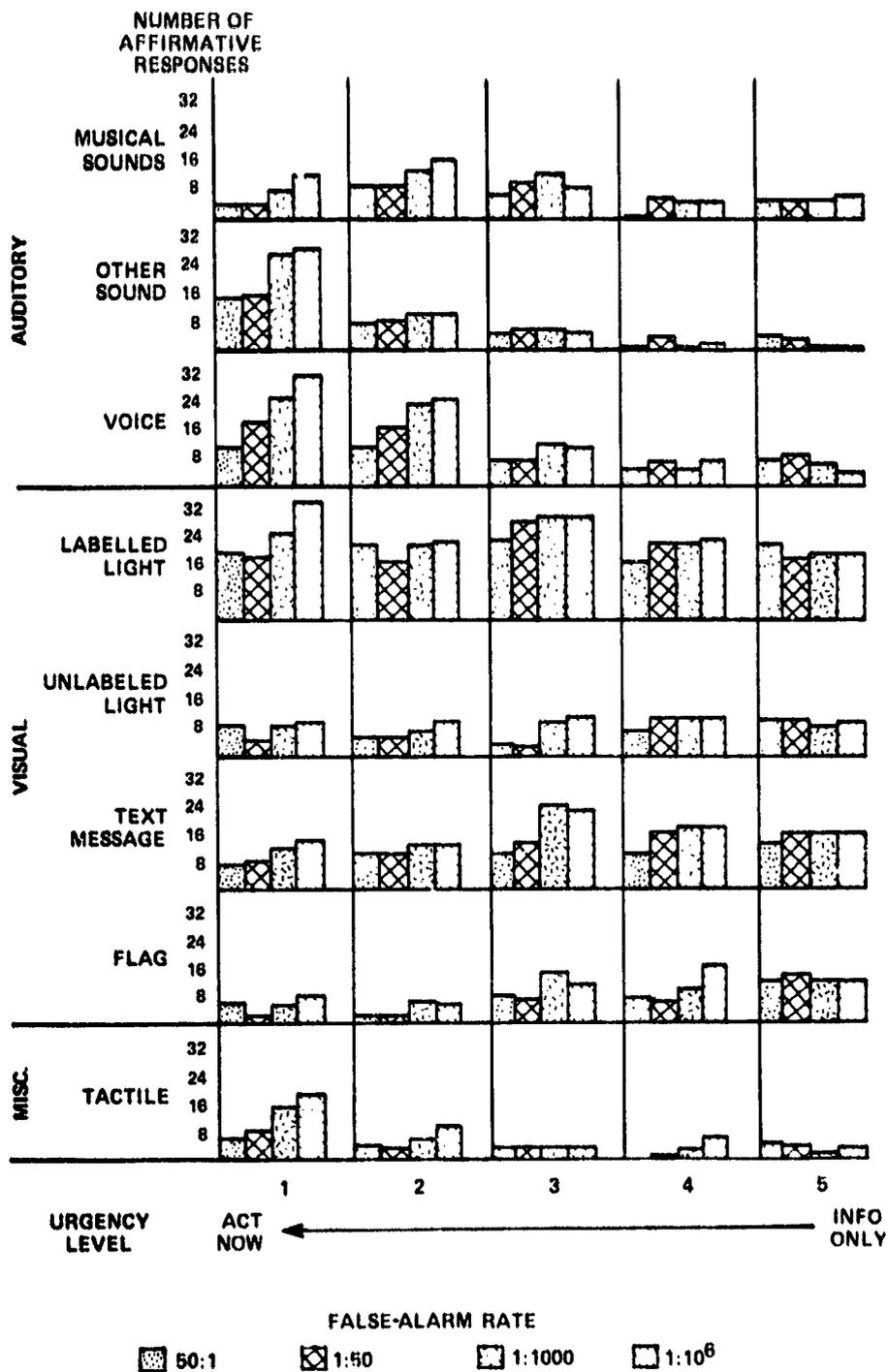


Figure 1.- Number of affirmative responses to warning methods as a function of warning urgency and false-alarm rate.

VOICE WARNING
PRIORITY 1 - ACTION NOW

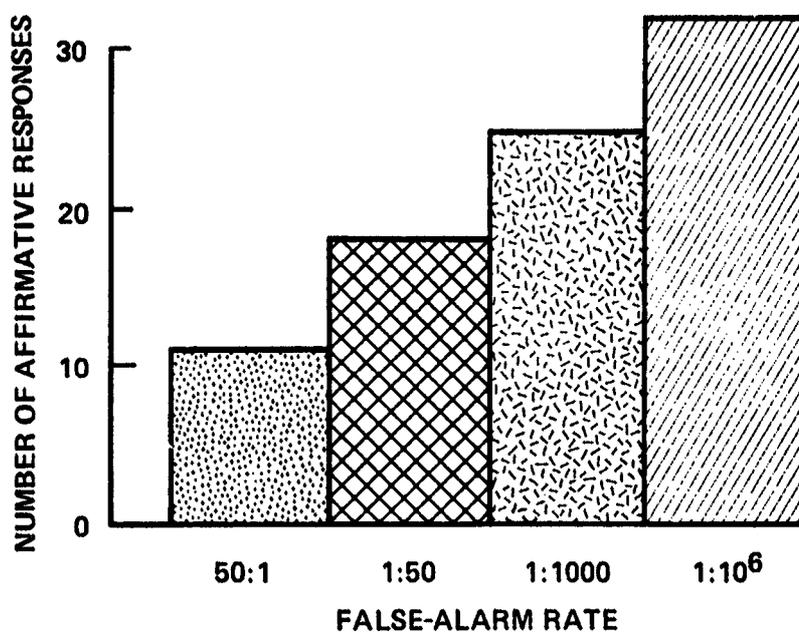


Figure 2.- Voice warning - priority 1 cell of warning method preferences figure.

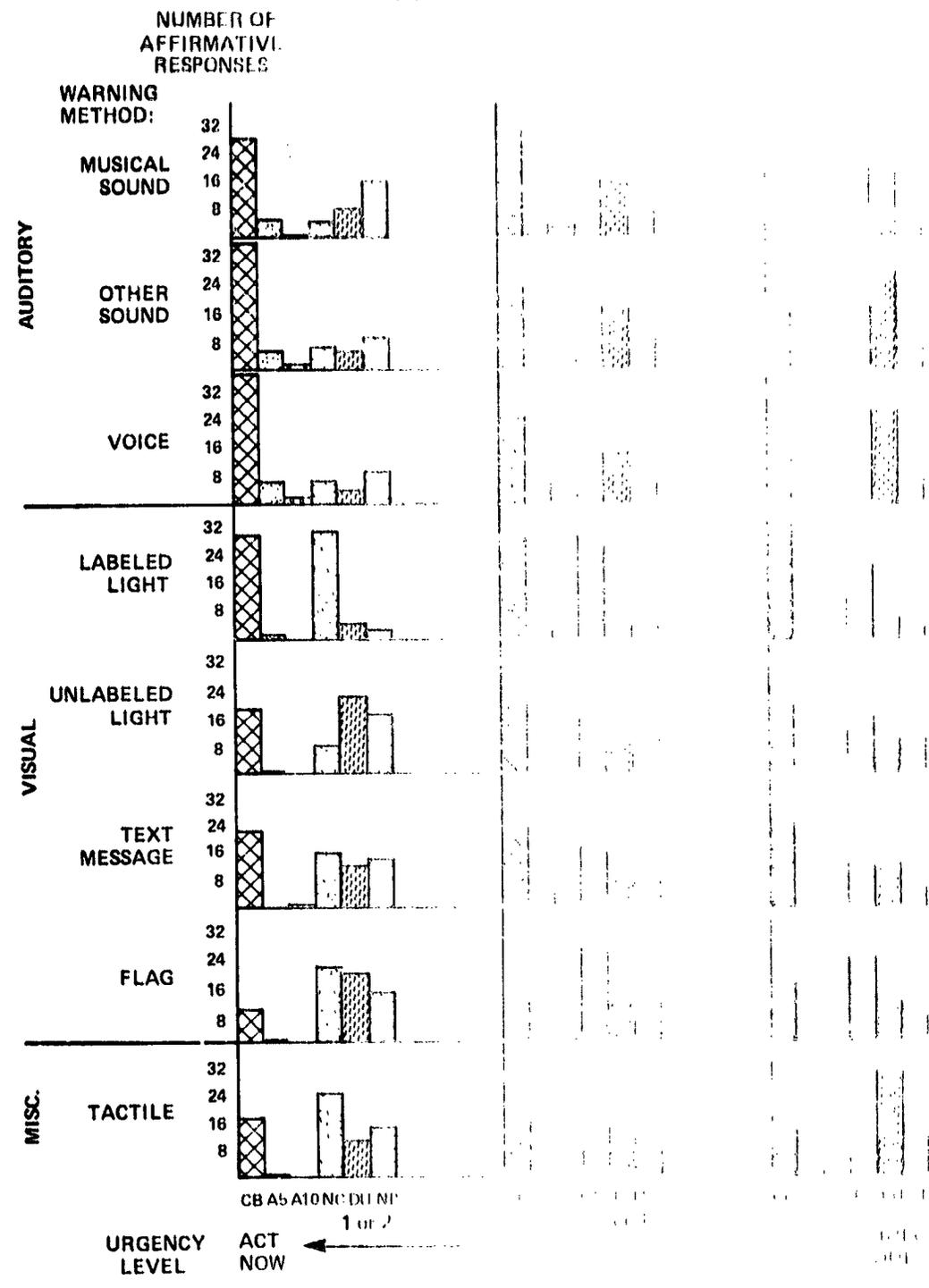


Figure 3.- Number of affirmative responses as a function of warning method and urgency level.

VOICE WARNING

PRIORITY 1 - ACTION NOW
PRIORITY 2 - ACTION NOW

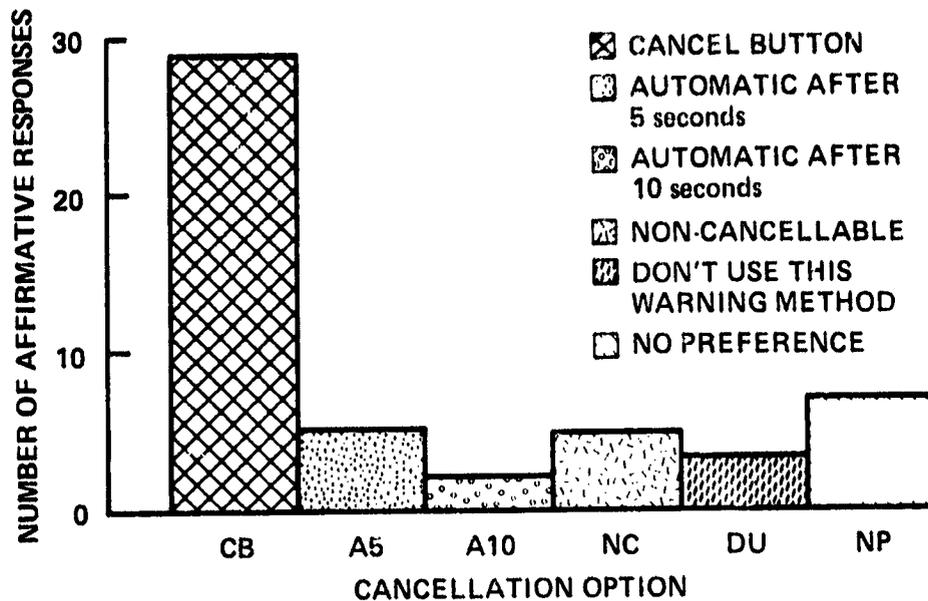


Figure 4.- Voice warning - priority 1 or 2 cell of cancellation options figure.

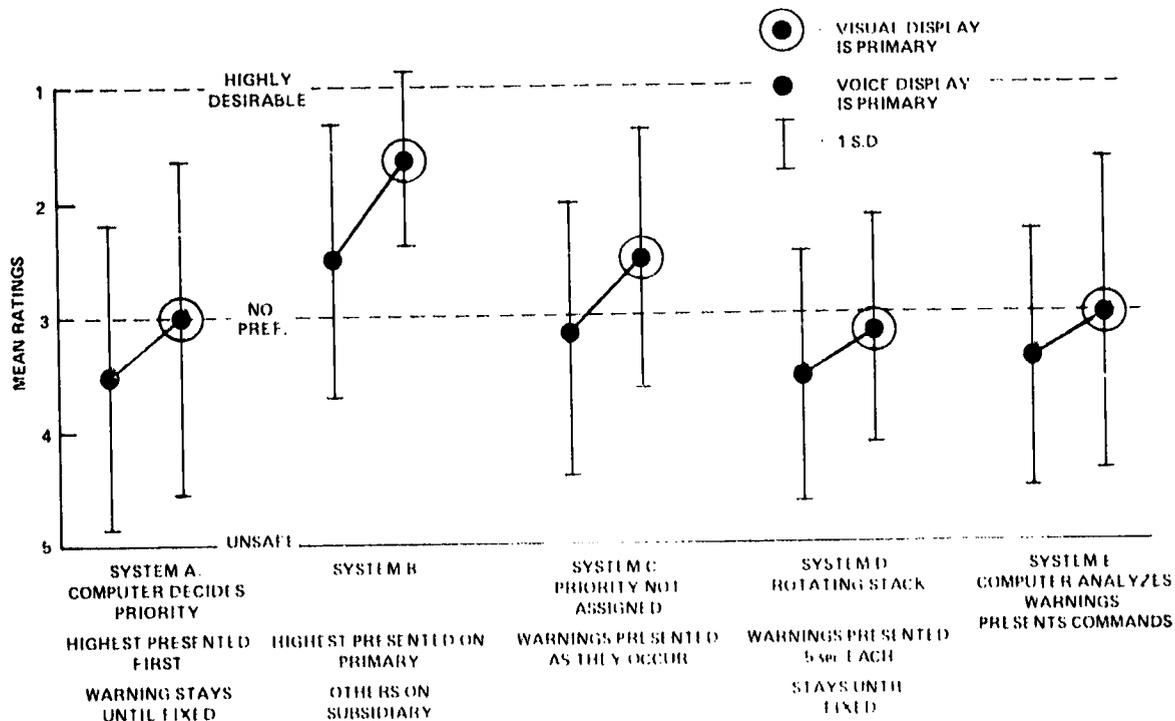


Figure 5.- System logic preferences.

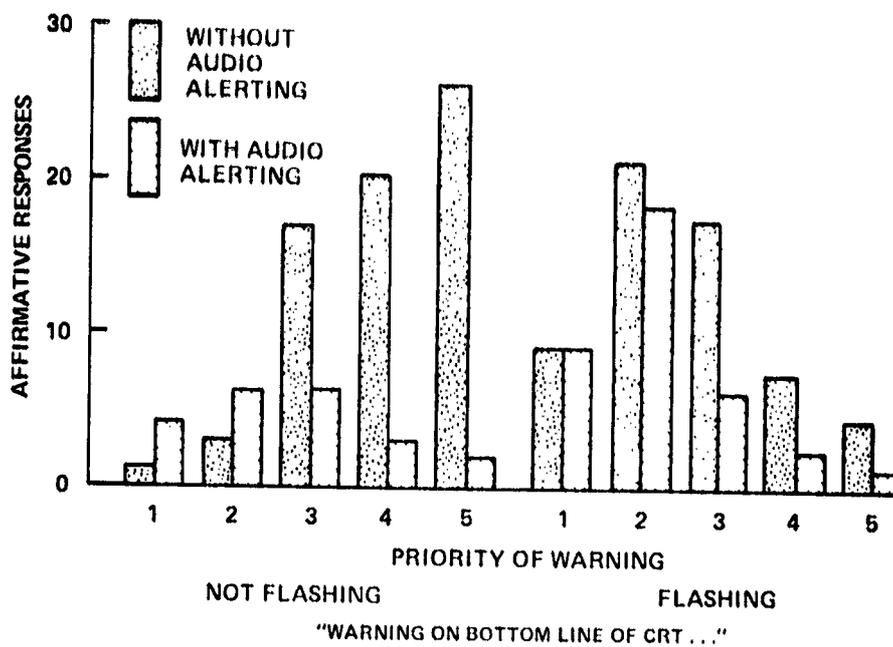


Figure 6.- Preferences for CRT line warning.

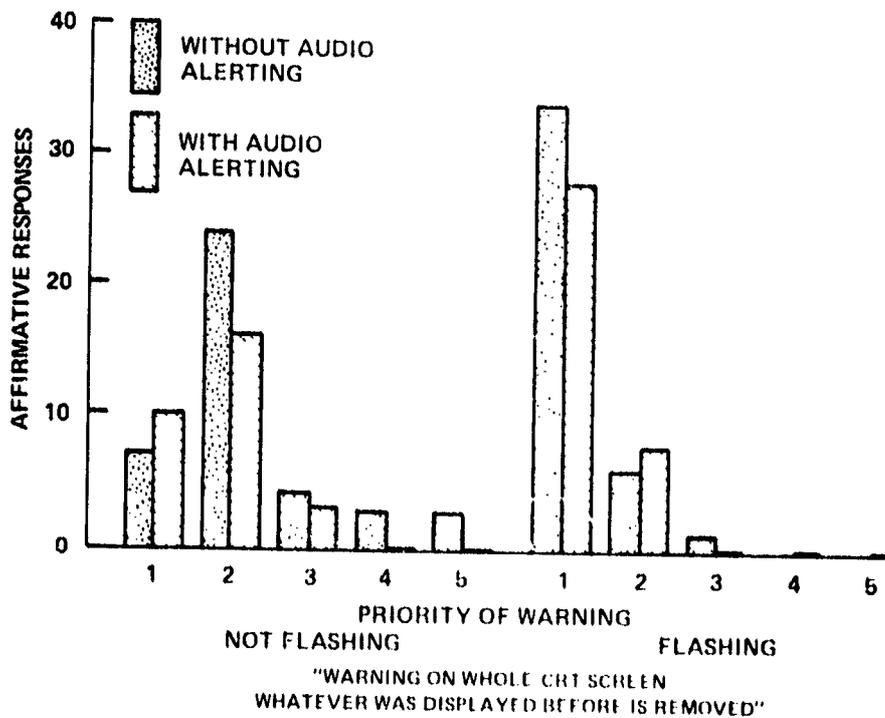


Figure 7.- Preferences for whole CRT screen warning.

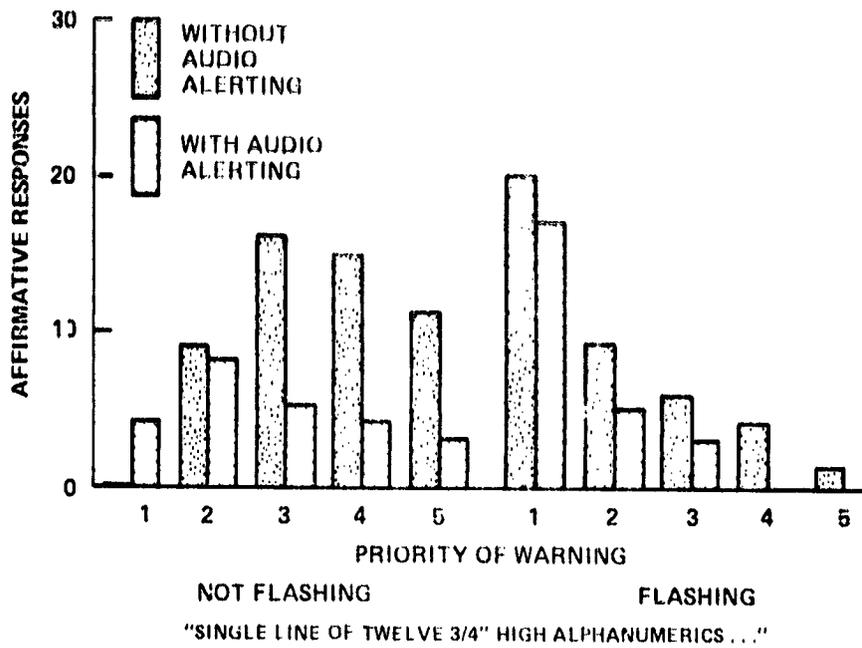


Figure 8.- Preferences for single line of alphanumerics.

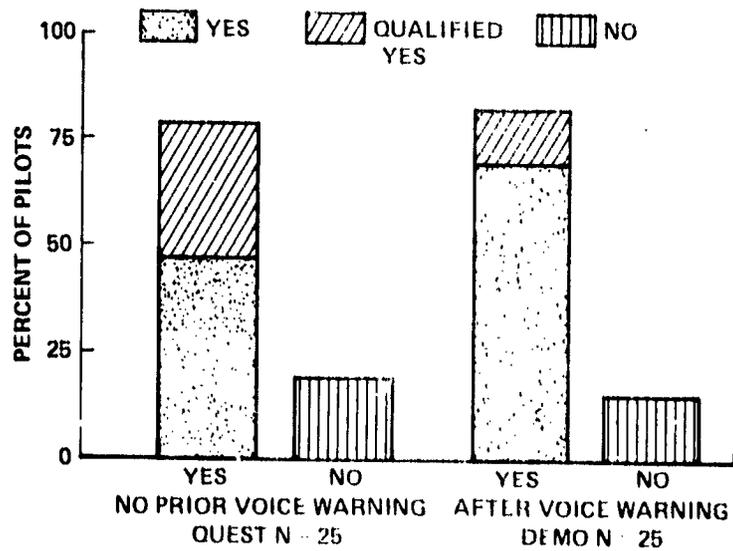


Figure 9.- Pilot acceptability of proposed voice warning system.

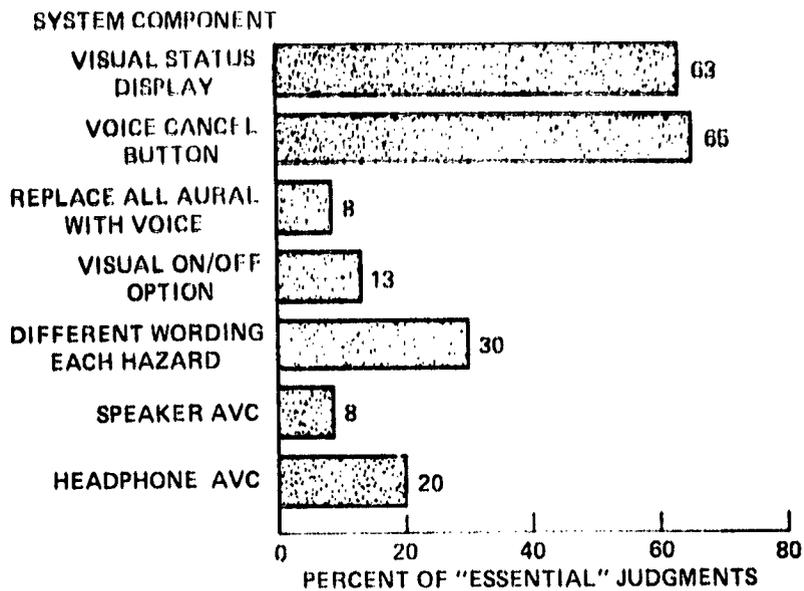


Figure 10.- Percent of "essential" judgments for each component of proposed voice warning system.

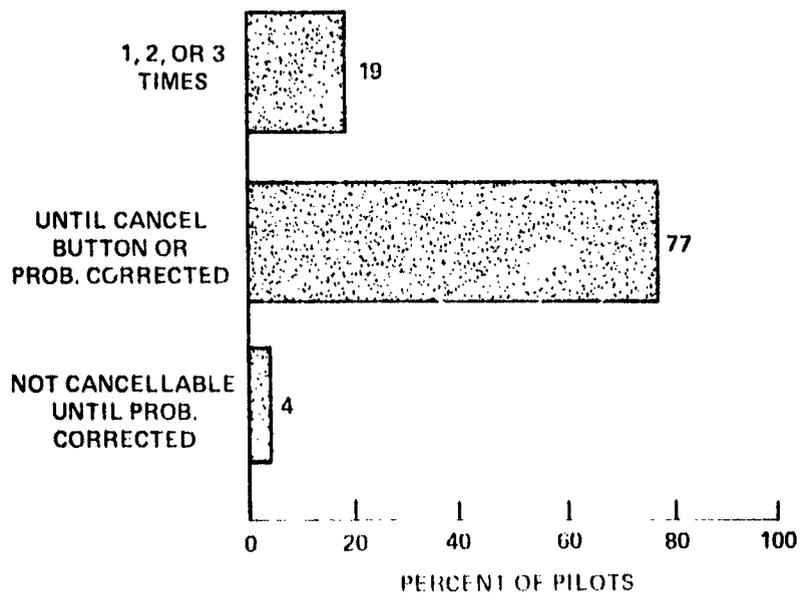


Figure 11.- Pilot preferences for repetition of voice warnings, N = 50.

REPETITION OF WARNINGS SHOULD DEPEND ON URGENCY-PRIORITY

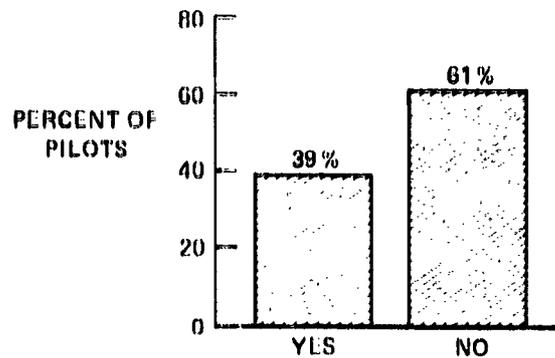


Figure 12.- Percent of "yes" and "no" responses for question regarding effect of warning urgency on preferred type of voice warning repetition.

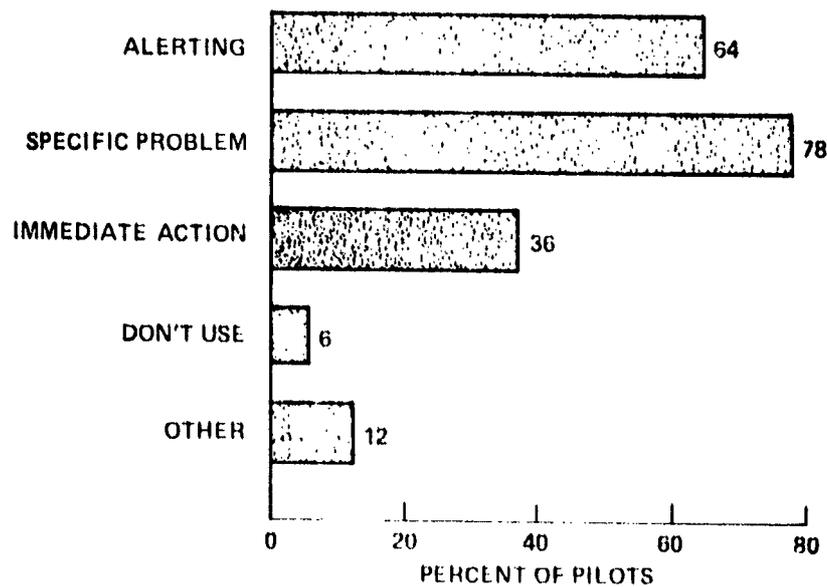
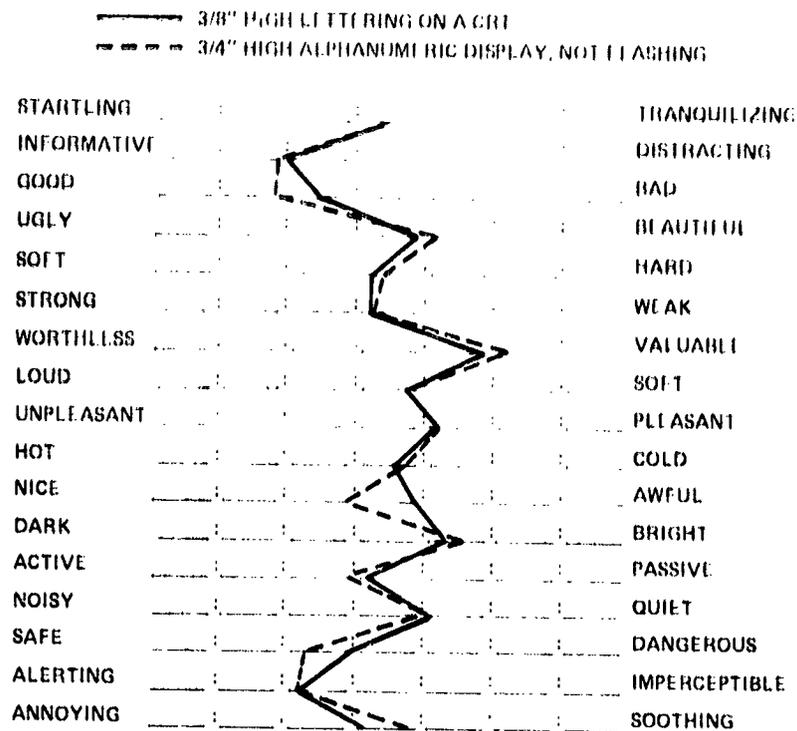
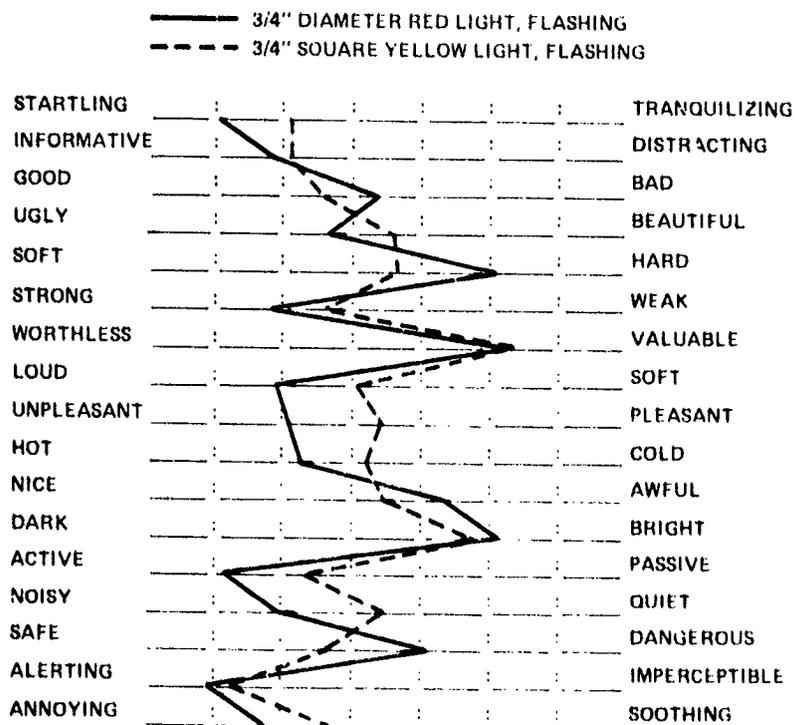


Figure 13.- Pilot preferences for uses of electronic voice in cockpit.



(a)



(b)

Figure 14.- Semantic differential average responses, N = 41.

CONTROLLED
WORDS

.....	TRANQUILIZING
.....	DISTRACTING
.....	BAD
.....	BEAUTIFUL
.....	HARD
.....	WEAK
.....	VALUABLE
.....	SOFT
.....	PLEASANT
.....	COLD
.....	AWFUL
.....	BRIGHT
.....	PASSIVE
.....	QUIET
.....	DANGEROUS
.....	IMPERCEPTIBLE
.....	SOOTHING

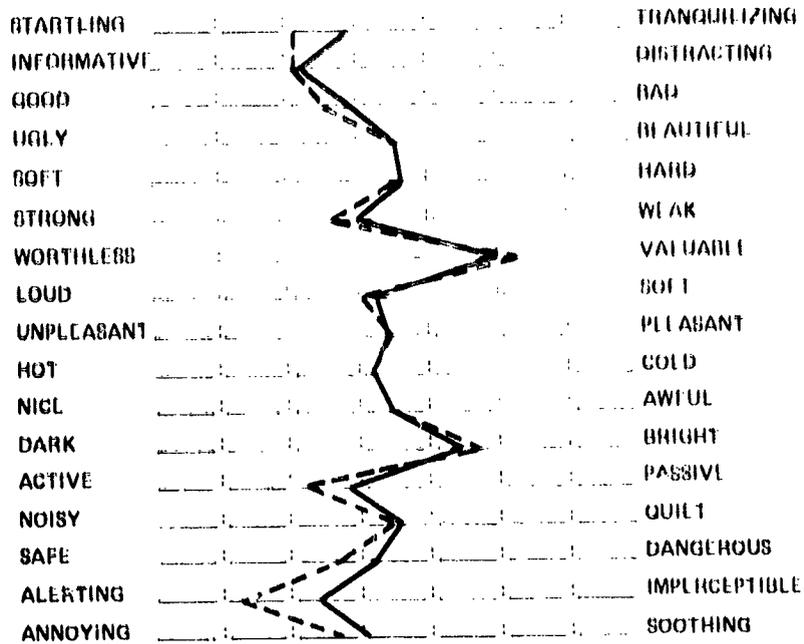
(c)

EMERGENCY BELLS
VIBRATION, WHOOOP, PULL UP, PULL UP

.....	TRANQUILIZING
.....	DISTRACTING
.....	BAD
.....	BEAUTIFUL
.....	HARD
.....	WEAK
.....	VALUABLE
.....	SOFT
.....	PLEASANT
.....	COLD
.....	AWFUL
.....	BRIGHT
.....	PASSIVE
.....	QUIET
.....	DANGEROUS
.....	IMPERCEPTIBLE
.....	SOOTHING

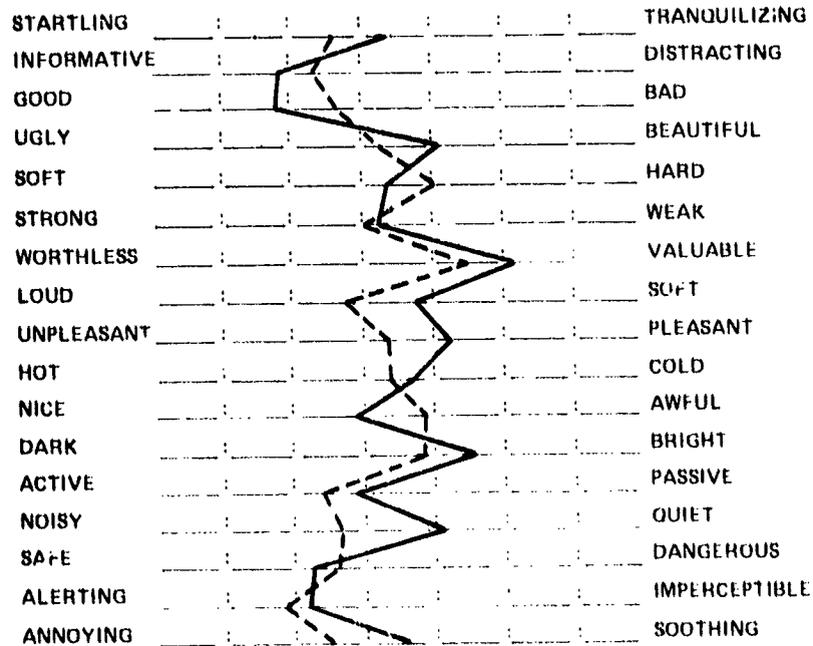
continued.

- - - - 3/4" SQUARE YELLOW LIGHT, FLASHING
 ——— 3/4" SQUARE YELLOW LIGHT, NOT FLASHING



(e)

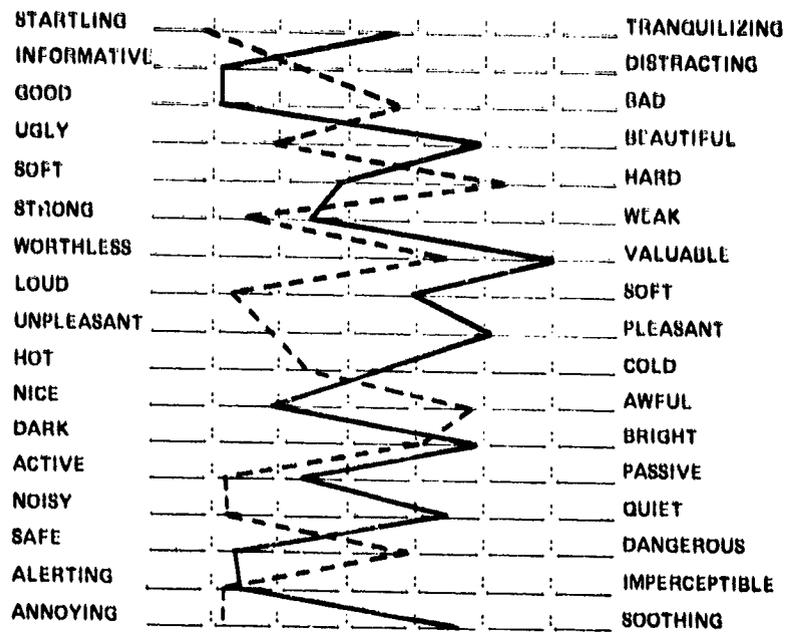
——— 3/4" HIGH ALPHANUMERIC DISPLAY, NOT FLASHING
 - - - - SYNTHESIZED SPEECH



(f)

Figure 14.- Continued.

— VABI LIGHTS
 - - - WHOOP, WHOOP, PULL UP, PULL UP



(g)

Figure 14.- Concluded.