

D46

N79-15634

TIME ESTIMATION AS A SECONDARY TASK TO MEASURE WORKLOAD:

SUMMARY OF RESEARCH

Sandra G. Hart*
University of Utah
Salt Lake City, Utah

Duncan McPherson**
University of California
Berkeley, California

Leslie L. Loomis**
San Jose State University
San Jose, California

Abstract

This paper outlines the results of a series of experiments designed to evaluate the utility of time estimation as a secondary measure of piloting workload. Actively produced intervals of time were found to increase in length and variability, whereas retrospectively produced intervals decreased in length although they also increased in variability with the addition of a variety of flight-related tasks. If pilots counted aloud while making a production, however, the impact of concurrent activity was minimized, at least for the moderately demanding primary tasks that were selected. The effects of feedback on estimation accuracy and consistency were greatly enhanced if a counting or tapping production technique was used. This compares with the minimal effect that feedback had when no overt timekeeping technique was used.

Actively made verbal estimates of sessions filled with different activities decreased in length as the amount and complexity of activities performed during the interval were increased. Retrospectively made verbal estimates, however, increased in length as the amount and complexity of activities performed during the interval were increased. These results support the suggestion that time estimation provides a useful index of the workload involved in performing concurrent tasks.

* Supported by NASA Grant NGR-45-003-108 to the University of Utah.

** Supported by NASA Grant NSG 2269 to the San Jose State University Foundation.

692
PAGE INTENTIONALLY BLANK

INTRODUCTION

The workload involved in performing different manual control and decision making tasks is often difficult to measure within a single task or to compare between different tasks. It is difficult to infer an operator's workload from his measurable performance because: 1) individuals may compensate for additional task load by working harder, resulting in little measurable variation in performance and 2) the total workload is composed of a variety of subtasks such that performance on any one may or may not reflect varying degrees of task load in the others. In addition, different measurement techniques may be required to determine subtask-specific variation in workload.

The purpose of this research program was to develop a battery of primary task indices and unobtrusive secondary tasks that would specifically measure the load imposed by different subtasks that make up the total piloting task in order to measure the overall workload in real and simulated flight. Performance on secondary tasks is often used as an index of primary task workload. Secondary tasks that are commonly used often load the operator to determine his remaining capacity to perform additional tasks while performing the primary task. However, it was decided that tasks selected for inclusion in the workload assessment battery should be unobtrusive and measure primary task load with minimal interference. The tasks also should be similar to tasks that are normally performed in flight, easily learned, implemented and scored.

The results of this research have suggested time estimation as one such secondary measure of the cognitive demands of piloting because it has been shown that an individual's ability to estimate intervals of time varies as a function of concurrent task load. Time estimation is a task that is normally performed in flight. It is unobtrusive, easily learned, implemented and scored and is not altered by repeated presentations unless knowledge of results is given.

Intra- and Inter-Subject Variability

Although individuals tend to be consistent in the length of their time estimates, there are large differences among different individuals. For this reason, each subject should be used as his own control: estimates obtained under different conditions of primary task load can be most easily and unambiguously analyzed by comparison with estimates obtained from the same subject in the absence of concurrent task demands. Individual estimation accuracy seems to be a less important measure than are the direction of change in the length of estimates and the increase in variability of estimates with the addition of a primary task.

Estimation Measurement Method

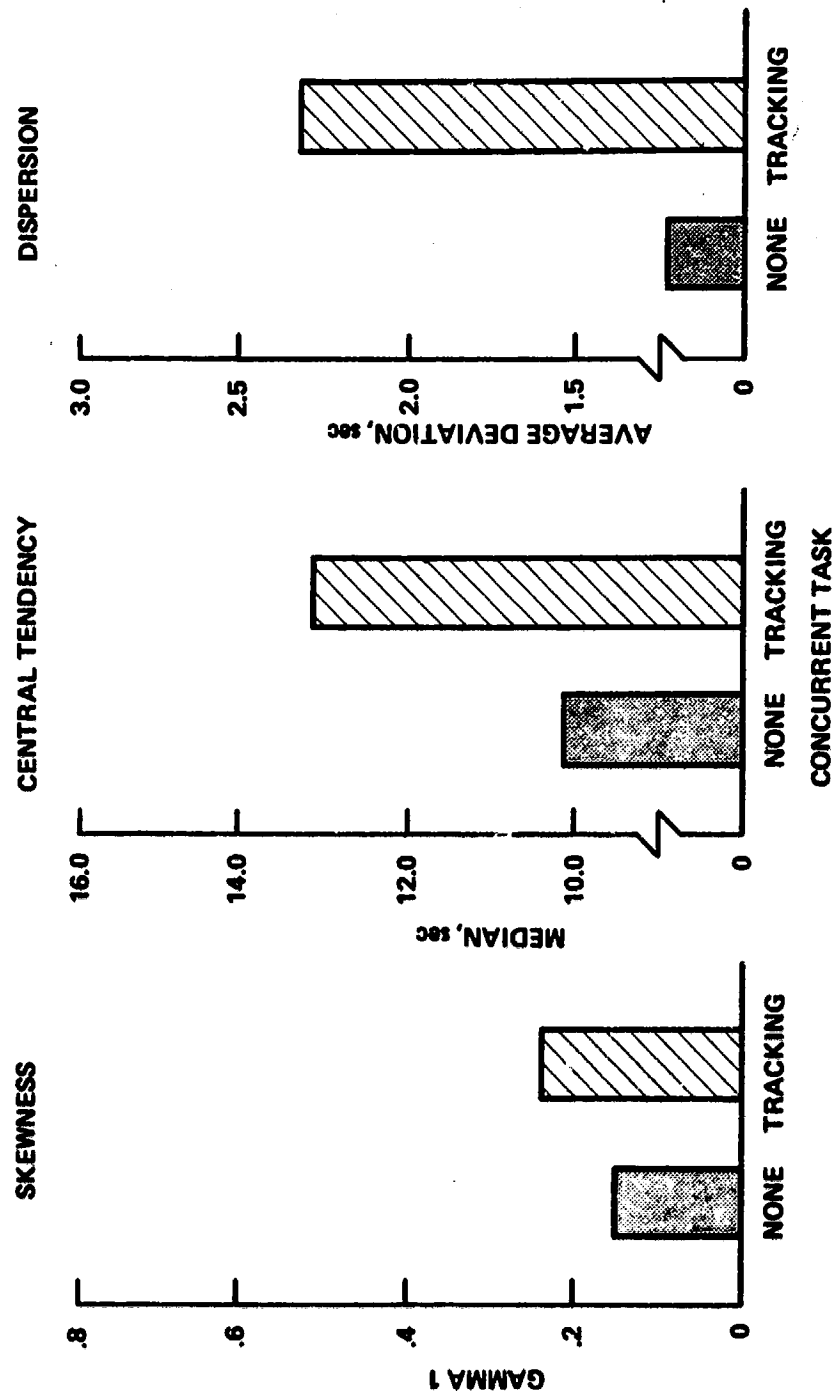
Four methods have been used extensively to measure an individual's ability to estimate or produce specified intervals of clock time. The verbal estimation method requires that individuals vocalize or record their judgement of the duration of an operationally presented interval. The production method requires that subjects physically generate an interval whose duration is specified by the experimenter. The reproduction method, which combines elements of verbal estimation and production, requires the operational production of an interval whose duration was presented operationally. The method of comparison involves a relative judgement between the durations of two or more operationally presented intervals.

Estimation Mode

Rather than being perceived directly, the temporal aspects of experiences are inferred or deduced from the events that occur in time. Man has adopted objective standards and labels to allow quantification of and communication about temporal experiences because of the difficulties involved in dealing with time in the abstract. Individuals represent durations subjectively by correlating personally

Figure 1.

BASELINE ESTIMATES COMPARED TO THOSE
PRODUCED DURING COMPENSATORY TRACKING



ORIGINAL PAGE IS
OF POOR QUALITY

experienced events with objective temporal standards or rules, such as clocks.

Active Mode

When individuals must produce a specific duration or verbally estimate the length of a presented interval unaided by an objective timing device, they may rely on impressions of past events or mentally or physically replay or generate a sequence of events that is believed to last a specific interval of time in order to make the temporal dimension of the interval concrete. This mode of estimation has been referred to as active estimation (ref. 1).

Retrospective mode

Individuals may also make temporal estimates without attending to time as it passes. They may estimate the duration of an interval at its conclusion by comparing the number and complexity of events that occurred during the interval with remembered durations of intervals similarly filled (ref. 2). This mode of estimation has been referred to as retrospective estimation (ref. 1).

Influence of Concurrent Activity on Active Estimation

The attention demanded by concurrent activity tends to interfere with active estimation. Whenever attention is diverted from active estimation, time passes unnoticed so that individuals may wait too long to terminate a production or verbally underestimate the length of the interval.

Active productions

Hart and McPherson (ref. 3) and Hart and Simpson (ref. 4) have shown that subjects do indeed wait too long to terminate their productions when distracted from active time estimation by competing simple compensatory tracking tasks (fig. 1) or speech recognition. A series of stylized representations of the mean length of 10 sec

Figure 2.

REPRESENTATIVE DISTRIBUTIONS OF 10-SEC PRODUCED DURATIONS: INFLUENCE OF CONCURRENT ACTIVITY

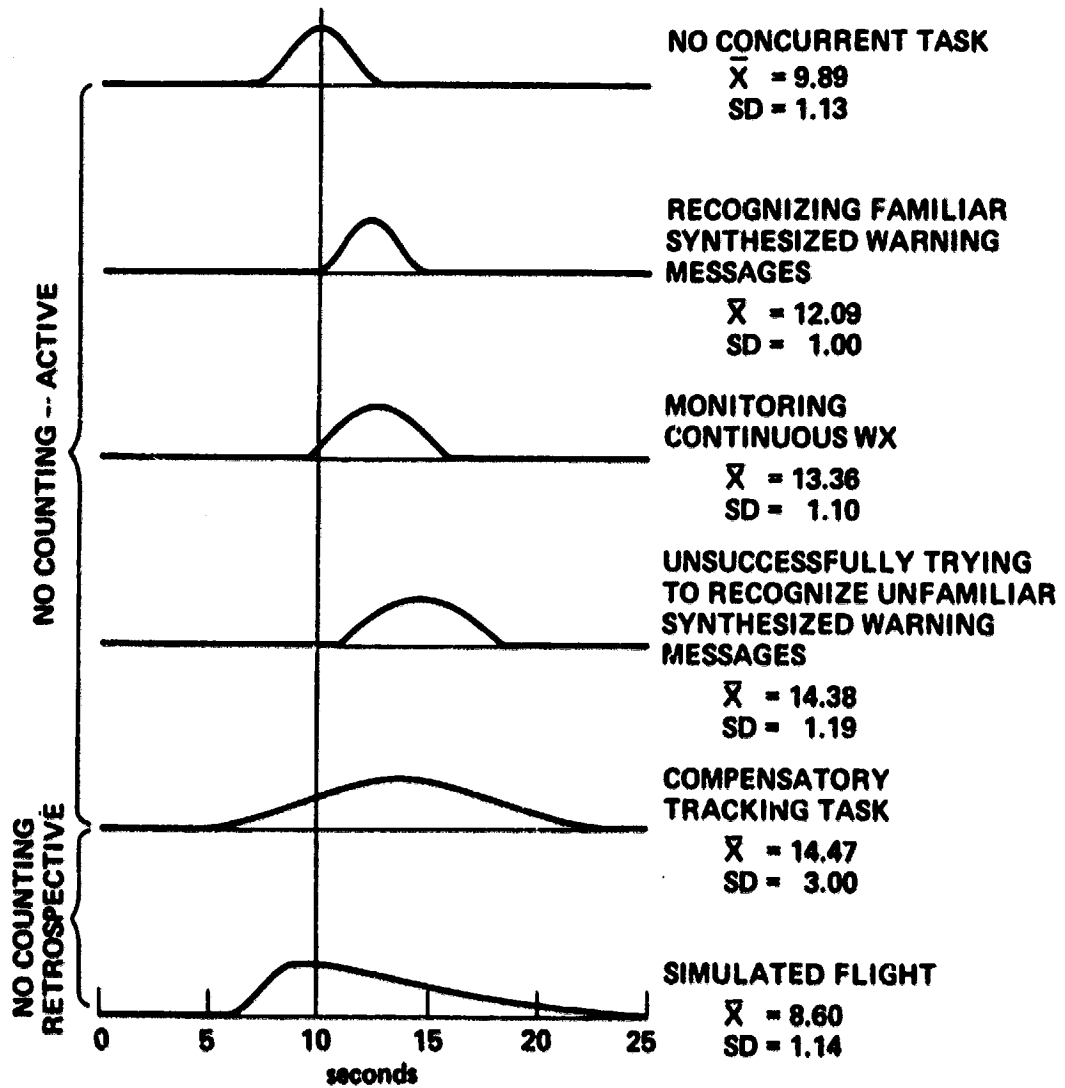
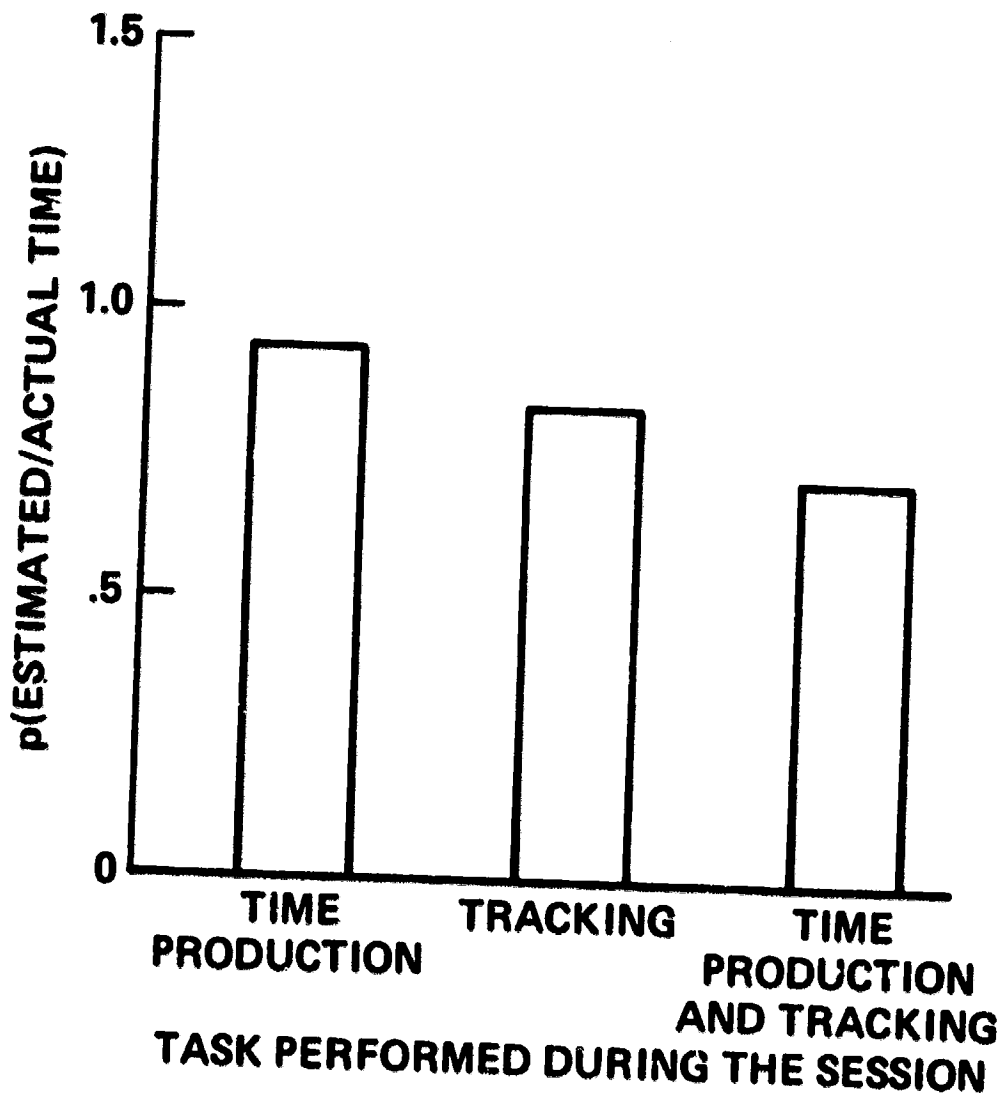


Figure 3.

RATIO OF VERBALLY ESTIMATED DURATION TO ACTUAL SESSION LENGTH (n=9)



productions obtained under different experimental conditions is given in figure 2. Each distribution's shape approximates that of actual data obtained and was drawn to include three standard deviations about the obtained mean. As the demands of the concurrent tracking and recognition tasks were increased, the length of produced durations increased by 4 sec or more and their variability more than doubled. Other, less demanding concurrent tasks, such as monitoring continuous aviation weather broadcasts, were also associated with an increase in the central tendency and variability of estimate distributions, but to a lesser degree, as one would expect from their less demanding nature.

Active verbal estimates

Hart (ref. 5) and Hart, McPherson, Kreifeldt, and Wempe (ref. 6) found that actively made verbal estimates decreased in length with the addition of either a simple compensatory tracking task (fig. 3) or a complex multi-manned flight simulation (fig. 4b). The more difficult levels of each task were associated with the shortest active verbal estimates. This is consistent with the finding that active verbal estimation and active production are reciprocally related, and the observed directions of change in estimated and produced durations are both the consequence of underestimation of the passage of time.

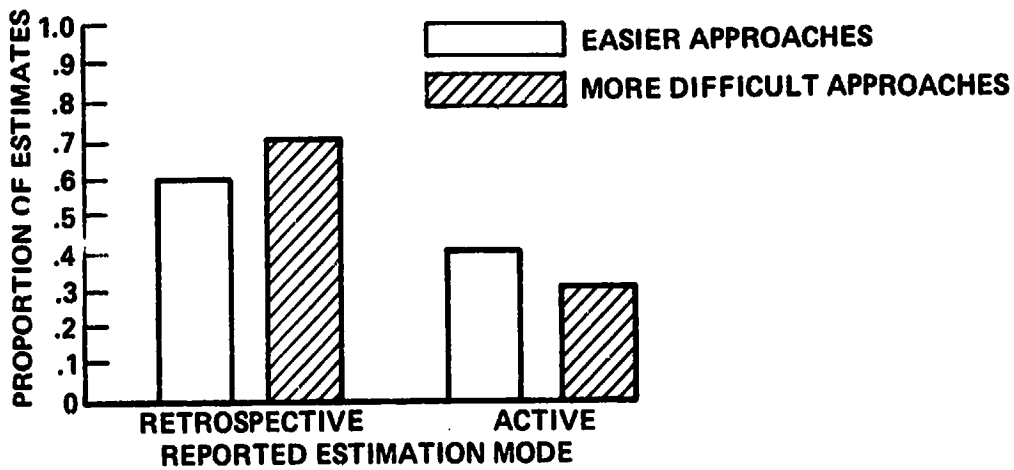
Influence of Concurrent Activity on Retrospective Estimation

As the attention demands of a primary task increase, there is less and less attention available for time estimation. When active estimation becomes impossible the retrospective mode of estimation becomes necessary. Here, one presumably remembers the events that occurred during the interval, compares them to other experiences with known duration, and then verbally estimates the duration of the interval or decides whether or not it is time to terminate a production. As the number and complexity of events

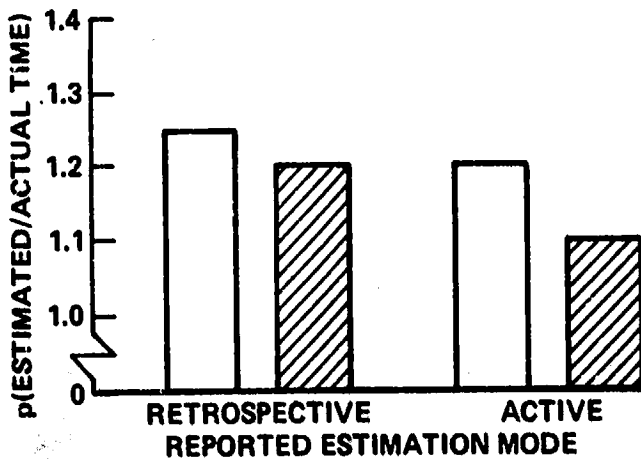
ORIGINAL PAGE IS
OF POOR QUALITY

Figure 4

A. PROPORTION OF VERBAL ESTIMATES OF THE DURATION OF FINAL APPROACHES THAT PILOTS REPORTED MAKING ACTIVELY AND RETROSPECTIVELY (n = 9)



B. AVERAGE RATIO OF ESTIMATED DURATION TO ACTUAL DURATION FOR ACTIVE AND RETROSPECTIVE ESTIMATES



that fill the interval are increased, there is a tendency toward over-estimation of the amount of time that has passed resulting in the termination of produced durations too soon or the verbal overestimation of elapsed time. Note that the directions of change in retrospectively verbally estimated and produced durations are the opposite of those obtained with active estimation and production and again the length of verbal estimates and productions are reciprocally related.

Retrospective productions

Hart and McPherson (ref. 3) have shown that the central tendency of 10 sec productions, obtained from pilots during simulated flight, decreased in length, as predicted, and the variability of the produced durations increased in comparison to estimates obtained with no competing activity. (fig. 5) Pilots reported that active estimation was difficult, resulting in their use of the retrospective mode. The distributions of retrospectively made productions were also positively skewed due to a few very long estimates which resulted from the estimation task occasionally being forgotten under conditions of high concurrent task load.

Retrospective verbal estimates

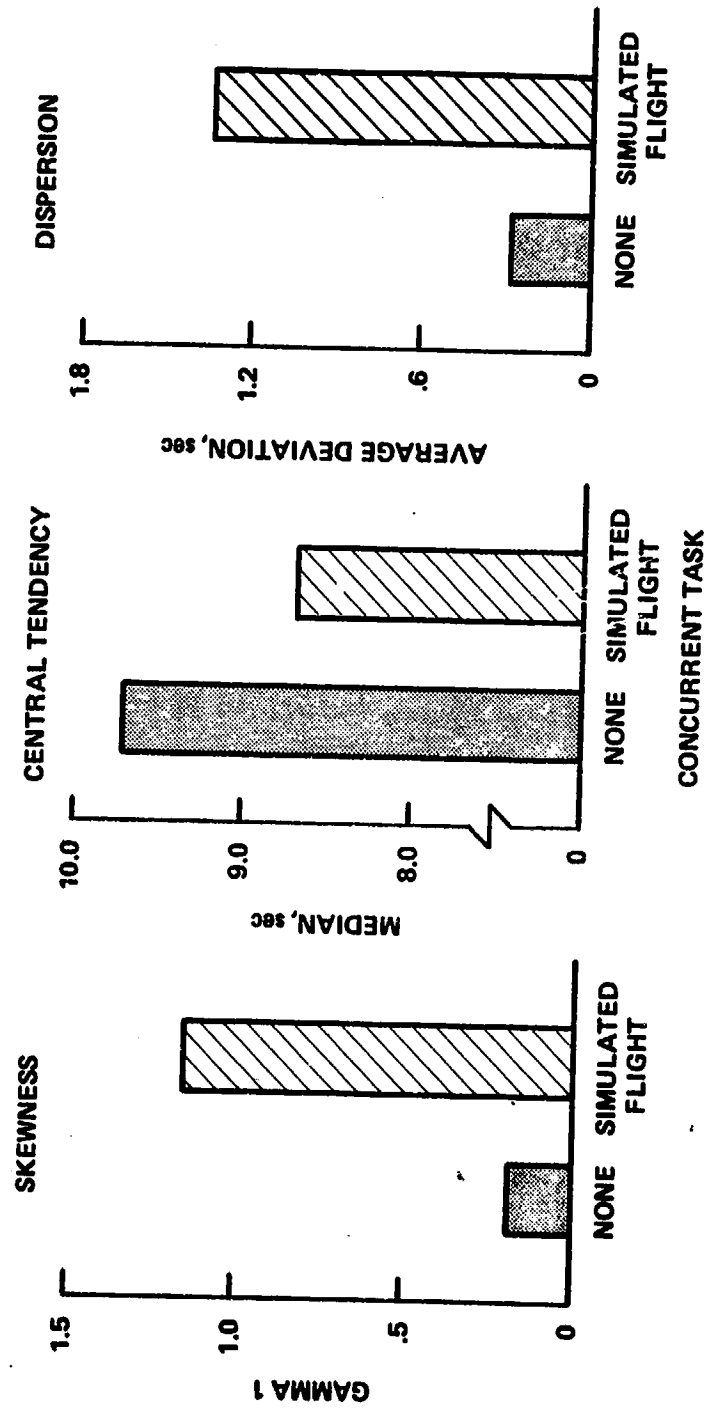
Following a complex multi-manned simulation flight, (ref. 6) pilot indicated that 66% of their estimates of the length of time taken to fly the final two miles of an approach were made retrospectively and that the proportion of retrospectively made estimates increased as the difficulty of the approach increased. (fig. 4a) Retrospectively made estimates were consistently longer than were estimates that pilots reported that they had made actively as predicted. (fig. 4b)

Interaction of Estimation Technique and Concurrent Task

Within the active mode of estimation there are many timekeeping techniques available. A standardized, rhythmic temporal metric (such as

Figure 5.

BASELINE ESTIMATES COMPARED TO THOSE PRODUCED DURING SIMULATED FLIGHT



tapping) not only fixes an individual's attention on the time estimation task, which is otherwise difficult to do for a task as abstract and stimulus-deficient as time estimation, but also provides a concrete, repeatable way to keep track of time. Timekeeping techniques that are not externalized, however, are more easily disrupted by additional, more compelling activities and are less stable across time. Some of the estimation techniques that subjects have reported using to keep track of time include counting, tapping, mentally replaying a phrase of music estimated to have the appropriate duration, mentally rehearsing the pre-flight checklist for a helicopter, counting heart beats or breaths, picturing the dial of a clock with a second hand moving around it, or "just waiting" for 10 sec. Of these techniques, those that are externalized, such as counting, provide standard, repeatable units with which to mark off intervals of time resulting in improved estimation stability. Mental rehearsal of remembered experiences judged to have the appropriate duration resulted in less stable productions, because the interval that was repeated may or may not have lasted the appropriate duration. Further, it is difficult to control the rate at which one's mind steps through a memory.

Hart, Loomis and Wempe (ref. 7) found that when attention was focused on a time production task by requiring subjects to rhythmically count aloud 1-sec intervals, production accuracy and consistency were not affected by the addition of a concurrent task. (fig. 6 and fig. 7) With no overt counting, however, the length and variability of produced durations increased significantly with the addition of a tracking task, replicating earlier results (ref. 3). Because performance on the tracking task was the same with both productions techniques, it appears that the shift in attention away from time production found with the no-counting technique was not because subjects could not innately perform both tasks but merely that they in fact did not. When attention was focused on the time production task by the counting technique, production accuracy was not degraded and there was no concomitant degradation of tracking task performance.

ORIGINAL PAGE IS
OF POOR QUALITY

Figure 6.

REPRESENTATIVE DISTRIBUTIONS OF PRODUCED DURATIONS: INTERACTION BETWEEN ESTIMATION TECHNIQUE AND INFLUENCE OF CONCURRENT ACTIVITY

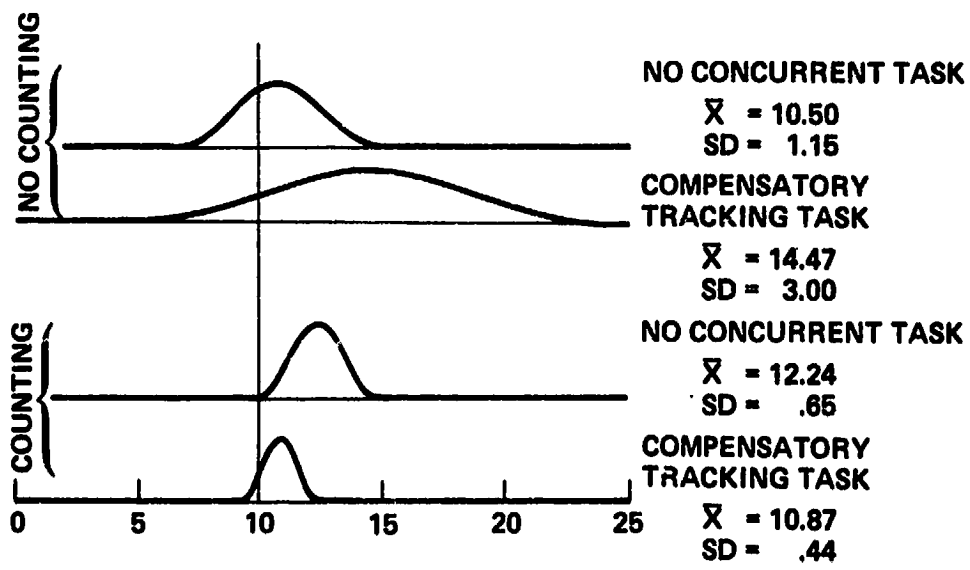
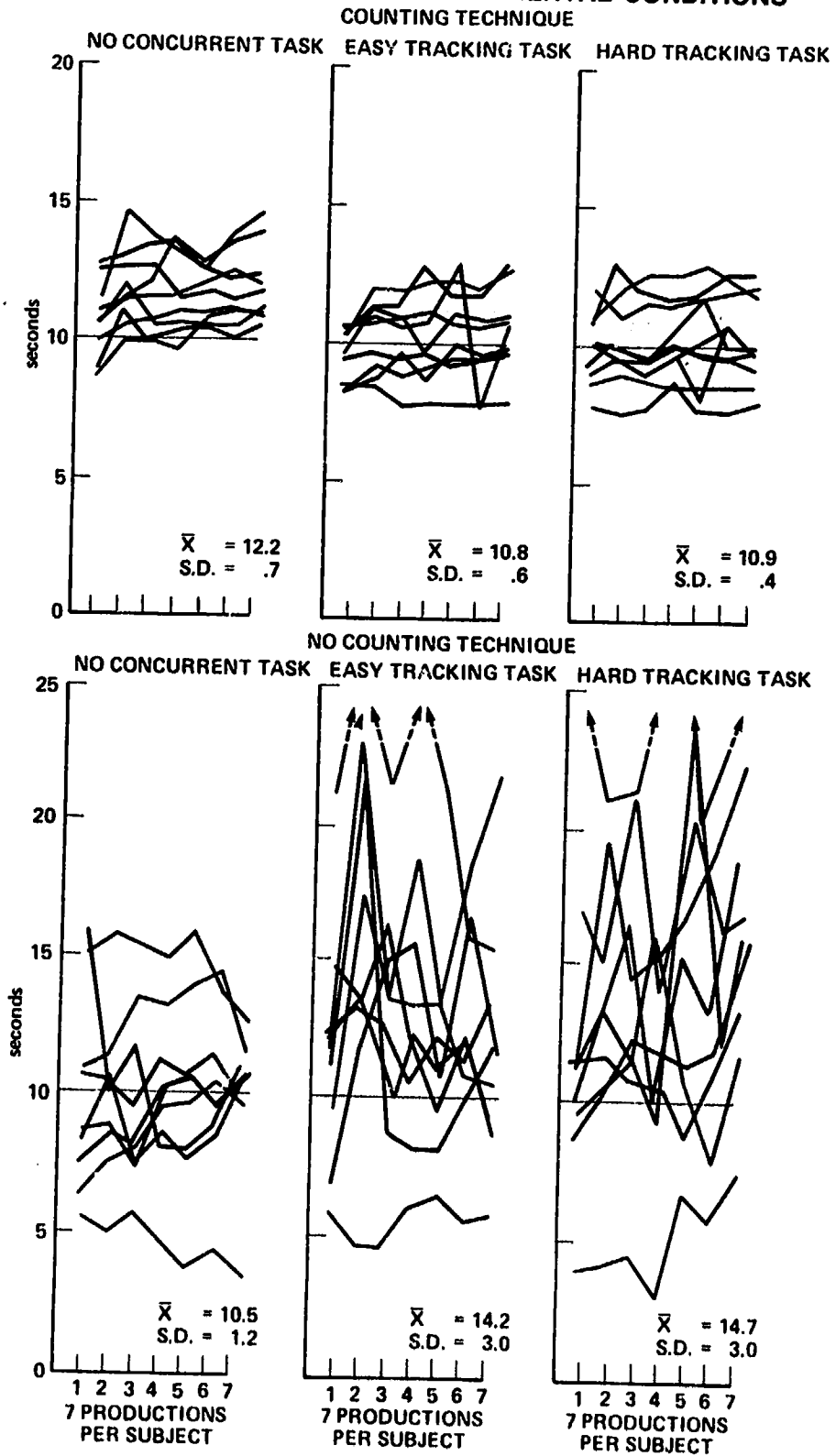


Figure 7.

DURATIONS OF THE 7 PRODUCTIONS MADE BY EACH SUBJECT UNDER SIX EXPERIMENTAL CONDITIONS



It is likely that more demanding concurrent activity, such as simulated flight, would also impact the consistency of durations produced with a counting technique. However, no such effects were found with the moderately demanding tracking tasks that were used.

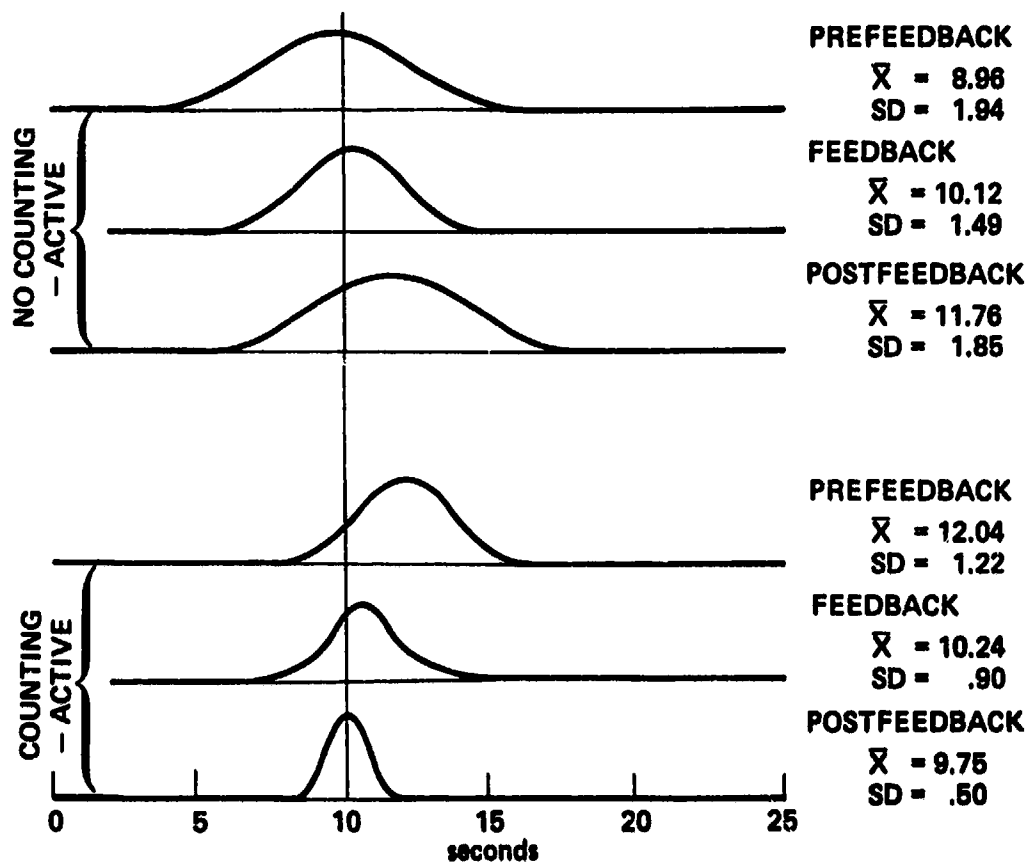
Interaction of Estimation Technique and Feedback

If the ability to estimate and produce intervals of time is learned, then it is likely that knowledge of results (feedback) should enhance timekeeping accuracy and consistency. In addition, the use of estimation techniques that provide rhythmic division of an interval into standard, repeatable units should focus attention on timekeeping and make the temporal dimension of the interval more concrete, thereby enhancing an individual's ability to take advantage of feedback.

In a recent study, Hart, Loomis and Wempe (ref. 8) found that individuals, using estimation techniques that did not involve some sort of overt counting, made less efficient initial use of feedback and did not experience any long term benefits from feedback. Overall accuracy of 10-sec productions, but not variability, was improved significantly by the presentation of feedback, with a rapid return to prefeedback performance levels when feedback was removed. (fig. 8 and fig. 9). During feedback, subjects repeatedly overcorrected. If told that one production was too long, the next production was typically too short and vice versa. Even after 30 trials with feedback following every production, subjects were unable to estimate accurately from trial to trial even though their estimate durations appeared to be accurate overall. If the subjects were instructed to rhythmically tap a button at 1-sec intervals in order to produce a series of 10-sec durations, both accuracy and variability were improved significantly by the addition of feedback. This improvement persisted for at least as long as 30 additional trials after feedback was removed. With this production technique, subjects were able to maintain consistent and accurate estimates from trial to trial, and did not overcorrect as they had

Figure 8.

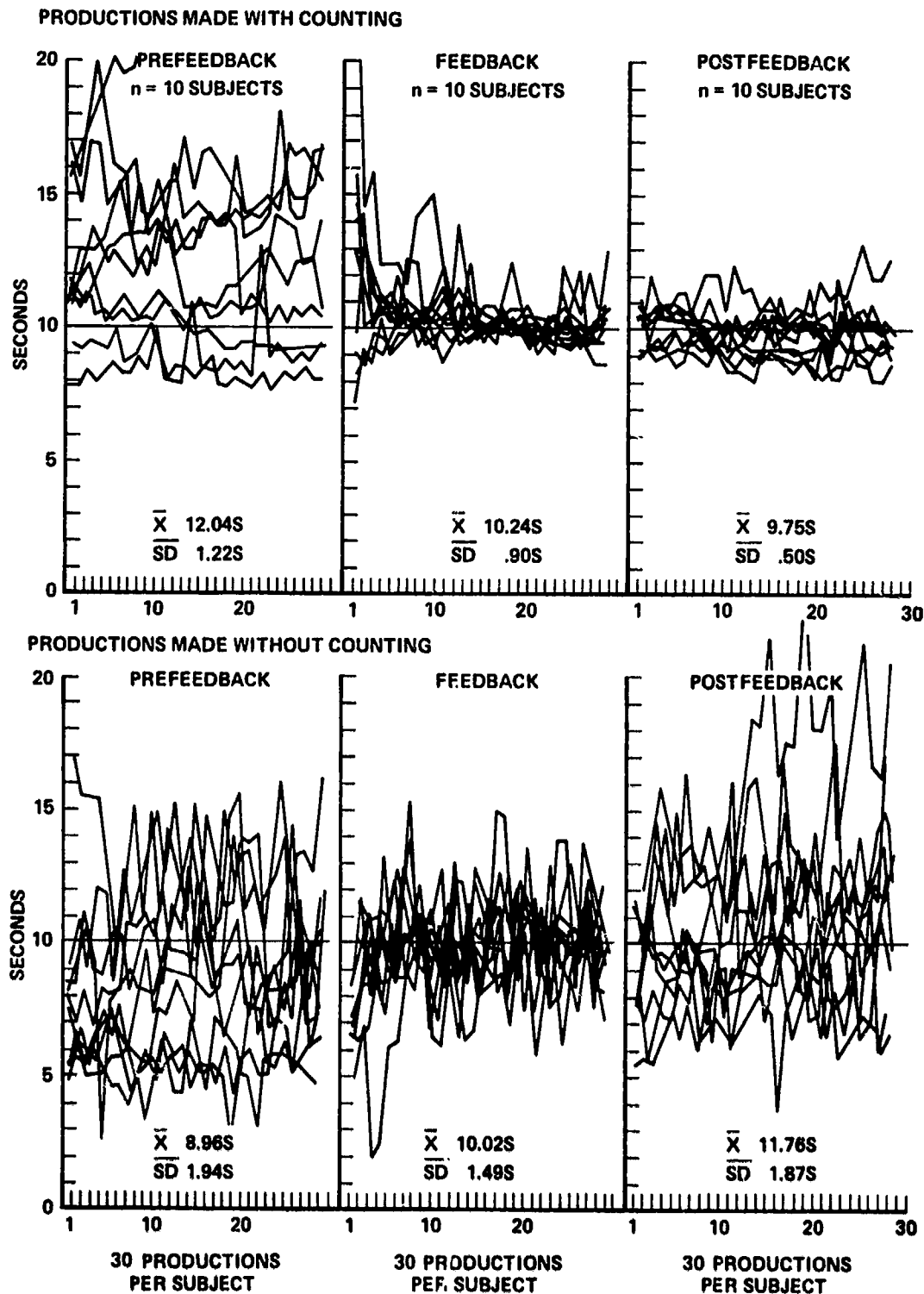
REPRESENTATIVE DISTRIBUTIONS OF 10-SEC PRODUCED DURATIONS: INFLUENCE OF KNOWLEDGE OF RESULTS (FEEDBACK)



ORIGINAL PAGE IS
OF POOR QUALITY

Figure 9.

COMPOSITE GRAPHS OF RAW SCORES OF 10 SUBJECTS



with the no-counting technique. The data suggest that tapping rhythmically not only provides a standardized repeatable temporal metric, but also fixes subject's attention on the time production task, which together combine to enable subjects to use feedback more effectively.

Conclusion

As a result of the foregoing research effort, several recommendations can be made concerning the use of time estimation as a secondary measure of the attention demands of a primary task.

Method

The production of brief intervals of time appears to be the most useful experimental method. The duration and variability of time productions in the range of 1 to 30 sec have been shown to reflect the attention demands of primary manual control, message recognition, and simulated flight tasks. Relatively brief intervals should be used so that the primary task load remains reasonably uniform and describable during the produced interval.

The verbal estimation method also shows some promise as a secondary measure of primary task workload. Its primary advantage over the production method is ease of implementation. Its primary disadvantage is that subjects tend to round off their estimates, thereby losing precision, and their responses tend to become stereotyped if a number of estimates are required. This method appears to have some value, but is less sensitive than the method of production.

Mode

Estimation mode (active or retrospective production or verbal estimation) must also be controlled or identified to obtain reliable and clear results with a time estimation task. Because retrospective productions decrease in length with increasing task load whereas active productions increase in length, care must be taken to identify the mode of production used. If retrospective and active productions are combined in an analysis, their direction of change with the addition of another

task would tend to cancel out masking detailed changes in the underlying processes.

Technique

Timekeeping techniques that are not externalized are most easily disrupted by concurrent task demands and thus provide the most useful measure of primary task demands. Thus, if time production is to be used as a measure of workload, subjects should not be allowed to use any overt time estimation technique such as tapping or counting. If estimation accuracy and consistency are required, however, an overt timekeeping technique should be used. Further research is required to determine at what level of concurrent task load the overt estimation technique would also be disrupted.

Feedback

If an overt timekeeping technique is used, feedback is effective in reducing both error and variability after only two or three repetitions, and the effects of feedback last long after it has been removed. With no overt timekeeping technique, however, estimation error is reduced only on the average, and variability remains high with a rapid return to pre-feedback error levels following removal of feedback.

Data Analysis

Time estimation performance is best evaluated relatively. That is, the amount and direction of change in estimation accuracy and consistency observed in the presence of additional primary tasks should be compared to estimates obtained from the same subject with no additional activity. Care should also be taken to select the appropriate measures of central tendency and variability as distributions of time productions are often positively skewed, particularly when obtained in the presence of competing concurrent activity.