THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, 11(1), 15-31

Vertical Navigation Displays: Pilot performance and workload during simulated constant-angle-of-descent GPS approaches

Charles M. Oman,[2] Andrew J. Kendra,[1] Miwa Hayashi,[2]
Mary J. Stearns,[1] and Judith Bürki-Cohen[1]

[1] Department of Transportation/Volpe National Transportation
Systems Center, DTS-79
Cambridge, MA

and

[2]Man-Vehicle Laboratory
Massachusetts Institute of Technology

Submitted June 28, 1999; Revised March 26, 2000

ABSTRACT

This study compared the effect of alternative graphic or numeric vertical navigation (VNAV) aircraft cockpit displays on horizontal and vertical flight technical error (FTE), workload and subjective preference. Displays included: a) a moving map with altitude range arc, b) the same format, supplemented with a push-to-see profile view, including a vector flight-path predictor, c) an equivalent numeric display, and d) a numeric non-VNAV display. Sixteen pilots each flew four different approaches with each format in a Frasca 242 simulator. Our VNAV displays reduced vertical FTE by as much as a factor of two without increasing workload. Relative advantages of the graphics formats are discussed.

Supported by FAA Office of the Chief Scientific and Technical Advisor for Human Factors, AAR-100 via DOT/Volpe National Transportation Systems Center, Cockpit Human Factors Program, DTS-79 Contracts DTRS-57-P-80766 and DTRS-57-97-C-00022-TTD02.

Keywords: cockpit displays, navigation displays, flight management systems, vertical navigation, VNAV

INTRODUCTION

Aircraft vertical navigation (VNAV) systems provide vertical guidance to a specified waypoint at a particular altitude, often along a path defined by a line joining two waypoints at specified altitudes. Alternatively the path may be defined by a vertical angle from a given waypoint, or be a fuel-use-optimized curved path between waypoints (FAA AIR-120, 1988; RTCA-SC-159, 1996). Most flight management system (FMS) equipped aircraft and some area navigation (RNAV) equipped aircraft have VNAV capability. VNAV displays allow the pilot to plan and check a VNAV route, monitor VNAV function when the autopilot or FMS is flying the aircraft, or manually fly the computed path. Graphic VNAV displays probably enhance vertical situation awareness. VNAV systems which incorporate some form of flight path predictor display have been popular, as pilots can confidently adjust power and drag to comply with air traffic control waypoint altitude constraints, and don't have to rely on fallible rule-of-thumb techniques (e.g. "Rule-of-Three"; Curry, 1979). Vertical navigation capability and map displays are planned for the next generation of GPS (global positioning system) Wide Area Augmentation System navigation receivers (RTCA-SC-159, 1996) to make it easier for pilots to fly complex FMS arrivals and approaches within narrow required navigation performance envelopes. Air carriers are also considering the use of constant-angle-of-descent procedures and VNAV equipment to reduce the likelihood of controlled flight into terrain (CFIT) on non-precision instrument approaches (Gregory, 1996). VNAV displays with flight path predictors will likely also appear in business aircraft, and even in GPS navigators used aboard low end general aviation aircraft. In such aircraft, VNAV displays may prove useful as supplementary navigation displays for both non-precision and emergency instrument approaches (e.g., managing a descent to the nearest airport after partial or complete engine failure).

Unfortunately, the basic human factors requirements for VNAV displays have so far received little research attention. A Federal Aviation Administration Human Factors Team report (FAA, 1996) noted that additional research on VNAV displays was needed, but nonetheless recommended reduction or eventual elimination of instrument approaches which lack vertical path guidance, in the interest of flight safety. FAA regulations and human factors guidelines for VNAV displays currently address only waypoint annunciation and the sensitivity of the vertical path deviation indicator (FAA AIR-120, 1988). The latter were based on empirical studies (Jensen & Roscoe, 1973; VanderKolk & Roscoe, 1973), but other than these there are no published data comparing pilot performance and workload among different VNAV display formats prior to the present investigation. An Society of Automotive Engineers aerospace behavioral engineering technology (G-10) subcommittee on vertical situation awareness displays was recently established to propose standards for performance/planning VNAV displays, and eventually also for VNAV displays used for flight path control. Loss of altitude and terrain awareness remain two of the most serious safety problems in air transport and general aviation (GA), causing many fatal accidents involving CFIT (e.g., Bud et al., 1997; Khatwa & Roelen, 1996). Nonfatal altitude deviation incidents have been reported to NASA's Aviation Safety Reporting System at an average rate of more than one per hour (Palmer et al., 1993). The complexity of FMS VNAV programming and poor FMS VNAV mode awareness have also been identified as contributing factors (Palmer et al., 1993).

The basic component of a VNAV display is a vertical course deviation indicator (CDI), analogous to an Instrument Landing System glideslope needle, which shows vertical path error. Many displays also include some form of flight-path prediction information in order to offset the workload increase associated with manually flying a glideslope needle through an entire departure or arrival. Perhaps the simplest method is to numerically display a recommended vertical speed which will keep the aircraft converging with the desired path. Some systems continuously display the distance to the planned top-of-climb or bottom-of-descent altitude intercept. On some VNAV equipment this is shown in numeric form. Alternatively, it is sometimes presented as symbol on a "moving map" display used for lateral navigation (LNAV). For example, on Boeing/Honeywell FMS displays, a moving green "altitude range arc" continuously shows where the aircraft will reach the pre-selected altitude. On Airbus aircraft, moving "Top-of-Climb" and "Bottom-of-Descent" bent-arrow symbols perform a similar function. Pilots adjust vertical speed to keep the altitude range arc on the display superimposed over the next waypoint.

Many altitude deviation incidents have been attributed to the lack of explicit vertical situation information in the cockpit (Vakil et al., 1996). Since moving map displays do not depict the aircraft's vertical situation explicitly, some manufacturers have experimented with supplemental "profile" displays, which graphically depict altitude vs. distance en route in a manner analogous to the profile (elevation) view on a paper instrument approach plate. One model includes a flight-path predictor vector (Gulfstream/Honeywell G4/5, Hughes, 1995a). Typically the profile view has been located in a narrow area beneath the moving map display, so altitude resolution has been a concern. Several recent simulator studies (Chappell et al., 1997; Hutchins, 1996; Prevot, 1998; Vakil et al., 1996) suggest that profile displays could increase vertical situation and FMS mode awareness. We believe that supplemental profile displays are also useful when planning and checking FMS and RNAV routes because vertical errors are relatively hard to notice using only a map view. Given these putative advantages, it is important to establish how a pilot's workload and performance are influenced during manually flown approaches when VNAV information is presented in different ways. For the purposes of the present experiment, we designed three generic VNAV display formats which present the same vertical information in physically and cognitively different ways (numeric, map, and map/profile formats). To provide a basis of comparison, we also asked our participants to use a numeric "LNAV" display, which had no VNAV vertical path deviation or flight path prediction information, but is representative of GPS displays currently used by general aviation pilots to fly non-precision (i.e. non-ILS) approaches. Although our displays were sized (and in the case of the map/profile display, overlaid) so as to fit a small general aviation cockpit, the basic non-precision approach flying task we studied is similar in air transport and business aircraft as well. Our goal was to answer the question: when pilots fly non-precision and rely on VNAV equipment or rule-of-thumb techniques to manage descents, how much improvement in vertical and horizontal flight technical error (FTE) and workload results ? We did not specifically measure vertical situation awareness, but expect that if pilots are able to fly non-precision approaches more easily using VNAV displays, spare attention could be more readily allocated to maintaining vertical situation awareness.

METHOD

Displays were generated by custom software on an Avidyne 5RR Multifunction Flight Computer. On the instrument panel of the Frasca 242 light twin flight simulator, the Avidyne's 5 inch diagonal (¼ VGA AMLCD) color display was mounted in the avionics stack to the right of the pilot's primary field of view (see Figure 1). One LNAV format and three VNAV formats described below were used in this simulation. For all three VNAV formats, vertical path deviation was displayed not only on the Avidyne display, but also concurrently in the pilot's primary field of view using the glideslope needle on the horizontal situation indicator (HSI).

(Figure 1)

Figure 2 shows the appearance of the LNAV and three VNAV formats used. In the VNAV formats, the minimum altitude on the current leg was shown numerically in the upper-right-hand corner of the display, and updated at waypoint passage. All displays updated at approximately 1 Hz. The detailed descriptions for each format are as follows:

(Figure 2)

Numeric LNAV format ("L"): Numeric LNAV format that displays only cross-track error, as it might generically appear on a GPS RNAV used for non-precision instrument approaches. Pilots had to manage their descent using rule-of-thumb techniques based on the altimeter and vertical speed indicator.

Numeric VNAV format ("N"): Similar to the L format, but with additional numeric VNAV data such as vertical path error, range to altitude intercept, instantaneous vertical speed, and recommended vertical speed to remain on the programmed flight-path.

Map VNAV ("A"): A track-up moving map navigation display with a green altitude range arc and a magenta line depicting the programmed approach route. The five-mile-long velocity vector in front of the triangular aircraft symbol fixed at the bottom part helped pilots judge track angle error and infer the display scale. Waypoints were shown using a simplified composite of Boeing (Honeywell, 1996) and SAE (Young, 1997) symbology. The active waypoint was always magenta, while all others were white. A magenta-colored diamond-shaped "football" repeated the vertical path error information available on the pilot's CDI. Map scale was 15 miles (vertical) × 20 miles (horizontal) before the final approach fix (FAF), and automatically zoomed by a factor of 3 after passing the FAF.

Map/Profile VNAV ("P"): A layered display consisting of the A-format moving-map-view window replaceable with a profile-view window by pushing a thumb switch on the yoke. These windows were layered because the small size and limited resolution of the Avidyne display prevented simultaneous presentation of both windows. The map view was the default display before the FAF, and the profile view became the default after the FAF. In the profile view, as the airplane progressed, the display scrolled from right to left. As the aircraft descended, the aircraft symbol moved down the left side of the display, but did not rotate even if the aircraft pitched. The aircraft's predicted flight-path was shown with the five-mile-long velocity vector. Waypoints appeared as vertical bars. The minimum altitude on the current leg was shown by a yellow horizontal line. Its intersection with the aircraft velocity vector was identified by a moving green cross which corresponded to a side view of the altitude range arc in the map view. Waypoint names were available only in the map view. The vertical range of the profile view was 9000 feet before the FAF, and 3000 feet afterward. The scale on the abscissa of the profile view was always equal to the ordinate scale of the map view.

Sixteen (960-18000 hour) multi-engine, instrument rated pilots each manually flew 5-7 practice approaches and 16 trial approaches (4 approach types × 4 display formats) with patchy turbulence and altitude dependent wind, using aircraft dynamics resembling a Piper Aztec. Approaches were blocked in sets of 4 by approach type crossed with display format type, so that each pilot flew all four approaches once with each of the four display formats. The order of display presentations varied across groups of four pilots, but in all cases the sequence of presentations during approaches 1-8 was reversed for approaches 9-16 to help balance learning effects.

A computer generated display of clouds and runway was visible out the front window. Prior to each approach, pilots were given the paper approach plate and Automated Terminal Information System message in written form for one of the four non-precision approach procedures (approach types) in Table 1.

(Table 1)

Pilots were instructed to fly the entire approach at 120 knots and in accordance with practical instrument flight test standards (i.e., plus or minus 10 knots airspeed, plus or minus 100 feet altitude; FAA, 1994), maintaining a constant angle of descent on all descending legs, all the way to the runway threshold. Each approach included an initial five-mile level-flight segment and up to two turns. As the aircraft approached the published minimum descent altitude, pilots had to make a land or go-around decision, depending on whether the runway was or was not visible. Air traffic control voice communications from approach and tower were simulated on each approach. When flying with the LNAV format, pilots had no VNAV information other than that available from the altimeter and vertical speed indicator, and had to manage their descent using rule-of-thumb techniques. After each approach, pilots rated their subjective workload before and after the FAF using a modified Bedford scale as a measure of spare attention

(Roscoe & Ellis, 1990; Huntley, 1993). At the end of the test session, pilots completed a 38-item post-session questionnaire consisting of three parts. The first part asked 28 specific questions about participants' opinions of the display formats. The six items in the second part asked the pilots to compare each of the formats on a "head-to-head" (HTH) basis, placing a mark on a visual analog "strength-of-preference" scale based on which format they would choose if they had to make an approach in marginal weather. The third part of the questionnaire contained four items asking pilots to rank the four formats in terms of ease of interpretation (EI), ability to fly accurately (FA), effect on overall workload (OW), and overall preference (OP), respectively.

RESULTS

Root Mean Square Altitude Error

The Root Mean Square (RMS) altitude error was computed for each of the 16 approaches flown as a measure of vertical flight technical error. As the four approach types used in the simulation had different path lengths and turning patterns, three two-mile segments from the path of each approach was extracted in order to compare flight technical error across different approach types. The three segments chosen were 1) a level segment, before the aircraft began to descend, 2) a descent segment, while the aircraft was descending at a nominally constant angle, and 3) a final approach segment, ending at the breakout from the cloud ceiling or go-around decision point.

Inasmuch as the distributions of RMS altitude errors of each segment from all of 256 trials were skewed to the right, the RMS altitude errors were natural-log-transformed to make the resulting distribution approximately normal. The log-transformed RMS altitude errors from the three segments were separately analyzed by univariate ANOVA (Systat v. 8.0, SPSS, Inc.) to compute mean square (MS) errors. The independent variables (main effects) were participant, approach type, and display format. Since participants and approaches were assumed to be sampled from larger populations, they were treated as random effects while display format was considered as a fixed effect.

Table 2 shows the F-ratio, F'-ratio, and corresponding p-values of the univariate mixed-model ANOVA hypothesis tests (see Appendix for detail) on the RMS altitude errors. The R-squared values appear on the top row. As expected, significant main participant effects were found on all three segments. The main approach effect was also significant except on the level segment, which is reasonable because the level segments were always on the beginning of each approach path, and were similar among approach types. The display effect, our main focus, also showed significant differences except for the level segment. The latter finding was expected because the VNAV features were important only during descending flight. A significant interaction effect of subject and approach was observed on the level segment. On the descent segment, a significant interaction of approach and display was found.

(Table 2)

The least-square estimates of the mean effect on RMS altitude error of each display type with error bars (plus or minus 1 SEM) are shown in Figure 3. Note that the vertical axes in Figure 3 Top, employ log scale. For instance, on the final segment, the least-square mean for the L format is approximately $\exp(4.8) \approx 122$ feet and for P format is approximately $\exp(4.2) \approx 67$ feet. Thus the altitude error associated with the P format was reduced by almost a factor of two as compared to the L format.

(Figure 3)

As shown in Figure 3, VNAV information significantly reduced final approach RMS altitude error for all participants and approaches. Tukey pairwise testing of the final segment display effects showed that the graphics VNAV formats (A and P formats) significantly reduced the altitude error as compared to the L format ($p < 0.001$ for A vs. L, and P vs. L). In addition, the P format significantly reduced the altitude error compared to the N format ($p < 0.001$).

The interaction effect of approach and display on the descent segment RMS altitude error was significant. Therefore, Tukey pairwise comparison tests on the display effect were made for each approach separately. Display effects were significant for the Los Angeles International (LAX) and Athens (ATH) approaches (Figure 4), which respectively had the steepest descent segment (see Table 1) and highest workload demand (see Workload, below). The VNAV formats helped pilots to reduce the altitude errors especially with the two relatively challenging approaches, and never hurt performance relative to the L display. With the LAX approach, the altitude error was reduced significantly when any of the VNAV formats was used comparing to L format ($p < 0.040$ for N vs. L, $p < 0.004$ for A vs. L, and $p < 0.001$ for P vs. L). With the ATH approach, the altitude error was reduced significantly when either of the graphical VNAV formats was used compared to the L format ($p < 0.001$ for A vs. L, and $p < 0.001$ for P vs. L).

(Figure 4)

Root Mean Square Cross-Track Error

In order to measure the effect on concurrent lateral flight technical error, analogous hypothesis tests were performed on the log-transformed RMS cross-track errors. The analysis showed significant main effects of participant and approach on all three segments,[1] but a main effect of display only on the descent segment. A significant interaction of approach and display effects was found on the final segment.

(Figure 5)

Figure 5 shows the least-square estimates of the mean effect of display type on the descent segment with plus or minus1 SEM bars. The graphical VNAV formats (A and P) reduced the cross track error significantly when compared to the L format (pairwise Tukey test, $p < 0.001$ for A vs. L and $p < 0.002$ for P vs. L). We suspect this is because the pilot could directly visualize track angle error when using the graphical A and P formats which provide map (plan) view. Analog track angle error information has been shown to significantly reduce horizontal flight technical error (Oman et al., 1995; 1996). The A format provided significant improvement in the cross-track error even when compared to the N format ($p < 0.015$ for A vs. N).

On the final segment, a significant interaction of approach and display was found. Tukey pairwise comparisons showed that for three of the approach types, including the two most challenging ones, there was no reliable effect of display on cross-track performance. For the approach to Boston/Logan Airport (BOS), however, the cross track performance on the final segment was better with the A format than with the N format ($p < 0.017$). None of the other comparisons were significant by a Tukey HSD within the BOS approach except only a trend by participants to perform better with the A format than with the P format ($p < 0.090$).

Workload

The average workload scores for all formats both before and after the FAF fell in the middle of the “tolerable” range on the Bedford Scale. An univariate ANOVA, this time including the trial number effect as a covariate in addition to the participant, approach type, and display format effects, showed a significant effect of the trial number for both before and after the FAF (the trial number effect was omitted from the RMS altitude error and RMS cross-track error analyses above because it did not show any significant difference). Furthermore, the mixed-model ANOVA hypothesis showed a significant main effect of participant for both before and after the FAF and an approach type effect only for before the FAF.[2] A Tukey pairwise comparison test on the approach effect for before the FAF showed that the ATH approach significantly increased pilots' workload scores compared to other approaches ($p < 0.018$ for ATH vs. JFK, $p < 0.001$ for ATH vs. LAX, and $p < 0.046$ for ATH vs. BOS).

No significant display effect was found for either before or after the FAF. This shows that the performance improvements associated with VNAV formats as compared to traditional LNAV/altimeter did not come at the expense of any consistent increase in workload.

Participants’ Preferences

To analyze the post-session questionnaire data, HTH preference indications were converted to a numeric score using a tournament scoring method, and then converted to ranks. Rank sums across all 16 pilots for each of the 5 preference measures were computed. Resulting preference ranks are shown in Table 3.

(Table 3)

Participants ranked the P format as best in both tournament (HTH) and direct (OP) measures. The A format ranked first in terms of ease of interpretation (EI) and effect in reducing workload (OW).

Participants also believed that they could fly most accurately using the P format (FA), and that was consistent with their altitude error data (Figure 3). In the questionnaire, all said that the map view should be the default prior to the FAF, but 9 of 16 participants believed that the map should remain as the default after passing the FAF, and the profile view should remain "push-to-see." Thirteen of the 16 participants agreed that the map/profile switch should be mounted on the yoke rather than on the instrument panel. Eleven of the 13 pilots said they never or rarely had any problems interrelating the map and profile views (this question was omitted for the first three participants). Several pilots suggested that waypoint names and numeric altitudes be shown in both map and profile views.

The N format was ranked third on all 5 preference scales. Ten of the 16 pilots reported they flew the N format by comparing the Instantaneous Vertical Speed (IVS) and Desired Vertical Speed (DVS) values rather than distance measures such as Distance to Waypoint and Range to Altitude Intercept. Several commented they found the IVS/DVS presentation more intuitive, and noted they could cross check IVS with their barometric instrument.

Friedman ANOVA showed that the five measures were significantly concordant (Friedman test statistic 13.6, df = 3, $p < 0.004$). As Table 3 indicates, most of the pilots preferred the graphics formats (A and P) over the numeric formats (L and N). The individual data, however, also showed that there were still some pilots who strongly preferred the numeric formats. The questionnaire asked each pilot whether they had prior experience with moving map displays. We noted that all 8 pilots who answered "yes" preferred one of the two graphic formats, whereas all 4 who ranked the numeric format (N) ahead of the graphic format (A or P) answered "no". However, we could not demonstrate a statistical relationship between vertical FTE and display experience.

Fifteen of the 16 pilots found it extremely useful to have the minimum altitude on each leg automatically displayed numerically. All of 16 pilots said they had enough practice approaches so they understood how to fly with each type of display. Nonetheless, 13 of 16 pilots recalled making significant blunders in vertical navigation; in most cases they recalled making several (2-8) of them.

CONCLUSION

The analysis of RMS altitude error data showed that the three VNAV formats we tested (N, A, and P formats) significantly reduced vertical flight technical error on descent and final segments of simulated constant-angle-of-descent approaches as compared to traditional methods employing LNAV data (L format), an altimeter, and rules-of-thumb. While precision was not formally required, we asked pilots to fly a stabilized approach, and due to the turbulence used in the simulation, the magnitude of the altitude deviations observed with some displays was large enough to be operationally unacceptable. The improvement in vertical FTE on the descent and final segments associated with the Map/Profile format (P format) approached a factor of two. Graphical formats (A and P) were particularly helpful on the descent segments in the two most challenging approaches of the four approaches flown. The vertical performance on the descent and final segments was slightly superior with the Map/Profile format as compared to the Map format (A format), but the difference was not statistically significant. On the final segment, the vertical performance with the Map/Profile format was significantly better than the numeric format.

Display format also affected lateral flight technical error. Cross track error was significantly reduced when the graphic formats during the descent segments was being used. This was expected, since these formats provided a moving-map plan view. With the Map/Profile format, the profile view became the default after the final approach fix, since horizontal maneuvers are not normally made on final approach. However, we noted that on one of the four types of approaches flown, final approach cross track error increased with the Map/Profile format. The majority of our pilots suggested that the Map/Profile format could be improved by keeping the map view as the default throughout the entire approach, with the profile view remaining “push-to-see.”

Our results provide quantitative support for the widely held view (e.g., Hughes, 1995b) that a moving map with altitude range arc allows performance equivalent in most respects to profile displays. However, supplementary profile views probably help improve vertical situation awareness. We demonstrated that a profile view can be layered in the same display space as a map view, usually without an increase in lateral or vertical FTE as compared to the map view alone.

No significant difference was found in workload score across display formats. We believe our pilots tried to work about as hard as they reasonably could. As a result, they may effectively have kept their workload constant, and it was their performance which varied across approach and display.

Our pilots preferred the Map/Profile format over the other formats tested. Eighty five percent of the pilots tested said they never or rarely had problems interrelating the map and profile views. It may have been important that the display scales zoomed together, the airplane symbol was consistently on the left, and the active waypoint was clearly distinguishable in both views.

The update rate of the displays was approximately 1 Hz, comparable to that of C129 TSOed GPS receivers, but probably at the low end of the useful range. In order to perform the flying task, pilots had to devote most of their instrument scan to the primary attitude instruments and CDI. Information from the Avidyne's numeric or graphic vertical flight path predictors was probably useful in helping pilots judge the rate of change of vertical and horizontal path error (i.e., develop manual control lead), so that flight technical error was reduced. Experiments with a faster display update rate may well show larger performance improvements.

In the future, if minimum performance standards for VNAV displays are to be developed on a scientific basis, additional research will be needed to evaluate other potential graphical VNAV display formats (e.g., perspective views), and the test paradigm expanded to assess not only flight technical error, but also altitude and terrain awareness in both manual and autoflight modes, and the ability to detect route errors in flight planning mode. For profile displays, important questions remain, such as: how essential is it that the altitude/range ratio remain constant? Should the aircraft symbol or the map move vertically? Should the profile view always be from the same side of the flight-path or match the view on the published approach plate?

Our Avidyne VNAV display software was subsequently successfully installed and informally evaluated in Volpe's Piper Aztec equipped with a KLN90B GPS receiver and Shadin Air Data computer. This proof-of-concept demonstration (Oman & Kendra, 1998) shows that useful graphical VNAV display features, including flight path predictors, can be added at relatively low cost to current-generation GPS navigators.

ACKNOWLEDGMENTS

This study was conducted for the FAA Office of the Chief Scientific and Technical Advisor for Human Factors, AAR-100. Supported by FAA Office of the Chief Scientific and Technical Advisor for Human Factors, AAR-100 via DOT/Volpe National Transportation Systems Center, Cockpit Human Factors Program, DTS-79 Contracts DTRS-57-P-80766 and DTRS-57-97-C-00022-TTD02.

We thank our 16 participants, and also Thomas McCloy, Eleana Edens, and Thomas Imrich of FAA, Don Sussman, Ray Shih, Melissa Bud, Frank Sheelen, Jack Giurleo, and John Turner of Volpe, and Alan Natapoff, Scott Rasmussen, Chris Carr, and Nick Patrick of MIT, and also Brian Kelly, Bruce Turner, and Tom Twiggs.

APPENDIX
F' Ratio Estimation Method

The experiment design includes three factors: participant (A), approach type (B), and display format (C). The participants and approach types were randomly picked up from a general set, and are considered random effects, while the display formats used in this experiment were considered fixed effects.

Our procedure for estimating F ratios was conservative: Following Winer (1971), if *p*, *q*, and *r* are the number of levels in factor A, B, and C, respectively, and *P*, *Q*, and *R* be the number of levels in factor A, B, and C in general, respectively. Then the expected value for Mean Squares (MS) of the main effect, for example A, is:

$$MS_A = \sigma_\varepsilon^2 + q(1-r/R)\sigma_{AC}^2 + r(1-q/Q)\sigma_{AB}^2 + qr\sigma_A^2 ,$$

where ε is the measurement error. A σ^2_{ABC} term (three-factor interaction) effect was omitted because the number of observations in each cell was unity. The terms, *p/P* and *q/Q*, were approximated as 0 (random effects), while *r/R* was 1 (fixed effect). The expected value for MS of two-factor interaction effect, for instance of AB, is

$$MS_{AB} = \sigma_\varepsilon^2 + r\sigma_{AB}^2 .$$

The expected MS, the null hypotheses under test, and the associated F-ratios are shown in Table 4. Since the appropriate denominator term for the F-ratio was not available for testing $\sigma_C^2 = 0$, an F'-ratio (Quasi-F Ratio) was used to approximate its value (Winer, 1971). The degrees of freedom for the denominator were approximated as

$$df_{denom} = Round\left(\frac{(MS_{BC} + MS_{CA} - MS_\varepsilon)^2}{MS_{BC}^2/df_{BC} + MS_{CA}^2/df_{CA} + MS_\varepsilon^2/df_\varepsilon} \right) ,$$

where *Round* rounds to the nearest integer.

REFERENCES

Bud, M. J., Mengert, P., Ransom, S., & Stearns, M. D. (1997). *General aviation accidents, 1983-1994: Identification of factors related to Controlled Flight Into Terrain (CFIT) Accidents* (Final Report DOT/FAA/AAR-100-97-2). Washington, D.C.: DOT Office of Aviation Research.

Chappell, A. R., Crowther, E. G., Mitchell, C. M., & Govidaraj, T. (1997). The VNAV Tutor: Addressing a mode awareness difficulty for pilots of glass cockpit aircraft. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 27(3), 372-385.

Curry, R. E. (1979, November). *Human factors of descent energy management*. Paper presented at the 18th IEEE Conference on Decision and Control, Ft. Lauderdale, FL.

FAA (1994). *Instrument Rating Practical Test Standards* (FAA-S-8081-4B). Washington, D.C.: Federal Aviation Administration Flight Standards Service. Available from FAA, the GPO, or online as a .pdf
ftp://ftp.metroflight.w1.com/metroflight.w1.com/pts_ir1.pdf.

FAA (1996). *The Interfaces Between Flightcrews and Modern Flight Deck Systems*. FAA Human Factors Team Report. Abbott, K., Slotte, S. M., & Stimson, D. K. eds. Federal Aviation Administration, Washington, DC.

FAA AIR-120 (1988). *AC 20-129 Airworthiness Approval of Vertical Navigation (VNAV) systems for use in the US National Airspace System (NAS) and Alaska* (Advisory Circular): Federal Aviation Administration. US Department of Transportation, Publications Section, Washington, DC.

Gregory, J. W. (1996). *CFIT - Procedure design considerations: use of VNAV on conventional non-precision approach procedures* (OCP-WG-WP 4.18). St. Petersburg, Russia: Obstacle Clearance Panel Working Group ICAO (International Civil Aviation Organization), Document Sales Unit 999 University Street, Montreal, Quebec H3C 5H7, Canada.

Honeywell (1996). *Boeing 747-400 Flight Management System Pilot's Guide*. (C28-3641-24-00 ed.). Honeywell Air Transport Systems, Phoenix, AZ.

Hughes, D. (1995a, February 6, 1995). Gulfstream using vertical profile display. *Aviation Week and Space Technology, vol. 142, no. 6*, p. 55.

Hughes, D. (1995b, January 30, 1995). Incidents Reveal Mode Confusion. *Aviation Week and Space Technology, vol. 142, no. 5*, p. 56.

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

Hutchins, E. (1996). *The Integrated Mode Management Interface Final Report* (Final Report), Hutchins, Edwin, California Univ. Dept. of Cognitive Science, San Diego, CA, USA, June 1996, 204p, NASA-CR-202220, NASA Scientific and Technical Information Program Office, Langley Research Center, VA.

Jensen, R. S., & Roscoe, S. N. (1973). *Flight tests of pilotage error in area navigation with vertical guidance* (Technical Report FAA-RD-72-126). Champaign, IL: Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-Champaign.

Khatwa, R., & Roelen, A. L. C. (1996). An analysis of controlled-flight-into-terrain (CFIT) accidents of commercial operators, 1988 through 1994. *Flight Safety Digest*, 15(4/5), 1-45.

Oman, C. M., & Kendra, A. J. (1998) e-publication, viewable with a web browser at http://www.volpe.dot.gov/opsad/vnd.html. Contains detailed photos of the simulator and displays used in experiment.

Oman, C. M., Huntley, M. S. Jr., Rasmussen, S. A., & Robinson, S.K. (1995). *The use of analog track angle error display for improving simulated GPS approach performance* (Report, DOT-VNTSC-FAA-95-29). Cambridge, MA: US DOT RSPA Volpe National Transportation Systems Center.

Oman, C. M., Huntley, M. S. Jr., & Rasmussen, S. A. (1996). Analog track angle error displays improve simulated GPS approach performance. In *Proceeding of the 40th Annual Meeting, Human Factors and Ergonomics Society,* (pp. 25-29). Philadelphia, PA, HFES. Proceedings are available from Human Factors and Ergonomics Society, Santa Monica, CA.

Palmer, E. A., Hutchins, E. L., Ritter, R. D., & van Cleemput, I. (1993). *Altitude deviations: Breakdowns of an error tolerant system.* (NASA Report TM 108788). Moffett Field, CA: NASA Ames Research

Center. A NASA Technical Memorandum (also available on the web at eos.arc.nasa.gov/ihi/papers/publications/palmer/AltDevTM/AltDevTM.html).

Prevot, T. (1998, May). *A display for managing the vertical flight path: an appropriate task with inappropriate feedback*. Paper presented at the International conference on Human Computer Interaction in Aeronautics (HCI Aero 1998), Montreal, Canada.

Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight* (Tech. Report TR90019). Farnborough, England: Royal Aeronautical Establishment.

RTCA-SC-159 (1996). *Minimum operational performance standards for global positioning system/wide area augmentation system airborne equipment* (Rep. No. MOPS RTCA/DO-229). Washington, DC: RTCA, 1140 Connecticut Ave, NW, Suite 1020 Washington, DC 20036 (www.rtca.orgdefault.htm).

Vakil, S. S., Midkiff, A. H., & Hansman, R. J. (1996). *Development and Evaluation of an Electronic Vertical Situation Display* (Aeronautical Systems Laboratory Research Report ASL-96-2): Massachusetts Institute of Technology.

VanderKolk, R. J., & Roscoe, S. N. (1973). *Simulator tests of pilotage error in area navigation with vertical guidance: effects of descent angle and display scale factor* (Research Report FAA-RD-73-202). Urbana-Champaign, IL: Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-Champaign.

Winer, B. J. (1971). *Statistical Principles in Experimental Design*, McGraw-Hill Book Company, New York, NY. Section 5.6

Young, T. (1997). *Electronic Aeronautical Symbols* (Aerospace Recommended Practice ARP5289). Warrendale, PA: SAE G-10 Aerospace Behavioral Engineering Technology Committee. Order through the World Wide Web: www.sae.org/servlets/productDetail?PROD_TYP=STD&PROD_CD=ARP5289.

FOOTNOTES

[1] RMS Cross-Track Errors -- Level segment: Participant F(15, 45) = 5.45, $p < 0.001$; Approach F(3,45) = 24.1, $p < 0.001$; Participant * Approach F(45,135) = 1.49, $p < 0.043$; display and all other interaction terms not significant.

Descent segment: Participant F(15, 45) = 4.47, $p < 0.001$; Approach F(3,45) = 40.2, $p < 0.001$; Display F'(3,10) = 7.39, $p < 0.007$; Participant * Approach F(45,135) = 1.84, $p < 0.004$; all other interaction terms not significant.

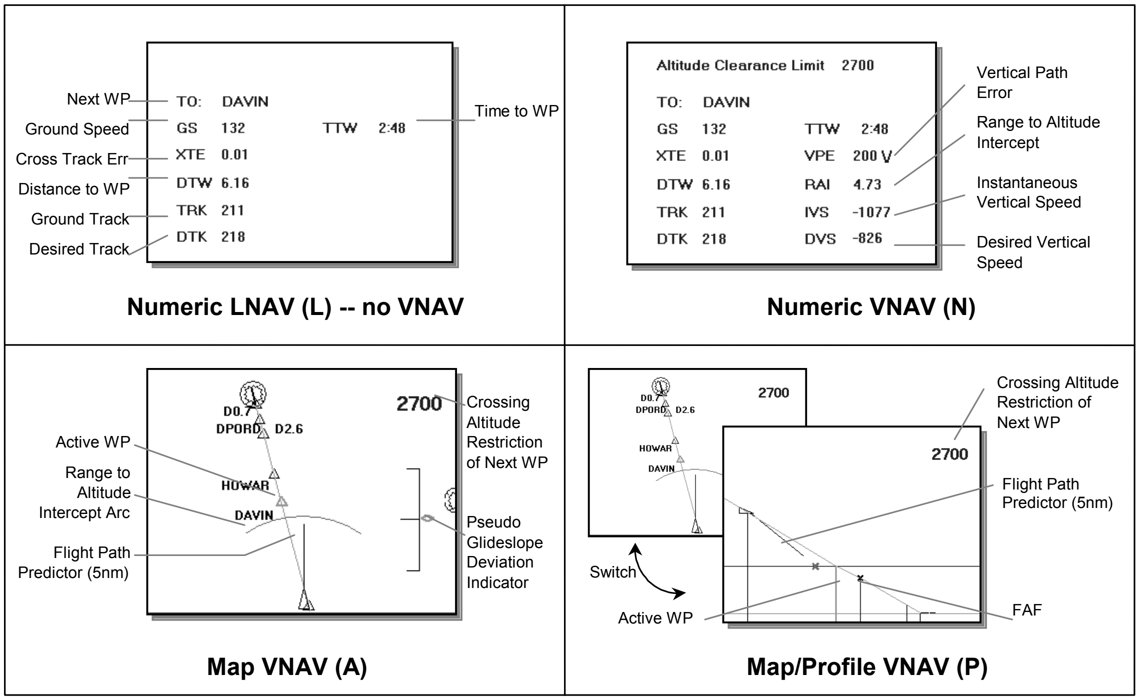
Final segment: Participant F(15, 45) = 7.80, $p < 0.001$; Approach F(3,45) = 49.8, $p < 0.001$; Participant * Approach F(45,135) = 1.65, $p < 0.015$; Approach * Display F(9,135) = 2.40, $p < 0.015$; Display and all other interaction terms not significant.

[2] Workload before FAF: Trial number F(1, 134) = 24.1, $p < 0.001$; Participant F(15,45) = 9.05, $p < 0.001$; Approach F(3,45) = 7.04, $p < 0.001$; Display and all interaction terms not significant.

Workload after FAF: Trial number F(1, 134) = 12.8, $p < 0.001$; Participant F(15,45) = 4.81, $p < 0.001$; Approach, Display and all interaction terms not significant.

FIGURES

Figure 1: Frasca 242 Simulator Instrument Panel
Avidyne 5RR display is located at the upper-right corner (Map VNAV format is in use).



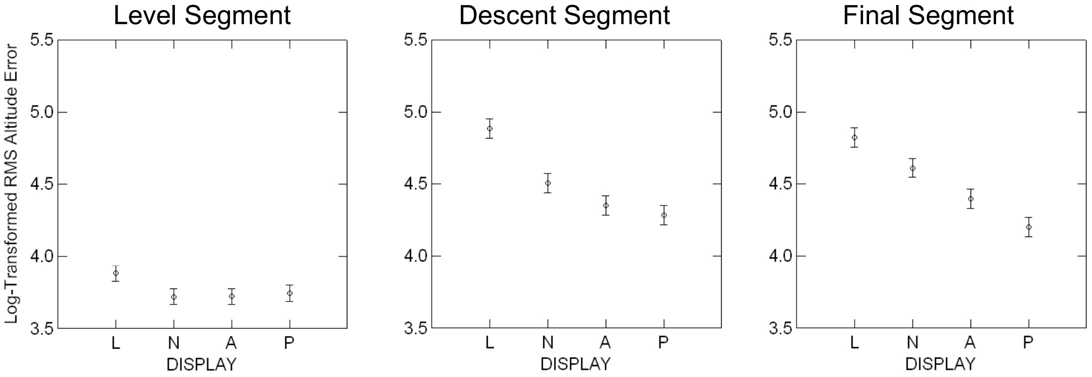
Numeric LNAV (L) -- no VNAV

Numeric VNAV (N)

Map VNAV (A)

Map/Profile VNAV (P)

Figure 2 revised caption: One lateral navigation and three vertical navigation display formats used in the simulation.
Actual display background was black. L and N display alphamerics were yellow. A and P display symbol colors as described in text.



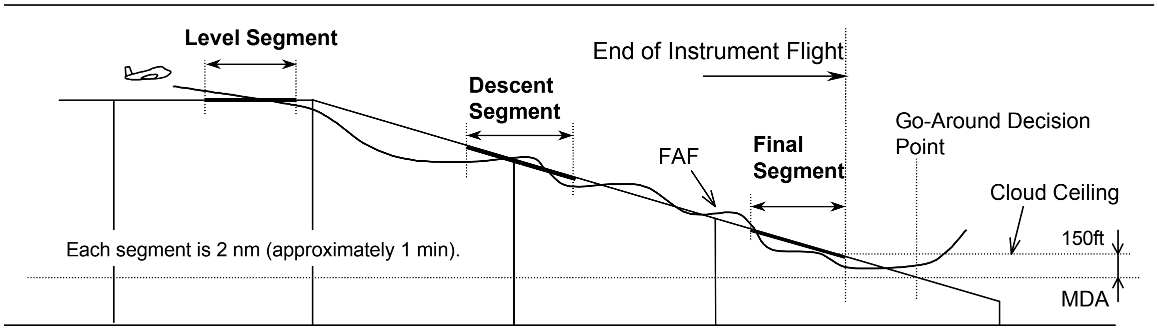


Figure 3, Top: Display Effect on Log-Transformed RMS Altitude Errors

Least-square means of RMS error with ±1 SEM for each display format.

Figure 3, Bottom: Approach Schematic (Side View)

Three 2-mile-long segments (level, descent, and final).

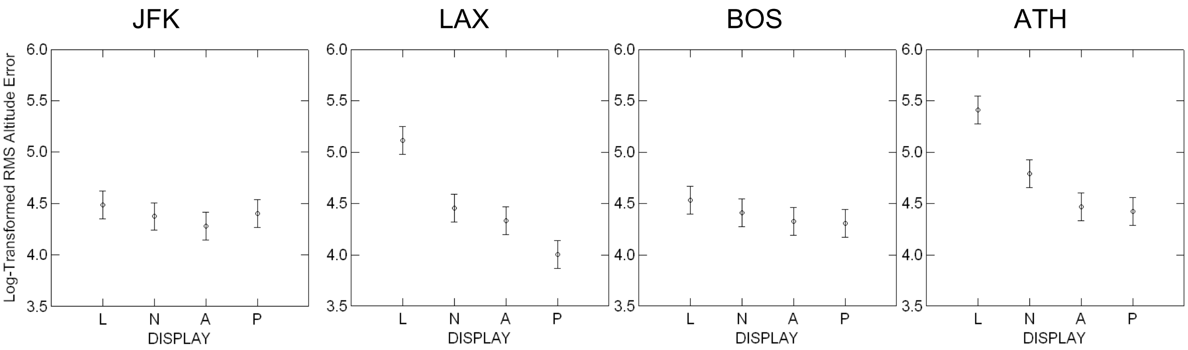


Figure 4: Display Effect on RMS Altitude Error on Descent Segment of Each Approach

Least-square means of RMS altitude errors of each display with ±1 SEM, computed within each approach type.

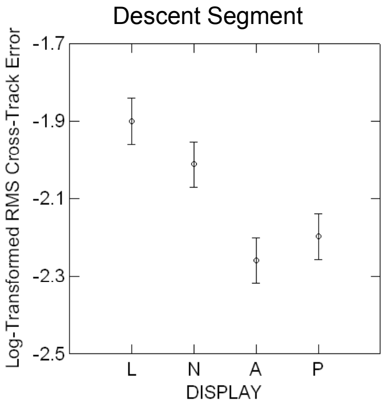


Figure 5: Display Effect on Log-Transformed RMS Cross-Track Error on Descent Segment

TABLE 1
Non-Precision Approach Procedures Used in the Simulation

Approach Types	*Runway*	*Description*
JFK	Kennedy 22L	12.8 nm. One left turn. The simplest approach.
LAX	Los Angeles 25L	17.8 nm. A right turn immediately followed by a left turn, with the steepest descent and final segments of the four approaches.
BOS	Boston 33L	15.7 nm. One right turn. A short level leg (0.5 mile) after the FAF was added so the constant angle approach cleared all intervening obstacles.
ATH	Athens 15L	23.2 nm. Two closely spaced right turns. The longest and most challenging approach.

Note: Airport designated by ICAO identifier: JFK= John F. Kennedy International, New York, NY; LAX=Los Angeles International , Los Angeles, CA; BOS=Boston/Logan International, Boston, MA; ATH=Athens International, Athens, Greece

TABLE 2
Results of Hypothesis Tests on RMS Altitude Errors

Effects	*Level Segment* ($R^2 = 0.706$)	*Descent Segment* ($R^2 = 0.654$)	*Final Segment* ($R^2 = 0.633$)
Participant	F (15,45) = 10.9 $p < 0.001$	F (15,45) = 5.37 $p < 0.001$	F (15,45) = 5.76 $p < 0.001$
Approach	---	F (3,45) = 8.15 $p < 0.001$	F (3,45) = 5.37 $p < 0.003$
Display	---	F' (3,9) = 5.65 $p < 0.019$	F' (3,8) = 17.4 $p < 0.001$
Participant * Approach	F (45,135) = 2.02 $p < 0.001$	---	---
Approach * Display	---	F (9,135) = 2.84 $p < 0.004$	---
Display * Participant	---	---	---

Note: Squared multiple (R^2), F-ratio (F), quasi F-ratio (F'), and corresponding p value (p). Only effects with $p < 0.05$ are shown.

TABLE 3
Participants' Rankings of Four Formats on Five Different Preference Scales

	5 Display Preference Scales (1=best, 4=worst)				
Formats	***HTH***	***OP***	***EI***	***FA***	***OW***
L	4	4	4	4	4
N	3	3	3	3	3
A	2	2	1	2	1
P	1	1	2	1	2

Note: Head to head comparison (HTH), overall preference (OP), ease of interpretation (EI), ability to fly accurately (FA), and effect on overall workload (OW).

TABLE 4
Expected Values of Mean Square and Test of Hypotheses

Effects	df	E(MS)	Null Hypotheses	F-ratios
A Main	15	$\sigma_{\varepsilon}^2 + 4\sigma_{AB}^2 + 16\sigma_{A}^2$	$\sigma_{A}^2 = 0$	$F = MS_A / MS_{AB}$
B Main	3	$\sigma_{\varepsilon}^2 + 4\sigma_{AB}^2 + 64\sigma_{B}^2$	$\sigma_{B}^2 = 0$	$F = MS_B / MS_{AB}$
C Main	3	$\sigma_{\varepsilon}^2 + 16\sigma_{BC}^2 + 4\sigma_{CA}^2 + 64\sigma_{C}^2$	$\sigma_{C}^2 = 0$	$F' = MS_C / (MS_{BC} + MS_{CA} - MS_{\varepsilon})$
AB Interaction	45	$\sigma_{\varepsilon}^2 + 4\sigma_{AB}^2$	$\sigma_{AB}^2 = 0$	$F = MS_{AB} / MS_{\varepsilon}$
BC Interaction	9	$\sigma_{\varepsilon}^2 + 16\sigma_{BC}^2$	$\sigma_{BC}^2 = 0$	$F = MS_{BC} / MS_{\varepsilon}$
CA Interaction	45	$\sigma_{\varepsilon}^2 + 4\sigma_{CA}^2$	$\sigma_{CA}^2 = 0$	$F = MS_{CA} / MS_{\varepsilon}$
Error	135	σ_{ε}^2	---	---