Hayashi, M., Oman, C. M., & Zuschlag, M. (2003). Hidden Markov Models as a Tool to Measure Pilot Attention Switching during Simulated ILS Approaches. Paper presented at the 12th International Symposium on Aviation Psychology, 502-507, Dayton, OH, Apr. 14-17.

HIDDEN MARKOV MODELS AS A TOOL TO MEASURE PILOT ATTENTION SWITCHING DURING SIMULATED ILS APPROACHES

Miwa Hayashi and Charles M. Oman
Massachusetts Institute of Technology
Cambridge, MA

Michael Zuschlag
Volpe National Transportation Systems Center
Cambridge, MA

The pilot's instrument scanning data contain information about not only the pilot's eye movements, but also the pilot's cognitive process during flight. However, it is often difficult to interpret the scanning data at the cognitive level because: 1) some instruments provide partially-redundant information, and 2) some instruments display more than one type of information. To avoid these problems, a new modeling and analysis technique is demonstrated that looks at the scanning data from the task attention-level as well as from the instrument-fixation level. Basic principles of attitude instrument flying were incorporated to construct a task-instrument framework, and a Hidden Markov Model (HMM) was used to estimate hidden task transitions. An ILS simulation experiment was conducted, and the task sequence estimated by the HMM matched 79-92% of the task epochs verbally reported by the pilot. The HMM and traditional instrument fixation analyses complement each other, and successfully detected effects of display format changes that did not produce significant changes in flight technical error and subjective workload. Potential advantages of this method for the analyses of advanced cockpit displays (e.g., primary flight displays, head-up displays) are discussed.

Introduction

The record of pilot's instrument scanning is not only a record of eye movements, but also a valuable reflection of the pilot's cognitive process during the flight. If the scanning data are analyzed only at the instrument-fixation level, however, it is often difficult to take advantage of this wealth of data. Two obstacles impede analyses beyond the fixation level. 1) Instruments sometimes provide partially redundant information. For example, rate-of-climb information is provided on the altimeter as well as on the vertical speed indicator (VSI). 2) Some instruments display more than one type of information. For instance, the attitude indicator (AI) displays both pitch and bank information. If the pilot is looking at the AI, the experimenter cannot immediately determine which information - pitch or bank - the pilot is attending to. The AI is especially problematic because it is looked at so frequently (e.g., 42-67% of total dwell time for various maneuvers, Pennington, 1979).

We propose a new analysis method that examines the scanning data from the task-attention level as well as from the fixation level. The following sections describe how this new modeling framework can be used to avoid these problems.

Task-Instrument Association Framework

Manually flying an aircraft involves multiple tasks. For example, manual instrument flight along a certain predetermined path requires simultaneous tracking of vertical, horizontal, and airspeed. In this paper, these tracking tasks are called vertical, horizontal, and airspeed task, respectively. Each tracking task requires a particular set of information displayed on the instrument panel. For instance, Table 1 shows an example of these task-instrument associations. The pilot's instrument scanning data are time series of the fixations that the pilot generated to collect information from these instrument. It is not possible to directly observe which task the pilot is attending to; however, based on associations such as in Table 1, one can infer it from the pilots' instrument scanning data.

Table 1: Task-Instrument Associations for Manual Instrument Flight

Task	Vertical	Horizontal	Airspeed
Instruments	• AI • Altimeter • VSI • CDI (glide slope)	• AI • T/S • HI • CDI (localizer)	• AI • ASI • Thrust

"AI" = Attitude Indicator, "VSI" = Vertical Speed Indicator, "CDI" = Course Deviation Indicator, "T/S" = Turn & Slip Indicator, "HI" = Heading Indicator, "ASI" = Airspeed Indicator

This modeling framework addresses the problem of the partial-redundancy of instruments by introducing a task-attention level viewpoint. The redundant instruments (e.g., altimeter and VSI) are now in the same group, i.e., in the same "task."

The second problem, where one instrument displays more than one type of information, is illustrated in Table 1, in which the AI and the course deviation indicator (CDI) appear in more than one task. To resolve the ambiguity this creates, the modeling framework exploits instrument cross-checking. In principle, the instrument set associated with each task in Table 1 can be further divided into control (e.g., AI), performance (e.g., altimeter, VSI), and navigation instruments (e.g., CDI). Pilots must continuously check across these three instrument categories to control the aircraft properly and ensure that it is following the desired path. Thus, given cross-checking, one can use the proximal fixations on other instruments to estimate the underlying task. For instance, if the AI fixations alternate with other vertical task instrument fixations (e.g., altimeter, VSI), then the AI is more likely being used for the vertical than for other tasks. In this way, by using the entire fixation sequence, one can estimate the most likely task sequence that underlies the observed fixation sequence. This estimation is carried out by fitting the fixation sequence with a Hidden Markov Model.

Hidden Markov Model (HMM)

The HMM structure has two levels of probabilistic processes. One is a hidden-state level (i.e., the task attention), which is not directly observable but assumed to follow a first-order Markov process transition rule. The other is the observation-symbol level (i.e., the fixation sequence data), which is observable and has a certain probability distribution that depends on the current hidden state. An HMM structure is completely describable by a set of three parameter matrices. The initial state probability distribution matrix, π, gives the probability that a given hidden state is the initial state; the state-transition probability distribution matrix, A, gives the transition probability from one hidden state to another; and, the observation symbol probability distribution matrix, B, gives the probability of each observation symbol within a given hidden state.

Given an observation-symbol sequence (i.e., the fixation sequence data), O, an Expectation-Maximization (EM) algorithm estimates the model parameter matrices, A, B, and π, that locally maximize the observation probability, $P(O|A,B,\pi)$ (There is no known analytic way to solve for the global maximum). Then, a dynamic programming algorithm is used to estimate the hidden-state sequence corresponding to the observed symbol sequence. For details of these algorithms, see Rabiner & Juang, 1993.

The following sections illustrate the application of the proposed modeling framework to a basic cockpit human factors display issue. This flight-simulator experiment examined the effects of an analog round-dial versus a digital display format. It appears that most of pilots prefer the round-dial format to digital format probably because the needles provide position and motion cues in peripheral vision (Bradley, 1996). This difference, however, is so subtle that the effect may not be reflected in the flight performance data or in the pilots' subjective workload scores. Even if effects are found, it is difficult to extract from the data the exact causal link between the missing needles and the performance change (e.g., Weinstein, et al., 1994). The following sections demonstrate how the HMM analysis revealed a subtle effect of format on fixation sequence and task attention switching. One pilot subject was used for this experiment; thus all effects examined were within-subject effects for this particular pilot.

Method

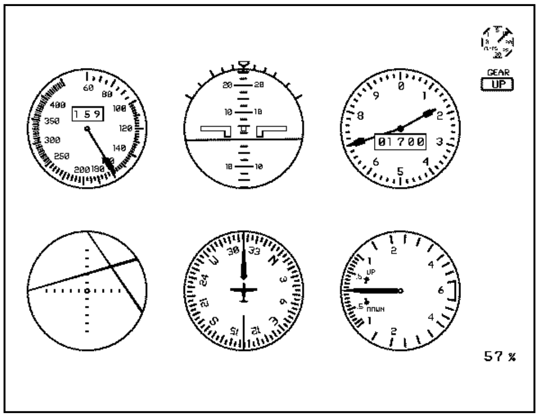
Flight Simulation Experiment

Simulator. This experiment used a fixed-base flight simulator configured with Boeing 757-200 flight dynamics.

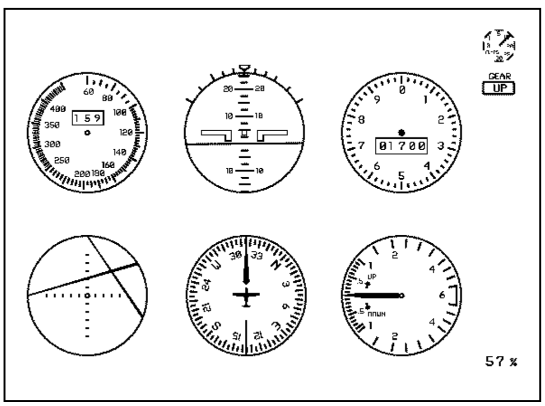
Display. Two display formats were used: Display 1 (D1) was a basic-T layout including a VSI and CDI (Fig. 1a). The airspeed indicator (ASI) and altimeter incorporated both analog needles and digital readouts. Display 2 (D2) was identical to D1, except that the needles on the ASI and altimeter were missing (Fig. 1b). The outer circles and scales of these instruments remained unchanged in order to keep the saliency comparable to that in D1. Both displays were depicted as green on a black background. The displays were presented on a 17-inch computer monitor, and subtended visual angle of 23.0° horizontally and 17.6° vertically at a viewing distance of 30 inches. The six large instruments were 5.2° in diameter, with 1.4° horizontal and 1.7° vertical separations.

Subject. The subject was a former military P-3 pilot with 1050 hours of flight time, of which 225 hours were in actual instrument condition.

Approach Scenario. Approaches were performed in simulated Instrument Meteorological Conditions with no out-the-window view. In this Instrument Landing System (ILS) approach scenario, the aircraft was initially positioned on the left side of the localizer course, as shown in Fig.2. This resulted in three flight segments with somewhat different task demands. Segment (i) involved straight-and-level flight. The subject was instructed to maintain 1700 ft and 160 kt. The lack of any assigned horizontal course suggests that this segment was the easiest of the three. Segment



(a) Display 1 (D1): The six large instruments are; (top, from left) airspeed indicator (ASI), attitude indicator (AI), altimeter, (bottom, from left) ILS course deviation indicator (CDI), heading indicator (HI), vertical speed indicator (VSI). Flap and gear indicators are at the upper right corner. Thrust indicator is at the lower right corner.



(b) Display 2 (D2): Identical to D1 except that the needles on the ASI and altimeter are removed.

Fig. 1: Displays Used in the Simulation

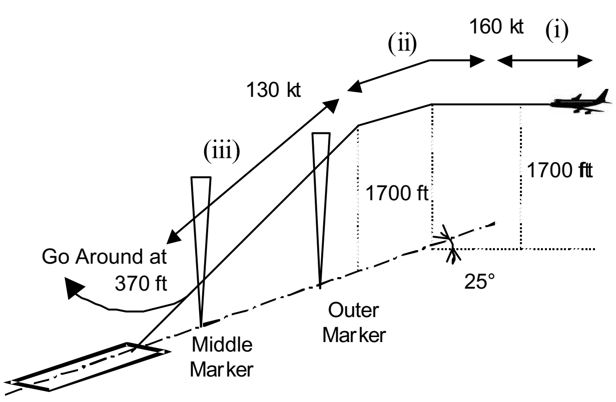


Fig. 2: ILS Approach Scenario

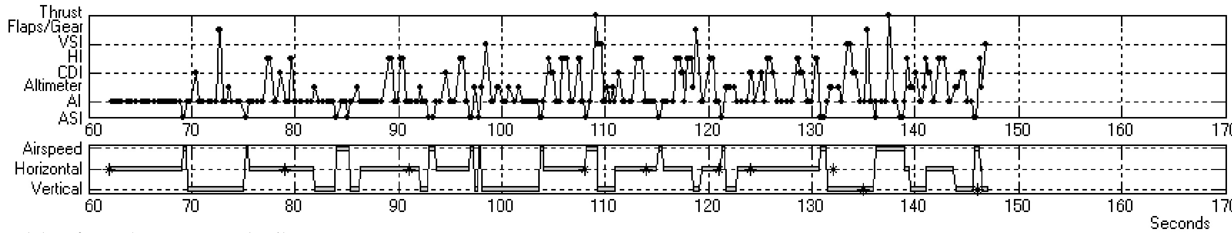
(ii) required a level turn to intercept and track the localizer. The segment started when the subject initiated the left turn. Altitude was to be maintained at 1700 ft. During this segment, the subject was expected to lower the gear and two notches of flaps, and then slow from 160 to 130 kt. This segment involved all three attention tasks (vertical, horizontal, and airspeed), and appeared to be the most demanding segment. The last segment (iii) was descent along the glide path. The segment started when the subject left 1700 ft to begin the descent. The instructions for the subject were to maintain 130 kt and keep both CDI needles at the center. If the aircraft was configured and stabilized well in the previous segment, this segment should be relatively easy to fly. After descent to 370 ft, the subject initiated a go around and the segment ended.

Data-Collection. After several practice approaches with each display, the subject flew six data-collecting approaches that alternated between D1 and D2 on successive approaches. The subject's eye-movements were measured by a head-mounted eye tracker (RK-726PCI/RK-620PC, ISCAN Inc., Burlington, MA) with a magnetic head tracker (InsideTRAK, Polhemus, Colchester, VT). In addition, the subject was asked to verbally report the instrument readings he attended to and his current flight objective (e.g., "localizer moving in," "too low, go up a little") whenever possible. Comments and activity were recorded on videotape. Only the reports referring to the AI, CDI, and any objectives related to them were used in later analyses. After each approach, the subject was asked to subjectively score his spare attention level using a modified Bedford workload scale (Roscoe & Ellis, 1990; Huntley, 1993). At the end of the experiment, the pilot was asked which display he preferred.

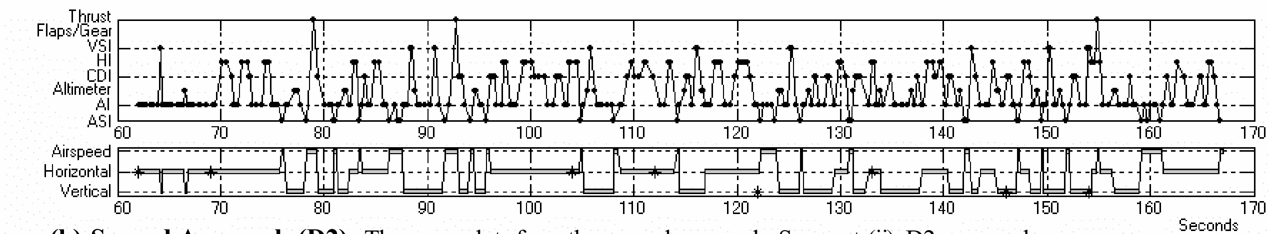
Data Analyses

HMM Analysis. The HMM had the structure described in Table 1, except that there was no turn and slip indicator (T/S), and flap and gear indicators were included in both the vertical and airspeed tasks. The EM and dynamic programming algorithms were run for observed fixation data from each approach, and the resulting estimated task attention sequences were compared with the pilot's verbal reports. The estimated task attention at a given point was considered to match the verbal report if there was a matching report within ±1 sec. The weights on the initial conditions of the *B* matrix were slightly modified and the parameters were re-estimated until the matches with the verbal reports showed no increase.

Flight Technical Error Analysis. The RMS altitude deviations from 1700 ft in Segments (i) and (ii) were computed from samples taken every second.



(a) First Approach (D1): The fixation sequence (top) and estimated task-attention sequence (bottom) from the first approach, Segment (ii). D1 was used. The "*" marks on the bottom plot indicate the tasks implied by the pilot's verbal reports.



(b) Second Approach (D2): The same plots from the second approach, Segment (ii). D2 was used.

Fig. 3: Instrument Fixation Sequence Data and HMM-Estimated Task Sequences

Analogously, the RMS airspeed deviations from 160 kt were measured both in Segment (i) and in Segment (ii) until flaps were lowered to the first notch. The RMS airspeed deviations were also measured from 130 knots both in Segment (ii) after the flaps were lowered to the second notch and in Segment (iii). The RMS localizer deviations in Segments (ii) and (iii), and the RMS glide slope deviations in Segment (iii) were also computed.

Experiment Results

For the six approaches, the HMM estimated tasks matched 79-92% of the subject's verbal task reports. Verifying that the match rate is high is important because that is the only way to assure the HMM estimation results agree with the tasks the pilot actually performed. Fig. 3 (a) and (b) show examples of the instrument fixation data and the corresponding estimated task attention with the tasks verbally reported by the subject.

The HMM analysis revealed interesting display effects on the subject's scanning strategy, which was not so apparent from the fixation data alone.

From the fixation data, first the fixation duration and interval for each instrument were analyzed. Since the distributions of the durations and intervals were positively skewed, the variables were transformed by taking natural logarithms. Successive approaches, testing D1 and D2 in sequence, were paired to constitute a single trial in order to eliminate possible learning or fatigue effects. General Linear Model (GLM) ANOVA (SYSTAT v.10, SPSS Inc.) of the main effect of segment, display, and trial, and of the segment$\times$display cross-effect was performed. The effects of segment and trial were, as expected, statistically significant but not relevant to this investigation. The analysis revealed that the fixation durations on the vertical- and airspeed-task instruments except the CDI (i.e., ASI, AI, altimeter, VSI, and throttle) were all significantly shorter for D2 than for D1[1]. The intervals for ASI and AI were significantly shorter for D2[2]. The segment$\times$display interaction revealed that the ASI intervals were significantly shorter in Segment (i) for D2 (see Fig. 4, $F(2,414) = 3.56$, $p = 0.030$).

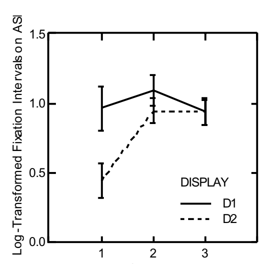


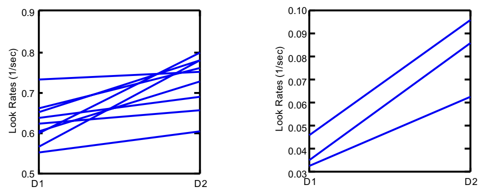
Fig. 4: Effects of Segment $\times$ Display on Least-Square Means of Log-Transformed Fixation Intervals on ASI, with ± 1 Standard Error

In addition, look rates, defined as the number of times per second the subject actually returned to a particular instrument, were computed and examined for display effect. Paired t-tests were performed for each trial to

[1] The display effects on the instrument fixation durations: ASI $F(1,384) = 9.84$, $p = 0.002$; AI $F(1,1027) = 8.66$, $p = 0.003$; altimeter $F(1,347) = 22.5$, $p < 0.000$; VSI $F(1,112) = 7.92$, $p = 0.006$; and throttle $F(1,92) = 4.24$, $p = 0.042$.

[2] The display effects on the instrument fixation intervals: ASI $F(1,414) = 4.22$, $p = 0.041$; AI $F(1,1111) = 5.93$, $p = 0.015$

test the within-trial display effect. The look rate on the AI was significantly higher for D2 than for D1 for all segments taken together (Fig. 5a, $t(8) = 4.38$, $p = 0.002$). The VSI also showed a higher look-rate for D2 in Segment (ii) (Fig. 5b, $t(2) = 6.36$, $p = 0.024$).



(a) AI in all segments (b) VSI in Segment (ii)

Fig. 5: Look Rates – Within Trial Display Effects

The above analyses indicate that, when there was no needle movement to watch on the ASI and altimeter, the subject fixated on the vertical- and airspeed-task instruments (except the CDI) only briefly, and visited the ASI (decreased interval), AI, and VSI (increased look rates) more often. Interestingly, the missing needles on the ASI and altimeter affected fixation durations and look rates on the other instruments whose needles were not missing. For the VSI, the increased look rate may indicate that the subject used it as a substitute for the missing needle of the altimeter to determine rate of descent.

What are not clear from these fixation-level analyses are the effects on task switching, which may affect more directly the overall flight performance. Despite of the existence of the display-related effects on the instrument fixation level, the GLM ANOVA revealed no significant main effect of display on the HMM-estimated task duration. Significant segment×display cross-effect on the airspeed task duration indicated that the airspeed task duration was longer for D2 in Segment (i) and shorter in Segments (ii) and (iii) than for D1 (Fig. 6, $F(2,376) = 5.25$, $p = 0.006$). There was no display-related effect on the task interval.

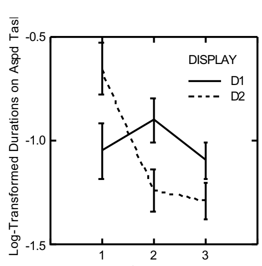


Fig. 6: Effects of Segment × Display on Least-Square Means of Log-Transformed Durations on Airspeed Task, with ±1 Standard Error

The task rates, defined as the number of times per second the HMM estimated the subject returned to a particular task, were also computed and examined for the display effects by paired t-tests. The rates were found to be significantly higher for D2 than D1 for the vertical and airspeed tasks taken together in Segment (ii) ($t(5) = 4.75$, $p = 0.005$).

These results indicate that with D2, which was missing indicator needles on the ASI and altimeter, the subject scanned the vertical-task instruments more rapidly but maintained about the same levels of the vertical-task durations and intervals as with D1. In Segment (ii), the decreased ASI duration, decreased airspeed task durations, and increased airspeed task rate imply that the vertical and horizontal tasks were frequently interrupted by brief sampling of ASI when D2 was used, but the samplings were brief enough not to change the durations and intervals of the vertical and horizontal tasks. In Segment (iii), the sampling of ASI was also brief with D2, but not as often as in the Segment (ii). The Segment (i), on the other hand, showed different trend from (ii) and (iii); The decreased ASI duration, decreased ASI interval, and increased airspeed-task duration mean that, when D2 was used, the subject performed brief and frequent scanning of the ASI, AI and thrust indicator, and stayed on the airspeed task longer than when D1 was used. Note that the Segment (i) was the least busy segment, and the subject had a plenty of time to attend the airspeed task.

These display effects were not reflected in any of the RMS measures of flight technical error and subjective workload scores. Paired t-test revealed no significant display effect in either of those variables. This could be explained by assuming that the task durations and intervals were not affected by the display. If the amount of time spent for each task was not changed by the displays, this may explain why the performances of each tracking task were approximately same. Although there was no significant difference in the flight performance and the subjective workload, the subject did prefer D1 to D2. This could be due to being forced to alter scanning strategy when D2 was used.

This analysis was based on the data from one subject. Clearly more subjects are needed to draw general conclusions relating to the specific display effects. However the analysis demonstrates the application and insights provided by HMM based task analysis.

Discussion

HMM task analysis appears to usefully complement traditional eye movement fixation analysis. The combined analysis revealed the subtle effects of display format that traditional measures such as the flight technical error and subjective workload score did not detect. The HMM analysis allowed investigators to estimate the relative attention devoted to the major cognitive tasks performed repetitively during an approach scenario, i.e. the "attention budget" of the pilot during the approach (Fig 3). This was not apparent from the fixation data alone. Results of the analysis provide insight on the origins of pilot mental workload, and its relationship to phase of flight. Since the HMM analysis technique is inherently a statistical identification method, it less useful for short records and identification of tasks performed only infrequently.

Many aircraft are now equipped with electronic flight information systems (EFISes) including primary flight displays (PFDs), and some have head-up displays (HUDs). It has been a challenge to analyze the effects of these advanced, highly integrated displays because multiple pieces of information tend to be shown at one place on these displays. The HMM analysis method demonstrated, however, may provide a key analysis tool to examine the effects of these advanced displays on pilot performance since it can accommodate the ambiguity of display information.

The experiment also demonstrated that using the HMM analysis, much more information can be derived even from a relatively small number of simulation trials. Currently, certifying a new cockpit display requires a large number of simulation and actual flight demonstrations to prove pilots' performance improvement. For instance, in the certification test of a HUD in Cat III operation, over 1000 simulated and actual touchdowns were performed (Desmond & Ford, 1984). The HMM analysis may help reduce the required number of flight demonstrations.

Conclusion

A new method of analyzing the pilots' instrument scanning using a Hidden Markov Model identification technique is described. Data from a flight simulator experiment was analyzed to demonstrate the advantages of the HMM analysis technique. The task attention sequence estimated by the HMM showed good matches with a pilot's verbal reports made during the simulation. Two slightly different displays were used in the experiment. Display effects were subtle, and produced no significant changes in flight technical error or subjective workload score. However the HMM analysis found significant changes in approach task performance metrics, and provided important insights on the pilot's attentional budget. The task results accounted for changes seen in the instrument fixation data. Results are preliminary, but HMM analysis has been demonstrated to provide an important adjunct for interpretation of pilot eye movement data, and empirical estimation of the moment-to-moment attentional shifts between the repetitive cognitive tasks associated with instrument flight. Because of these added insights, the proposed method may also provide a useful analysis tool to examine the advanced cockpit displays.

Acknowledgement

This study was supported by FAA Office of the Chief Scientific and Technical Advisor for Human Factors, AAR-100, working in cooperation with the FAA Transport Airplane Directorate, AMN-111. We thank our subject, and also Prof. Thomas Sheridan, Prof. James Kuchar, Dr. Alan Natapoff of MIT, and Andrew Kendra of Volpe Center.

References

Bradley, P. (1996, March). Dials Versus Tapes. *Business & Commercial Aviation,* 70-74.

Desmond, J. P., & Ford, D. W. (1984). Certification of a Holographic Head-Up Display System for Low Visibility Landing. *Proc. AIAA/IEEE 6th Digital Avionics Systems Conference*, 441-446.

Huntley, M. S. (1993). *Flight Technical Error for Category B Non-Precision Approaches and Missed Approaches Using Non-Differential GPS for Course Guidance* (DOT-VNTSC-FAA-93-17). Cambridge, MA: US Department of Transportation, Volpe National Transportation Systems Center.

Pennington, J. E. (1979). *Single Pilot Scanning Behavior in Simulated Instrument Flight* (NASA-TM-80178). Hampton, VA: NASA Langley Research Center.

Rabiner, L., & Juang, B. H. (1993). *Fundamentals of Speech Recognition.* Englewood Cliffs, NJ: Prentice Hall.

Roscoe, A., & Ellis, G. (1990). *A Subjective Rating Scale for Assessing Pilot Workload in Flight* (TR90019). Farnborough, UK.: RAE.

Weinstein, L. F., Gillingham, K. K., & Ercoline, W. R. (1994). United States Air Force Head-Up Display Control and Performance Symbology Evaluations. *Aviation, Space, and Environmental Medicine, 65*(5), A20-A30.