Hidden Markov Models for Analysis of Pilot Instrument Scanning and Attention Switching

by

Miwa Hayashi

M.S. Aerospace Engineering
The University of Texas at Austin, 1997

SUBMITTED TO THE DEPARTMENT OF AERONAUTICS AND ASTRONAUTICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN HUMANS AND AUTOMATION
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2004

Signature of Author: _____
Department of Aeronautics and Astronautics
July 30, 2004

Certified by: _____
Charles M. Oman, Senior Lecturer
Committee Chair, Department of Aeronautics and Astronautics

Certified by: _____
Thomas B. Sheridan, Professor Emeritus
Engineering Systems Division

Certified by: _____
Laurence R. Young, Apollo Program Professor of Astronautics
Department of Aeronautics and Astronautics

Certified by: _____
Michael Zuschlag, Research Scientist
U.S. DOT Volpe National Transportation Systems Center

Accepted by: _____
Jaime Peraire, Professor of Aeronautics and Astronautics
Chair, Committee on Graduate Students

HIDDEN MARKOV MODELS FOR ANALYSIS OF PILOT INSTRUMENT SCANNING AND ATTENTION SWITCHING

by

MIWA HAYASHI

Submitted to the Department of Aeronautics and Astronautics
on July 30, 2004, in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy in
Humans and Automation

ABSTRACT

Pilots' eye movements provide researchers rich information about the pilots' cognitive process during flight. Indeed, many researchers have included pilots' eye-movement measures in their flight simulator experiments. Currently, however, due to the lack of a reasonable model of pilots' scanning process, most researchers must rely on simple statistical analysis of eye-movement data, such as mean fixation durations on each instrument. The problem is that such statistical analyses often involve time-averaging operations, and so the information regarding the sequence of instrument scanning, the richest part of the data that reflects the pilot's moment-to-moment thought and attention processes, often has been lost or not fully utilized.

The thesis proposes a new analysis tool based on Hidden Markov Models (HMMs). This analysis exploits pilots' instrument-crosschecking eye movements within an instrument group related to the vertical-, horizontal-, or airspeed-tracking task. From the pilots' eye-movement data, the HMM estimates the most likely sequence of underlying tracking tasks that the pilot attended to. HMM analysis is especially useful when some instruments overlap among multiple tracking tasks (e.g., the attitude indicator overlaps among all three tracking tasks) because it can utilize the sequential information from the instrument scanning to compute the likelihood of each of the possible tracking tasks.

The actual pilot eye-movements data collected during ILS approach simulation experiments indicated that some experienced pilots may attend to more than three tasks during flight, with the fourth one being a monitoring task, while some inexperienced pilots may attend to only two, dropping one of the tracking tasks probably due to high workload. The results of another flight simulation experiment demonstrated how the pilots' attention budgeting among these tasks estimated by HMM analysis, combined with the pilots' eye-movement statistical results, could enhance a cockpit display format study. The experiments demonstrated what additional insights can be obtained by incorporating HMM analysis into the analysis of pilots' eye movements.

Thesis Supervisor: Charles M. Oman

Title: Senior Lecturer in Aeronautics and Astronautics

Acknowledgement

The author would like to express her great appreciation to Dr. Charles Oman, Professor Thomas Sheridan, Professor Laurence Young, and Dr. James Kuchar of MIT and Dr. Michael Zuschlag of the Volpe Center for their supervision and guidance during the writing of this thesis. The author is also grateful to Professor Nicholas Roy and Professor Eric Feron of MIT for their valuable comments and suggestions for the thesis. Her thanks also go to Dr. Andrew Liu of MIT and Dr. Dario Salvucci of Nissan Cambridge Basic Research for their advice, especially during the initial stage of the HMM analysis concept development, and to Dr. Alan Natapoff of MIT for his assistance in the statistical analysis. Many thanks go to Andrew Kendra of the Volpe Center and Rikki Razdan of ISCAN, Inc., for their technical assistance during the flight simulator experiment setup, and Purdy Ho and Tao Yue of MIT and Aidan Collins of Tufts University for their programming assistance. Great appreciation also goes to all the pilots who volunteered their valuable time to participate in the simulator experiments.

The author is grateful to Dale Dunford of the Federal Aviation Administration ANM-111 and Tom McCloy of AAR-100 for supporting this work. The research presented in this thesis was supported and funded by the U.S. Federal Aviation Administration through the John A. Volpe National Transportation Systems Center under DTRS57-01-D-30043.

Last but not least, the author would like to express special personal thanks to Carl Quesnel for his editorial assistance, as well as his wonderful friendship, throughout the development of this thesis, and to her loving parents, Dr. Toshiaki Hayashi and Kazue Hayashi, for their ceaseless faith in their daughter, for which the author will be eternally thankful.

Table of Contents

List of Abbreviations

AI	Attitude indicator
AOI	Area of interest
ASI	Airspeed Indicator
CAP	Captain
CDI	Course Deviation Indicator
D	Digits Only
DMM	Double Markov Model
FAA	Federal Aviation Administration
F/O	First Officer
GLM	Generalized Linear Model
HI	Heading Indicator
HMM	Hidden Markov Model
HUD	Head Up Display
ILS	Instrument Landing Systems
MM	Markov Model
PD	(Rotating) Pointers with Digits
PGD	(Rotating) Pointers with Gradation Marks with Digits
RMS	Root Mean Square
USAF	United States Air Force
VDI	Vertical Deviation Indicator
VSI	Vertical speed indicator

Chapter 1:

Introduction

Flying an aircraft is perhaps one of the most complex and demanding manual control tasks that human operators have ever been required to perform, and, because of that, a number of flight simulation experiments have been conducted to investigate pilots' performance. In the last decade or so, there has been a notable trend in these flight simulator experiments; more and more researchers started including measures of pilots' eye movements.

Measuring eye movements requires special precision equipment and typically generates huge amounts of data, which are typically both noisy and complex, and thus, quite frankly, are tedious to analyze. So, why have researchers suddenly started to include these measurements in their flight simulator experiments? Of course, one of the reasons must be the rapid advance of eye-tracking technology in the recent years, which has made eye-tracking systems more reliable and affordable. However, there are also important theoretical motivations; since flying tasks are accomplished with the assistance of specialized flight instruments, a better understanding of pilots' scanning process may help engineers design better cockpit displays. Furthermore, studying pilots' scanning process may enable researchers to infer part of pilots' thought processes. Therefore, for instance, by understanding the differences between novice and expert pilots' thought processes, one may be able to develop more effective pilot training programs that teach novice pilots to assimilate expert pilots' scan and thought patterns.

Note that the nature of flight tasks also offers some additional convenience for eye-movement analysis. Compared to automobile driving or many other human-operator manual control tasks, flying tasks, especially when referring only to the flight instruments (i.e., instrument flying), are extremely busy tasks, to the extent that it is often safe for researchers to assume that pilots are actually attending to the instruments during almost all of the flight time. In fact, Clement et al. calculated the required sampling frequencies for all the necessary flight variables to perform an Instrument Landing Systems (ILS) approach, the type of instrument approach most frequently performed these days, and concluded that pilots would have barely sufficient time to sample all

these variables (Clement, Jex, & Graham, 1968). This is a convenient assumption, because it allows researchers to assume that what the pilots look at reasonably represents what they are thinking about. Notice that such an assumption may not hold in other cases. For instance, in the case of driving a car, drivers often have more than enough time to control the vehicle, and, thus, their thoughts may occasionally drift away from the driving task even when they are still staring at an instrument or an outside scene. Therefore, what drivers are looking at may not always represent what they are thinking about.

Before introducing some of the past studies that have been completed in this field, let us briefly note the roles of foveal and peripheral vision. Most commercially available eye-tracker systems are based on the direction of the eyeball measured by electrooculography (EOG), pupil center, corneal reflex, and so on (Young & Sheena, 1988). Therefore, by definition, these account for only the foveal fixation, leaving out the other half of the story—peripheral vision. To illustrate the importance of peripheral vision, Levison and Elkind (1967) reported the interesting fact that, in their informal experiment of concurrently tracking two tasks on separate displays, where the participant had to move their fixation point from one display to the other, they blanked the display not being looking at by the participant. They reported that the participant quickly became unable to maintain tracking performance on both displays. In general, peripheral vision has been considered to help increase scanning efficiency by guiding the foveal fixation points (Levison & Elkind, 1967) and by forming an expectancy of the approximate instrument readout even before the foveal fixation gets there (Senders, Webb, & Baker, 1955). Thus, even though peripheral vision cannot be directly measured by any existing eye-tracking systems so far, if the analysis includes the fixation transitions among instruments, it still may indirectly take peripheral vision into account.

Now, the following section presents a brief overview of selected past studies of pilots' scanning process. Moray has also provided an excellent review of pilots' eye-movement modeling research up to the mid 1980s (Moray, 1986). After the overview of past studies, exactly what causes the difficulties in modeling and analyzing pilots' eye-movements, as well as the shortcomings that the models used in past studies had, will be explained. Then, the main purpose of the current thesis to address some of these problems will be stated.

1.1 Past Studies

The earliest and largest pilots' eye-movement database was established by Fitts and his colleagues in the late 1940s to early 1950s (Fitts, Jones, & Milton, 1950). Their data were recorded during actual flights, not flight simulations. Forty Air Force pilots participated in the study, and their eye movements while performing instrument approaches were recorded in 35-mm motion-picture films. The films were later analyzed frame by frame to determine the instrument the pilot fixated on at each instance. The results of the analysis of the eye-movement statistics revealed that the instruments most frequently looked at were the cross pointer (similar to the modern course deviation indicator), directional gyro (heading indicator but in a magnetic compass format), and gyro horizon (similar to a modern attitude indicator (AI)). Their experiment results were used to redesign instrument panel arrangements.

The first quantitative model of pilots' scanning process was proposed by Senders (Senders, 1955). For a pilot to reconstruct the signal by sampling, where the signal bandwidth is denoted as W, the signal has to be sampled at least at the frequency of 2W, also known as the Nyquist frequency. Later, Carbonell extended Senders's idea and applied queuing theory (Carbonell, 1966; Carbonell, Ward, & Senders, 1968). The basic idea was that a queue is associated with each instrument, and each instrument has a risk of not being looked at. Furthermore, the risk also grows as time goes on. Thus, the pilot's task is to make an intelligent decision each time before looking at an instrument to minimize the overall risk. A series of flight simulator experiments demonstrated the model's ability to accurately represent percent times of pilots' fixation on each instrument during flights, although there was also an issue of parameter tuning for each pilot and each flight phase. The model also does not predict the scanning sequence itself.

Weir and McRuer conducted crossover model analysis (McRuer & Jex, 1967) on the data of two air transport pilots flying instrument approaches on a flight simulator (Weir & McRuer, 1972). The data of these two particular pilots were selected for analysis because of their quite different multiloop control strategies: one pilot was flying predominantly by referencing pitch angle (inner loop), while the other pilot tended to attempt to directly minimize glide slope deviations (outer loop). Their eye-movement data were compared with their inner- and outer-loop crossover frequencies. The analysis results showed positive correlations between the pilots' crossover frequencies and their look rate (looks per unit of time) or dwell fractions (fraction of the time

spent on each instrument). The results suggested that pilots' workload increases with scanning requirements and task complexity.

Tole et al. examined the potential use of eye-movement data as a measure of pilots' workload and skill (Tole, Stephens, Vivaudou, Ephrath, & Young, 1983). In their flight simulator experiment, they measured eye-movement data for 11 pilots who had different skill levels. While maintaining a straight-and-level path, pilots were also asked to respond to certain mental-algorithm secondary tasks, where the questions were given auditory and the pilots responded verbally in order not to disturb the visual aspect of the main task. The results showed that the dwell time on primary instruments increased as the secondary task difficulty was increased. This effect was more noticeable for the novice pilots than for the expert pilots. The entropy, or disorderliness, of the instrument fixation sequence was also computed, and the results indicated a tendency for entropy to decrease as the secondary task difficulty increased, probably due to the increased dwell time (i.e., decreasing the disorderliness).

A slightly different approach was taken more recently by Doane and Sohn (2000). They used a computational cognition model called ADAPT to explain and predict the scanning and control behavior of 25 pilots during various simulated flight maneuvers. ADAPT's knowledge database consisted of hundreds of knowledge elements, which were classified into three categories: world knowledge (e.g., "current airspeed is 100 knots"), general knowledge (e.g., "power controls airspeed"), and plan element knowledge (e.g., "decrease airspeed"). Then ADAPT constructed a knowledge association network among these knowledge elements, and the most relevant plan element whose preconditions existed in the world was activated. The model predictions and human-pilot simulation results showed good matches. However, mismatches, such as altimeter fixation versus vertical speed indicator (VSI) fixation, also often occurred.

Finally, some researchers also conducted Markov Model (MM) analysis of various human operators' eye-movement data, such as pilots' eye-movement data while monitoring the Cockpit Display of Traffic Information (CDTI) (Ellis & Stark, 1986), or radar operators' eye movements while monitoring military aircraft intercepts (Moray, Neil, & Brophy, 1983). These studies were not about the pilots' flight instrument scanning process, but they are mentioned here because their MM analyses approach has an important connection to the Hidden Markov Model analysis that will be introduced later. The first-order Markov model is a process such that the transition

probability, $p_t(i,j)$, which is the probability to transition from the state, *i*, to the state, *j*, at time, *t*, is independent of *t*. In MM analyses, each area of interest (AOI) serves as each state, and the fixation transition probabilities between each pair of AOI, i.e., $p(i,j)$ for each *i* and *j*, are computed. The MM is a well-developed mathematical model, with which the researchers can perform various sophisticated analyses, such as the mean first passage time (MFPT)—the average time from the time the fixation left a particular AOI to the time it comes back to the same AOI (Moray et al., 1983; Kemeny & Snell, 1960). The weakness of the MM analysis approach is that if multiple different scanning strategies occur from time to time, the MM analysis returns the transition probabilities averaged over them, not the real transition probabilities of each scanning strategy.

1.2 Difficulties in Modeling Pilots' Eye Movements

As described above, Fitts's group's study computed pilots' eye-movement statistics, such as average fixation durations and look rates. In fact, the majority of pilots' eye-movement study literature, even from the most recent studies, relys on this type of data analysis. Such statistical analysis is usually simple to conduct. It is also the only way to analyze if a researcher does not have any reasonable quantitative model of pilots' scanning process. The statistical analysis provides researchers valuable information on how each instrument was actually used by the pilots throughout the total duration of the flight. However, the statistical analysis also inevitably involves time averaging, and so it loses information regarding the sequence of instrument scans, which contains valuable information about pilots' thought process from moment to moment.

About the fixation transition among the instruments, Fitts's group computed the link values, which are the relative frequencies of eye movements from one instrument to another instrument. One link value is assigned for each pair of instruments (i.e., the same value accounting for both directions). However, the concept of link values was later questioned by Ellis et al., whose MM analysis on eye-movement data revealed a statistical dependency of the transition probabilities on the previous fixation point (i.e., different directions had different transition probabilities). Ellis's group's study also indicated apparent randomness in the transition probabilities, which might suggest the existence of multiple underlying scanning strategies.

The other quantitative modeling studies described in the previous past-studies section may also

have incorrectly assumed that the pilot had only a single scanning strategy throughout the flight. As a result, the studies that defined a single "optimal" or "normative" scanning behavior model, which the pilots are supposed to follow all the time (e.g., sampling theory, crossover modeling), often failed to fit the model to the actual data. In other words, these models did not have enough parameters to account for the time-varying nature of the real pilots' scanning behavior (i.e., an under-fitting problem). On the other hand, the studies that depended more on the actual data to construct a time-invariant scanning behavior model (e.g., queuing theory, computational cognitive modeling) tended to have to keep adding more and more model parameters to describe the scanning behaviors of the actual pilots, which may be actually time-varying. As a result, such models risked ending up with much more model parameters than actually necessary (i.e., an over-fitting problem).

Furthermore, attempts to accurately predict pilots' fixation sequence are often hampered by certain features of modern aircraft instruments that were intentionally provided to facilitate pilots' scanning. One such feature is information redundancy. For example, the altimeter and the VSI both provide rate of climb/descent information, and therefore are partially redundant. As a result, accurate prediction of which instrument, the altimeter or the VSI, the pilot would look at next is usually difficult, as seen in the Doane & Sohn study. Another feature is information integration; some instruments provide more than one piece of information, such as the AI displaying both pitch and bank angle information as well as their higher-order derivatives. Thus, when a pilot is looking at the AI, the researcher cannot tell which information the pilot is actually attending to.

1.3 Purpose of This Thesis

To address the above difficulties, this thesis proposes the use of a new method for analyzing pilots' eye-movement data—Hidden Markov Model (HMM) analysis. HMMs have been widely used in speech recognition and other pattern recognition applications (Rabiner & Juang, 1993), but this is the first time HMMs have been applied to the analysis of aircraft pilots' eye-movement data. The HMM analysis proposed here takes advantage of information from pilots' instrument crosscheck. That is, from a pilot's eye-movement data, it estimates the pilot's attention switching among different tracking axes, each of which corresponds to a particular set of instruments

within which the pilot crosschecks. As readers will see in the next chapter, many difficulties confronted by previous models can be resolved or lessened by the use of HMM analysis. For instance, HMM analysis utilizes the sequential information of pilots' instrument fixation data, and provides more information than fixation statistics alone can offer. Moreover, the analysis is based on the HMM, a time-varying stochastic process model. This ability to accommodate the time-varying characteristics of the pilots' eye movements, which none of the previous models had, may be the biggest advantage of using HMM analysis. Because of this ability, the problem of under-fitting or over-fitting is largely reduced. Furthermore, the issues of instrument redundancy and information integration are greatly lessened by taking advantage of instrument crosscheck. More details of HMM analysis are provided in Chapter 2.

In addition to proposing the new concept of HMM analysis, the goals of the thesis also included establishing an implementable procedure of HMM analysis of pilots' eye-movement data. The thesis also sought to demonstrate the usefulness of the extra information that HMM analysis results provide for advancing the knowledge essential for designing better cockpit displays and improving pilot training. It is also part of the thesis' goals to observe the capacity and the limitations of HMM analysis for fitting actual pilots' eye-movement data. Since actual pilots' instrument scanning process must be much more complicated than the simplified model used in the HMM, observing the capacity and limitations of HMM analysis may help illuminate some important aspects of real human operators' cognition system.

To achieve these goals, the thesis takes the following steps: In Chapter 2, the basic concept of HMM analysis is introduced. Then, brief explanations of the original HMM algorithms and their adaptation for being implemented into the pilots' eye-movement analysis are described. Chapter 3 presents a flight simulator experiment conducted as a proof-of-concept of HMM analysis, which also can be used as a walk-through example of an actual implementation of HMM analysis. This flight simulator experiment examined differences in the scanning processes among four pilots who had different levels of flight expertise, and the HMM analysis results, which may have important implications for pilot training, are discussed. In Chapter 4, HMM analysis is incorporated into the actual analysis process of a head-up display format study. This study also can serve as another walk-through example. The result demonstrates how the pilots' attention sequence estimated by HMM analysis helped the researchers to obtain further insight into the

experiment data by bridging the pilots' eye-movement data and their flight performance. In Chapter 5, some issues of HMM analysis, such as implementation issues related to verbal-report requirements and theoretical issues related to the serial assumption and the Markov process assumption of the attention process, are discussed. Some suggestions for future work are also presented in this chapter. Finally, Chapter 6 presents the summary and contributions of the thesis.

Chapter 2:

Hidden Markov Model Analysis

As mentioned in the previous chapter, HMM analysis utilizes the sequential information of pilots' instrument crosscheck eye-movement data. In this chapter, first the characteristics of instrument crosscheck are explained. Second, original HMM estimation algorithms and its estimation process, also often used in speech recognition applications, are briefly reviewed. Third, how this technique can be adapted to estimating pilots' hidden attention-switching process from their eye-movement data is described. Then, finally, the differences between the proposed use of HMM for pilots' eye-movement data analysis and other conventional HMM applications, such as speech recognition, are discussed.

2.1 Instrument Crosscheck

In most modern airplanes, cockpit instrument panels present at least the following basic four instruments: the airspeed indicator (ASI), attitude indicator (AI), altimeter, and heading indicator (HI). These instruments are often arranged as the so-called "basic-T instruments" or "basic-T arrangement," where the ASI appears to the left of the AI, the altimeter to the right of the AI, and the HI below the AI. The basic-T instruments have been required for all transport-type airplanes since 1957 (CAR 4b (CAB, 1953) later replaced with FAR §25.1321 (FAA, 1970)). In addition to these four, most cockpits are also equipped with a turn-and-slip indicator, vertical speed indicator (VSI), power indicators, course deviation indicators combined with vertical deviation indicator (CDI/VDI, or just CDI for short in this thesis), and so on.

Since none of these instruments alone completely describes the entire behavior of an aircraft, pilots must crosscheck among multiple instruments. For example, if the altimeter indicates higher altitude than the desired value, then the pilot may also want to crosscheck with the VSI to determine if the aircraft is still climbing or not, because the appropriate actions are different for each case. As this example illustrates, some instruments are more closely related than others in this instrument crosscheck process. For instance, the AI, the altimeter, and the VSI each indicate

a different aspect of the information related to vertical-axis tracking, and pilots have to crosscheck among them if they are to fly at the intended altitude. Table 2-1 lists three groups of instruments characterized by tracking task: vertical tracking, horizontal tracking, and airspeed tracking. These groups are consistent with the instrument groups defined in the Instrument Flying Handbook (FAA, 2001), except that the ASI is not in the vertical group and the AI is included in the airspeed group. These particular groups in Table 2-1 were chosen to explicitly separate the airspeed-tracking task from the vertical-tracking task, because that fits the purpose of our later experiments and analysis better. (If, for some reason, it is believed that airspeed tracking is not being performed separately, then the airspeed-tracking instrument group could be merged into the vertical-tracking instrument group in Table 2-1. More about selecting proper instrument groups will be discussed in Chapter 3.) The pilot's task is to perform multi-axis tracking along the three axes shown in Table 2-1 by crosschecking the appropriate instruments associated with each tracking axis.

Table 2-1. Three instrument groups of each tracking axis

	Vertical-Tracking Instrument Group	Horizontal-Tracking Instrument Group	Airspeed-Tracking Instrument Group
Instruments	AI (Pitch) Altimeter VSI CDI (Glide Slope)	AI (Bank) HI Turn and Slip CDI (Localizer / Course)	AI (Pitch) Power ASI

Although there has been some controversy over whether human operators attend to multiple tasks in a serial or parallel manner, in the case of multi-axis tracking during instrument flight, the adequate working assumption is that pilots attend to each tracking axis in a serial manner rather than in a parallel manner. One of the justifications for the serial-processing assumption is the complexity of the task. Pilots use their mental model of the aircraft's dynamics along each axis or cross-coupled dynamics among different axes to interpret the multiple instrument readings and eventually determine the appropriate amount and direction of input to adjust the aircraft's behavior (Bellenkes, Wickens, & Kramer, 1997). Such a complex process requires a certain amount of cognitive effort, and thus the tasks are assumed more likely to be processed in a serial manner. Another justification for this assumption is the nature of vision—although there is

evidence that pilots use peripheral vision, accurate instrument reading requires fixation. Because fixations occur one at a time, it is an adequate approximation that pilots attend to tasks in a serial manner as well (Moray, 1986). This serial-processing assumption is the basis of the concept of this analysis and will be revisited and discussed again in Chapter 5.

This serial-processing assumption now leads to a new model of pilots' scanning process. In this model, pilots' eye-movement patterns depend on their internal state, i.e., the tracking axis they are attending to. The model focuses more on the underlying process, the serial switching among multiple tracking tasks, that drives pilots' eye movements rather than the eye movements themselves. Understanding how the pilot switched scan patterns during the flight, which in turn shows how the pilot switched among these tracking tasks, may provide researchers additional information to bridge the gap between pilots' eye movements and pilots' actual tracking performance.

Note that in this new model framework, the problem of instrument information redundancy is greatly reduced. For instance, the altimeter and the VSI both provide rate of climb/descent information, and the HI and the turn-and-slip indicator both show rate of turn information. These information redundancies sometimes caused problems in previous models. In this new model framework, however, notice that all the redundant instruments belong to the same tracking-axis group. (In fact, the redundancy is more or less equivalent to membership in the same group in Table 2-1.) The focus of this new model framework is the tracking-axis group the pilots were attending, rather than which particular instrument the pilots looked at, and that is why instrument redundancy becomes much less of an issue in this model framework.

The attention-switching process among the tracking tasks is, of course, not directly observable. Only the pilots' eye movements are observable. So, the question is, is it possible to estimate the sequence of the pilots' attention switching from their instrument fixation sequences? Such estimation would be trivial if the instrument and the tracking axis had a one-to-one relationship. Unfortunately, that is not the case, as Table 2-1 indicates. In Table 2-1, the AI and the CDI—notice both of them are integrated displays—are overlapping in more than one tracking task. Therefore, when one of these overlapping instruments is fixated on, investigators cannot determine which task the pilot is attending to. This is particularly a problem for the AI, because this is usually the instrument most looked at. The average percent times on the AI in the total

flight time that literature reported range from 15% to as much as 59%, depending on the flight segments and the instrument arrangement (Fitts et al., 1950; Weir & Klein, 1971; Pennington, 1979). This indicates that the underlying tracking task could remain ambiguous for a rather long period of the flight time. Hidden Markov Model (HMM) analysis provides a solution for this ambiguity that is created by these overlapping instruments. It exploits pilots' instrument crosscheck in Table 2-1; i.e., if an overlapping instrument is scanned along with other instruments that belong to a particular tracking task, then that instrument is more likely being used for that task.

2.2 Hidden Markov Models (HMMs)

2.2.1 Overview

The basic theory of the HMM was first introduced by Baum and Petrie in 1966 (Baum & Petrie, 1966). It has been since the late 1980s, however, that HMMs have become widely applied to various pattern recognition applications, including speech recognition (Rabiner & Juang, 1993, and numerous other references), face recognition (e.g., Oliver, Pentland, & Berard, 2000), sign language recognition (e.g., Starner, Weaver, & Pentland, 1998), protein structure analysis (e.g, Krogh, Brown, Mian, Sjolander, & Haussler, 1994), and so on. The nearly twenty-year gap between the introduction of the theory and actual applications was probably due to the fact that the HMM theory was first published in mathematical journals seldom read by pattern recognition engineers (Rabiner, 1989). Additionally there were no sufficient tutorial materials available. Now some good tutorials are available (e.g., see Rabiner & Juang, 1986; Rabiner, 1989; Rabiner & Juang, 1993, Chapter 6). The following presents brief descriptions of HMMs. For more details, the reader should consult the tutorial materials listed above.

The HMM framework consists of two stochastic layers. One is a hidden-state process, which is not directly observable but is assumed to follow first-order Markov process transition rules. The other layer is an observation-symbol process, which is physically observable and has a certain probability distribution depending on the current hidden state. In the case of pilots' scanning process, the tracking task was modeled as the hidden-state process, and the instrument fixation corresponded to the observation process. Figure 2-1 illustrates the transitions among the hidden states (depicted as arrows) and corresponding instrument fixations, given each hidden state.

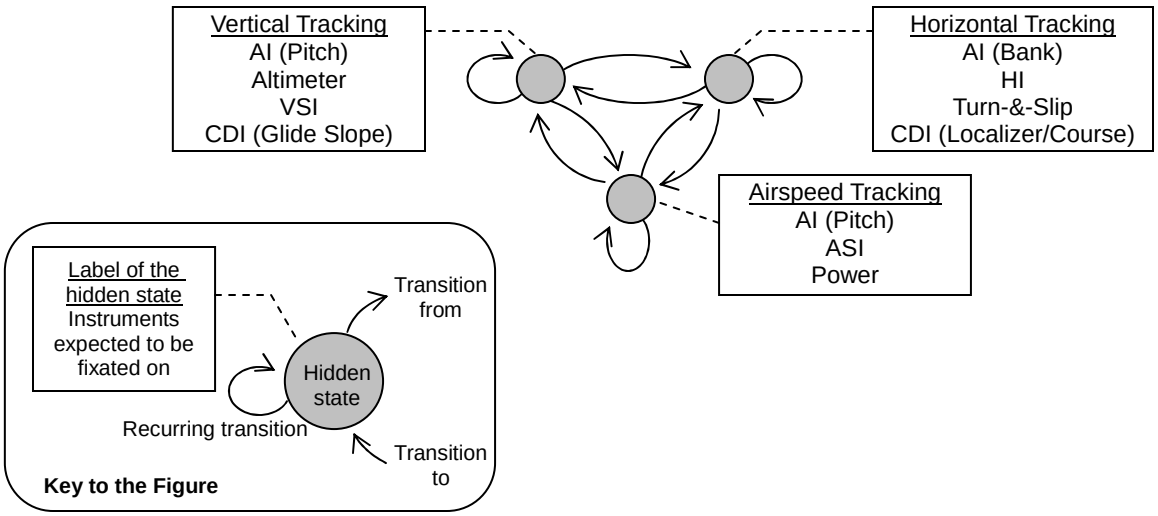


Figure 2-1. HMM with three states

A model structure for an HMM can be completely described by a set of three parameter matrices, $\lambda = (A, B, \pi)$. The descriptions of each parameter are as follows. The initial state probability distribution matrix,

$$\pi = \{\pi_i\} \qquad 1 \le i \le N \tag{2.1}$$

gives the probability that the hidden state, i, is the initial state; the state-transition probability distribution matrix,

$$A = \{a_{ij}\} \qquad 1 \le i, j \le N \tag{2.2}$$

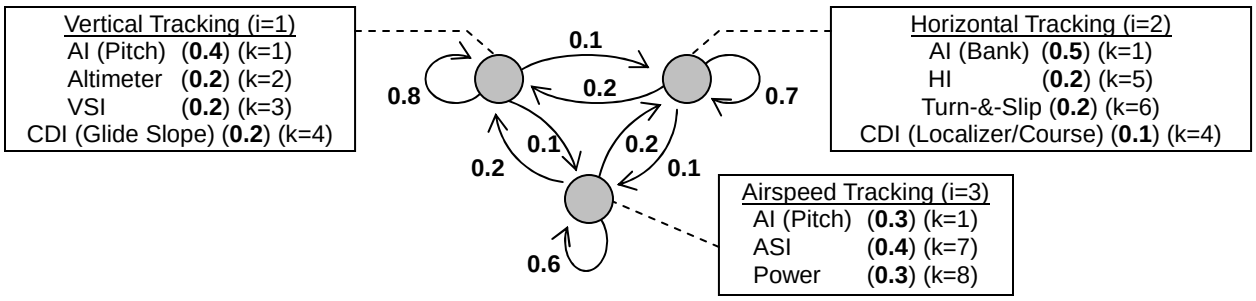
gives the transition probability from the hidden state, i, to state, j; and the observation symbol probability distribution matrix,

$$B = \{b_j(k)\} \qquad 1 \le j \le N, 1 \le k \le M \tag{2.3}$$

gives the probability of the observation symbol, v_k, within the given hidden state, j.

2.2.2 Illustrative Example

In order to illustrate the meanings of these model parameters, an example is presented in Figure 2-2. (Readers already familiar with HMMs can skip this section.) In this figure, hypothetical numbers of the state-transition probabilities and the observation-symbol probabilities and the corresponding A and B matrices are shown.



Corresponding Parameter Matrices

$$A = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}$$

Vert. (i=1) | Horiz. (i=2) | Arspd. (i=3)

Vertical (i=1)
Horizontal (i=2)
Airspeed (i=3)

$$B = \begin{bmatrix} 0.4 & 0.2 & 0.2 & 0.2 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.2 & 0.2 & 0.1 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0.3 \end{bmatrix}$$

AI (k=1) | Altim. (k=2) | VSI (k=3) | CDI (k=4) | HI (k=5) | T&S (k=6) | ASI (k=7) | Powr. (k=8)

Vertical (j=1)
Horizontal (j=2)
Airspeed (j=3)

Figure 2-2. Illustrative example of HMM parameters with hypothetical probabilities

The diagram in Figure 2-2 shows that the probability to transition from the vertical-tracking task ($i = 1$) to the horizontal-tracking task ($i = 2$) is *0.1*, and the transition probability to the airspeed-tracking task ($i = 3$) is also *0.1*. The recurring transition probability (i.e., the probability to stay within the same state) of the vertical-tracking task is *0.8*. Note that all these transition probabilities from the vertical-tracking task add up to *1*, and that is true for the other tracking tasks as well. In the *A* matrix in this example, the first row, i.e., $\{a_{1j}\}$ $(1 \leq j \leq 3)$, represents the three transition probabilities from the vertical-tracking task ($i = 1$), including the recurring transition probability, a_{11}. Likewise, the second and third rows represent all the transition probabilities from the horizontal-tracking and airspeed-tracking tasks.

Note that the diagonal elements of the *A* matrix represent the recurring transition probabilities of each hidden state. In fact, these recurring probabilities may play a key role in pilots' eye-movements data analyses. Since humans generally do not switch their thoughts around every half second or so, it is presumed that each task tends to have a certain duration of time. The

HMM is capable of reflecting this kind of assumption by including higher recurring probabilities (the diagonal elements of the *A* matrix) than the other transition probabilities (the off-diagonal elements in the *A* matrix). It is effectively equivalent to imposing certain costs on switching tasks to ensure certain durations for each tracking task.

Regarding the instrument-fixation level, the example in Figure 2-2 shows that if the pilot is attending to the vertical-tracking task, the probability that he/she fixates on the AI is *0.4*. Likewise, the probabilities that he/she fixates on the altimeter, the VSI, and the CDI are *0.3*, *0.2*, and *0.1*, respectively. Notice that all these probabilities within the vertical-tracking task also add up to *1*, and the same is true for the other tracking tasks as well. In the *B* matrix, the first row represents the probabilities of each instrument fixation given the vertical-tracking task ($j = 1$). The second and third rows represent those given the horizontal-tracking task ($j = 2$) and the airspeed-tracking tasks ($j = 3$), respectively. Each column represents each instrument ($1 \le k \le 8$). Note that if there is more than one non-zero entry in a single column, the instrument is an overlapping instrument (i.e., the first column representing the AI, and the fourth representing the CDI in Figure 2-2). Also notice that the HMM does not take the transition probabilities among the instruments into account.

In this example, the rows of the *A* and *B* matrices represented the current hidden states, and the columns represented the next hidden state or the instruments fixated on. By the way, this is not the only way to set up the *A* and *B* matrices. For instance, one can also reverse the row and column configurations without loss of generality. In fact, as long as it complies with the official definitions presented in Eq. (2-1)-(2-3), what constitutes the row or column is not important. In the following sections, all algorithms are described using on the official definitions.

2.2.3 Three Problems of HMMs and Their Solutions

Let us denote an observation-symbol sequence (i.e., the instrument fixation sequence data) as $O = \{o_1, \ldots, o_T\}$. Then, given an observation-symbol sequence, *O*, the goal is to estimate the hidden-state sequence, $q = \{q_1, \ldots, q_T\}$, which is "optimal" in some sense. To achieve this goal, the so-called "three problems" have to be solved (Rabiner & Juang, 1986). In the following text, these three problems and solutions are briefly described. The entire set of HMM analysis algorithms, resulting from these solutions, are tabulated later in Table 2-2.

Problem 1 is, given the observation sequence, O, and a model $\lambda = (A, B, \pi)$, how can the conditional probability, $P(O|\lambda)$, be computed efficiently? Here, the point is computational efficiency because, in theory, the probability can be directly computed by

$$P(O|\lambda) = \sum_{all\ q} P(O|q,\lambda)P(q|\lambda) \tag{2.4}$$

and

$$P(O|q,\lambda)P(q|\lambda) = \{b_{q_1}(o_1)b_{q_2}(o_2)...b_{q_T}(o_T)\}\{\pi_{q_1}\ a_{q_1q_2}\ a_{q_2q_3}...\ a_{q_{T-1}q_T}\} \tag{2.5}$$

The number of multiplications required in Eq. (2.5) is $2T - 1 \approx 2T$, where T is the length of the observation sequence, O. Then, as indicated in Eq. (2.4), the computation results of $P(O|q,\lambda)P(q|\lambda)$ for all possible q have to be summed. That is, the summation is repeated N^T times, where N is the number of hidden states. For instance, even for relatively small N and T, such as $N = 3$ and $T = 100$, the required number of computations is $2T \times N^T = 2 \times 100 \times 3^{100} \approx 10^{49}$. This already seems computationally infeasible. Clearly a much more efficient algorithm is needed.

Fortunately, there is a solution. The algorithm called "Forward Procedure" utilizes induction to greatly streamline the above computation process. Notice that there are only N hidden states, and, say, for an arbitrary hidden-state sequence, q^{arb}, there are also many other qs, which are just slight variations of the q^{arb}. So, why repeat the same computations over and over? To take advantage of repeated computations, first the forward variable,

$$\alpha_t(i) = P(o_1 o_2 ... o_t, q_t = i \mid \lambda) \tag{2.6}$$

is defined. This is the probability of the partial observation sequence, $\{o_1, \ldots, o_t\}$ ($1 \le t \le T$) and the hidden state, i, at time, t, given the model λ. If the values of $\alpha_t(i)$ for all i ($1 \le i \le N$) are known at a given time, t, then the values of $\alpha_{t+1}(j)$ for all j ($1 \le j \le N$) can be computed by induction using the values of a_{ij} and $b_j(o_{t+1})$. Each induction step from t to $t+1$ requires $N \times N$ computations; thus, the total number of computations required is N^2T—much smaller than $2T \times N^T$. Finally, the probability in which we were originally interested is computed as

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{2.7}$$

It is also possible to define the "Backward Procedure"—the counterpart of the Forward Procedure—which computes the backward variable, $\beta_t(i)$. Both the forward variable and the backward variable will be needed to solve problem 3. The backward variable, $\beta_t(i)$, is defined as follows:

$$\beta_t(i) = P(o_{t+1} o_{t+2} \ldots o_T \mid q_t = i, \lambda) \tag{2.8}$$

This is the probability of the partial observation sequence from $t+1$ to T, given hidden state, i, at time, t. This time, the values of $\beta_t(i)$, for all i ($1 \le i \le N$) are induced backward from $t = T$ to $t = 1$. The required number of computations is, again, N^2T.

Note that in the actual computation of the Forward and Backward Procedures, scaling technique is used to prevent numerical underflow problems (Juang & Rabiner, 1985). Since computation of the original α and β involves a large number of multiplications of a_{ij} and $b_j(o_t)$ terms (see Eq. (2.5)), whose values are generally significantly less than 1, the resulting values of α and β would quickly exceed the precision range of any machine if proper scaling were not applied. Table 2-2 presents both the original and the scaled forms of the Forward and Backward Procedures so the reader can compare them.

Problem 2 is: given the observation sequence, O, and the model, λ, how can an optimal hidden-state sequence, q, be chosen? Although there are several possible optimal criteria to apply, the most widely used criterion is to find the path that maximizes $P(q|O,\lambda)$, which is also equivalent to finding the path that maximizes $P(q,O|\lambda)$. To solve this maximization problem, a dynamic programming method called the Viterbi algorithm is employed (Viterbi, 1967). In this algorithm, a new variable, $\delta_t(i)$, is defined as follows:

$$\delta_t(i) = \max_{q_1, q_2, \ldots, q_{t-1}} P(q_1, q_2, \ldots, q_{t-1}, q_t = i, o_1, o_2, \ldots, o_t \mid \lambda) \tag{2.9}$$

This variable, $\delta_t(i)$, indicates the highest probability among all possible paths, $\{q_1, \ldots, q_{t-1}\}$, as well as having $q = i$ at time t and the first t observations, $\{o_1, \ldots, o_t\}$, given the model, λ. Again, $\delta_{t+1}(j)$ can be computed from the $\delta_t(i)$ by the following induction:

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i) a_{ij} \right] b_j(o_{t+1}) \tag{2.10}$$

To retrieve the hidden-state sequence at the end, the argument that maximizes Eq. (2.10) for each *t* and *j* is recorded by another variable,

$$\psi_t(j) = \arg\max_{1 \le i \le N} \left[\delta_{t-1}(i)\, a_{ij} \right] \tag{2.11}$$

As shown in Table 2-2, the Viterbi algorithms have two versions: the original form and a logarithm form. In the logarithm form, all multiplications in the original form are converted into additions, and, as a result, the numerical underflow problem can be avoided. Since no scaling is necessary for the logarithm form, often this form is used in the actual HMM estimation process rather than the original form.

Problem 3 is that, given an observation sequence, *O*, how can the model parameters, $\lambda = (A, B, \pi)$, be adjusted to maximize $P(O|\lambda)$? This is the most difficult to solve of the three problems. There is no known way to analytically solve for the global maxima of the model parameters in a closed form. However, there is still a way to locally maximize the likelihood, $P(O|\lambda)$, using an iterative procedure called the Baum-Welch method, also known as the EM (expectation-maximization) method (Dempster, Laird, & Rubin, 1977).

To start, a new variable, $\xi_t(i,j)$, is defined as follows:

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j \mid O, \lambda) \tag{2.12}$$

This variable indicates the probability that the hidden state is *i* at time *t*, and then *j* at time *t+1*, given the observation sequence and the model parameters. Then, the values of this variable, $\xi_t(i,j)$, can be computed as

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \tag{2.13}$$

Furthermore, another new variable, $\gamma_t(i)$, is defined as follows:

$$\gamma_t(i) = P(q_t = i \mid O, \lambda) \tag{2.14}$$

That is, $\gamma_t(i)$ is the probability of being in the hidden state, *i*, at time, *t*, given the observation sequence and the model parameters. Thus,

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j) \tag{2.15}$$

Now, if these new variables are summed over time, they can be interpreted as

$\sum_{t=1}^{T} \gamma_t(i)$ = Expected number of transitions from state *i* in *O*

$\sum_{t=1}^{T} \xi_t(i,j)$ = Expected number of transitions from state *i* to state *j* in *O*

Therefore, from the current model parameters, $\lambda = (A, B, \pi)$, the re-estimation of the model parameters, $\lambda' = (A', B', \pi')$, can be computed as

$$\pi_i' = \gamma_1(i) \tag{2.16}$$

$$a_{ij}' = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{2.17}$$

$$b_j'(k) = \frac{\sum_{t=1,\, s.t.,\, o_k=v_k}^{T-1} \gamma_t(i)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{2.18}$$

It has been proven that, when Eqs. (2.16)-(2.18) are used to compute the model parameters re-estimation, $\lambda' = (A', B', \pi')$, from the original parameters, $\lambda = (A, B, \pi)$, then, either (1) $\lambda = \lambda'$, or (2) $P(O|\lambda') > P(O|\lambda)$ (Rabiner & Juang, 1993).

By combining the above algorithms, i.e., the Forward Procedure, the Backward Procedure, the Viterbi algorithm, and the Baum-Welch method, one can construct a generic HMM estimation process, which takes the initial model parameters, $\lambda_0 = (A_0, B_0, \pi_0)$, and the observation sequence, *O*, as inputs, and returns the estimated model parameters, $\lambda' = (A', B', \pi')$, and the estimated hidden-state sequence, q^* (and the estimated probability of the observation sequence, P^*, an important value in the case of speech recognition and other pattern recognition applications, but not used in pilots' eye-movement data analysis). Table 2-2 lists the entire HMM algorithms, and Figure 2-3 summarizes the overall implementation of these HMM estimation algorithms. The MATLAB® codes for the HMM analysis used in the experiments are shown in Appendix A.

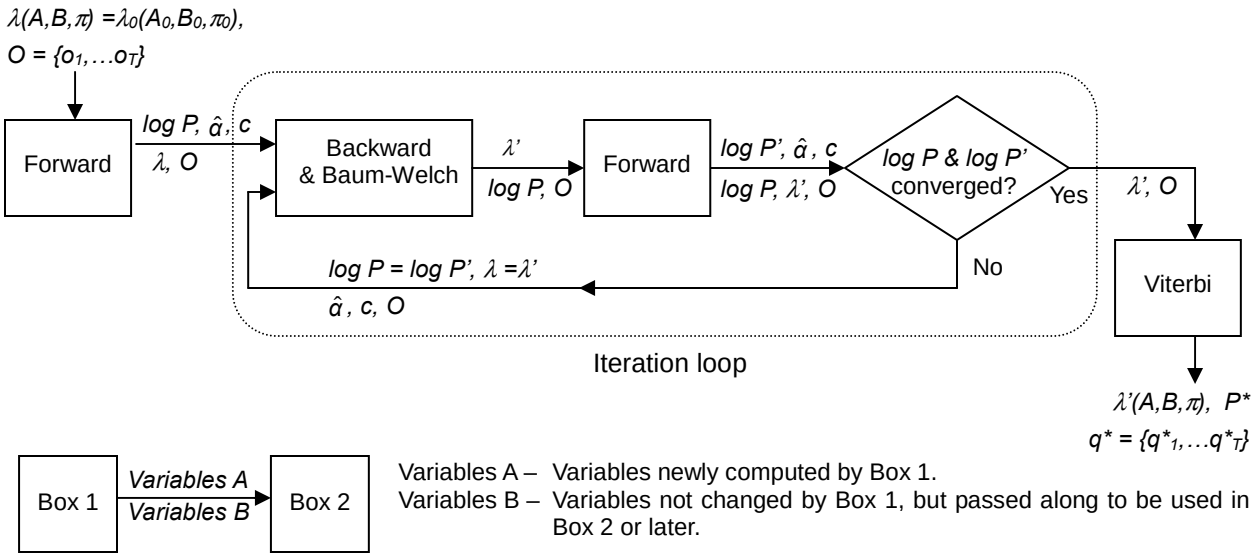


Figure 2-3. HMM algorithm implementations

Table 2-2. HMM algorithms

Forward Procedure	***Original Form:***	
	1. Initialization: $\alpha_1(i) = \pi_i \; b_i(o_1) \,, \quad 1 \le i \le N$	(2.A1)
	2. Induction: $\alpha_{t+1}(i) \;=\; \left[\sum_{j=1}^{N} \alpha_t(j) \, a_{ji} \right] b_i(o_{t+1}), \quad 1 \le t \le T-1, \; 1 \le i \le N$	(2.A2)
	3. Termination: $P(\, O \mid \lambda \,) \;=\; \sum_{i=1}^{N} \alpha_T(i)$	(2.A3)
	Scaled Form:	
	1. Initialization: $\hat{\hat{\alpha}}_1(i) = \pi_i \; b_i(o_1) \,, \quad 1 \le i \le N$	(2.B1)
	$c_1 = \left[\sum_{i=1}^{N} \hat{\hat{\alpha}}_1(i) \right]^{-1}$	(2.B2)
	$\hat{\alpha}_1(i) = c_1 \, \hat{\hat{\alpha}}_1(i)$	(2.B3)
	2. Induction: $\hat{\hat{\alpha}}_{t+1}(i) \;=\; \left[\sum_{j=1}^{N} \hat{\alpha}_t(j) \, a_{ji} \right] b_i(o_{t+1}) \,, \quad 1 \le t \le T-1, \; 1 \le i \le N$	(2.B4)
	$c_t = \left[\sum_{i=1}^{N} \hat{\hat{\alpha}}_t(i) \right]^{-1}$	(2.B5)
	$\hat{\alpha}_t(i) = c_t \, \hat{\hat{\alpha}}_t(i)$	(2.B6)
	3. Termination: $\log \left[P(\, O \mid \lambda \,) \right] \;=\; -\sum_{t=1}^{T} \log \, c_t$	(2.B7)
Backward Procedure	***Original Form:***	
	1. Initialization: $\beta_T(i) \;=1 \,, \quad 1 \le i \le N$	(2.C1)
	2. Induction: $\beta_t(i) \;=\; \sum_{j=1}^{N} a_{ij} \; b_j(o_{t+1}) \, \beta_{t+1}(j), \quad t = T-1, \ldots, 1, \; 1 \le i \le N$	(2.C2)
	Scaled Form:	
	1. Initialization: $\beta_T(i) \;=1 \,, \quad 1 \le i \le N$	(2.D1)
	$\hat{\beta}_T(i) \;= c_T$	(2.D2)
	2. Induction: $\beta_t(i) \;=\; \sum_{j=1}^{N} a_{ij} \; b_j(o_{t+1}) \, \hat{\beta}_{t+1}(j), \quad t = T-1, \ldots, 1, \; 1 \le i \le N$	(2.D3)
	$\hat{\beta}_t(i) \;=\; c_t \beta_t(i)$	(2.D4)

CONTINUED TO THE NEXT PAGE.

Table 2-2. HMM algorithms (Continued)

<table>
<tr><td>Viterbi Algorithm</td><td>

Original Form:

1. Initialization: $\delta_1(i) = \pi_i b_i(o_1), \quad \psi_1(i) = 0, \quad 1 \le i \le N$ (2.E1, E2)

2. Recursion: $\delta_{t+1}(j) = \max_{1 \le i \le N} \left[\delta_t(i) a_{ij} \right] b_j(o_{t+1}), \; 1 \le t \le T-1, 1 \le j \le N$ (2.E3)

$$\psi_{t+1}(j) = \arg \max_{1 \le i \le N} [\delta_t(i) a_{ij}]$$ (2.E4)

3. Termination: $$P^* = \max_{1 \le i \le N} [\delta_T(i)]$$ (2.E5)

$$q_T^* = \arg \max_{1 \le i \le N} [\delta_T(i)]$$ (2.E6)

5. Backtracking: $q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \ldots, 1$ (2.E7)

Logarithm Form:

1. Preprocessing: $\tilde{\pi}_i = \log(\pi_i), \; \tilde{a}_{ij} = \log(a_{ij}), \; 1 \le i \le N, 1 \le j \le N$ (2.F1,F2)

$$\tilde{b}_i(o_t) = \log [b_i(o_t)], \quad 1 \le t \le T$$ (2.F3)

2. Initialization: $\tilde{\delta}_1(i) = \tilde{\pi}_i + \tilde{b}_i(o_1), \quad \psi_1(i) = 1, \quad 1 \le i \le N$ (2.F4, F5)

3. Recursion: $\tilde{\delta}_{t+1}(j) = \max_{1 \le i \le N} \left[\tilde{\delta}_t(i) + \tilde{a}_{ij} \right] + \tilde{b}_j(o_{t+1}), \quad 1 \le t \le T-1, \; 1 \le j \le N$ (2.F6)

$$\psi_{t+1}(j) = \arg \max_{1 \le i \le N} [\tilde{\delta}_t(i) + \tilde{a}_{ij}]$$ (2.F7)

4. Termination: $$\tilde{P}^* = \max_{1 \le i \le N} \left[\tilde{\delta}_T(i) \right]$$ (2.F8)

$$q_T^* = \arg \max_{1 \le i \le N} \left[\tilde{\delta}_T(i) \right]$$ (2.F9)

5. Backtracking: $q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \ldots, 1$ (2.F10)

</td></tr>
<tr><td>Baum-Welch Method</td><td>

$$\xi_t(i,j) = \frac{\hat{\alpha}_t(i) a_{ij} b_j(o_{t+1}) \hat{\beta}_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\alpha}_t(i) a_{ij} b_j(o_{t+1}) \hat{\beta}_{t+1}(j)}, \quad \begin{matrix} 1 \le t \le T-1 \\ 1 \le i \le N, \; 1 \le j \le N \end{matrix}$$ (2.G1)

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$$ (2.G2)

Parameter Re-estimation:

$$\pi_i' = \gamma_1(i), \qquad a_{ij}' = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \qquad b_j'(k) = \frac{\sum_{t=1,\, s.t.,\, o_k = v_k}^{T-1} \gamma_t(i)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$ (2.G3,G4,G5)

</td></tr>
</table>

2.3 HMM Analysis of Pilots' Eye Movements

The HMM algorithms described in the previous section are now adapted to estimate pilots' attention-switching process from their eye-movement data. This section provides step-by-step descriptions of the analysis process. Some important details, such as assumption of instrument fixation durations, effects of the initial conditions of model parameter matrices, and the requirement for pilots' verbal reports, are also discussed.

Modern eye-tracking systems typically record eye-movement data at the rate of 60Hz or more, which is far more frequent than is needed for the purpose of HMM analysis. Therefore, the first task for researchers is usually data reduction. First, the saccade data points (i.e., the data points indicating rapid eye movements from one fixation point to the other) are eliminated so that the remaining data points are all considered fixations. Blinking should also be accounted for; for instance, if the pilot looked at the same instrument before and after a blink, then it was considered as if the pilot kept looking at the instrument even during the blink. On the other hand, if the pilot switched to a different instrument after a blink, then the fixation on the first instrument was considered to have been terminated right before the blink, and the fixation on the second instrument was considered to have been started right after the blink, and so on.

When is a fixation considered consecutive fixations? This question is related to how long it usually takes for pilots to read an instrument. Fitts and his colleagues reported that, in their in-flight experiment, the average instrument fixation durations were about 0.6 seconds (Fitts et al., 1950). On the other hand, Carbonell and his colleagues presumed in their queuing theory model that pilot fixations longer than 0.4 seconds indicated consecutive selection of the same instrument (Carbonell et al., 1968). In this thesis, the 0.5-second threshold was chosen to identify consecutive fixations. Based on this threshold, the number of consecutive fixations were counted as follows: fixations staying on the same instrument for less than 0.5 seconds were considered a single fixation; fixations remaining for more than 0.5 seconds were counted as two fixations; those staying for more than one second were three fixations; and so on.

The resulting instrument fixation sequence, $O = \{o_1, \ldots, o_T\}$, and the initial conditions for the HMM parameters, $\lambda_0 = (A_0, B_0, \pi_0)$, are sent to the HMM estimation process as shown in the left part of Figure 2-3. To determine an adequate λ_0 at the beginning, researchers are forced to use

their engineering judgment. As described in the previous section, the Baum-Welch method computes the local maximums, not the global ones. Thus, it is critical to pick adequate initial conditions so that the resulting local maximums are as close to the global maximums as possible. Experience has shown that the estimation results are usually less sensitive to the choice of A_0 or π_0 than that of B_0 (Rabiner & Juang, 1993). Even random or uniform A_0 or π_0 may be adequate for obtaining useful re-estimation results (subject to the constraints, $\Sigma_j\, a_{ij} = 1$, and $\Sigma_i\, \pi_i = 1$). The diagonal elements of A_0, a_{ii}, represent the recurrent probability of the state, i, and in the experiments in this thesis, these were set to *0.7* or *0.8* in most cases to slightly bias toward staying in the same state rather than moving out to another state. Then the remaining *0.3* or *0.2* was distributed uniformly among the rest of the elements of the row, a_{ij} $(i \neq j)$, each of which represents the transition probability from state i to state j. Uniform π_0 was used in all cases.

On the other hand, as previously mentioned, the choice of B_0 seems to have relatively more influence on both the Baum-Welch and Viterbi estimation results than that of A_0 or π_0. Especially the balance of the elements of B_0 among overlapping instruments seems to determine the bias on the Viterbi algorithm outcomes, and therefore should be chosen carefully. For example, if $b_{vertical}(AI) >> b_{horizontal}(AI)$, the Viterbi algorithm tends to pick the vertical task for AI fixations. Knowing these characteristics, researchers still have to rely on engineering judgment to pick the initial B_0 (also subject to the constraint, $\Sigma_k\, b_j(v_k) = 1$). The actual initial conditions, A_0, B_0, π_0, used in the experiments in Chapter 3 and 4 are listed in Appendix C.

When the backward procedure and the Baum-Welch algorithm are completed, $log[P'(O|\lambda')]$ is calculated by the forward procedure with the new λ', and checked for convergence with the original $log[P(O|\lambda)]$. The iteration loop is repeated until the difference between *log P* and *log P'* becomes less than the user-defined tolerance threshold—1% was used for the experiments in this thesis—of *log P'*. In these experiments, the number of iterations rarely exceeded three. In fact, if *log P* does not reach convergence within the first couple of iterations, that may be a sign that the HMM structure needs to be modified. The issue of choosing a proper HMM structure will be discussed more in Chapter 3. After *log P* converges, the Viterbi algorithm computes the estimated hidden-state sequence, $q^* = \{q_1,\ldots,q_T\}$.

To verify this estimated hidden-state sequence, q^*, it is compared with the pilots' own verbal

reports. During the simulation flights, the pilots' verbal reports of the tracking tasks to which they are currently attending are collected. This can be done by asking them to verbalize either their current intentions or the instrument information they are attending to. In the first experiment, described in Chapter 3, the pilots were given no specific format for this verbal reporting, and they freely used their own words to describe their current flight intentions (e.g., "climbing," "left turn for localizer") or the instrument information they were attending to (e.g., "2.5 degree," "169 knots"). This type of verbal report format will be called "free mumbling" in this thesis. With this format, the researcher can see the context of the flight. Also this format is convenient when the proper verbal report formats to be instructed to the pilots are unknown because the HMM structure is not initially known. In that case, the researcher can also use the pilots' free-mumbling verbal report results to construct the initial HMM structure after some preliminary experiments. The drawback of this type of format is that the free generation of the verbal reports may increase the pilots' cognitive workload, which in turn may affect their flight performance.

In the effort to reduce the pilots' cognitive workload, a different reporting format was attempted in the next experiment, described in Chapter 4. In this experiment, the pilots were instructed to say "pitch" when they were attending to the pitch information on the AI, and "bank" when attending to the bank information. The use of predetermined words eliminated the need of free generation of the verbal reports, and was expected to reduce the pilots' cognitive workload. To use this reporting format, the researcher already had to have a good idea of what the HMM structure should be. Then, the reporting words were chosen to reduce ambiguity about tracking tasks when the pilots looked at overlapping instruments (in the experiment in Chapter 4, the AI was the only overlapping instrument). This type of format will be called "structured reporting" in this thesis. Although the structured reporting was originally implemented to address a certain disadvantage of the free mumbling, it turned out that both formats had their own pros and cons. Further discussion about the verbal-report format issues will be given in Chapter 5.

Then, these verbal reports were converted to the corresponding tasks so that the reports could be used as the indicators of their true hidden states in the verification process. If an HMM estimated task matched the verbal report at the reported time, the report was considered to "match" the estimate. If the free-mumbling format was being used, the verbal reports regarding the

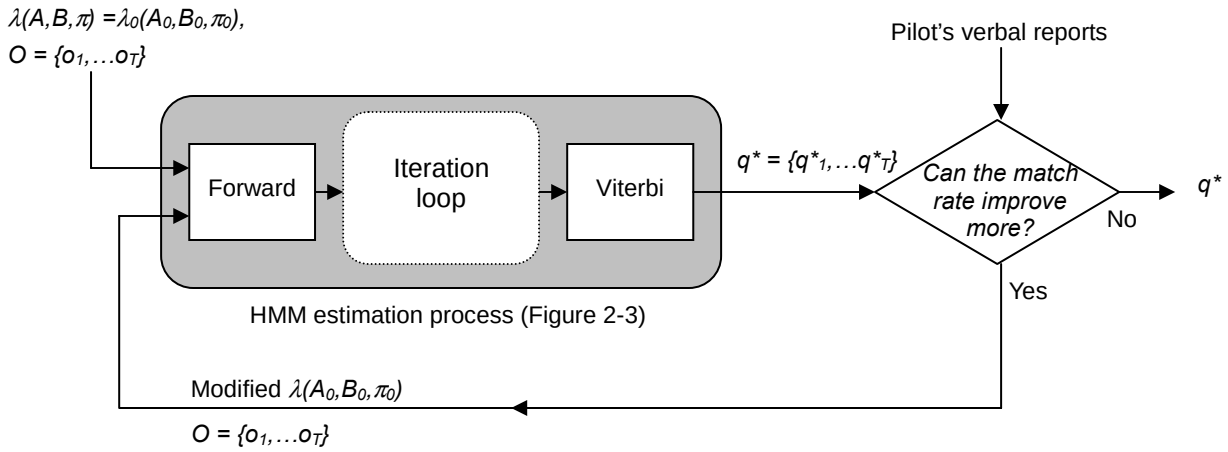


Figure 2-4. HMM analysis process for pilots' eye movements

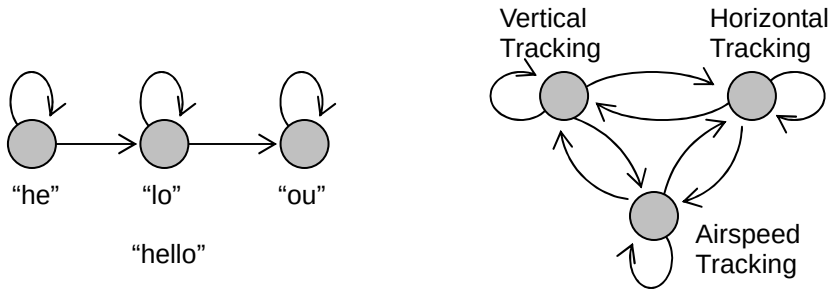
instruments included in only one tracking task, such as the ASI (airspeed tracking), the altimeter (vertical tracking), and the HI (horizontal tracking), was omitted from the analysis because they usually do match and easily increase the overall rate of match. Eliminating them kept the verification process conservative.

If the match rate between the pilots' verbal reports and the HMM estimation results did not turn out very good, and also there were certain trends or patterns in the way the mismatches occurred, then the researcher might be able to improve the match rate by adjusting the initial conditions of the HMM parameters, $\lambda_0 = (A_0, B_0, \pi_0)$, especially the elements of the B_0 matrix corresponding to the overlapping instruments. For instance, if missed detections repeatedly occurred when the pilot generated verbal reports related to the pitch-related task (vertical-tracking task) while the HMM estimation resulted in the horizontal-tracking task, then the $b_{vertical}(AI)$ in the B_0 matrix could be increased in comparison with the $b_{horizontal}(AI)$ to increase bias toward the vertical-tracking task when the AI was looked at.

The entire HMM estimation process was repeated until the resulting rate of match showed no further improvement. The Figure 2-4 illustrates the overall iteration process described above.

2.4 Comparison with Speech Recognition HMM

The way the HMM was utilized in the analysis of pilots' eye movements above is fundamentally different in several ways from the conventional way that HMMs have been used in speech



(a) Three-state left-right model (e.g., Speech recognition)

(b) Three-state ergodic model (Pilots' eye-movement analysis)

Figure 2-5. HMM types

recognition and other pattern recognition applications. So far, most of the conventional HMM applications have been about searching for a particular hidden-state sequence based on noisy or probabilistic observation symbol sequences. For instance, in speech recognition cases, their main interest is to see if a sequence of phonemes in the sound data of a spoken word fits the sequence of any particular word in their vocabulary database. Analogously, the main interest in sign-language recognition is to look for a particular sequence of hand movements, and in protein structure analysis it is to seek similarities in amino acid sequences, and so forth.

To further explain the difference from these conventional HMM applications, let us use the speech recognition application as a representative example. Such phoneme sequence recognition can be done with HMMs by defining the phonemes of a target word as the hidden states, and the actual spoken-sound data as the observation symbol sequence. For instance, the HMM structure for the word, "hello," may appear as shown in Figure 2-5 (a), consisting of three hidden states corresponding to the three phonemes, "he," "lo," and "ou," respectively. The state transition probability values are assigned so that only one-directional transitions from left to right are allowed. In other words, $a_{ij} = 0$ for $i > j$. Such HMMs are called left-right type models (Rabiner & Juang, 1993).

Unlike the left-right type HMMs for the speech recognition, the HMM for pilots' eye-movement process has the hidden states all connected to each other, and, thus, every hidden state can be reached from every other state within one step. In other words, and $a_{ij} \neq 0$ for all i and j (Figure 2-5 (b)). Such models are called ergodic type models.

By the way, the ergodic type models, as well as the left-right type models, assume that the transition probabilities, including the recurring probabilities, do not change over time. This assumption may be a little oversimplified for the case of pilots' eye-movement data analysis, though these HMMs still would work fine regardless of whether this assumption is true or not. (In case the probabilities are time-varying, the probabilities estimated by the HMM will be simply the average over time.) This point will be discussed again in Sections 3.4.3 and 5.3.

Another major difference between pilots' eye-movement data analysis and speech recognition is that there is no training data (or labeled data) available indicating the pilots' true hidden states. For this reason, the pilots' own verbal reports have to be collected for verification. For speech recognition applications, there are spoken-word sound databases where the true states are known. Speech recognition engineers can use these databases to train their HMMs until correct recognition rates reach the maximum. Then, the engineers create a bank of HMMs whose model parameters are trained for different words. The spoken sound data of an unknown word are sent to each HMM and processed by the Viterbi algorithm to compute the probability of the model likelihood, P^*. If the spoken word is the word that one of the HMMs was trained for, then the Viterbi algorithm of the corresponding HMM should return significantly higher P^* than other HMMs. That HMM indicates the recognized word.

This brings us to our final point. The ultimate purpose of speech recognition is to build a robust classifier that can recognize the spoken words that are yet to come. On the other hand, the goal of pilots' eye-movement data analysis is to find out the model parameters and a hidden-state sequence that best describe the pilots' eye-movement data already obtained. For the speech recognition case, to achieve robust performance, the engineers often use cross-validation for training (Bishop, 1996). In this validation method, the data are partitioned into S segments, and the HMM is trained using only the $S-1$ segments. Then, the omitted segment is used to test the trained HMM. This process is repeated S times so that all segments are assigned as the test data once. Then, the HMM that results in the highest correct recognition rate of the S HMMs is chosen as the best classifier. This approach was not taken for the pilots' eye-movement data analysis in this study, and all data always were used to compute the model parameters. However, if one were interested in building a pilots' eye-movement data classifier that could classify new data in the future, this cross-validation approach might prove to be useful.

Chapter 3:

Flight Simulator Experiment:

Differences Among Individual Pilots

The HMM analysis described in the previous chapter is now applied to actual pilots' eye-movement data collected in a flight simulator experiment. The experiment constitutes a proof of concept of the use of HMM analysis for investigating attention switching sequences. However, more importantly, this experiment also explores a potential HMM structure(s) for the actual pilots' instrument fixation data. Although the three-state HMM introduced in the previous chapter mostly agrees with the instrument categorization used in instrument flight textbooks (e.g., FAA, 2001), that does not necessarily guarantee that this is the only possible HMM structure. If there is any other HMM structure, how can one identify it from the eye-movement data? This fundamental issue of HMM analysis is examined in this chapter.

3.1 Differences Among Individual Pilots

It has been reported that there are large differences among the instrument scanning patterns of different pilots. For instance, Fitts and his colleagues noted marked differences in the eye-movement characteristics of 40 Air Force pilots (Fitts et al., 1950). During their in-flight instrument approach experiment, for example, one pilot made 154 fixations per minute on average, while another pilot averaged only 65 fixations per minute. Obviously, different pilots have different levels of agility of eye motion. In addition, it was also observed that one pilot made 41 fixations on the artificial horizon (i.e., the AI) out of 140 fixations per minute, while another pilot hardly used this instrument at all. While the varying eye-movement agility may be simply due to different scanning speeds of individuals, the different fixation frequencies on the AI may reflect more basic control strategy difference, such as the ones found in the crossover model analysis by Weir and McRuer (Weir & McRuer, 1972). They analyzed the control inputs of two airline pilots during instrument approaches on a simulator, and found quite different

longitudinal loop closing strategies between them: one pilot emphasized pitch response (inner-loop response), whereas the other emphasized glide slope deviation (outer-loop response). The eye-movement data of the pilot who controlled primarily based on pitch response indicated more frequent fixations on the AI than the other pilot (when flight director bars were off).

Although these pilots' personal scanning styles or control strategy differences are important for researchers to understand, and will certainly affect the parameters of the HMM, it is unlikely that they affect the HMM structure itself. Whether the pilot fixated on the AI and the altimeter (both in the vertical-tracking group) back and forth 5 times in 5 seconds or only 2 times in 5 seconds, or whether the pilot used the AI (pitch) more often than the CDI (glide slope deviation) for vertical axis control, as long as these are in the same vertical-tracking instrument group, it does not affect the HMM structure itself. The framework of the HMM structure provides robustness for these kinds of differences among pilots' styles.

What is being sought here is any new set of instrument groups different from the vertical-, horizontal-, and airspeed-tracking grouping regime. Even though these three instrument groups are taught to pilots during their initial stage of instrument flight instructions, their scanning patterns are mostly developed through their own flight experience. The large variety of scanning patterns among pilots reported in the literature supports this point as well.

Therefore, the suspicion here is that the pilots' levels of flight expertise may be a major factor in causing different HMM structures. Hence, in the following flight simulator experiment, the eye-movement data of four pilots who had different levels of flight expertise were analyzed to examine the following null hypothesis:

***H_0*:** There is no other instrument scanning HMM structure than the three-state HMM structure.

This null hypothesis is rejected if at least one HMM structure different from the original three-state HMM structure is found. This may sound simple. However, the real challenges in this investigation lie in how to identify an HMM structure from given data and how to justify it. Note that finding an exhaustive list of existing HMM structures is not the purpose of this experiment.

3.2 Method

3.2.1 Pilot Participants

Four pilots of different levels of expertise participated in the simulation. The pilots were numbered in order of experience. Table 3-1 lists these pilots.

Table 3 -1. Pilot participants

Pilot #	Pilot Types	Total Hours	Instrument Hours	Major Aircraft Flown
1	Private Pilot, Instrument Rated	250	60	C-172, PA-28
2	Commercial Pilot, Certified Flight Instructor-Instrument	700	100	C-172, AA-5B, PA28, PA28R
3	Former Navy Pilot, Commercial Pilot, Instrument and Multi-Engine Rated	1050	225	P-3
4	Airline Transport Pilot	3500	350	C-5, B-707, B-757, B-767

All participants read and signed an informed consent form complying with the MIT COUHES Protocol 3021 (Appendix B) before the experiment started. All pilots were employees of Volpe Center in Cambridge, MA, and participated in the experiment as part of their normal work (i.e., no additional compensation was offered).

Although Pilot 4 was mostly flying a C-5 at the time of experiment, this pilot also had previous experience of flying the Boeing 757, the flight dynamics the flight simulator used, and may have had an advantage in this experiment. Of the four pilots, Pilot 2 and 4 were current pilots as of the day of the experiment.

3.2.2 Flight Simulator

The experiment used a fixed-base flight simulator at Volpe National Transportation Systems Center located in Cambridge, MA. The simulator used the Boeing 757-200 flight dynamics model provided in Microsoft® Flight Simulator 1998, enhanced by the Precision Manuals Development Group 757-200 add-on. A custom OpenGL® program generated the instrument panels, which are described more in the next paragraph. In the simulator cockpit, a control column with a control wheel was provided in front of the pilot. A single throttle lever that controls both engines of the aircraft was at the right side of the pilot. The flaps, gear, and trim

thumb switches were provided on the control wheel, so that the pilots could change these settings without looking away from the display.

3.2.3 Display

The Basic-T instrument (Figure 3-1) was presented on a 17-inch computer monitor in front of the pilot, approximately 30 inches away from the pilot's eyes. The instrument images in the display were generated by a custom OpenGL® program and depicted in bright green on a black background. The instruments included were the airspeed indicator (ASI), attitude indicator (AI), altimeter, course deviation indicator (CDI), heading indicator (HI), and vertical speed indicator (VSI). The power indicator (displaying the engine "N1" turbine RPM value) was presented at the lower right corner in a digital format, and the flaps-and-gear indicator was placed in the upper right corner. A turn-and-slip coordinator was not included in this display due to space limitations. The display area subtended a visual angle of 23.0° horizontally and 17.6° vertically at a viewing distance of 30 inches. The six large instruments subtended a visual angle of 5.2° in diameter, with 1.4° horizontal and 1.7° vertical separations. The visual angles for the instrument diameters were approximately comparable to those of actual aircraft gauges, but the visual angles for the

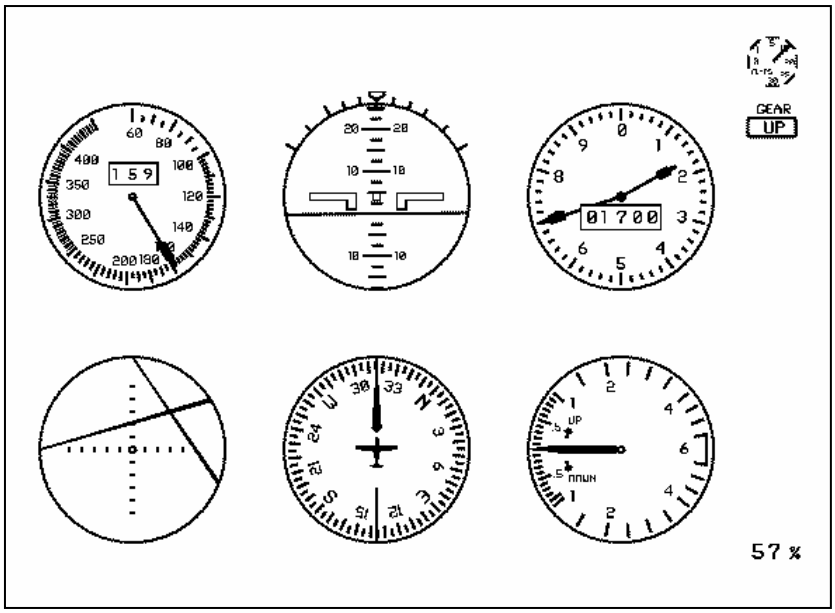


Figure 3-1. Display used in the simulation: The six large instruments are (top, from left) airspeed indicator (ASI), attitude indicator (AI), altimeter; (bottom, from left) course deviation indicator (CDI), heading indicator (HI), vertical speed indicator (VSI). The flaps-and-gear indicator is at the upper right corner. The power indicator is at the lower right corner.

instrument separations were slightly smaller than those of actual aircraft instrument panels, due to the size of the monitor used in the experiment.

3.2.4 Simulation Scenario

Figure 3-2 shows a schematic of the simulated ILS approach. No out-the-window view was presented; i.e., all flights were flown referencing only the instruments. The aircraft was initially positioned on the left side of the localizer course. The approach can be divided into three flight segments:

(i) Straight and level at 1700 ft, 160 knots
(ii) Level turn to intercept the localizer, flaps down by two notches, gear down, and slow to 130 knots
(iii) Final descent along the glide path to 1000 ft

In segment (i), pilots were instructed to maintain 1700 ft and 160 knots. When the localizer needle became active, the pilots started a left turn and this started segment (ii). The angle of intercept was approximately 25°. This segment involved configuration changes, such as lowering gear as well as lowering flaps by two notches, and slowing to 130 knots. When the localizer was centered and the glide slope needle started to move in, the pilots began the descent, and this

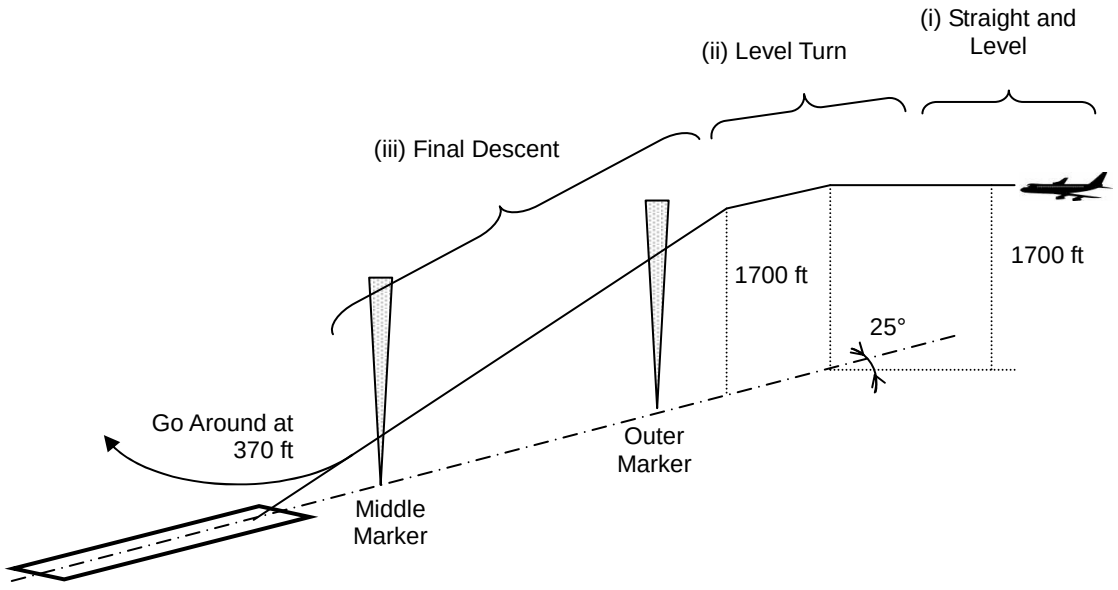


Figure 3-2. Three flight segments in the ILS scenario

started segment (iii). The pilots were instructed to keep the localizer and glide slope needles centered and maintain 130 knots. When the aircraft reached 370 ft, the pilots initiated a go-around, and data collection ended for the approach. It was decided not to add any wind in any segment of this simulation because the disturbance induced by the pilots themselves already made the approaches hard enough to fly.

3.2.5 Data Collection

During each approach, the pilots' eye-movement data were collected with a head-mounted video eye-tracker (RK-726PCI/RK-620PC, ISCAN, Inc., Burlington, MA) and a magnetic head tracker (InsideTRAK, Polhemus, Colchester, VT). The accuracy of the eye-tracker is approximately 1°. The static accuracies of the head tracker are 0.05" root mean square (RMS) for X, Y, or Z position, and 2° RMS for orientation.

The eye-movement data were collected at the rate of 60 Hz. Flight variables were also recorded, at 1 Hz. In addition, the pilots' verbal reports of their current intention or instrument reading were recorded on a videotape for later HMM analysis. The free-mumbling format (i.e., the pilots freely generate sentences using their own words) was used for the pilots' verbal reports in this experiment. Each pilot flew three data-collection approaches. Before the data-collection approaches, each pilot received a briefing (Appendix B) and made several practice approaches, at least one of which was made while wearing the eye-tracking headgear and making verbal reports for familiarization.

3.2.6 Data Analysis

From the eye-movement and head-movement data, intersection coordinates of the pilots' line-of-sight and the display plane were computed. All computations for the eye-movement data listed in this section (preprocessing, K-means segmentation, and HMM analysis) were performed by MATLAB® programs coded by the author. The MATLAB® codes for the HMM analysis are listed in Appendix A.

First, the preprocessing for the saccade and blink data points was performed as follows. The data associated with saccades were eliminated by omitting the data points that failed to meet the fixation criteria (i.e., staying within an ellipse of 5 pixels horizontally, 3 pixels vertically, for 2 consecutive samples). Blinking was also handled in the manner described in Section 2.3.

Second, the K-mean clustering segmentation was applied as follows to map each data point to a corresponding instrument. The line-of-sight intersections for each instrument were usually well separated, as shown in Figure 3-3 (a), and, thus, they could be easily clustered for the associated instrument. For instance, in the experiments in this thesis, a K-mean clustering segmentation algorithm (Bishop, 1996, pp.187-188) was employed to cluster them. In this algorithm, the n intersection coordinates vectors, $\{x_1,\ldots, x_n\}$, were divided into K clusters, where each cluster was represented by the mean vectors, μ_j, $j = 1, \ldots, K$. The j-th cluster contained N_j intersection vectors. First, arbitrary mean vectors, μ_j, were picked as the initial conditions, and distance to each x_i,

$$d_{ij} = \left\| x_i - \mu_j \right\|^2 \tag{3.1}$$

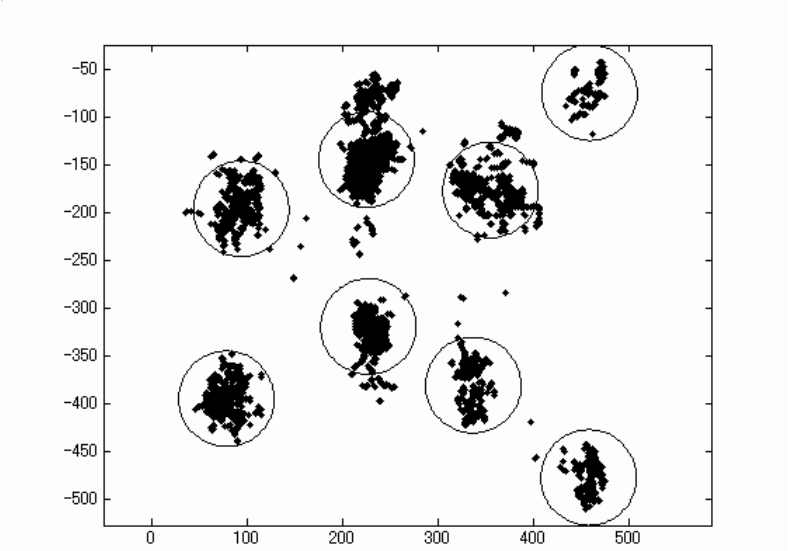
was computed. Then, each x_i was assigned to the nearest cluster, S_j, whose mean vector, μ_j, had the shortest distance d_{ij}. Then, the new mean vectors were recomputed by

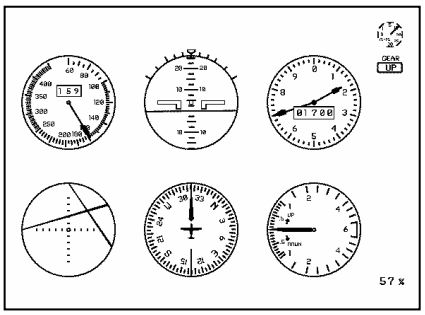
$$\mu_j = \frac{1}{N_j} \sum_{x_i \in S_j} x_i \tag{3.2}$$

Then, each x_i was reassigned to new cluster, S_j, whose new mean vector was the closest to x_i. This process was repeated until there was no further change in the grouping of x_i. This could be assessed by monitoring the increments of the sum of distances,

$$J = \sum_{j=1}^{K} \sum_{x_i \in S_j} \left\| x_i - \mu_j \right\|^2 \tag{3.3}$$

Figure 3-3 (a) shows an example of the K-mean segmentation clustering results of actual line-of-sight intersection coordinates of a pilot scanning the display arranged as in Figure 3-3 (b). The centers of circles in Figure 3-3 (a) indicate the resulting representative mean vectors of the clusters. (Note that circles do not represent the boundaries of clusters.) Each x_i was assigned to its closest cluster. Since all of the major six instruments had identical size and shape, and the two smaller instruments at the upper right and the lower right were sufficiently far apart from the major six instruments even though they were slightly smaller in size, no weight was imposed for each instrument when the distance from each x_i to each mean vector, μ_j, was computed in Eq. (3.1). If each instrument had a different shape and size, weights could be applied for different instruments in this equation—as seen in the HUD experiment case described in the next chapter.





(b) Display

(a) Intersections of the pilot's line-of-sight and the display plane (the first approach of Pilot 2, saccade points have been removed) and K-mean segmentation results: the centers of circles indicate the centers of the clusters that the K-mean segmentation algorithm estimated. The units are pixels.

Figure 3-3. K-mean segmentation

Once clustering was finished, the instrument fixation sequences were computed. As explained in Section 2.3, the intersections staying within the same instrument for less than 0.5 seconds were considered a single fixation. Likewise, intersections remaining for more than 0.5 seconds were counted as two fixations; those staying for more than one second were three fixations; and so on.

Finally, HMM analysis was applied separately to each pilot's data. In addition, since each flight segment had different task requirements, each segment was analyzed separately. Let $O^{s,r,p} = \{o_1^{s,r,p}, \ldots, o_T^{s,r,p}\}$ denote the fixation sequence data from the segment, s $(1 \le s \le 3)$, of the r-th approach $(1 \le r \le 3)$, of the Pilot, p $(1 \le p \le 4)$. The simplest way to estimate the parameters that maximize $P(O^{s,1,p}, O^{s,2,p}, O^{s,3,p} | \lambda)$ for a given segment for a specific pilot was to concatenate $O^{s,1,p}$, $O^{s,2,p}$, $O^{s,3,p}$ to make one long data string, $O^{s,p} = \{o_1^{s,1,p}, \ldots, o_{T1}^{s,1,p}, o_1^{s,2,p}, \ldots, o_{T2}^{s,2,p}, o_1^{s,3,p}, \ldots, o_{T3}^{s,3,p}\}$ $(T1 = T^{s,1,p}, T2 = T^{s,2,p}, T3 = T^{s,3,p})$, and then estimate the parameters that maximize $P(O^{s,p} | \lambda)$. This process connected two data points that were not originally connected (i.e., $\{o_{T1}^{s,1,p}$ and $o_1^{s,2,p}\}$ and $\{o_{T2}^{s,2,p}$ and $o_1^{s,3,p}\}$), but the effects of it were considered minimal because the model structure was an ergodic type, which did not have any prohibited hidden-state transitions (see Section 2.4 for model types). Finally, the estimated attention sequences, $Q^{s,p}$ $(1 \le s \le 3, 1 \le p \le 4)$, are divided back to the partial sequences corresponding to each approach, $Q^{s,r,p}$ $(1 \le s \le 3, 1 \le$

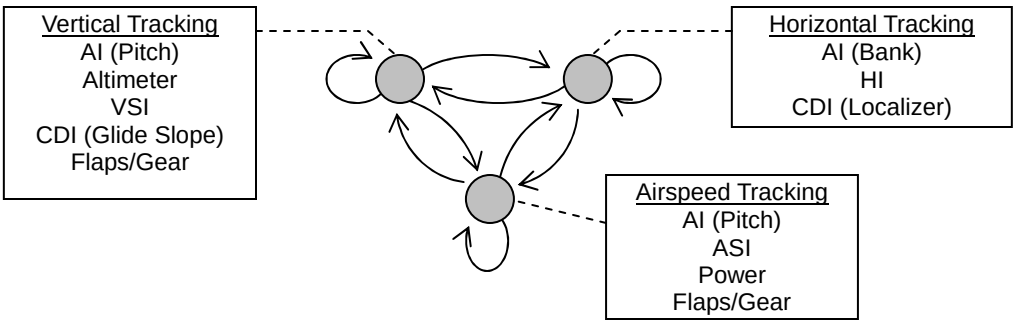


Figure 3-4. Three-state HMM

$r \leq 3, 1 \leq p \leq 4$).

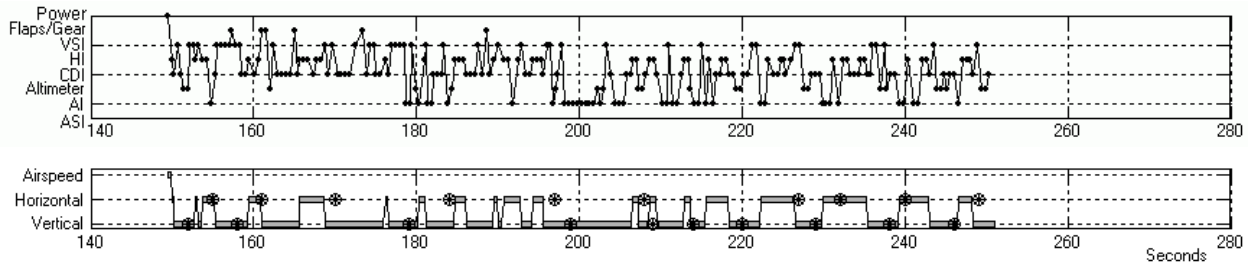
3.3 Results

First, the three-state HMM shown in Figure 3-4 was applied to all twelve combined fixation sequences of all three segments and all four pilots, i.e., $O^{s,p}$ ($1 \leq s \leq 3, 1 \leq p \leq 4$). The three-state HMM applied (Figure 3-4) was similar to the one in Figure 2-1 but slightly adjusted for the specific display used in this experiment by adding the flaps/gear indicator in both the vertical-tracking group and the airspeed-tracking instrument group, and removing the turn-and-slip indicator from the horizontal-tracking instrument group. Appendix C lists all the initial conditions and the estimation results of these model parameters.

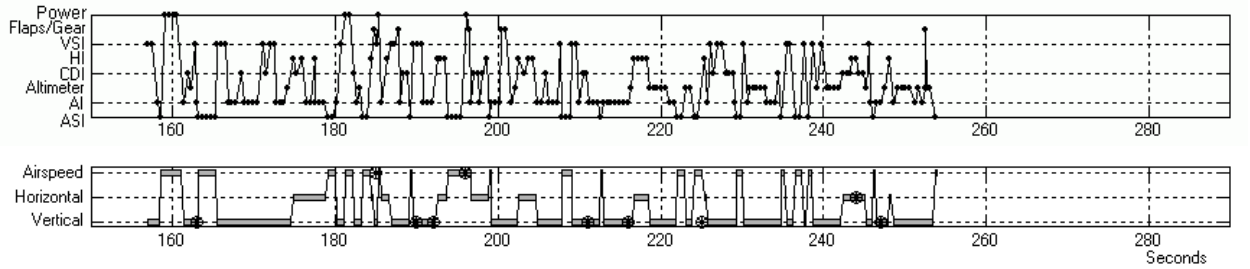
Because the segment (iii) indicated the most drastic differences among the pilots, the results of this segment are presented first.

3.3.1 Segment (iii) – Final Descent

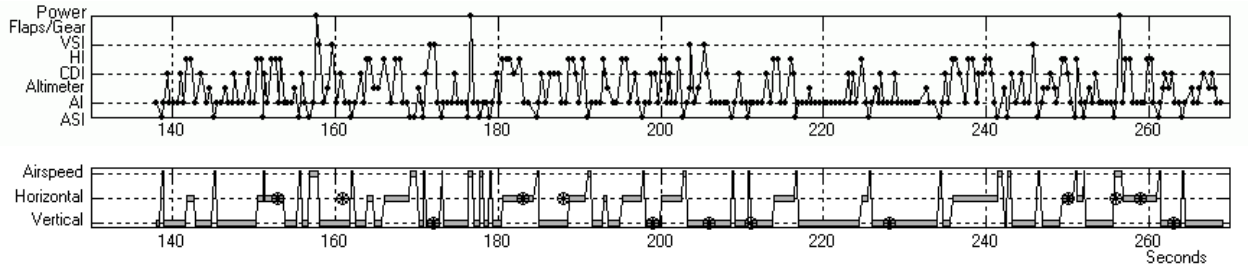
Figure 3-5 (a)-(d) plot fixation sequences and the corresponding attention sequence estimated by the three-state HMM in segment (iii), final descent, of the last approach each pilot flew, i.e., the fixation sequence, $O^{3,3,p}$ ($1 \leq p \leq 4$), and the corresponding estimated attention sequence, $Q^{3,3,p}$ ($1 \leq p \leq 4$). In this figure, the last approaches ($s = 3$) were selected to be plotted for illustration purposes, but the actual analysis included the other two approaches ($s = 1, 2$) as well.



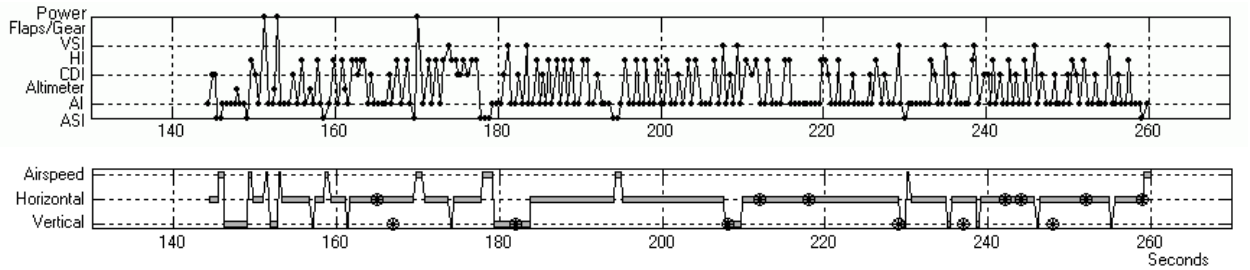
(a) Pilot 1



(b) Pilot 2



(c) Pilot 3



(d) Pilot 4

Figure 3-5. The fixation sequence (top) and estimated task sequence (bottom) from segment (iii), final descent segment, in the last approach each pilot made. The three-state HMM was used for all pilots. The circles with asterisk marks inside shown on the bottom plots indicate the tasks implied by the pilot's verbal reports.

The circles with asterisk marks inside shown in the attention sequence plots (bottom plots) in Figure 3-5 (a)-(d) indicate the tasks implied by the pilot's verbal reports, $R_\tau^{3,3,p}$ $(1 \le \tau \le T^{3,3,p})$. The time of each verbal report, τ, is the time the verbal report sentence started. As mentioned in the previous chapter, the verbal reports regarding instruments included in only one tracking task group (e.g., the ASI, the altimeter) were omitted, and only the reports related to the overlapping instruments (i.e., the AI and CDI) or the pilot's current intentions (e.g., "climb up") are shown in Figure 3-5.

Each verbal report was considered "matched" if the estimated task attention within ±1 second of the time of the report coincided with the task of the verbal report. In other words, the $R_\tau^{s,r,p}$ is considered "matched" if

$$\exists t \left[\tau - 1 \le t \le \tau + 1 \quad \wedge \quad R_\tau^{s,r,p} \;=\; Q_t^{s,r,p} \right] \tag{3.3}$$

The ±1-second buffer was included for several reasons: 1) the time point of the verbal reports may not exactly coincide with any of the time points of the fixations; 2) it is natural to assume that there are delays in the pilots' verbal responses simply due to their internal information processing time; and 3) it was sometimes difficult for the experimenter to pinpoint exactly when the verbal report sentence started from the videotape records, and, the ±1-second buffer covers the potential margin of the experimenter's error.

Table 3-2 lists the match rates of the verbal reports, i.e., the numbers of the matched reports divided by the total number of the reports × 100 (%), in this segment. The table shows that, while Pilot 1, 2, and 3 all resulted in relatively higher match rates (around 90%), Pilot 4, the most experienced pilot, resulted in a relatively low rate (66.7%). What happened with Pilot 4?

Table 3-2. Match rates of verbal reports in segment (iii)

Pilots	Total Number of Reports	Number of Matched Reports	Match Rates
Pilot 1	67	61	91.0%
Pilot 2	41	37	90.2%
Pilot 3	47	42	89.4%
Pilot 4	36	24	66.7%

Indeed, Figure 3-5 (a)-(d) suggest Pilot 4 had characteristics that were very different from those of the other pilots. Figure 3-5 (d) shows that Pilot 4 spent most of the time on the horizontal-tracking task, which is, frankly, not realistic for the following reason. In this segment, the aircraft had to descend precisely, tracking the glide slope, and this required the pilot to pay close attention to the vertical-tracking task as well as the horizontal-tracking task, as the other three pilots did. Most of the missed detections (i.e., the HMM estimates failed to match the verbal reports) occurred when Pilot 4 was looking at the AI for a relatively longer duration of time (typically 2 to 7 seconds), and this AI fixation was either preceded or followed by a CDI fixation (e.g., in Figure 3-5 (d), the reports at τ = 167, 237, and 248). Since Pilot 4 also looked at the HI frequently, the combination of viewing those three instruments tended to make the HMM estimate these portions as a horizontal-tracking task, which conflicted with the pilot's vertical-tracking-related verbal reports.

The post-experiment interview with Pilot 4 revealed that, at these points, the pilot checked both pitch and bank to maintain stable path tracking on the final descent. The description did not fit either the vertical-tracking or the horizontal-tracking task. Therefore, a new hidden state, namely an "attitude-monitoring" task associated with longer fixation on AI with occasional looks at the CDI was created and added to make an alternative 4-state HMM (Figure 3-6). This fourth state may appear to be a subset of the vertical-tracking or horizontal-tracking task, but is different due to the much longer fixation on the AI. The fact that this state does not include performance instruments (e.g., altimeter, VSI, HI, ASI), which are necessary to directly check the aircraft's response to the pilot's input, implies that in this fourth state, the pilot was not actively controlling the aircraft. The aircraft was presumably already stabilized, and the pilot's task was simply to

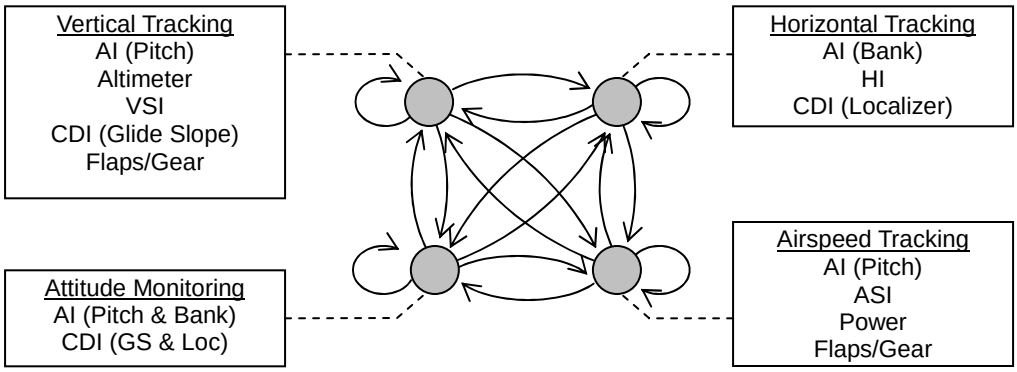


Figure 3-6. Four-state HMM

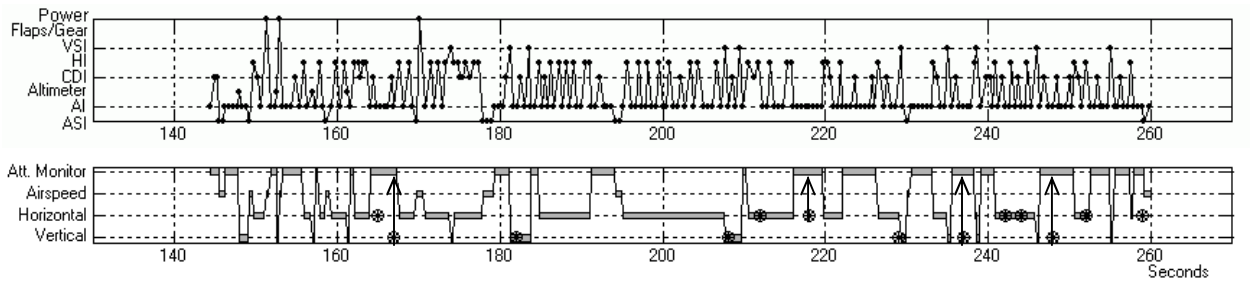
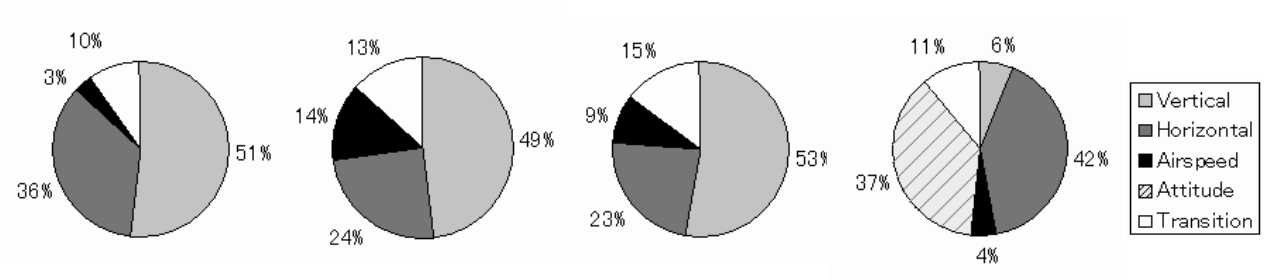


Figure 3-7. The fixation sequence (top) and estimated task sequence (bottom) from segment (iii), final descent segment, in the last approach of Pilot 4. The four-state HMM was used.

monitor for any deviations from the "target" pitch and bank angles that maintain this stable tracking. Figure 3-7 plots the re-estimation results using the four-state HMM with the same fixation sequence data in Figure 3-5 (d). As a result, the match rate of the verbal report in this segment increased from 66.7% to 94.4% (= 34/36).

From the HMM-estimated attention sequences, the durations of each tracking task were computed. Figure 3-8 (a)-(d) shows the percentages of the total duration time of each tracking task within the total flight times in segment (iii). The times in between two tasks that did not belong to any task were categorized in the "Transition" in Figure 3-8. The figure shows that Pilot 4 spent only 6% of the time on the vertical-tracking task. However, the attitude-monitoring task, on which 37% of the time was spent, also watched for any pitch deviations that affect vertical tracking, and, together, the vertical-tracking performance should have been well covered. The same can be said for the airspeed-tracking task, on which only 4% of the time was spent. The aircraft probably had been already well-stabilized on the course, and the pilot did not have to control the aircraft actively. To examine this inference, RMS errors, attitude angle statistics, and



(a) Pilot 1 (b) Pilot 2 (c) Pilot 3 (d) Pilot 4

Figure 3-8. Percentages of time spent for each tracking task within the total flight times of segment (iii), final descent.

instrument fixation percentages in this segment were computed and shown in Table 3-3. The relatively low RMS flight technical errors from the glide slope (vertical tracking) and 130 knots (airspeed tracking) and standard deviations of the pitch and bank angles of Pilot 4 shown in this table support this inference. It may be arguable that the vertical-tracking task and/or the airspeed-tracking task in the four-state HMM could be removed based on their infrequent occurrences. However, here, the fact that the pilot still kept checking the vertical speed, altitude, and airspeed occasionally is important, and thus these states were left in the model.

Table 3-3. Flight-data and fixation-data statistics in segment (iii)

	RMS flight technical error			Attitude angle statistics		Instrument fixation percentages					
	From glide slope [deg]	From loc. [deg]	From 130 knots [knots]	Pitch mean (std.) [deg]	Bank mean (std.) [deg]	ASI [%]	AI [%]	Alt [%]	CDI [%]	HI [%]	VSI [%]
Pilot 1	0.44	0.74	14.9	4.13 (2.06)	0.08 (4.42)	1.29	9.80	5.62	21.6	19.5	10.7
Pilot 2	0.31	0.27	12.6	3.85 (1.55)	-0.11 (1.29)	8.94	24.3	9.98	7.67	7.89	5.17
Pilot 3	0.29	0.54	6.60	4.41 (1.98)	0.24 (3.00)	5.93	34.0	4.93	6.42	8.63	1.53
Pilot 4	0.11	0.13	6.94	3.76 (0.64)	0.001 (1.79)	2.78	47.0	1.04	5.16	12.3	1.42

Besides Pilot 4, Figure 3-8 also shows that Pilot 1 spent only 3% of the segment (iii) flight time on the airspeed task. It was, however, a different story from Pilot 4's case. The larger pitch and bank standard deviations and larger RMS flight technical errors shown in Table 3-3 indicate that the flight required more effort. The RMS airspeed error from 130 knots was relatively large, but the ASI was looked at for only 1.3% of segment (iii). Pilot 1 used the throttle several times in this segment, but almost always without an accompanying fixation on the ASI. This implies that the pilot was using the throttle mainly for rate-of-descent control rather than for the airspeed control. This suggests merging the airspeed task into the vertical task to make a two-state HMM

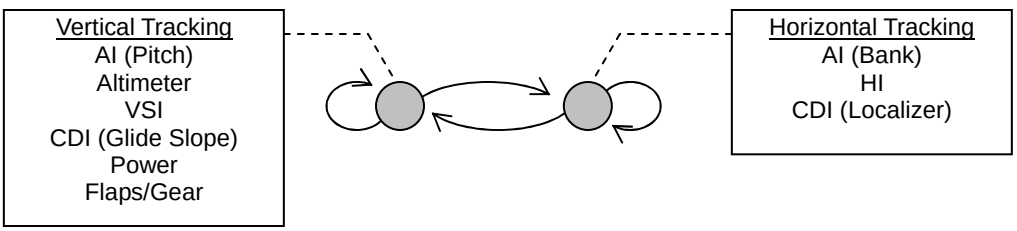


Figure 3-9. Two-state HMM

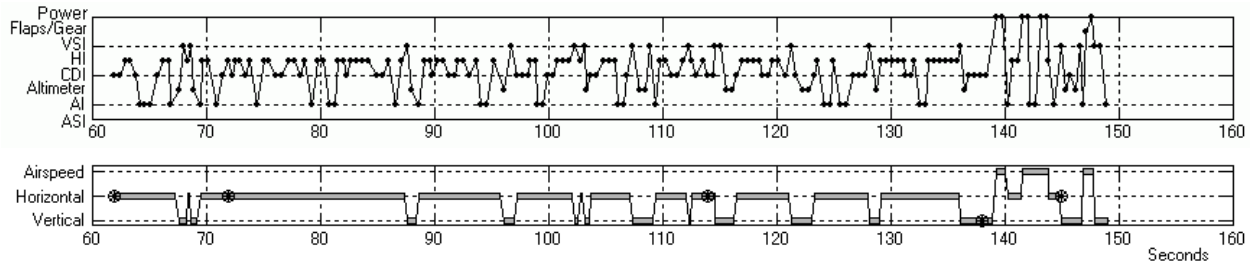
(Figure 3-9). Basically the reduction of the number of hidden states did not change the rate of verbal report matches, i.e., the rates were 91% for both the three-state and the two-state HMM. In that case, the model with a lower number of hidden states was favored by parsimony. The principle of parsimony in model selection will be further discussed in the Discussion section in this chapter.

3.3.2 Segment (ii) – Level Turn

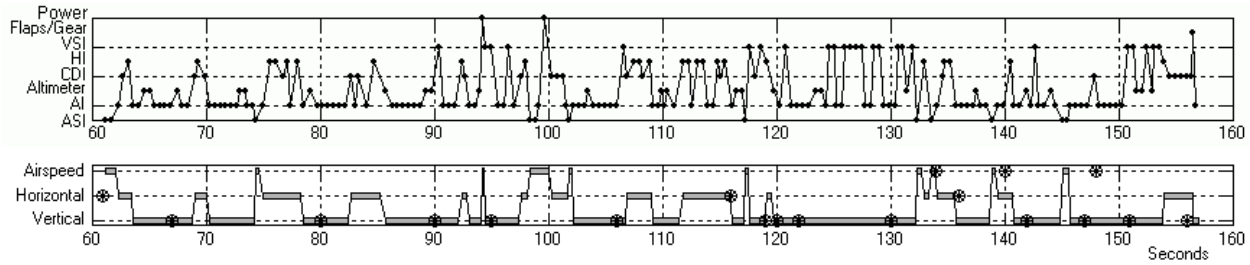
Analogous to segment (iii), the original three-state HMM structure shown in Figure 3-4 was applied in segment (ii). The HMM model parameters were re-estimated for segment (ii). The fixation sequence of the last approach each pilot made, $O^{2,3,p}$ $(1 \leq p \leq 4)$, and the corresponding estimated attention sequence, $Q^{2,3,p}$ $(1 \leq p \leq 4)$ are plotted in Figure 3-10 (a)-(d). The rates of verbal report matches are listed in Table 3-4. High match rates of verbal reports were observed for all pilots. Thus, the three-state HMM generally worked very well for all of the pilots in this segment.

Table 3-4. Match Rates of verbal reports in segment (ii)

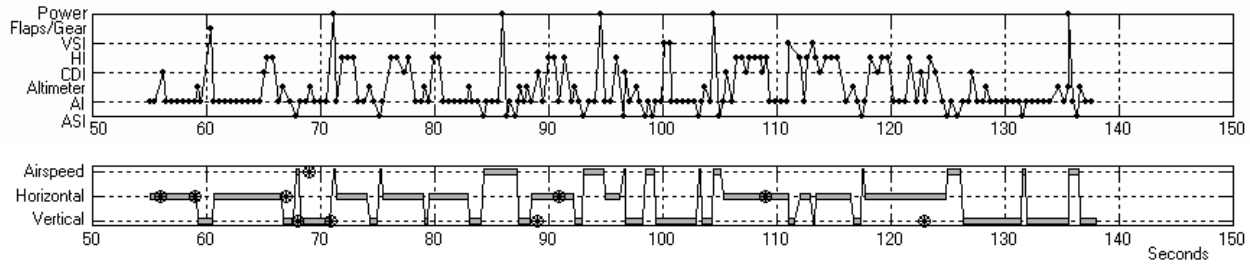
Pilots	Total Number of Reports	Number of Matched Reports	Match Rates
Pilot 1	19	17	89.5%
Pilot 2	42	40	95.2%
Pilot 3	31	27	87.1%
Pilot 4	30	27	90.0%



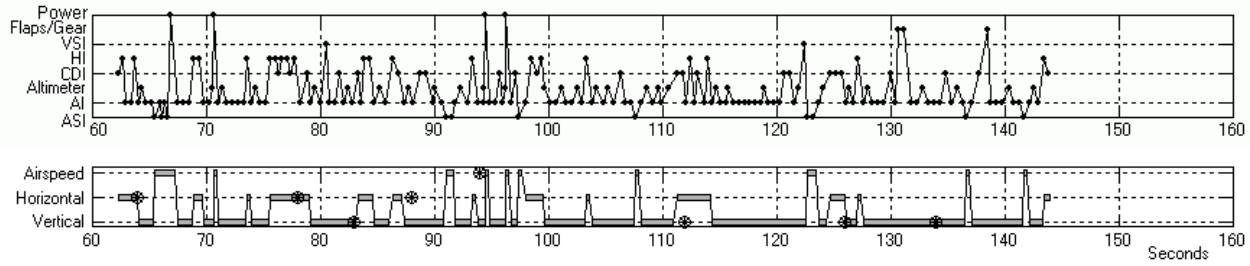
(a) Pilot 1



(b) Pilot 2



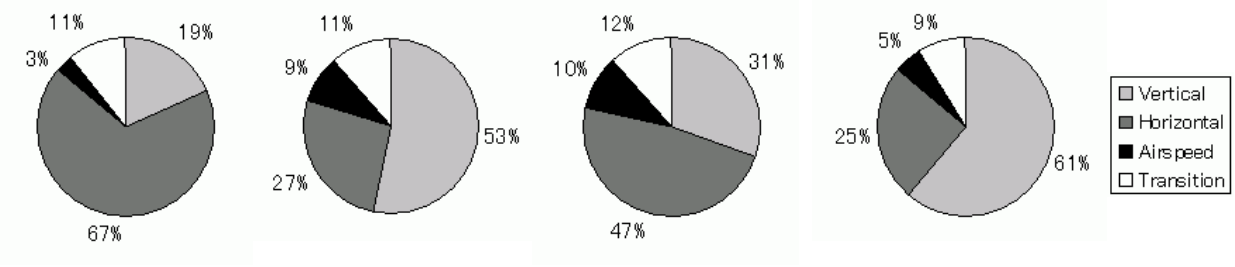
(c) Pilot 3



(d) Pilot 4

Figure 3-10. The fixation sequence (top) and estimated task sequence (bottom) from segment (ii), level turn segment, in the last approach each pilot made. The three-state HMM were used for all pilots. The circles with asterisk marks inside shown on the bottom plots indicate the tasks implied by the pilot's verbal reports.

Figure 3-11 illustrates the percentages of the total duration time of each tracking task within the total flight times in segment (ii), and Table 3-5 shows RMS errors, attitude angle statistics, and instrument fixation percentages in this segment. The figure and the table indicate that for Pilot 1, attention characteristics similar to those that appeared in segment (iii) also appeared in segment (ii); the airspeed tasks accounted for only 3% of the total time in segment (ii), and the ASI was fixated less than 1%. Thus, the same two-state HMM (Figure 3-9) was applied for Pilot 1 in this segment. As in segment (iii), the resulting rates of verbal report matches were 89.5% for both the three-state and the two-state HMM.



(a) Pilot 1 (b) Pilot 2 (c) Pilot 3 (d) Pilot 4

Figure 3-11. Percentages of the time spent for each tracking task within the total flight times of the segment (ii), level turn.

Table 3-5. Flight-data and fixation-data statistics in segment (ii)

	RMS flight technical error		Attitude angle statistics		Instrument fixation percentages					
	From 1700 ft [deg]	From loc. [deg]	Pitch mean (std.) [deg]	Bank mean (std.) [deg]	ASI [%]	AI [%]	Alt [%]	CDI [%]	HI [%]	VSI [%]
Pilot 1	69.8	0.88	7.08 (1.33)	2.96 (5.23)	0.11	13.1	5.17	15.5	29.2	7.24
Pilot 2	61.2	0.37	5.20 (1.38)	3.01 (6.20)	5.04	34.6	7.10	6.01	9.18	4.89
Pilot 3	58.7	1.28	6.39 (1.77)	2.12 (6.53)	4.39	41.2	5.90	4.48	9.21	1.19
Pilot 4	40.7	0.49	5.56 (1.04)	2.44 (4.99)	2.83	37.2	9.68	6.54	9.13	0.39

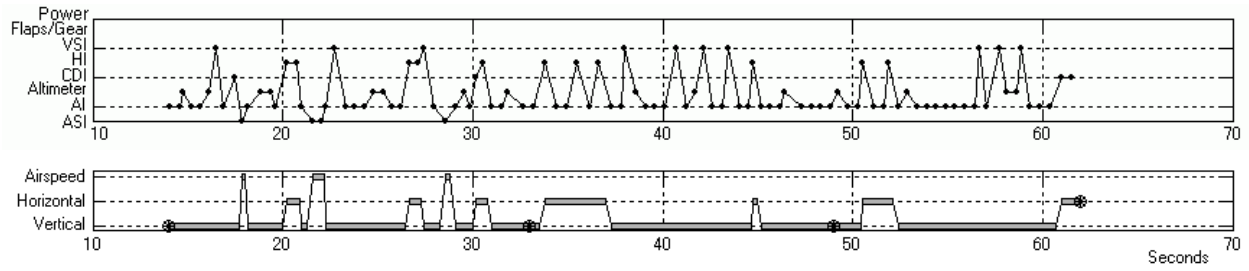
3.3.3 Segment (i) – Straight and Level

Once again, Figure 3-12 (a)-(d) plot fixation sequences and the corresponding attention sequence estimated by the original three-state HMM (Figure 3-4) in segment (i), straight and level, of the last approach each pilot flew, i.e., the fixation sequence, $O^{1,3,p}$ $(1 \leq p \leq 4)$, and the corresponding estimated attention sequence, $Q^{1,3,p}$ $(1 \leq p \leq 4)$.

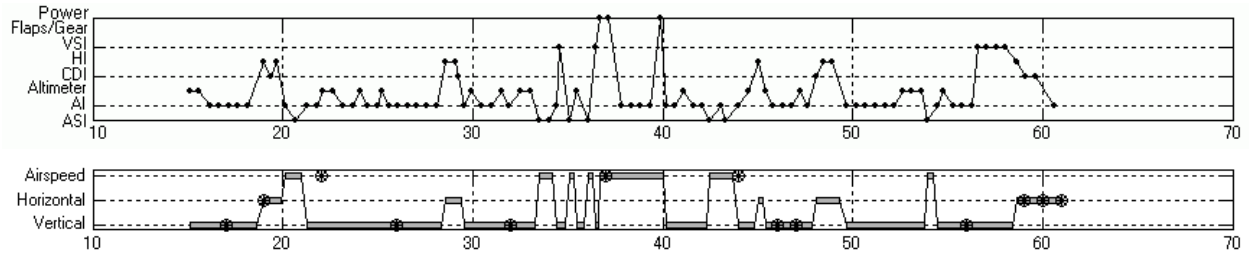
The match rates of verbal reports are listed in Table 3-6. All pilots indicated high match rates. Again, the three-state HMM generally worked very well in this segment.

Table 3-6. Match Rates of verbal reports in segment (i)

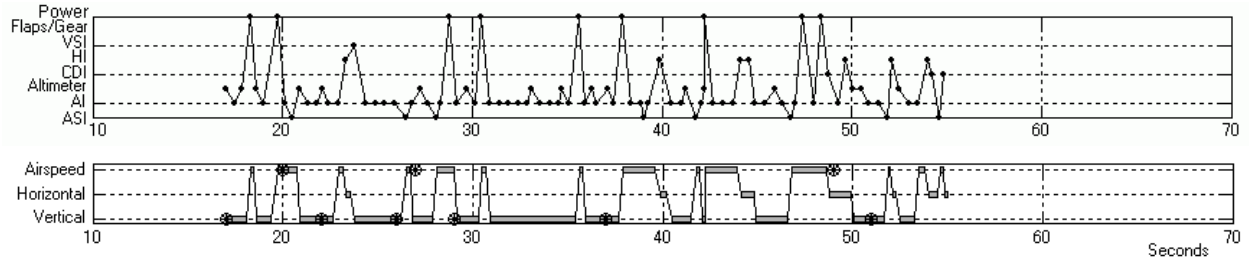
Pilots	Total Number of Reports	Number of Matched Reports	Match Rates
Pilot 1	14	14	100%
Pilot 2	26	26	100%
Pilot 3	24	24	100%
Pilot 4	15	14	93%



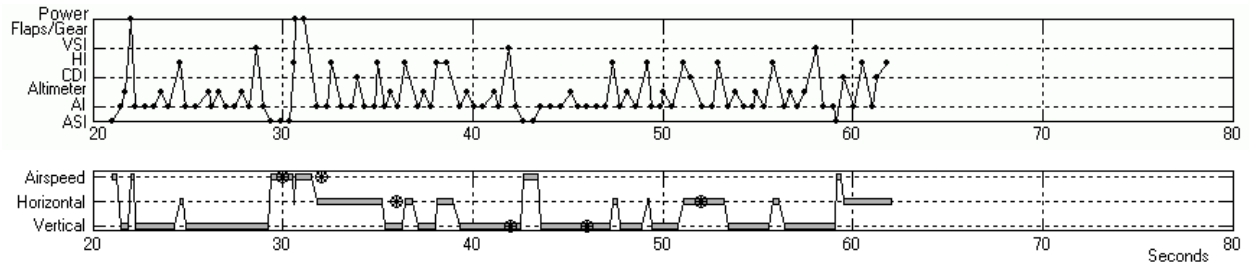
(a) Pilot 1



(b) Pilot 2



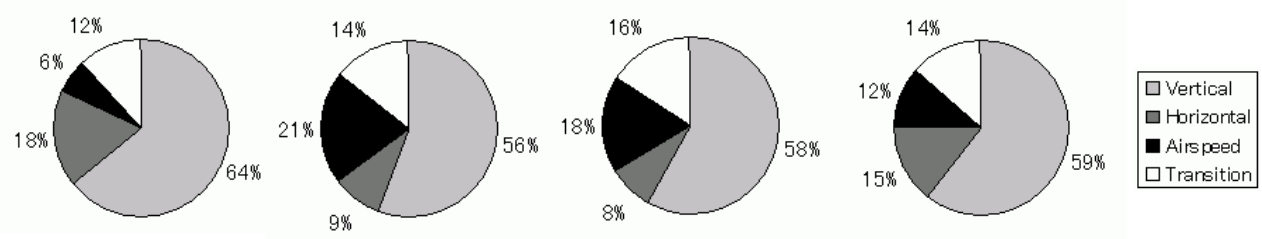
(c) Pilot 3



(d) Pilot 4

Figure 3-12. The fixation sequence (top) and estimated task sequence (bottom) from segment (i), straight and level segment, in the last approach each pilot made. The three-state HMM were used for all pilots. The circles with asterisk marks inside shown on the bottom plots indicate the tasks implied by the pilot's verbal reports.

Figure 3-13 illustrates the percentages of the total duration time of each tracking task within the total flight times in segment (i). Notice that the time percentages of the horizontal task were decreased compared to segments (ii) and (iii). This is because the HMM learned during the parameter estimation that the pilots tended to look at vertical- and airspeed-tracking instruments more often than the horizontal-tracking instruments in this straight-and-level segment, and adjusted the parameters by itself to bias toward the vertical- and airspeed-tracking tasks when the overlapping instruments were looked at.



(a) Pilot 1 (b) Pilot 2 (c) Pilot 3 (d) Pilot 4

Figure 3-13. Percentages of the time spent for each tracking task within the total flight times of the segment (i), straight and level segment.

Table 3-7 shows RMS errors, attitude angle statistics, and instrument fixation percentages in segment (i). In this segment, there was no assigned heading (although it was suggested that pilots maintain the heading of 320° to keep the flight paths consistent from approach to approach). Thus, the workload in this segment was likely lower than in the other two segments, where all three tracking axes had assigned values to track. Perhaps because of that, Pilot 1 looked at the ASI more often (2.6%) than during the other segments. As in segments (ii) and (iii), this pilot used the throttle for rate-of-descent control rather than for airspeed control. Therefore, a slightly altered three-state HMM in Figure 3-14, which included the power indicator in the vertical-tracking instrument group rather than in the airspeed-tracking instrument group, was applied to Pilot 1's fixation sequence in segment (i). The airspeed task was not removed from the model this time because keeping this task conveys the important fact that, in this segment, the pilot occasionally looked at the ASI. The resulting match rate of verbal reports was 100%, the same as the original three-state HMM. Since the numbers of hidden states are the same for both the original and alternative HMMs, it could not be concluded which model was better than the

other based on analysis of this segment only. However, if the analyses of the other two segments were taken into account, the alternative HMM may have had a slight advantage, since it was more consistent with the two-state HMM used in the other two segments.

Table 3-7. Flight-data and fixation-data statistics in segment (i)

	RMS flight technical error		Attitude angle statistics		Instrument fixation percentages					
	From 1700 ft [deg]	From 160 knots [deg]	Pitch mean (std.) [deg]	Bank mean (std.) [deg]	ASI [%]	AI [%]	Alt [%]	CDI [%]	HI [%]	VSI [%]
Pilot 1	42.7	8.30	7.61 (1.30)	0.24 (0.46)	2.58	32.2	11.4	3.23	10.8	6.93
Pilot 2	62.6	6.85	6.31 (1.40)	-0.15 (0.69)	8.03	30.9	10.7	2.49	6.17	3.71
Pilot 3	54.4	4.13	6.24 (1.13)	0.28 (0.78)	3.37	35.2	11.4	2.87	3.55	3.15
Pilot 4	38.5	5.48	7.25 (0.92)	0.46 (2.02)	4.53	38.2	10.7	1.58	7.93	1.30

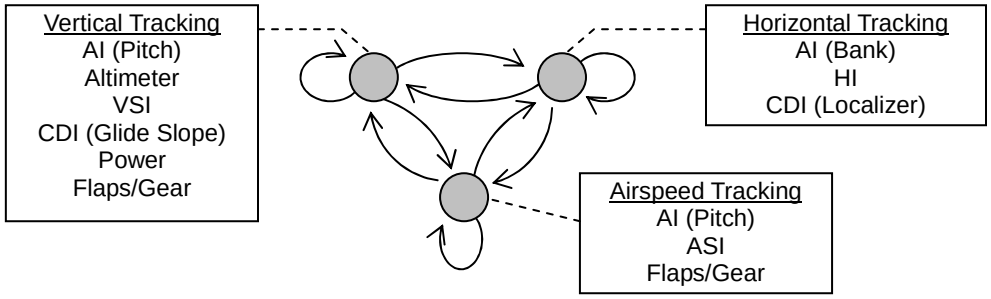


Figure 3-14. Alternative three-state HMM for Pilot 1 in segment (i): the power indicator is included in the vertical-tracking instrument group

3.4 Discussion

3.4.1 Selection of HMM Structure

The HMM analysis was started with the three-state HMM structure shown in Figure 3-4, and some variations of the model were derived that described the data of each pilot during the course of the analysis. In choosing the HMM structures, caution was used in selecting the proper number of hidden states, especially when a new hidden state was added. It is generally true that,

when maximum-likelihood principles are applied for selecting a model, the models that have more parameters, and thus provide more degrees of freedom in fitting the data, tend to result in higher probability than those that have fewer parameters, and therefore are always favored. This problem occurs wherever a "right" model has to be chosen from multiple alternative models with different dimensions, such as Markov chain analyses or Autoregressive Moving-Average (AR-MA) analyses. In fact, some criteria, such as Akaike Information Criterion (AIC) (Akaike, 1974) or Bayesian Information Criterion (BIC) (Schwarz, 1978), have been introduced in these areas to avoid opting for the models with the highest possible dimensions. These criteria basically penalize the models for the number of model parameters.

This experiment did not use these formal criteria, but the principle of parsimony employed had the same effect. In this experiment, adding a new state was justified only when the addition substantially improved the match rate between the HMM-estimated attention and the pilots' own verbal reports. For instance, the fourth state was added to the HMM of Pilot 4 in segment (iii) because that improved the match rate considerably. No new hidden state was added to the original three-state HMM for Pilot 4 in the other segments (i) and (ii) or for Pilot 2 and Pilot 3 in any of the three segments, where few conflicts between the verbal report and the three-state model estimation results were found. For these cases, no hidden state was removed either, because all three hidden states occurred reasonably often. On the other hand, in the HMM of Pilot 1 in segments (ii) and (iii), the airspeed-tracking state was removed, since having this state or not did not affect the match rate with the pilot's verbal reports.

In addition to the match rate of verbal reports, sometimes the number of the HMM estimation loops required to achieve the convergence of *logP* may indicate the quality of the model selection. If *logP* did not converge within the first couple of loops, that would mean that the HMM estimation algorithms could not fit the fixation data to the given HMM very well. Thus, revising the model, including adding a new hidden state, may be needed. Note that the fast convergence of *logP* may be a necessary condition for a good model selection, but not a sufficient one. Therefore, it should be used when accompanied by other criteria.

The other criterion sometimes used in the experiment was based on the researcher's discretion to preserve certain characteristics of the data. It may be also seen as a criterion for construction validity. If a hidden state is removed from the model, the model will not be able to re-create the

actual data, where the removed state occurred. In any case, the decision for this criterion is inevitably more subjective than that for the parsimony-based criterion, and, therefore, when this criterion is being applied, the reason why a certain state was kept should be clearly stated. For instance, the airspeed tasks in the HMM for Pilot 1 in segment (i) and the HMM for Pilot 4 in segment (iii) were kept in the model even though the occurrence of this state was relatively rare, because it was an important fact that the pilots still looked at the ASI occasionally.

3.4.2 Effects of Pilots' Expertise Level

Based on the above criteria, the variations of the original three-state HMM (Figure 3-4) were found in pilots who have different expertise levels and also were flying different types of maneuvers. Thus, the null hypothesis,

***H_0*:** There is no other instrument scanning HMM structure than the three-state HMM structure

was rejected. As mentioned at the beginning, it was not the purpose of this experiment to find an exhaustive list of existing HMM structures.

However, the experiment did find an interesting positive correlation between the number of hidden states and the pilots' expertise level (Figure 3-15), and that was probably not a coincidence. The two-state HMM, exhibited by Pilot 1, the least experienced pilot in this experiment, indicated that one of the tracking tasks was dropped because of the high workload, while the added attitude-monitoring state in the four-state HMM, which appeared in the data of

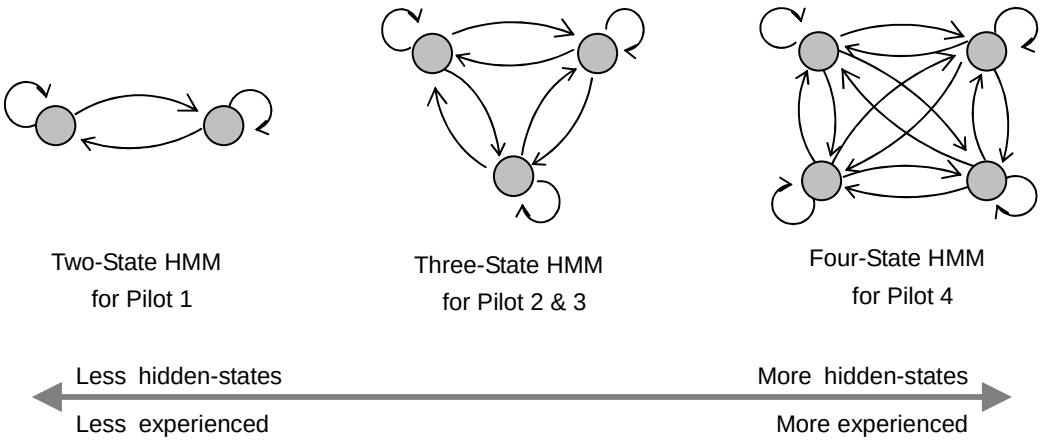


Figure 3-15. Number of hidden states and pilots' expertise levels

Pilot 4, who was the most experienced pilot in this experiment and also had experience flying an actual Boeing 757, indicated extra time the experienced pilot had due to the well-stabilized aircraft behavior during the final descent segment. This characteristic of this expert pilot implies an interesting circular effect that skilled pilots fly well not only because they track and stabilize well, but also because the extra time this creates gives them more attentional resources so that they can respond more quickly. Both intermediate-level pilots, Pilot 2 and Pilot 3, indicated the three-state HMM.

The attitude monitoring state of Pilot 4 may have a connection to the anecdotal scanning behavior that is said to exist among highly experienced pilots; that is, some highly experienced pilots look at the center of the instrument panel (about the point beneath the AI), and use peripheral vision to monitor all the instruments at once (Machado, 1999, p. 135). Indeed, it was possible that Pilot 4 also used his peripheral vision to scan the adjacent instruments, even though there was no evidence of that because the eye-tracking system cannot measure peripheral vision.

Bellenkes and his colleagues reported that, in their flight simulator experiment, the flight instructors (i.e., expert pilots) made more fixations on the flight variables that were not being maneuvered ("minding the store") than novice pilots, which may indicate a sign of good attentional flexibility (Bellenkes et al., 1997). Actually, in our experiment, the minding-the-store trend did not necessarily occur in proportion to the pilots' expertise levels. For instance, Figure 3-13 and Table 3-7 indicate that, in the straight-and-level segment, the pilot who attended the horizontal tracking task and also scanned the HI and CDI (i.e., not currently maneuvered) for the longest time period was Pilot 1. Also, Figure 3-11 and Table 3-5 show that, in the level-turn segment, the pilot who attended the vertical-tracking task or scanned the altimeter (i.e., not currently maneuvered) for the longest time period was Pilot 4, who was followed by Pilot 2, the third most experienced pilot in this experiment. Although the minding-the-store trend may generally exist for the experienced pilot, this phenomena itself may not necessarily be guaranteed to occur all the time. Yet, the expertise levels of all the pilots were correctly reflected in their HMM structure, which indicates the robustness of the HMM analysis.

Note that it was not this study's intention to declare that all expert pilots have this particular fourth state. In fact, the reader will see in the next chapter that that is not the case—some highly experienced pilots exhibit only the three-state HMM, but still fly very well. Obviously, the

three-state HMM is sufficient to fly approaches properly. The real message of the current experiment results is that expert pilots may have extra internal states, or "extra tricks," up their sleeves that they can pull out whenever they need them. Indeed, there is no evidence that Pilot 4 has only four states; there may be a fifth or sixth. The eye-movement data are only 2D coordinates, and there is a limit to how much information these data can convey. That is, if the pilot has multiple hidden states, each of which tends to generate similar eye movements, then it will be very difficult for the HMM analysis to detect them—or, for any other pattern recognition tools based only on the eye-movement data. Thus, in general, one should expect that the more experienced the pilot is, the more difficult the estimating of hidden states becomes. Nevertheless, HMM analysis has worked well for the eye-movement data of the moderately-experienced or less-than-moderately experienced pilots, where there data indicated the characteristics of the three-state or two-state HMM.

While the scan patterns indicating the three-state HMM may be sufficient for flying instrument approaches, those indicating the two-state HMM may not be. Thus, detecting which pilots show a two-state HMM scan pattern may provide some benefit in pilot training and evaluation.

3.4.3 Scanning Behavior Changes Within a Segment

As mentioned in section 2.4, the HMM used in the pilots' eye-movement analysis assumes that the probabilities do not change over time. In order to assure the validity of this assumption, the experiment divided the entire approach into three segments: (i) straight and level, (ii) level turn, and (iii) final descent. Each of these segments requires different flight maneuvers and flight task priorities, so separating these segments was presumed to help the pilots' scanning behavior stay relatively consistent within each segment.

However, the actual eye-movement data suggested that that may not be always the case. For instance, in the straight-and-level segment (i), as the aircraft approaches the localizer, the pilots often started to shift their attention to the horizontal task more frequently. In the level-turn segment (ii), as the aircraft became established and stabilized on the localizer course, the pilots tended to start shifting their attention to the vertical-tracking task slightly more often. In the final-descent segment (iii), for instance, Pilot 4 showed the scan pattern associated with the fourth state more frequently as the aircraft became more stable along the glide path.

These transitions between scanning behaviors within a single segment are perhaps inevitable. So, does that mean each segment should be further divided into sub-segments? In theory, yes. However, exactly when each sub-segment starts within a segment would not be known until the HMM analysis were run with the regular three segments first. Some pilots may adjust and stabilize the aircraft behavior relatively quickly, and transition to the next sub-segment, while other pilots may take more time to do so, or even never transition within the segment. If the length of the time the pilot needed to stabilize on the course is the interest, the researcher can run the HMM analysis within one of the regular three segments first, and then divide each segment into sub-segments according to the result of the first HMM analysis to run the second HMM analysis. Yet, it also should be advised that dividing the flight segment into too many small sub-segments may also cause insufficient data for the estimation of reasonably accurate model parameters.

3.5 Conclusion

The eye-movement data from four pilots flying simulated ILS approaches were collected. The analysis of the data served as the proof of concept of the HMM analysis. The analysis also revealed variations of the HMM structures in the real pilots' eye-movement data aside from the original three-state HMM introduced in the previous chapter. In fact, there was an interesting positive correlation between the number of hidden states and the pilots' expertise level. The data from two intermediate-level pilots were described well by the three-state HMM, comprising vertical-, horizontal-, and airspeed-tracking tasks, for all segments. The level turn and final descent segment data of the least experienced pilot indicated that the pilot dropped the airspeed-tracking task, and thus the pilot's data fit well with the two-state HMM, including vertical- and horizontal-tracking tasks. The data from the final descent segment of the most experienced pilot were best described by the four-state HMM, which includes vertical-tracking, horizontal-tracking, airspeed-tracking, and attitude-monitoring tasks. The fourth task is consistent with a flight technique well known among experienced instrument pilots. The criteria to select the right number of hidden states for each pilot, as well as the possibility to divide each flight segment into sub-segments for further analysis, were also discussed.

Chapter 4:

Application: HUD Airspeed Indicator and Altimeter Symbology Formats

This chapter demonstrates how HMM analysis can be used for the study of aircraft cockpit displays. This example application investigated the effects of the rotating pointers and gradation marks of head-up display (HUD) airspeed indicator (ASI) and altimeter symbology formats.

4.1 Issues of Head-Up Display (HUD) Symbology Format

HUDs are transparent displays that provide flight information in the pilot's primary field of view, superimposed on the outside scene. The HUD's transparency offers various operational advantages that would not be attained by conventional head-down instrument panels, such as the ability to land and take off under very low visibility conditions (e.g., CAT IIIa) (Kaiser, 1994).

At the same time, the characteristics of the HUD—the transparency and the criticality of operating conditions—raise unique design tradeoff questions. One important constraint related to transparency is that the number of symbology pixels must be reduced as much as possible so that the total area of the outside scene occluded by the HUD symbology is minimized. On the other hand, in the critical flight phases, where little or no margin of error exists (e.g., low-visibility landing or takeoff), any symbology pixels essential to accomplish precise path tracking should not be compromised. A number of human-factors design issues of the HUD are listed by categories in Zuschlag & Hayashi (Zuschlag & Hayashi, 2003). Among them, this study focuses on the issues regarding the airspeed indicator (ASI) and altimeter formats. (Conventionally, these two indicators are always presented in the same format.)

There is a preceding study of HUD ASI/altimeter formats: Prior to developing a military standard for HUD symbology, the US Air Force (USAF) conducted a flight simulator study to investigate five ASI/altimeter formats, which included a) rotating pointers with gradation marks, b) rotating

pointers without gradation marks, c) vertical moving tapes, d) digits only, and e) digits with vertical trend bars (Ercoline & Gillingham, 1990; Weinstein, Gillingham, & Ercoline, 1994). Their study found both rotating-pointer formats (a and b) significantly reduced RMS airspeed and altitude errors in comparison to the other formats. Both rotating-pointer formats also resulted in better subjective ratings than the others. However, no difference was found between the two rotating-pointer formats—with and without gradation marks—in terms of RMS airspeed or altitude errors or subjective ratings. Accordingly, the resulting military standard, MIL-STD-1787B (USDOD, 1996), requires the rotating-pointer format to be used for ASI/altimeter symbology. In addition, the standard also requires the gradation marks to be present despite a formal lack of experimental evidence supporting their inclusion, as they are still believed to provide additional advantages.

This study investigated this blunt "belief" in the advantages of the gradation marks. As seen in the USAF study, the effects of the gradation marks could be too subtle to be captured by the resolutions of the traditional simulator-study measures, such as the flight technical error or subjective rating. However, in addition to these traditional measures, the current study also incorporated eye-movement analysis, including HMM analysis. Thus, the purpose of this study was to see if these additional analyses could prove the benefits of the gradation marks, as well as confirm the advantages of rotating pointers over digits-only display formats. For this purpose, three ASI/altimeter formats similar to the ones used in the USAF study were compared in the current study: the two rotating pointers (with and without the gradation marks) and digits only.

4.2 Method

4.2.1 Flight Simulator

The experiment used the same fixed-base flight simulator at Volpe Center that was used in the first experiment in Chapter 3. As in the previous experiment, the pilots were provided with a single throttle lever at the right side, and the flaps, gear, and trim thumb switches on the control wheel. In this experiment, however, the Boeing 737-400 flight dynamics model provided in Microsoft® Flight Simulator 2002 was used. A custom OpenGL® program generated the HUD symbology (Figure 4-1), which was projected on a screen mounted on a front wall, approximately 180 inches from the pilot's eyes. The projection area had a subtended visual angle

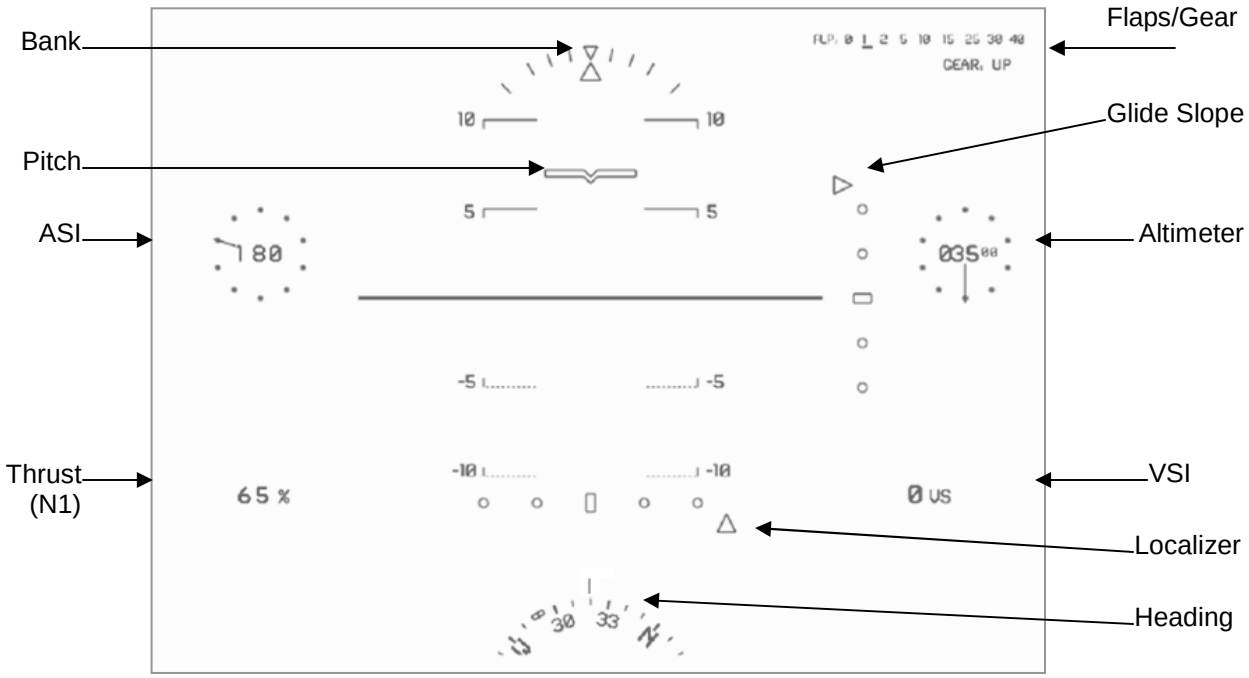


Figure 4-1. HUD Symbology (with PGD)

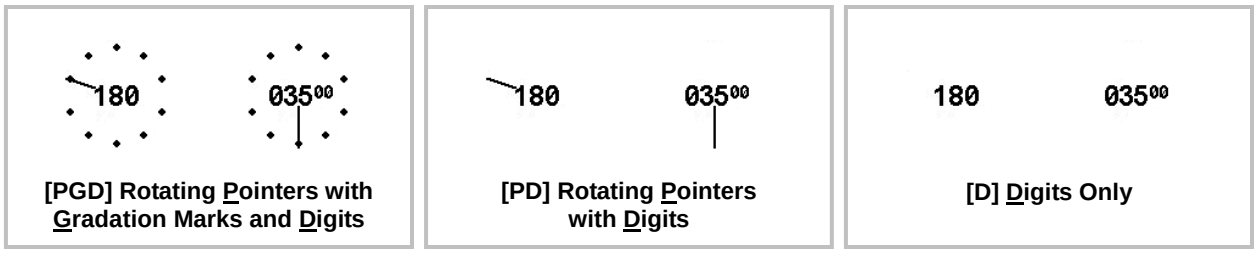


Figure 4-2. Three ASI/altimeter formats

of 21° horizontally and 16° vertically. The symbology was depicted in bright green with a black background.

4.2.2 HUD ASI/Altimeter Format

Three ASI/altimeter formats similar to the ones used in the USAF study, rotating pointers with gradation marks and digits readout (PGD), rotating pointers with digits readout but no gradation marks (PD), and digits only (D) were used (Figure 4-2).

4.2.3 Pilot Participants

Six airline transport pilots participated in the experiment, including 3 captains (CAPs), 2 first officers (F/Os), and 1 former first officer. Table 4-1 lists all the pilot participants. The individual

pilots' total flight time as of the date of the experiment ranged from 4,000 to 17,500 hours. One of the captains had previous experience flying approaches with HUD-equipped aircraft. All participants had flown Boeing type aircraft before, but only one had flown the Boeing 737, the flight dynamics of which were used in this experiment.

Table 4 -1. Pilot participants

Pilot #	CAP or F/O	Total Hours	Major Aircraft Flown	HUD Experience
1	F/O	4000	C-5, B-707, B-757, B-767	No
2	CAP	11000	DC-9, MD-80, A-300, B-767	Yes
3	CAP	11500	B-727, B-757, B-767, A-300, DC-9, MD-11	No
4	F/O	5600	B-767, B-757, MD-80, B-1, KC-135	No
5	F/O	4400	B-757, B-767, C-141	No
6	CAP	17500	A-320, B-767, B-757, B-737, B-727	No

All participants read and signed an informed consent form complying with the MIT COUHES Protocol 3021 before the experiment started (see Appendix B). The pilot who was also a Volpe employee participated in the experiment as part of his normal work. The other participants received a $100 gift certificate as compensation for their participation.

4.2.4 Simulation Scenario

The pilots were asked to fly simulated ILS approaches (Figure 4-3). The aircraft was initially positioned on either the left or right side of the localizer course. The approach was divided into 5 flight segments:

(i) Straight and level at 3500 ft, 180 knots for 1 minute
(ii) Constant-airspeed descent at 180 knots to 2000 ft, thrust set idle
(iii) Straight and level at 2000 ft, flaps down to approach configuration, gear down, and slow to 150 knots
(iv) Level turn to intercept the localizer
(v) Final descent along the glide path to 1000 ft

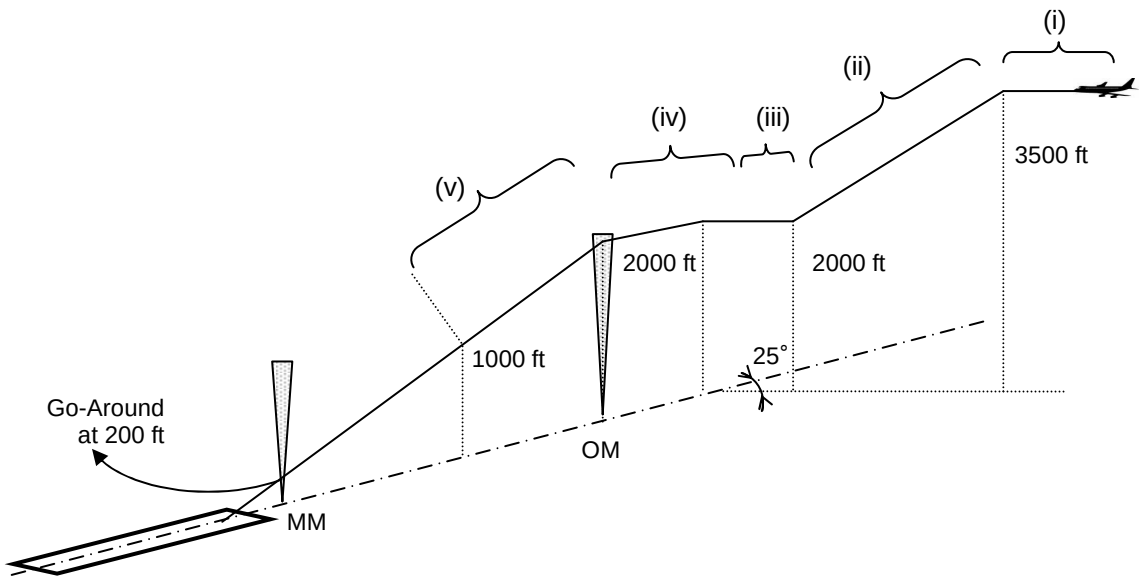


Figure 4-3. Five flight segments in the ILS scenario

After passing 1000 ft, the flight continued until reaching the decision height (200 ft), and then the pilot initiated a go-around. The data from segments (i) through (v) were used in this study. The flight segment lengths were (i) 2.3, (ii) 4.5, (iii & iv combined) 5.5, and (v) 3.2 nautical miles, respectively. Each approach took approximately 7 minutes to complete.

4.2.5 Data Collection

As in the experiment described in Chapter 3, pilots' eye-movement data were collected with a head-mounted eye camera (RK-726PCI/RK-620PC, ISCAN, Inc., Burlington, MA) and a magnetic head tracker (FasTRAK, Polhemus, Colchester, VT). The accuracy of the eye-tracker is approximately 1°. The static accuracies of the head tracker are 0.03" root mean squares (RMS) for X,Y or Z position, and 1.5° RMS for orientation.

The eye movements were collected at the rate of 60 Hz. Flight variables were also recorded, at 1 Hz. For the verbal reporting collection, the participants were given a choice between the free-mumbling format (i.e., the pilots freely generate sentences using their own words, which is the same format used in the previous experiment in Chapter 3) or the structured-reporting format (i.e., the pilots reported the attitude indicator readings using the predetermined words, "pitch," "bank," or "both") for the verbal report format. All participants except Pilot 2 used the structured-reporting format. Their verbal reports were recorded on videotape.

Each pilot flew 9 data-collection approaches: 3 approaches for each format in an order counter-balanced within and between the participants. The sides of the localizer for the approach started were also mixed and balanced to prevent the flight task becoming completely repetitive. Before the data collection approaches, each pilot received a briefing (Appendix B) and made several practice approaches. After all the approaches were completed, the pilots were asked to provide their subjective preferences between each pair of symbology formats by marking them on continuous preference scales (an example in Figure 4-4).

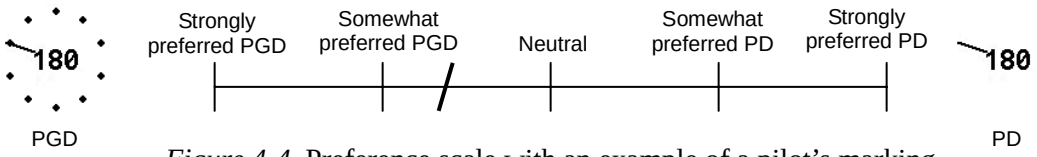


Figure 4-4. Preference scale with an example of a pilot's marking for the preference between "PGD vs. PD."

4.3 Results

In this section, first the analysis results of the traditional measures, such as RMS flight technical errors and subjective ratings, are presented to demonstrate the consistency with the previous USAF study. Then, the statistical analysis results of the eye-movement data and the HMM anlalysis are presented to demonstrate what additional information can be provided with these analyses.

4.3.1 Root Mean Square Airspeed and Altitude Errors

Root Mean Square (RMS) airspeed error was computed for segments (i), (ii), (iv), and (v) from the assigned airspeeds of 180, 180, 150, and 150 knots, respectively. Since the distributions of the RMS error were positively skewed, the values were transformed by taking natural logarithms, and a generalized linear model (GLM) repeated measures analysis (SYSTAT 10, SPSS, Inc.) was applied, with the participant as the random effect. The main effect variables included in the model were Segment, Format, and Trial Block (block 1 included the first three approaches, block 2 the second three, and block 3 the last three). A separate analysis indicated that the side of the localizer the approach started from did not have any effect, thus it was not included in the analysis. The results showed significant effects of Segment (df = 3, F = 11.45, $p < 0.001$), Trial

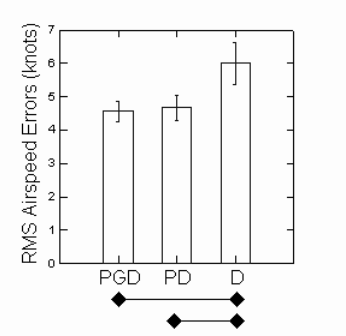


Figure 4-5. Grand means and standard errors of RMS airspeed error (the values before taking logarithms). Diamonds connected by a line indicate a significant difference between the two formats ($p < 0.05$) computed by pairwise comparisons.

(df = 2, F = 10.55, p = 0.003), and Format (df = 2, F = 5.955, p = 0.020) in the RMS airspeed error. No interaction effect (i.e., Segment × Trial, Segment × Format, Trial × Format, and Segment × Trial × Format) was found. Further pairwise comparison analysis regarding the Format effect indicated that the airspeed error was significantly reduced when PGD was used versus when D was used (df = 1, F = 8.848, p = 0.031), and also when PD was used compared to when D was used (df = 1, F = 8.522, p = 0.033). Figure 4-5 plots the grand means for each format with the pairwise comparison results (connected diamonds indicating the significant difference). The result was consistent with that of the USAF study.

RMS altitude error was also computed for segments (i), (iii), and (iv) from the assigned altitudes of 3500, 2000, and 2000 ft, respectively, and transformed by taking natural logarithms. The same GLM repeated measures analysis was performed. Unlike the USAF study, no significant format effect was found in the RMS altitude error in this study.

4.3.2 Subjective Preference Rankings

The positions of the pilots' markings on the preference scales (Figure 4-4) were converted to preference scores by measuring the distance from the opposite side of the scale. Scores for the same format were added within each pilot and ranks within each pilot were assigned (3 for the most preferred, 2 for the second most, and 1 for the least preferred). The rank sum of all pilots indicated that the most preferred format was PGD (rank sum = 16), the second most preferred was PD (12), and the least preferred was D (8) (Friedman test statistic = 5.33, df = 2, p = 0.070). The result was consistent with that of the USAF study.

Therefore, so far, the RMS airspeed error results and the subjective preference ranking results were consistent with the USAF study. These results indicated that the pilots preferred PGD over

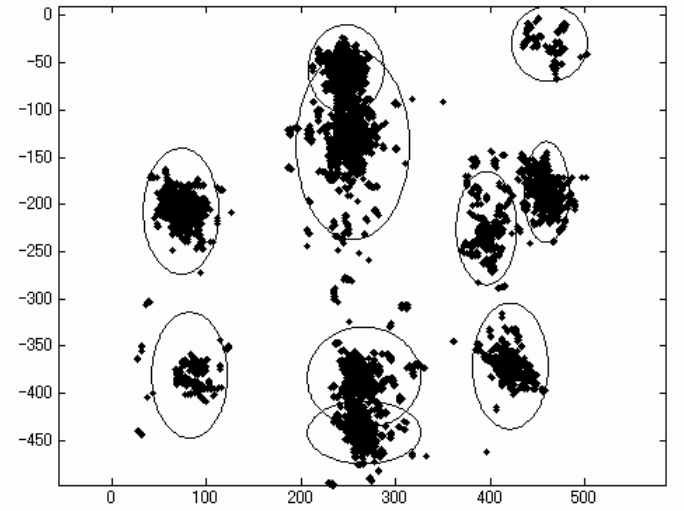
PD, but there was no detectable difference in the RMS airspeed and altitude errors between these two formats. Next, the eye-movement data are analyzed to see if they show any difference between PGD and PD.

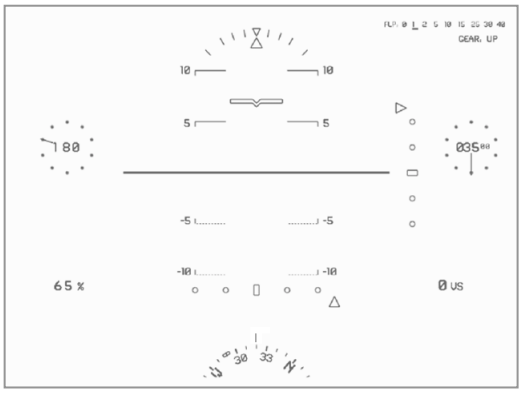
4.3.3 Indicator Fixation Durations and Look Rates

From the intersection coordinates of the pilots' line-of-sight and the display plane, the saccade points were eliminated and blinking was handled in the same way as in the previous experiment to extract the fixation points (see Section 3.2.6 for details). Then, each fixation point was associated with each indicator by K-mean clustering segmentation, as described in section 3.2.6. One difference from the case in Chapter 3 was that the sizes of indicator areas were not equal. Therefore, weights were applied in the distance computation; i.e., Eq. (3.1) was modified to

$$d_{ij} \quad = \quad \left\| \left(x_i - \mu_j\right) W \right\|^2 \quad , \tag{4.1}$$

where W is a 2×2 diagonal matrix, whose diagonal elements were w_{11} = (width of the j-th indicator)$^{-1}$, and w_{22} = (height of the j-th indicator)$^{-1}$. (Note the weight was the inverse of the physical size of the indicator width or height so that the larger indicator will result in closer weighted distance). Figure 4-6 (a) shows example clustering results using a K-mean clustering segmentation algorithm. Note that the ellipses in this figure are not boundaries, but their centers





(b) HUD Symbology

(a) Intersections of the pilot's line-of-sight and the display plane (the first approach of Pilot 3; saccade points have been removed) and K-mean segmentation results: the centers of ellipses indicate the centers of the clusters that the K-mean segmentation algorithm estimated.

Figure 4-6. K-mean segmentation with weights

indicate the center of the clusters, and their width and height indicate the weights. Figure 4-6 (b) is the geometric arrangement of the display. As seen in this figure, the AI area cluster was initially divided into two elliptical areas, a large one for the entire AI area cluster, and a small one for the bank indicator area cluster. This was done because otherwise the mean vector for the AI area cluster would have shifted upward due to the density around the bank area cluster. After the clustering was done, the data points in the two AI clusters were merged together to make one large cluster. One may wonder if the bank area could be separated from the pitch area, then the HMM analysis might not be necessary because there would no longer be an overlap any more. However, the pilots can still extract the bank information from the artificial horizon line, and the line between the bank area and the other areas of the AI was often not clear. Thus, the data points in these two clusters were eventually merged, and submitted together to the HMM analysis.

Due to the positively skewed distributions of the fixation durations, the values were transformed by taking natural logarithms before any statistical analyses were performed. The sizes of the fixation duration data were too large to be processed all together at once; thus, the data for each indicator format were further divided into each flight segment, and analyzed separately. Segment (iii) was not included in this analysis because of the large variation in the duration of this segment in the experiment (depending on how the descent in segment (ii) was performed, some approaches had a very short segment (iii), or sometimes even no segment (iii) at all). Since each format had a different number of fixations within each segment (i.e., “unbalanced” data), mixed regression repeated measures analysis (SYSTAT 10, SPSS, Inc.) was applied instead of a GLM. In this repeated measures analysis, the pilot was treated as the random effect. The model used was

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_{Trial} X_{Trial}, \tag{4.2}$$

where Y was the fixation durations on each indicator, X_1 and X_2 were dummy variables for the indicator format, i.e.,

$$(X_1, X_2) = \begin{cases} (1, 0) & \text{if PGD} \\ (0, 1) & \text{if PD} \\ (-1, -1) & \text{if D}, \end{cases} \tag{4.3}$$

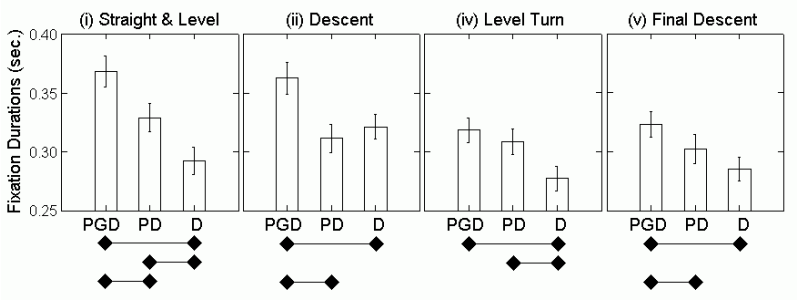
and the last variable, X_{Trial}, was the trial block number. The estimated intercept term, α,

represented the grand mean of each format within each segment, and each β_* was the estimated coefficient of corresponding variables, X_*. After the coefficients, β_1 and β_2, in Eq. (4.2) were estimated in each segment, the null hypotheses, H_0: $\beta_1-(-\beta_1-\beta_2) = 2\beta_1+\beta_2 = 0$, H_0: $\beta_2-(-\beta_1-\beta_2) = \beta_1+2\beta_2 = 0$, and H_0: $\beta_1-\beta_2 = 0$, were tested within each segment to assess the differences between PGD and D, PD and D, and PGD and PD, respectively. The test results are listed in Table 4-2. The *p*-values shown are 2-tail *p*-values. The same results are also shown graphically in Figure 4-7, where fixation duration grand means of each format in each segment are plotted along with the hypothesis test results tabulated in Table 4-2. Note that the grand means plotted in Figure 4-7 are not the log-values, but the values before taking logarithms for easier reference. The actual regression analyses and hypothesis tests were conducted on the log-values. In addition, the bar graphs plotted in Figure 4-7 are the simple grand means, suppressing all pilot and trial effects, and not exactly the same values that the mixed regression repeated measures analyses were performed on. However, these values usually correspond and, thus, are still useful for quickly grasping the results. For the exact estimated values, see Table 4-2.

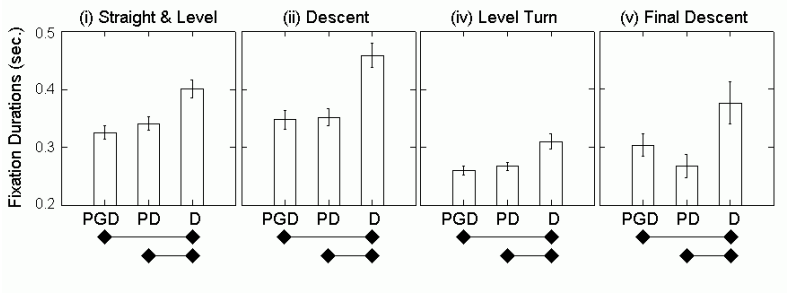
Table 4-2. Indicator fixation durations: Mixed regression repeated measures analysis, indicator format effect hypothesis test results (only $p < 0.05$ are shown)

Segment	Effect Tested	Estimate (log)	Standard Error	z	p
	ASI				
(i)	PGD-D	0.2064	0.0400	5.1586	0.0000
	PD-D	0.105	0.0398	2.6357	0.0086
	PGD-PD	0.1014	0.0326	3.1106	0.0018
(ii)	PGD-D	0.1072	0.0348	3.0789	0.0020
	PD-D	-0.022	0.0352	-0.6256	---
	PGD-PD	0.1292	0.0286	4.5222	0.0000
(iv)	PGD-D	0.1536	0.0361	4.2565	0.0000
	PD-D	0.1095	0.0363	3.0151	0.0026
	PGD-PD	0.0441	0.0296	1.4920	---
(v)	PGD-D	0.1076	0.0417	2.5813	0.0098
	PD-D	0.0205	0.0416	0.4925	---
	PGD-PD	0.0871	0.0340	2.5609	0.0104
	Altimeter				
(i)	PGD-D	-0.1705	0.0466	-3.6592	0.0002
	PD-D	-0.1127	0.0462	-2.4368	0.0150
	PGD-PD	-0.0578	0.0379	-1.5249	---
(ii)	PGD-D	-0.253	0.0498	-5.0776	0.0000
	PD-D	-0.2042	0.0501	-4.0746	0.0000
	PGD-PD	-0.0488	0.0408	-1.1960	---
(iv)	PGD-D	-0.1324	0.0419	-3.1625	0.0016
	PD-D	-0.1064	0.0413	-2.5770	0.0098
	PGD-PD	-0.026	0.0339	-0.7659	---
(v)	PGD-D	-0.1625	0.0766	-2.1207	0.0340
	PD-D	-0.2677	0.0775	-3.4520	0.0006
	PGD-PD	0.1052	0.0629	1.6714	---
	VSI				
(i)	PGD-D	-0.1582	0.0379	-4.1765	0.0000
	PD-D	-0.0455	0.0375	-1.2142	---
	PGD-PD	-0.1127	0.0308	-3.6635	0.0002
(ii)	PGD-D	-0.0464	0.0474	-0.9789	---
	PD-D	-0.0319	0.0472	-0.6763	---
	PGD-PD	-0.0145	0.0386	-0.3756	---
(iv)	PGD-D	-0.0164	0.0443	-0.3705	---
	PD-D	0.0686	0.0429	1.5976	---
	PGD-PD	-0.085	0.0356	-2.3873	0.0168
(v)	PGD-D	0.0285	0.0434	0.6573	---
	PD-D	0.0759	0.0431	1.7622	---
	PGD-PD	-0.0474	0.0353	-1.3433	---

(a) Airspeed Indicator (ASI)



(b) Altimeter



(c) Vertical Speed Indicator (VSI)

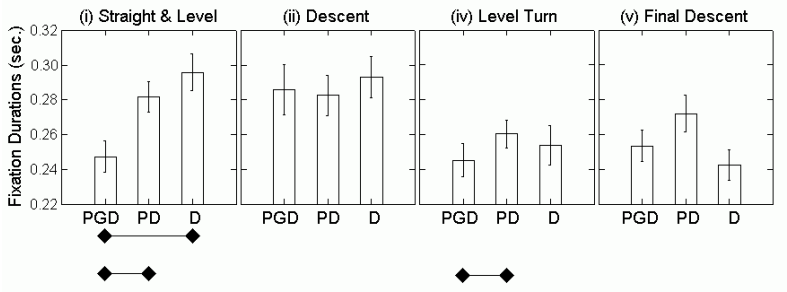
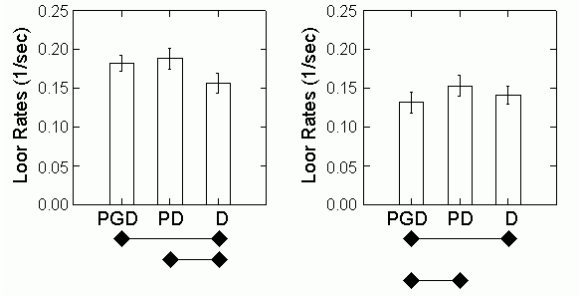


Figure 4-7. Grand means and standard errors of fixation durations (the values before taking logarithms) on (a) ASI, (b) altimeter, and (c) VSI in segments (i), (ii), (iv), and (v). Diamonds connected by a line indicate a significant difference between the two formats ($p < 0.05$) computed by hypothesis tests.

As seen in Figure 4-7, the fixation durations on the ASI and altimeter showed opposite trends; the durations on the ASI tended to be longer when PGD or PD was used than when D was used, while those on the altimeter tended to be shorter when PGD or PD was used than when D was used. Differences between the two rotating-pointer formats (PGD and PD) were found in the ASI fixation durations except on (iv) the level turn segment. No difference between the two rotating pointers was found in the altimeter fixations, but, interestingly, a difference between them appeared in the fixations on the VSI during segments (i) and (iv). In those segments, the durations on the VSI were significantly longer when PD was used than when PGD was used. Note that the format of the VSI was always digits only and did not change.

The look rates (i.e., frequencies of visits per second) in segments (i), (ii), (iv), and (v) were also examined. To analyze the format effects, GLM repeated measures analysis with the random effect being the Subjects, and the main effect variables being Segment and Format, was applied. In the look rates for the altimeter, the results revealed significant Segment effect (df = 3, F = 38.71, $p < 0.001$) and Format effects (df = 2, F = 8.495, p = 0.007). In the look rates for the VSI, the result showed significant Format effects (df = 2, F = 8.657, p = 0.007). Further pairwise comparison with respect to the Format effects showed that the altimeter look rates were significantly higher when either one of the rotating-pointer formats (PGD or PD) was used compared to when D was used (df = 1, F = 12.605, p = 0.016 for PGD vs. D; df = 1, F = 16.638, p = 0.010 for PD vs. D). On the other hand, the VSI look rates were significantly higher when PD or D was used than when PGD was used (df = 1, F = 16.129, p = 0.010 for PGD vs. D; df = 1, F = 10.872, p = 0.022 for PGD vs. PD) (Figure 4-8).



(a) Altimeter (b) VSI

Figure 4-8. Grand means and standard errors of (a) altimeter and (b) VSI look rates. Diamonds connected by a line indicate a significant difference between the two formats ($p < 0.05$) computed by pairwise comparison.

4.3.4 Tracking Task Durations (HMM Analysis)

In addition to the fixation durations and look rate analyses, HMM analysis was also applied to see the effects on the pilots' attention management among tracking tasks in segments (i), (ii), (iv), and (v). The indicator-task associations are shown in Table 4-3. Unlike the Basic-T instrument used in the previous chapter, the HUD symbology used in this simulation had separate localizer and glide slope deviation indicators (as shown in Figure 4-1). Therefore, if the three-state HMM is used, the only portion that overlapped among the multiple tasks was the AI (and also the flaps/gear indicator, but this indicator is used much less frequently compared to the AI).

Table 4-3. Indicator-task associations

Flight Tasks	Vertical Tracking	Horizontal Tracking	Airspeed Tracking	Attitude Monitoring
Associated Indicators	AI (Pitch) Altimeter VSI Glide Slope Flaps/Gear	AI (Bank) Heading Localizer	AI (Pitch) ASI Thrust (N1) Flaps/Gear	AI (Pitch, Bank) Heading* Localizer* Glide Slope*

* Included when they apply

In the attitude monitoring task, the localizer and glide slope deviation indicators were included during segments (iv), level turn, and (v), final descent. Some pilots looked at the localizer and/or glide slope deviations even in the earlier segments, such as segments (i), straight and level, and (ii), constant airspeed descent. For these pilots, these indicators were also included in the attitude monitoring task during these earlier segments as well. In addition, for some pilots, the small probability of heading indicator fixations (e.g., ≤ 0.05) was also included in the attitude monitoring task if that improved the HMM estimation results.

Analogous to Chapter 3, HMM analysis was applied separately to each pilot's data. Let $O^{s,r,f,p} = \{o_1^{s,r,f,p}, \ldots, o_T^{s,r,f,p}\}$ denote the fixation sequence data from the segment, *s* $(s = \{(i), (ii), (iv), (v)\})$, of the *r*-th approach $(1 \le r \le 3)$, of the format, *f* $(f = \{PGD, PD, D\})$, and of the Pilot, *p* $(1 \le p \le 6)$. Then, the fixation sequence of the same segment, format, and pilot were concatenated together to make one long sequence, $O^{s,f,p} = \{o_1^{s,1,f,p}, \ldots, o_{T1}^{s,1,f,p}, o_1^{s,2,f,p}, \ldots, o_{T2}^{s,2,f,p}, o_1^{s,3,f,p}, \ldots, o_{T3}^{s,3,f,p}\}$ $(T1 = T^{s,1,f,p}, T2 = T^{s,2,f,p}, T3 = T^{s,3,f,p})$, and the model parameters that maximize $P(O^{s,f,p}|\lambda)$ were estimated.

First, a three-state HMM (vertical-, horizontal-, and airspeed-tracking tasks) was applied for all cases, then the estimated task sequence was compared with the pilot's verbal reports. If mismatches with the pilot's verbal reports occurred, especially when the pilot was looking at the AI for a longer period of time, then a four-state HMM (vertical-tracking, horizontal-tracking, airspeed-tracking, and attitude-monitoring tasks) was applied. Example initial conditions of the model parameters and the estimation results of them are listed in Appendix C. Table 4-4 lists the resulting match rates of the pilots' verbal reports. The numbers of the pilots' verbal reports obtained vary largely from pilot to pilot. For instance, Pilot 1 generated plenty of verbal reports, while Pilots 2 and 5 tended to provide much fewer. Note that Pilot 2 used the free-mumbling

format for verbal reports, and many of his reports that were readouts of indicators involved in only one tracking task had to be discarded, in the same manner used in the previous experiment in Chapter 3. Generally, the more verbal reports, the more reliable the HMM estimates are (see Chapter 5 for further discussion). However, the fixation data of Pilot 2 and Pilot 5 seemed relatively orderly, and, in most of the cases, the Baum Welch method convergence was fast, which is a good sign of the data orderliness with respect to the given model. Since these pilots' HMM estimates appeared reasonably reliable, it was decided to include their data in this study.

Table 4-4. Match Rates between the verbal reports and the HMM hidden-state estimates
(the number of matches/number of total verbal reports × 100 (%))
when the 3-State or 4-State HMM was applied.

Pilot	PGD		PD		D	
	3-State HMM	4-State HMM	3-State HMM	4-State HMM	3-State HMM	4-State HMM
Segment (i) Straight & Level						
Pilot 1	41/74 (55.4%)	64/74 (86.5%)	45/73 (61.6%)	66/73 (90.4%)	46/71 (64.8%)	70/71 (98.6%)
Pilot 2	5/6 (83.3%)	---	8/8 (100%)	---	9/9 (100%)	---
Pilot 3	13/13 (100%)	---	13/14 (92.9%)	---	16/16 (100%)	---
Pilot 4	29/36 (80.6%)	33/36 (91.7%)	25/32 (78.1%)	27/32 (84.4%)	25/29 (86.2%)	26/29 (89.7%)
Pilot 5	7/10 (70.0%)	---	6/7 (85.7%)	---	6/9 (66.7%)	~~6/9 (66.7%)~~
Pilot 6	9/11 (81.8%)	---	8/10 (80.0%)	---	9/12 (75.0%)	---
Segment (ii) Constant Airspeed Descent						
Pilot 1	59/102 (57.8%)	91/102 (89.2%)	32/53 (60.4%)	50/53 (94.3%)	65/94 (69.2%)	86/94 (91.5%)
Pilot 2	9/11 (81.8%)	---	15/15 (100%)	---	13/13 (100%)	---
Pilot 3	18/18 (100%)	---	19/19 (100%)	---	16/16 (100%)	---
Pilot 4	51/70 (72.9%)	---	37/67 (55.2%)	51/67 (76.1%)	46/66 (69.7%)	~~46/66 (69.7%)~~
Pilot 5	6/8 (75.0%)	---	8/9 (88.9%)	---	8/10 (80.0%)	---
Pilot 6	15/16 (93.8%)	---	9/9 (100%)	---	14/18 (77.8%)	---
Segment (iv) Level Turn						
Pilot 1	50/72 (69.4%)	63/72 (87.5%)	49/69 (71.0%)	59/69 (85.5%)	47/58 (81.0%)	55/58 (94.8%)
Pilot 2	12/16 (75.0%)	14/16 (87.5%)	14/16 (87.5%)	---	11/15 (73.3%)	14/15 (93.3%)
Pilot 3	24/35 (68.6%)	28/35 (80.0%)	39/47 (83.0%)	43/47 (91.5%)	33/41 (80.5%)	35/41 (85.4%)
Pilot 4	41/55 (74.6%)	51/55 (92.7%)	40/47 (85.1%)	42/47 (89.4%)	24/33 (72.7%)	29/33 (87.9%)
Pilot 5	16/20 (80.0%)	17/20 (85.0%)	13/17 (76.5%)	---	4/5 (80.0%)	---
Pilot 6	26/33 (78.8%)	30/33 (90.9%)	25/33 (75.8%)	30/33 (90.9%)	28/42 (66.7%)	30/42 (71.4%)
Segment (v) Final Descent						
Pilot 1	43/55 (78.2%)	54/55 (98.2%)	31/44 (70.5%)	39/44 (88.6%)	40/56 (71.4%)	51/56 (91.1%)
Pilot 2	3/6 (50.0%)	~~4/6 (66.7%)~~	7/8 (87.5%)	---	3/3 (100%)	---
Pilot 3	20/30 (66.7%)	23/30 (76.7%)	25/29 (86.21%)	---	25/33 (75.8%)	27/33 (81.8%)
Pilot 4	42/61 (68.9%)	54/61 (88.5%)	37/52 (71.2%)	50/52 (96.2%)	25/29 (86.2%)	27/29 (93.1%)
Pilot 5	9/9 (100%)	---	10/12 (83.3%)	---	11/11 (100%)	---
Pilot 6	29/35 (82.9%)	---	22/36 (61.1%)	~~24/36 (66.7%)~~	31/40 (77.5%)	---

As seen in Table 4-4, most of the HMM state estimates, with either the three-state or the four-state HMM analysis, indicated at least 70% match rates with the verbal reports. Note that increasing the number of states did not always improve the match rate, such as in the four cases in Table 4-4 that have numbers in the four-state HMM column struck out. In these cases, the match rates with the original three-state HMM estimation results were kept, even though these were below 70%. As Table 4-4 shows, better match rates for Pilot 1 always resulted when the four-state HMM was applied than when the three-state HMM was applied. Some of Pilot 4's data also resulted in better match rates when the four-state HMM was applied. In segment (iv), all pilots showed longer fixation durations on the AI, the signature scan pattern of the attitude monitoring task, which thus resulted in better match rates when the four-state HMM was applied with at least one indicator format. In segment (v), only Pilots 1, 3, and 4 showed that signature AI fixation pattern, which thus resulted in better match rates when the four-state HMM was applied.

Once the task sequences were estimated by HMM analysis, the sequences, $Q^{s,f,p}$, were divided back into the original three approaches, $\{q_1^{s,1,f,p}, \ldots, q_{T1}^{s,1,f,p}\}$, $\{q_1^{s,2,f,p}, \ldots, q_{T2}^{s,2,f,p}\}$, and $\{q_1^{s,3,f,p}, \ldots, q_{T3}^{s,3,f,p}\}$. Then, the task durations for each segment were computed. Again, since the task durations had positively skewed distributions, the values were transformed by taking natural logarithms before statistical analyses were performed. As in the fixation duration analyses, the task duration data were analyzed separately for each format in each segment. The same mixed regression repeated measures analysis was applied with the same model as in Eq. (4-2) and (4-3).

Analogous to the fixation duration analysis, the null hypotheses, $H_0: \beta_1-(-\beta_1-\beta_2) = 2\beta_1+\beta_2 = 0$, $H_0: \beta_2-(-\beta_1-\beta_2) = \beta_1+2\beta_2 = 0$, and $H_0: \beta_1-\beta_2 = 0$, were tested to examine the differences between PGD and D, PD and D, and PGD and PD, respectively. The hypothesis tests for the attitude-monitoring task were not performed for segments (i), (ii), and (v), due to the lack of sufficient data. The test results are listed in Table 4-5. The *p*-values shown are 2-tail *p*-values. Figure 4-9 plots the grand means along with the hypothesis test results.

Table 4-5. Flight task durations: Mixed regression repeated measures analysis, indicator format effect hypothesis test results (only $p < 0.05$ are shown)

Segment	Effects Tested	Estimates (log)	Std. Error	z	p
Vertical Tracking					
(i)	PGD-D	-0.1700	0.0831	-2.0448	0.0414
	PD-D	0.0083	0.0833	0.0996	---
	PGD-PD	-0.1783	0.0680	-2.6239	0.0088
(ii)	PGD-D	0.0041	0.0861	0.0476	---
	PD-D	0.2398	0.0866	2.7689	0.0056
	PGD-PD	-0.2357	0.0705	-3.3432	0.0008
(iv)	PGD-D	0.0422	0.0672	0.6284	---
	PD-D	0.2554	0.0666	3.8364	0.0002
	PGD-PD	-0.2132	0.0546	-3.9052	0.0000
(v)	PGD-D	0.0428	0.0693	0.6172	---
	PD-D	-0.0815	0.0692	-1.1773	---
	PGD-PD	0.1243	0.0566	2.1973	0.0278
Horizontal Tracking					
(i)	PGD-D	-0.1893	0.1118	-1.6932	---
	PD-D	-0.0744	0.1134	-0.6560	---
	PGD-PD	-0.1149	0.0919	-1.2497	---
(ii)	PGD-D	0.2904	0.1103	2.6334	0.0086
	PD-D	-0.0600	0.1106	-0.5427	---
	PGD-PD	0.3504	0.0902	3.8866	0.0002
(iv)	PGD-D	0.1609	0.0725	2.2187	0.0264
	PD-D	0.0566	0.0719	0.7867	---
	PGD-PD	0.1043	0.0590	1.7685	---
(v)	PGD-D	0.2828	0.0819	3.4542	0.0006
	PD-D	0.2761	0.0813	3.3963	0.0006
	PGD-PD	0.0067	0.0666	0.1006	---
Airspeed Tracking					
(i)	PGD-D	0.2325	0.0772	3.0120	0.0026
	PD-D	0.1746	0.0770	2.2687	0.0232
	PGD-PD	0.0579	0.0629	0.9200	---
(ii)	PGD-D	-0.1686	0.0809	-2.0828	0.0376
	PD-D	-0.3273	0.0815	-4.0147	0.0000
	PGD-PD	0.1587	0.0663	2.3926	0.0168
(iv)	PGD-D	-0.0597	0.0662	-0.9023	---
	PD-D	-0.2160	0.0662	-3.2646	0.0012
	PGD-PD	0.1563	0.0540	2.8932	0.0038
(v)	PGD-D	-0.5161	0.0766	-6.7363	0.0000
	PD-D	-0.2639	0.0765	-3.4497	0.0006
	PGD-PD	-0.2522	0.0625	-4.0347	0.0000
Attitude Monitoring					
(iv)	PGD-D	-0.0823	0.0750	-1.0972	---
	PD-D	-0.3071	0.0759	-4.0475	0.0000
	PGD-PD	0.2248	0.0616	3.6495	0.0002

(a) Vertical-Tracking Task

(b) Horizontal-Tracking Task

(c) Airspeed-Tracking Task

(d) Attitude-Monitoring Task

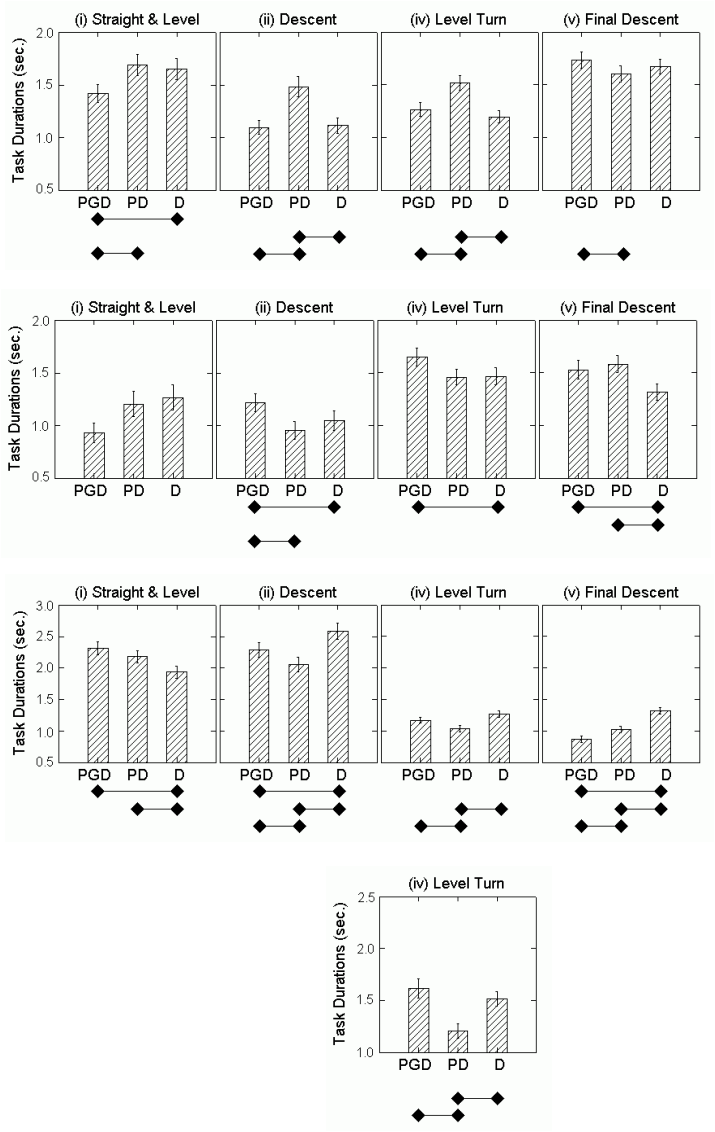


Figure 4-9. Grand means and standard errors of task durations (the values before taking logarithms) on the (a) vertical-tracking, (b) horizontal-tracking, and (c) airspeed-tracking tasks, in segments (i), (ii), (iv), and (v), and the (d) attitude-monitoring task, in segment (iv). Diamonds connected by a line indicate a significant difference between the two formats ($p < 0.05$) computed by hypothesis tests.

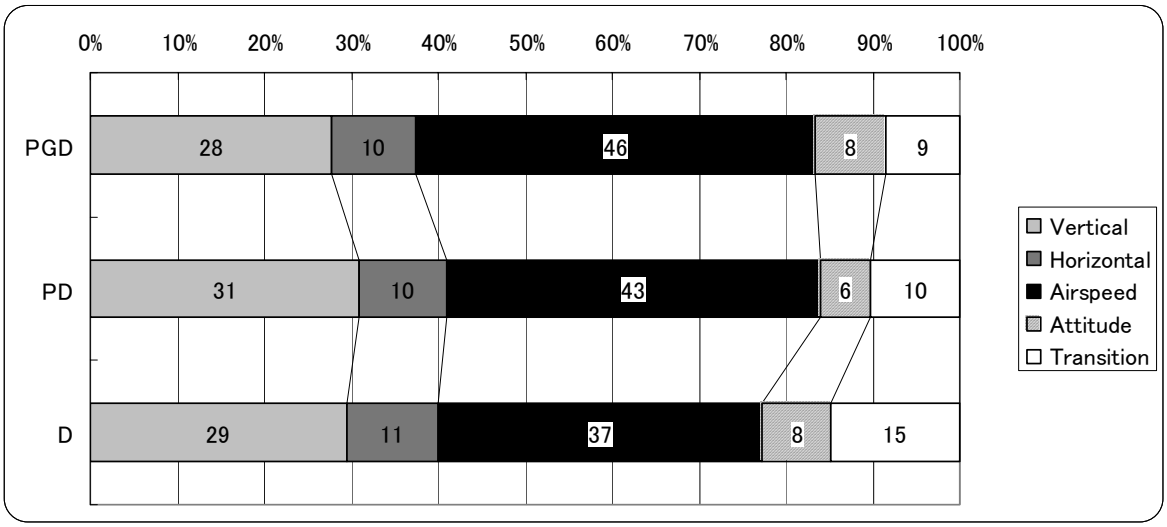
Figure 4-9 shows the switching of the pilots' attention management strategies from one segment to another. For instance, in segment (i), straight and level, the vertical tracking and airspeed tracking had longer durations, and thus were the dominant tasks, while the airspeed tracking was the single dominant task in segment (ii), constant airspeed descent, and the vertical- and horizontal-tracking tasks were the dominant ones in segment (v), final descent.

Significant differences between PGD and PD were observed in the task durations in many

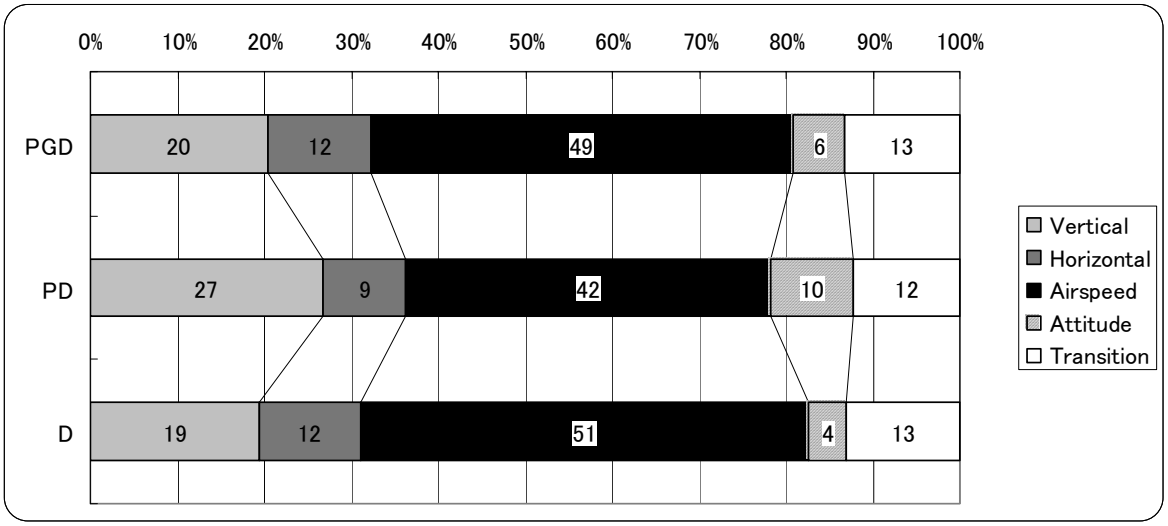
segments. Especially, in segment (ii), constant airspeed descent and (iv), level turn, the vertical-tracking task durations significantly increased when PD was used compared to when PGD or D was used.

The frequencies of visits to each task per second—"task rates" (as the counterpart of "look rates")—were also examined for segments (i), (ii), (iv), and (v). The GLM repeated measures analysis with the random effect being Subjects and the main effect variables being Segment and Format was applied. The results found no significant format effect on any of the four task rates.

Now, to summarize the above results, Figure 4-10 shows the percent times of each tracking task within the total time of each flight segment by each indicator format. In these bar plots, all the pilots' data are added together. Since there was no Format effect in the task rates, the percent times mostly reflect the Format effects on the task durations shown in Figure 4-9. For instance, in segment (i) (Figure 4-10 (a)), the vertical-tracking task percent time was larger when PD or D was used than when PGD was used. Also, the airspeed-tracking task percent time was smaller when the D format was presented compared to when the other formats were presented, which is consistent with the results in Figure 4-9. In segment (ii) and (v), the vertical-tracking task percent time increased when the PD format was used compared to when one of the other two formats was used. The extra time spent on the vertical task when PD was used was apparently taken from the horizontal-tracking (in segment (ii)), airspeed-tracking (in both segments), and attitude-monitoring (in segment (iv)) tasks. These are also consistent with Figure 4-9.

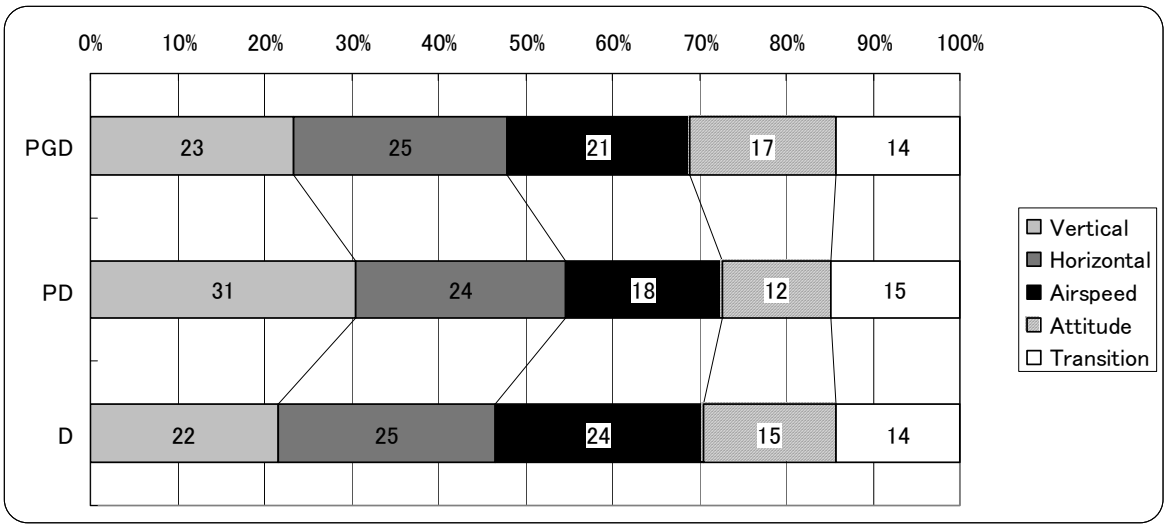


(a) Segment (i): Straight & Level

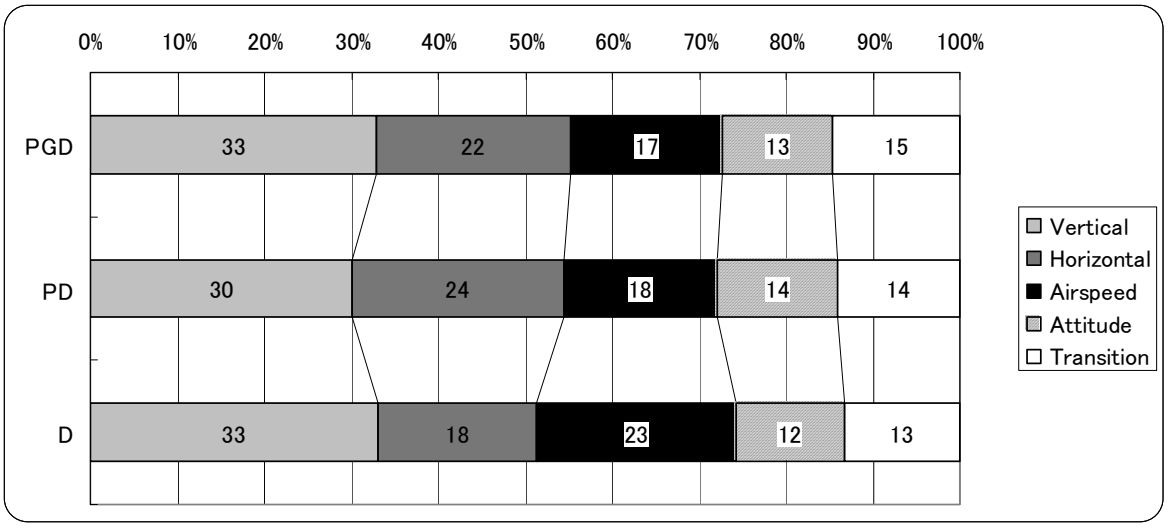


(b) Segment (ii): Constant Airspeed Descent

Figure 4-10. Total percent times (%) of each tracking task for each flight segment and for each indicator format. All pilots' data were added together.



(c) Segment (iv): Level Turn



(d) Segment (v): Final Descent

Figure 4-10. (Continued) Total percent times (%) of each tracking task for each flight segment and for each indicator format. All pilots' data were added together.

4.4 Discussion

4.4.1 Segment (i) – Straight and Level

As Figure 4-7 (a) shows, the ASI fixation durations were significantly longer when either one of the rotating pointer formats (PGD or PD) was used than when the digit-only format (D) was used. Figure 4-9 (c) shows that the same trend was observed in the airspeed-tracking task durations as well. On the other hand, Figure 4-7 (b) indicates that the altimeter fixation durations were significantly shorter when PGD or PD was used than when D was used. However, Figure 4-9 (a) shows that, at the tracking task level, while the vertical-tracking task durations were significantly shorter when PGD was used, the vertical-tracking task durations for the PD were actually more similar to the D format than the PGD format despite the shorter trend observed in the altimeter fixation durations. The longer vertical-tracking task durations for the PD were probably due to the VSI fixations; the VSI fixation durations were significantly longer when PD or D was used compared to when PGD was used. The trends at the task level may be easier to see in Figure 4-10 (a).

Note that the VSI format was always digits-only, and did not change throughout the experiment. The ASI and the altimeter were the only indicators that changed formats. Therefore, the increased fixation durations on the VSI when PD was used compared to when PGD was used may imply that, when the gradation marks were not present, the pilots did not utilize much altitude rate information in the altimeter, even though the rotating pointers were present.

Remember that the RMS flight technical error analysis results indicated that the RMS airspeed error was significantly reduced when either the PGD or the PD format was used compared to when the D format was used, but no significant format effect was observed in the RMS altitude error. These flight technical error results and the results from the eye-movements analysis and HMM analysis together mentioned above can provide an interesting insight into the attention budgeting of the pilots in this segment. When the gradation marks were not present (PD or D), the pilots attended the altimeter for a longer time than when they were present (PGD), and, as a result, they managed to maintain the same level of altitude accuracy as when the gradation marks were present (PGD). The extra time spent for the vertical-tracking task when the gradation marks were not present (PD or D) was taken from the airspeed-tracking task, resulting in larger airspeed

error when D was used. The airspeed error, however, did not increase when PD was used, despite the increased demand for the vertical-tracking task.

These observations suggest that the pilots gave priority to altitude accuracy over airspeed accuracy. One reason for this could be that the airspeed is a derivative value of the aircraft position, and may be less conspicuous to the pilots than more direct values such as altitude. Another possible reason is that many flight segments were defined by attributes of the altitude rather than the airspeed (e.g., straight and level, level turn, descent), and thus the pilots may have unconsciously prioritized the altitude over the airspeed. Also, in real-world flights pilots are monitored by air traffic control by altitude, not airspeed. Thus, it was also possible that the pilots were more sensitive about errors in altitude than errors in airspeed. Such operational common sense may have affected their attention budgeting strategy.

4.4.2 Segment (ii) – Constant Airspeed Descent

In this segment, the pilots set the engine idle, and controlled the airspeed by pitch angle to maintain 180 knots. Thus, the ASI, the AI, and the VSI were the main indicators to look at (for the VSI, the pilots were instructed not to exceed the 2000 ft/min rate of descent). As Figure 4-7 (a) shows, the ASI, one of the main indicators, was looked at for significantly longer when the gradation marks were present (PGD) compared to when they were not present (PD or D). As seen in Figure 4-7 (c), there was no Format effect in the VSI fixation durations. Figure 4-7 (b) indicates that the altimeter—mostly used near the end of the descent—was fixated for significantly longer when the digits-only format (D) was used than when one of the rotating pointer formats (PGD or PD) was used.

According to the above observations, one would expect to see longer airspeed-tracking task durations with PGD and/or longer vertical-tracking task durations with D. However, the task duration analysis results showed that the longest airspeed-tracking task was with D, and the longest vertical-tracking task was with PD. Since the fixation duration results of the ASI, the altimeter, and the VSI did not directly reflect onto the task duration results, the results must have been coming from how the AI (the other main instrument in this segment) was used in combination with these indicators. In any case, as Figure 4-9 (a), (b), and (c) show, the vertical-tracking task durations were significantly longer when the gradation marks were not present (PD) compared to the other format (PGD or D), and the extra time spent for the

vertical-tracking task was seemingly taken from the horizontal- and airspeed-tracking tasks times. Again, these trends can be also observed in the percent times for each format in Figure 4-10 (b).

4.4.3 Segment (iv) – Level Turn

In this segment, the main task for the pilots was to capture the localizer course and stabilize along it. Thus, from this segment, the horizontal-tracking task starts demanding its share in the pilots' attention budgeting. Indeed, as Figure 4-9 (b) and (c) indicate, the overall horizontal-tracking task durations jumped while the overall airspeed-tracking task durations went down. However, Figure 4-9 (a) shows that the vertical-tracking task durations did not change much compared to those in the other segments. This observation supports the speculation stated in the segment (i) section; that is, the pilots may have had the tendency to give priority to the altitude over the airspeed.

At the fixation duration level, as Figure 4-7 shows, the basic trends in the fixation durations are mostly similar to those in segment (i), except that the absolute values for the fixation durations became smaller due to the increase for the horizontal-tracking instrument fixations. For the tracking task durations, as Figure 4-9 (a) shows, the vertical-tracking durations were significantly longer when the gradation marks were not present (PD) compared to when one of the other formats (PGD or D) was used. Again, as in segment (i), the increased vertical-tracking task duration is probably due to the increased demand for VSI fixations, which also implies that, when the gradation marks were not present, the pilots did not extract as much altitude rate information from the altimeter as when they were present (PGD), despite the presence of the rotating pointers. As shown in Figure 4-10 (c), the increased vertical-tracking task durations resulted in a decrease in the airspeed-tracking task and attitude-monitoring task durations in this segment.

4.4.4 Segment (v) – Final Descent

In this segment, the aircraft is supposed to be already stabilized at the assigned airspeed, and descending along the glide slope. Since the ASI and altimeter are usually not looked at as often as in the other segments, it was expected that the effects of the ASI and altimeter would be minimal. However, some format effects were still observed. As Figure 4-7 (a) shows, the ASI fixation durations were significantly longer when the gradation marks were present (PGD) than

when one of the other formats was used. Figure 4-7 (b) indicates that the altimeter fixations were significantly longer when the digits-only format (D) was used compared to when one of the other formats was used. At the tracking task durations level, as seen in Figure 4-9 (c), the airspeed-tracking task durations were significantly longer when the digits-only format (D) was used compared to when one of the other formats was used.

4.4.5 So, Are Gradation Marks Recommended or Not?

So far, a number of differences between PGD (with gradation marks) and PD (without gradation marks) were found, but, in general, it is difficult to tell which one is better than the other based on these differences. Are the longer fixation durations or longer tracking task durations good or bad? Longer fixations may mean the indicator is hard to read at a glance, but also shorter fixations may mean it does not provide much information. A longer tracking task duration may mean the attended task will be performed well, but it may also mean cognitive tunneling or attention trapping to a single tracking task is occurring while other tracking tasks are being neglected.

Although it may not be clear whether some fixation durations or tracking task durations are too long or too short, some of the findings of the characteristics that seemed to be caused by lack of the gradation marks may raise the eyebrows of some people. For instance, the increased VSI fixation durations and look rates observed when PD was being used in segments (i) and (iv) may suggest that the pilots were not utilizing the gradation marks of the altimeter well, and, thus, were compensating for the lack of the rate of climb/descent information by fixating on the VSI for longer and/or more often. This observation implies that the lack of gradation marks in the altimeter is a disadvantage. Again, note that the indicators that changed formats were the altimeter and the ASI, not the VSI itself.

In addition, knowing that the PGD format has been accepted as a standard Air Force HUD symbology for quite a while already, and that the D format has been regarded as insufficient for their operations, one may be able to assess if the PD format is acceptable or not by simply comparing it to PGD or D. For example, Figure 4-7 (a) indicates that, when PD was used, the fixation durations on the ASI in both segment (ii) and (v) were significantly shorter than those observed when PGD was used and, as a result, rather similar to those observed when D was used (see the connected diamond marks at the bottom). At the task level, as indicated in Figure 4-9,

the vertical-tracking task durations in segment (i), the horizontal-tracking task durations in segment (ii), and the airspeed-tracking task in segment (v) resulted in similar effects; that is, when PD was used, these tracking task durations were significantly different from those that resulted when PGD was used, while being rather similar to those observed when D was used.

These observations suggest that in certain flight segments, when the gradation marks were not present, the rate information of the ASI and the altimeter were not utilized by the pilots very well despite the presence of the rotating pointers. This also may explain the pilots' slight preference for PGD over PD. The difference between them may be subtle—in fact the experiment did not find any difference in the RMS flight technical errors, the ultimate performance measure criterion—but these slight differences may prevail someday after tens of thousands of operational runs. Therefore, it would be safer to recommend including gradation marks in the HUD ASI/altimeter rotating pointer formats.

4.5 Conclusion

The study showed that using either one of the rotating pointers (PGD and PD) resulted in smaller airspeed error and higher pilot preference ratings. These results were consistent with the USAF study, although, unlike their study, this study did not find any significant format effect in the altitude error.

In addition to these conventional measures analyses, the study also incorporated eye-movement data analysis and HMM analysis, which provided further insight into the effects of the rotating pointers and gradation marks. These analyses revealed that the pilots seemed to have given higher priority to altitude over airspeed, and, consequently, when the digits-only (D) format was used in segment (i) or (iv), where the horizontal-tracking task demand increased, they tended to sacrifice airspeed-tracking performance for the vertical-tracking performance. This agrees with the RMS airspeed and altitude error analysis results. In addition, fixation analysis results showed that, when the gradation marks of the rotating pointer format were not present (PD), the VSI fixation durations and look rates both increased, suggesting that the pilots could not use the rate of climb/descent information in the altimeter when gradation marks were not present, despite the presence of the rotating pointers. This increased fixation demand on the VSI in turn resulted in the vertical-tracking task durations also tending to increase when the gradation marks were not

present. Furthermore, in both descent segments (segment (ii), constant airspeed descent, and segment (v), final descent), the pilots' ASI fixation durations became significantly shorter when the gradation marks were not present (PD) compared to when they were present (PGD), resembling more those observed when the digits-only format (D) was used. When the gradation marks were not present (PD), the vertical-tracking task duration in segment (i), the horizontal-tracking in segment (ii), and the airspeed-tracking in segment (iii) also showed more resemblance to those observed when digits-only (D) was used, rather than those observed when the gradation marks were present (PGD).

The study not only confirmed the superior scanning efficiency of the rotating pointer formats (PGD and PD) over the digits-only format (D), but also found a number of differences between the two rotating pointer formats (PGD vs. PD). Considering these results as well as the RMS flight technical error and the pilots' subjective preference rating analysis results, the study concluded that PGD has slight advantages over PD, and thus including the gradation marks in the rotating-pointer HUD ASI/altimeter indicator format is recommended. Thus, this study provided empirical support for the common belief of many human factors practitioners about the advantages of including gradation marks.

Chapter 5:

Further Discussions about HMM Analysis

As described in Chapter 2, some aspects of the HMM analysis of pilots' eye-movement data are distinct from such analysis in other HMM applications, e.g., speech recognition. One of the most characteristic differences is the verbal-report requirement. In this chapter, various issues raised by this requirement are discussed. Furthermore, the basic assumptions of the HMM analysis made at the beginning, such as the serial attention assumption and the Markov process assumption, are revisited. Finally, some areas for future study are suggested.

5.1 Issues Related to the Verbal-Report Requirement

5.1.1 Different Pilot Types

The first concern about asking pilots to speak during flight was whether this added workload would affect the way they scan and fly. In the experiments in this thesis, it seemed that it depends on the type of pilot. Some pilots seemed not to have any problem with speaking while flying. Pilots who regularly talk to themselves even in real flights may have felt more ease about the verbal-report requirement. Talking to himself or herself about the instrument readings or what to do next is an effective way to utilize one's verbal short-term memory in addition to regular short-term memory to help increase alertness. In fact, in a preliminary study conducted before the experiments described in this thesis, there was one pilot who liked to speak during the flight and took advantage of it so much that, in the trials where he was prohibited to speak, his flight performance actually got noticeably worse.

Other pilots, however, seemed to feel somewhat awkward about the verbal-reporting task at the beginning. In the experiments in this thesis, all pilots eventually got used to the verbal-report requirement after some practice. Still, many pilots stopped generating reports for a moment when

the flight workload became high, and had to be reminded by the experimenter to resume reporting. This implies that the verbal-report requirement did add to the workload, especially when the pilot was busy, and could have affected flight performance.

In order to reduce the extra cognitive workload caused by the verbal-report requirement, two different reporting formats, the free-mumbling format and the structured-reporting format, were tried in the two experiments in this thesis. Although the structured-reporting format was supposed to cause less workload and therefore be more beneficial, it turned out that both of the formats had their advantages and disadvantages.

5.1.2 Free-Mumbling Format

In the first experiment (Chapter 3), the free-mumbling format was used, which meant that the pilots generated reporting sentences using their own words. The advantage of this format is that it provides context for the flight and helps the experimenter understand the pilots' intentions. In cases where the HMM structure is initially unknown, free mumbling may be used as a starting point to construct an initial HMM structure. Another advantage is that, because the pilots have to speak to the radio or to the other crewmember during real flights, the free mumbling can actually be a realistic secondary task if the researcher wants to see the eye-movement data in a realistic setting. However, if the researcher does not want the pilots to be distracted, then forcing them to generate verbal sentences could be a problem, because that may require extra attentional resources, which may have to be drawn from the flight task.

In addition, for the free-mumbling-type data, the researcher has to interpret the reports to identify corresponding tracking tasks, and, thus, it not only may substantially increase the data analysis workload of the researcher, but also may create an additional potential source of error. If the researcher has instrument flight experience himself/herself, identifying the task is mostly straightforward; thus, in the case of such researcher, the chance of errors caused by misinterpretation should be relatively small. However, in general, a more objective way to conduct such a process involving subjective judgments may be to have multiple researchers interpret the verbal data independently, and then compare the results. In this way, the chance of misinterpretation can be minimized.

5.1.3 Structured-Reporting Format

In the second experiment (Chapter 4), in the effort to reduce the pilots' workload, the structured-reporting format was introduced. In this format, the pilots specifically reported which information on the AI—"pitch," "bank," or "pitch & bank (or both)"—they were attending to. Note that the only overlapping area was the AI in the HUD experiment, because the glide slope and localizer deviations were displayed separately on this display. Using predetermined words instead of self-generated sentences should reduce the cognitive workload and amount of attention required. In the experiment, the pilots were given options to use either the free-mumbling or the structured-reporting format, and all pilots except Pilot 2 chose to use the structured-reporting format. This suggests that the structured-reporting format imposed less workload than the free-mumbling format.

The structured-reporting format also greatly reduces workload on the experimenter's side because interpreting the verbal report data and converting the reports to the corresponding tasks are minimized. One disadvantage of the structured-reporting format is that the context of the flight is lost. For instance, if the pilot said "stop ballooning," then the researcher can know that the pilot probably just had lowered the flaps, and the pilot was trying to reduce the pitch angle and/or trimming to stop the momentary climbing trend caused by the suddenly increased lift. However, if the pilot just said "pitch," such context information is lost. Another disadvantage of structured reporting is that, when the pilot said "pitch," the experimenter cannot distinguish if the pilot meant the vertical-tracking or airspeed-tracking task anymore, because both tasks involve the "pitch" angle. That means the existence of the vertical- or airspeed-tracking task cannot be assessed based on this format. For the HUD experiment, this was not a significant issue, because the approximate underlying HMM structures were known, and all the pilots were highly experienced and expected to have at least the three states; i.e., the vertical-, horizontal-, and airspeed-tracking tasks. However, for the cases in which the initial model selection was the main interest, such as the experiment in the Chapter 3, the structured-reporting format may not have been appropriate.

5.1.4 Verbal Report Frequency

Another concern would be how frequently the verbal reports should be generated. The verbal reports are used to verify the HMM-estimated attention sequence; therefore, basically the more

verbal reports, the better. Yet, generating too many verbal reports may distract the pilot from the flying task. In addition, if the pilot generates too many verbal reports, it also becomes a concern that the verbal reports may start leading the eye-movements, not the other way around. In terms of the pilots' workload level, the comfortable reporting frequency seemed to be around once every five to ten seconds for the current experiment settings.

In terms of reliable model parameter verification, it may be useful to think in terms of degrees of freedom. For simple L-th order polynomial curve-fitting cases, for instance, at least L data points are required to prevent under-fitting (i.e., the coefficients cannot be specified from the given data). If a similar reasoning is applied to the HMM analysis, it is desirable to have at least the number of the model parameters being estimated, say L. For example, for the original three-state HMM in the first experiment (Figure 3-4), $L = L_A + L_B + L_\pi = (3\times3) + (5+3+4) + (3) = 24$. Here, note that the verbal reports are not being used to compute the parameter estimates, but to verify the attention sequence estimated by the Viterbi algorithm, which is based on the model parameters estimated by the Baum-Welch algorithm. Thus, even if the number of the verbal reports is less than L, a single set of the model parameters still can be estimated, although the resulting sequence estimated by these algorithms would never be fully verified.

In practice, however, first of all, obtaining L independent verbal reports may be difficult, especially if the pilot has a tendency to generate the verbal reports at similar attentional instances. If that is the case, some of the L verbal reports are actually duplicates. Second of all, there is no guarantee that the pilots generate the L full-spectrum verbal reports that are required to fully verify the L model parameters. For example, if a pilot almost always reported vertical-tracking related reports, then it is not clear whether the pilot was always in the vertical-tracking task,or just skipped reporting all other tracking tasks. In such a case, one can actually attain the maximum match rate of the verbal reports by heavily weighting the model parameters toward the vertical-tracking task so that the estimation results would result in the vertical-tracking task almost all the time. However, the legitimacy of such a modification is questionable because there is no real reason to believe that the other tracking tasks did not happen.

To minimize the above problems, the experimenter should instruct the pilots to generate verbal reports regarding not just one particular tracking task, but all the tracking tasks. Such instruction would be easier to give if the structured-reporting format were used ("Please try to report both

“pitch” and “bank.””), than if the free-mumbling format were used (How can one instruct pilots to report all tracking tasks without exposing what tracking tasks are being investigated?).

Even if instructions are given, the resulting verbal reports still may be insufficient. This could be another limitation of HMM analysis. Note that since the problem of the verbal reports is shortage, not excess, it is still reasonable to use the verbal reports as a basis to justify adding new hidden states. When the verbal reports are used to justify removing a hidden state, however, care should be taken. In Chapter 3, for instance, when the airspeed-tracking task was removed from the HMM structure for Pilot 1, additional observation that the throttle was used mainly to control the rate of descent provided supporting evidence to confirm that it was not just that the pilot skipped reporting the airspeed-tracking task. One suggestion for supplementing this shortage of verbal reports is to incorporate the pilots’ flight control input data in this verification process. This option will be discussed further in the future work section.

5.2 Pilots’ Attention Models—Serial vs. Parallel Assumptions

As mentioned in Chapter 2, the HMM analysis is based on the assumption that the pilots attend to each tracking task in a serial manner rather than in a parallel manner. Indeed, the pilots’ eye-movement data collected in the current experiments indicated that pilots tend to crosscheck among the related instruments, and that resulted in a serial-like attention sequence. Even though it is unknown whether the real attention process is serial or not, as long as the pilots’ eye-movement data indicate existence of a serial-like underlying process, the HMM analysis will work. Thus, for the purpose of the HMM analysis, the serial assumption appeared to be adequate.

In general cases, readers should bear in mind that there is still controversy over whether people process information in a serial manner or a parallel manner. In fact, there were some signs even in this study that might suggest that the process was not entirely serial, although they did not seriously affect the analysis outcomes of the current experiments. These are described in this section to remind readers of the limitation of the serial assumption.

One of these signs was that, as seen in the current experiments, some verbal reports did not fit the HMM-estimated attention sequence and, therefore, were labeled as missed detections. Taking a closer look, one may find that some of these reports described tasks that could not be

associated with any of the instruments the pilot was fixating on around the time of the report. Exactly what is happening in these instances? Of course, they could be simply the pilot's mistakes. However, these reports also could be delayed reports, which the pilot held in short-term memory when the flight workload was too high to perform the additional verbal report task, and executed later when the workload became moderate. Or, the pilot could have been actually processing in parallel; that is, he/she was thinking about one tracking task while fixating on instruments of the other tracking group. The possibility of that being the case indicates that a certain degree of parallel processing may also be going on.

The other sign was that there was also a slight trend for the pilots to make verbal reports at the same time they were switching the tracking task as well (see Figure 3-5, 10, and 12). This trend is, actually, not a desirable tendency for HMM analysis purposes, since such verbal reports can cause false positives in the verification process because of the ±1-second buffer. Nonetheless, it seemed to be consistently occurring. Generating verbal reports while flying certainly involves multitasking, and the trend that the verbal reporting occurred when the pilot switched to another tracking task may not be a coincidence. Whatever the neurophysiological function is that governs attention management (such as the central executive of Baddeley's working memory model (Baddeley & Hitch, 1974)), the time when this function is activated to switch the tracking task may be the easier time for the pilot to initiate the verbal reporting as well. Such an attention management function is another kind of parallel process.

These signs that suggest parallel processing may be accounted for with Wickens's multiple-resource theory framework (1980). In this theory framework, there are three attentional resource reservoirs: stage of processing (perceptual encoding vs. central processing vs. responding), hemisphere of processing (verbal working memory vs. spatial working memory), and modalities of processing (auditory vs. visual, manual vs. vocal, etc.). Each tracking task requires a visual channel for sampling. Thus, the visual channel of the perceptual encoding stage has to be time-shared, and this makes the pilots' eye-movement process appear to be driven by an underlying serial-like process. The pilots' verbal reports not related to the instruments that they were currently looking at may indicate the parallel processing between all of the three stages, perceptual encoding (the fixation point), central processing (a delayed or parallel cognitive processing), and responding (verbal report). The central processing also includes the

working memory component, which is thought to govern attention management among multiple tasks. Thus, pilots' tendency to generate verbal reports at the time they also switch tracking tasks may be caused by this capacity of central processing.

5.3 Markov Process Assumption

As described in section 3.4.3, the HMM-estimated attention sequences indicated that the transition probabilities sometimes appeared to have changed within a segment. For instance, in the level-turn segment (ii) in the experiment in Chapter 3, the pilots usually attended to the horizontal-tracking task very often at the beginning of the segment, but as the aircraft finished turning, became stabilized straight on the localizer course, and approached the top of the glide slope, they tended to start attending to the vertical-tracking task slightly more often than before (Figure 3-10). Strictly speaking, this violates the assumption of the first-order Markov process (i.e., the transition probability from the state, i, to the state, j, at time, t, $p_t(i,j)$, is independent of t). Nevertheless, as pointed out in section 3.4.3, the estimated task sequence matched well with the pilots' verbal reports, and, thus, the HMM analysis seemed to have worked fine, at least in the experiments in the current thesis.

When the transition probabilities changed within a flight segment, the estimated hidden-state transition probabilities (i.e., the A matrix) basically represented the averages of these transition probabilities over time. Even though the Viterbi algorithm used the single, averaged hidden-state transition probabilities to estimate the most-likely attention sequence throughout the segment, the attention sequence estimated by this Viterbi algorithm showed the changes of the frequency trends of each tracking task as the time progressed. For instance, in Figure 3-10, the Viterbi algorithm plots show the trend that the pilots started attending to the vertical-tracking task more often right before they reached the glide slope. This was because the Viterbi algorithm could also utilize the information that the pilots started to look at the altimeter and the VSI, the instruments included only in the vertical-tracking task instrument group, right before they reached the top of the glide slope. Thus, even though the pilots' attention sequence may violate the definition of the Markov process, eventually, the resulting estimated attention sequences may be relatively insensitive to this problem.

The changes of the hidden-state transition probabilities within each segment may not be a big

problem for the purpose of task sequence estimation, but it may cause trouble with regards to generating synthetic eye-movement data. (The generation of synthetic eye-movement data will be discussed more in section 5.4.3.) If generating the data from the model parameters is the interest, then those segments where the transition probabilities appeared to change in the middle should be divided into sub-segments, so that the characteristics of the hidden-state transition probabilities in the generated eye-movement data correctly reflect the real trends of the hidden-state transition probabilities within each sub-segment.

5.4 Future Work

Finally, here are suggestions for future work.

5.4.1 Incorporating Pilots' Control Input Data

In addition to the pilots' verbal reports, it may be possible to incorporate their control inputs, such as elevator angle, aileron angle, rudder angle, and throttle lever position, in the analysis. For instance, if the pilot is controlling the aileron (by turning the wheel to left or right), he/she is more likely attending to the horizontal-tracking task. If the pilot is moving the throttle lever, then the pilot is more likely attending to the airspeed-tracking task. If there is any reasonably objective way to map the pilots' control input movements to their corresponding attentional states, then it may be possible to incorporate the information of their input data to assist the HMM analysis of the pilots' eye-movement data.

There are two ways to incorporate these inputs. One is to incorporate them as another set of observation data for the HMM analysis. In this case, the control input sequence data and the fixation sequence data together form joint observation probability distributions. The other way, which is probably much simpler, is to incorporate them in the verification process. In this case, the HMM analysis is performed as described in this thesis, and then both the control input data and the verbal reports are used to verify the estimated attention sequence. This could be useful when the verbal reports collected are insufficient. Furthermore, there also may be potential that the verbal-report requirement can be completely eliminated and, if combined with remote eye-tracking technology (typically utilizing the facial-feature pattern recognition), the whole process can become completely unintrusive.

Some words of caution are that there are fundamental differences between the fixation sequence data and the control input data. The instrument fixations happen only one at a time; thus, this is a serial process. On the other hand, the control inputs can occur in parallel; that is, for instance, the pilot can pull the yoke (vertical tracking), turn the yoke (horizontal tracking), and push the throttle lever forward (airspeed tracking) all at the same time. Furthermore, again, borrowing the terminology of Wickens's multiple-resource theory (1980), the fixation sequence is in the perception stage, while the control data are in the responding stage. Thus, whether the fixation sequence and the control inputs have common hidden states or not is unknown, and further study is required.

5.4.2 Implementing Unsupervised Learning

In the current study, the selection of HMM structures for each pilot, including the selection of the number of hidden states, was performed manually by the investigator based on prior knowledge and simple observations. This is called supervised learning (Bishop, 1996). It would be convenient, as well as interesting from a theoretical standpoint, if the model selection could be performed in the manner of unsupervised learning; that is, a clustering of a given data set is applied automatically to seek all the potential HMM structure candidates. Of course, there is a risk that such automatic model selection will return totally unexplainable sets of hidden states. However, there is also a chance that the automatic protocol will return a similar or the same HMM structure as the one manually derived. Then, that would provide powerful evidence to support the HMM structure manually selected based on supervised learning.

5.4.3 Generating Synthetic Eye-Movement Data—Double Markov Process

Aside from HMM structure selection, model parameter estimation, and hidden-state sequence estimation, the HMM algorithm is also capable of generating synthetic eye-movement data using the model parameters. With the model parameters and two random number generators independent from each other, one for the hidden-state transition process and the other for the observation symbology sequence, one may be able to generate reasonably realistic synthetic eye-movement data. (Of course, how *realistic* such data can be is still a question to be explored.) This capacity could be useful for, for example, assessing prototype displays before calling in human subjects for a formal experiment. Also note that, as mentioned in section 5.3, if the human pilots tend to change the hidden-state transition probabilities within a segment, then the segment

should be divided into sub-segments, within each of which, the transition probabilities stay relatively the same, so that the generated synthetic eye movements would behave in a manner similar to that of the human pilots' eye movements.

In the illustration example of the HMM in section 2.2.2, it was briefly mentioned that the model that the HMM analysis is based on does not take the transition probabilities among the instrument fixations into account. This may be fine for estimation purposes, but may not be so for generation purposes. Say there are two instruments, *a* and *b*, on which the pilot would fixate when he is attending to a certain tracking task, *i*. Then, if the pilot looked at each instrument, alternatery every half a second while he is attending the tracking task, *i*, the total percent time of each instrument within the tracking task, *i*, would become 50%. By comparison, suppose the pilot looked at instrument *a* for the first half of the segment, and then switched only once to instrument *b* and looked at it for the rest of the time while attending the tracking task, *i*. If he repeated this behavior every time he attended the tracking task, *i*, then the percent time of each instrument within this tracking task, *i*, again would be 50%. Thus, these two very different cases would result in the same model parameters. More specifically, the row elements of the *B* matrix associated with this tracking task, *i*, would be the same. In other words, the same model parameters can generate two extremely different eye-movement behaviors. This was caused because there was no mechanism to handle the duration times of the fixations at the observation level. (Note that, at the hidden-state level, the recurring probabilities of the Markov process can control the probabilistic characters of the hidden-task durations.) Also, at the observation level, there was no mechanism to capture the probabilities of transitions from one instrument to the other.

To generate synthetic eye-movement data reasonably similar to the eye movements of real human pilots, one solution could be to make the observation process a Markov process as well. Let us call such a model a Double Markov Model (DMM). In this model, the observations (i.e., the instrument fixations) have transition probabilities from one instrument to the other, or the recurring probabilities of each instrument within each hidden state (i.e., the tracking task). Thus, in the DMM, the observation symbol probabilities are no longer represented by a single matrix, such as the B matrix, but by multiple square matrices, B_1, B_2, …, B_n—one matrix for each hidden state.

Generating synthetic eye-movement data with the DMM would be actually quite simple. One just needs, again, two independent random number generators, one for the hidden-state process and the other for the observation process. To estimate the DMM model parameters to be used for this generation process, one can first run the regular HMM analysis on the actual pilots' eye-movement data to obtain the most likely attention sequence, and, then count the numbers of transitions among the instruments for each hidden state to compute the probabilities of transitions among the instruments and the recurring probabilities within each instrument.

5.4.4 Other Applications

HMM analysis of eye movements was applied to research both the difference among pilots who had different expertise levels and the effects of different HUD formats. HMM analysis can be applied to many other investigations of the effects of different flight conditions, such as different weather conditions, different workload/stress/fatigue levels, different automation levels, etc., on pilots' scanning and attention patterns

HMM analysis can also be applied outside the airplane cockpit. Any human operator work environments in which the operators have multiple internal states, each of which is associated with a different set of fixation points and also occurs frequently enough to be meaningfully statistically characterized, can utilize this analysis technique. Such environments may include air traffic control, driving, medical or health monitoring, complex system fault diagnosis, military command and control, unmanned aerial vehicle (UAV) control, and so on.

As described in section 5.4.3, the HMM algorithm can be used to generate synthetic eye-movement data using the model parameters. The potential practical usages of such an application were also mentioned in that section. In addition, the HMM algorithm may be used to recognize particular patterns of an individual pilot's eye-movement data (as in speech recognition applications). This could be used for, for instance, identifying a pilot from his/her eye-movement data. Such applications may seem somewhat bold, but if there is any practical interest, their potential can be further explored as well.

Chapter 6:

Summary

6.1 Summary of Each Chapter

Chapter 1 described how the previously existing quantitative models of pilots' instrument scanning behavior tended to suffer either an under-fitting (too few model parameters to describe the large variety of the actual pilots' eye-movement data) or an over-fitting problem (too many model parameters). The lack of reasonable models has forced many researchers to rely on simple statistical analysis of eye-movement data, such as average fixation durations, average look rates, and so on. Such statistical analysis often involves a time-averaging operation that eliminates information about the sequence of instrument scans, which, if kept, can provide valuable information about pilots' moment-to-moment thought processes.

In order to address these problems, a new HMM-based analysis of pilots' eye-movement data was proposed in Chapter 2. HMM analysis exploits pilots' instrument crosscheck within the vertical-, horizontal-, and airspeed-tracking instruments. In the HMM proposed, these tracking tasks constitute hidden states, and the pilots' instrument fixation sequence corresponds to an observation sequence. The pilots' switching among the tracking tasks is a hidden stochastic process. HMM algorithms can then be used to estimate the most likely hidden-state sequence of attentional states from a given instrument fixation sequence.

Although HMMs have been used for speech recognition and other pattern recognition applications, this is the first time HMMs have been applied to the analysis of pilots' eye-movement data. There are also several differences between the HMM for the analysis of pilots' eye-movement data and the HMM for speech recognition applications. A pilot eye-movement HMM is based on the ergodic model, while the speech recognition HMM is based on the left-right model. Also, the pilots' eye-movement HMM does not have labeled training data like the speech recognition HMM does, and, therefore, it is necessary to collect pilots' own verbal reports for verification.

Chapter 3 presented a flight simulator experiment conducted to explore various potential HMM structures for the actual pilots' eye-movement data. The experiment also served as a walk-through implementation example, or a proof-of-concept, of the HMM analysis. Four pilots with various flight expertise levels participated in the experiment, and each flew three simulated ILS approaches on a flight simulator. The pilots' eye-movement data and the flight variables were recorded. The pilots' verbal reports were also collected using a "free-mumbling" format.

To select a proper HMM structure for each pilot in each flight segment (straight and level, level turn, and final descent), first the three-state HMM was applied for the instrument fixation sequences of all pilots and all flight segments. Then, the principle of parsimony was applied in deciding on the addition or removal of a hidden state; that is, a new hidden state was added only if doing so considerably improved the match rate of the verbal reports to the estimated hidden-state sequence, and a hidden state was removed only if the removal did not affect the match rate of the verbal reports. As a result, an interesting positive correlation between the number of hidden states and the pilots' expertise levels was found. The eye movements of the most-experienced pilot flying the final segment were best described by the four-state HMM, while the eye movements of two intermediate-level pilots were best described by the three-state HMM. The eye movements of the least-experienced pilot indicated that the pilot dropped the airspeed-tracking task during the level turn and final descent segments, and thus the eye-movement data for that pilot were well described by the two-state HMM. While the three- and four-state HMMs appeared to be sufficient for flying the ILS approaches, the two-state HMM appeared not to be.

The results of a second flight simulator experiment were presented in Chapter 4 to demonstrate how HMM analysis results can be helpful in a display study. Three head-up display (HUD) airspeed indicator (ASI) and altimeter symbology formats—rotating pointer with gradation marks (i.e., "PGD"), rotating pointer without gradation marks (i.e., "PD"), and digit only (i.e., "D")—were examined. A previous study conducted by the U.S. Air Force found the difference between the digit-only format (D) and the rotating pointer formats (PGD and PD) in terms of the magnitude of the airspeed and altitude deviations and pilots' subjective preferences, but could not find any difference between the two rotating pointers themselves. Thus, the challenge was to see whether eye-movement analysis and HMM analysis together could detect any difference

between these two rotating-pointer formats.

Each of six Air Transport Pilots flew nine approaches; i.e., three approaches for each ASI/altimeter format. In this experiment, the pilots' verbal reports were collected using the structured-reporting format. As in the USAF study, the difference between the two rotating pointer formats (PGD vs. PD) did not appear in the flight technical error. However, the eye-movement data analysis and the HMM analysis revealed significant differences between these two formats. When the gradation marks were removed (PD), the pilots tended to fixate on the ASI for significantly less time than when the gradation marks were present (PGD) in the straight-and-level, constant-airspeed-descent, and final descent segments. Although there was no significant difference in the altimeter fixation durations, the VSI fixation durations tended to be significantly longer when the gradation marks were removed (PD) than when they were present (PGD). In the straight-and-level and level-turn segments, the HMM analysis results showed that the vertical-tracking task duration tended to become significantly longer when the gradation marks were removed (PD) than when they were present (PGD), probably due to the increased demand for the VSI fixation durations (the VSI is one of the vertical-tracking task indicators). In addition, when the gradation marks were removed (PD), the horizontal tracking in the level-turn segment, and the airspeed tracking in final-descent segment also showed more resemblance to those observed when the digits-only (D)format was used, rather than those observed when the gradation marks were present (PGD).

These results provided the researchers evidence that when the gradation marks were removed (PD), in flight segments involving level fight maneuvers, the pilots may have tended to spend more time for the vertical-tracking task by sacrificing time that would otherwise be used for the airspeed-tracking task. Moreover, if the gradation marks were removed (PD) in the final descent segment, the pilots may not have extracted the rate information on the ASI as much as they would if the gradation marks were present. The levels of the airspeed and altitude deviations were equivalent for these two rotating-pointer formats, but the scanning strategy changes may explain the difference in the pilots' subjective preference scores between these formats (the pilots preferred PGD over PD). The trends found in the pilots' eye movements and their attention patterns also suggested that the pilots tended to give priority to the altitude performance over the airspeed performance. In conclusion, the experiment in this chapter demonstrated how the HMM

analysis results could bridge between the eye-movement data and the pilots' performance or preference data, and help the researchers to interpret the data from these experiments.

In Chapter 5, some issues regarding the HMM analysis that were raised during the experiments were discussed. First, the issues related to the verbal-report requirement were discussed. Different pilots seemed to have different reactions to this verbal-report requirement. Pilots who habitually speak to themselves even during real flights may have had less problem with this requirement. Other pilots may feel some awkwardness at first. These pilots may need some practice to get used to this requirement.

Two different formats to collect the verbal reports—free mumbling (pilots generate the reports using their own words) and structured reporting (pilots report using predetermined short words representing different information from the overlapping instruments)—were tried in the first experiment (Chapter 3) and second experiment (Chapter 4), respectively. It was found that these formats had different advantages and disadvantages. The free-mumbling format provided the researchers context for the flight. This format could also be useful when the underlying HMM structure is initially unknown. However, this format also may cause higher cognitive workload for the pilots. The structured-reporting format, on the other hand, seemed to require less cognitive workload for the pilots. Yet, this format loses the context of the flight and cannot be used to uncover new HMM structures.

In the HMM analysis, the verbal reports were used to verify the HMM-estimated sequence of the attended tracking tasks. Therefore, the more verbal reports, the better. However, generating too many verbal reports may also distract pilots from the flight task. For the current experiment settings, the comfortable verbal reporting frequency seemed to be around once every five to ten seconds. How frequent verbal reporting would be sufficient or exactly when during the flight the verbal reports should be generated to fully verify the model parameters estimated by the Baum-Welch algorithm is not known at this point. The only way to ensure sufficient verification of the model parameter estimation each time may be to incorporate other measures, such as the pilots' control input data.

In addition to the issues related to the verbal-report requirement, the assumption of the serial attention process, on which the HMM analysis is based, was reexamined. The eye-movement

data collected indeed indicated patterns of frequent crosschecking among the related instruments, which made the fixation sequence data at least appear to be driven by an underlying serial process. Though it is not known whether the real attentional process is serial or not, as long as the actual eye-movement data indicated serial-like characteristics, the serial assumption would be adequate for the purposes of the HMM analysis. However, some signs of parallel attentional processing observed during the data analysis were also mentioned so as to remind the readers of the potential limitation of the serial assumption.

Finally, some suggestions for future work were presented, such as incorporating pilots' control input data in the HMM analysis, implementing unsupervised learning to automatically search for all candidate HMM structures, generating synthetic eye-movement data to assess prototype displays, and extending the HMM analysis to other application areas.

6.2 Contributions of the Thesis

The thesis makes a number of practical and theoretical contributions to various academic and professional communities. First of all, the thesis provides a new analysis tool for pilots' eye-movement data that is applicable in the human factors research domains, where tasks involve repetitive scanning of displays. This new analysis tool adds another dimension—the attention sequence level—to the analysis of pilots' eye-movement data, which currently relies largely on simple statistical analyses at the instrument-fixation-sequences level only. This is an important breakthrough because eye-movement tracking technology is rapidly progressing. As eye-tracking equipment is becoming more reliable and more affordable, more research laboratories are equipping their simulators with eye-tracking devices, and, so, the amount of eye-movement research is expected to grow rapidly. The thesis provides researchers with more options to choose from for eye-movement data analysis methods.

Second, the thesis introduces to the human factors research community the concept of the HMM, which already has been well developed in the pattern-recognition engineers' community, but is less well known with human factors researchers. The concept of the HMM is very flexible and is versatile enough to be able to be applied to other eye-movement research and beyond.

Third, the thesis also brings a new HMM application area to the pattern-recognition engineers'

community. As described in Chapter 2, speech-recognition and other pattern-recognition HMM applications look for a particular hidden sequence based on the left-right model. HMM analysis of pilots' eye-movement data, on the other hand, estimates recurring hidden-state transitions based on an ergodic model. The example demonstrated by the thesis may open up new possibilities for using the HMM technique.

Fourth, the thesis reveals the differences among the eye movements of pilots who have different levels of expertise, and thus provides the pilot community with important implications for improving pilot training and instruction programs. The fourth state, the attitude-monitoring task, found in the most experienced pilot in the first experiment, is consistent with a flight technique well known among experienced instrument pilots. Now, the existence of this flight technique has been scientifically demonstrated. In addition, the two-state HMM found in the least experienced pilot seemed insufficient for flying the approach. Therefore, HMM analysis may provide a tool to detect pilots who exhibit such a pattern of instrument scanning and encourage them to take re-currency training.

Fifth, the thesis also provides the display manufacturer and certifier community results showing the value of gradation marks in rotating pointers in HUD's airspeed-indicator (ASI) and altimeter formats. Even though removing the gradation marks has not yet been found to degrade pilots' flight performance, the experiments in this thesis revealed significant changes in pilots' fixation and attention processes compared to when the gradation marks were present. Therefore, there is a chance in a long operational run that such effect may prevail and result in a catastrophic accident. Thus, even though the effect is very subtle, to be safe, presenting gradation marks in the HUD ASI and altimeter symbology is recommended. Of course, this should not be the only display format issue to concern the display manufacturer and certifier community, and for those who desire, the thesis also provides research means to investigate other similar display-format issues.

Some contributions of this thesis are purely academic, such as the following:

Sixth, the thesis provides the psychologist community with actual pilots' eye-movement data that implies a serial, or at least a serial-like, underlying attentional process. The data are consistent with Wickens's multiple resource theory (1980); that is, all the tracking tasks required the visual channel to sample the instruments, and, thus, the channel was time-shared. However, if

consideration is extended to include the central processing in addition to the perception process, then the experiment data also indicated some signs of potential parallel or hierarchical processes; the analysis results indicated that the pilots occasionally reported a task not related to any instruments they were looking at, and that the pilots tended to generate verbal reports when they switched the tracking task. As such, HMM analysis can be used to further investigate information processing in human cognition.

Last of all, the observed trend that the pilots generated verbal reports when they switched tracking tasks may imply the existence of some neurophysiological function that governs attention management, such as the central executive of Baddeley's working memory model (Baddeley & Hitch, 1974), which is considered to be associated with the prefrontal cortex. Therefore, this thesis also provided new research incentives for the cognitive neuroscientist community.

To follow through on these contributions to these academic and professional communities, the thesis provides some suggestions for future work in section 5.4.

Appendix A:

MATLAB Codes

MATLAB codes used for HMM analysis are listed. The HMM_main is the main routine, and all other codes are subroutines called from HMM_main.

- HMM_main: The main routine. It takes a T×1 observation sequence vector, a T×1 time stamp vector corresponding to the observation sequence, and the initial model parameters (*A*, *B*, *C*) as inputs (the character "*C*" is used for "π" vector in the codes). The two input vectors can be read in from a file by user-defined codes. The model parameters are directly input from within the routine file. The sizes of A and B matrices automatically define the numbers of hidden states and observation symbols. At each iteration loop, the program outputs on the screen the previous and new values of *A* and *B* matrices and *log_P*, and asks the user whether to continue the iteration or not. Hit "y" to continue, or "n" or any other key to stop. When the iteration is finally stopped, the program shows the final estimation values of *A*, *B*, and *C*, and the history of the values of *log_P*. Also, it plots the observation vector and the estimated state sequence (state_seq).
- forward_scale: A subroutine for the Forward Procedure (scaled form).
- backward_scale: A subroutine for the Backward Procedure (scaled form).
- ParaEst: A subroutine for the Baum-Welch method.
- viterbi: A subroutine for the Viterbi algorithm (logarithm form)

HMM_main

```
clear

%%%%%%%%%%%%%%%%%%
Some additional codes to read in a Tx1 observation sequence vector (Observs) is
required here (e.g., by using “textread” or “load” command).  If one wants to plot the
estimation results, a Tx1 time vector (Observs_time) for the observation sequence
is also required.  Basic plotting codes are shown at the bottom of the code,
which can be customized in any way to suit one’s purpose.
%%%%%%%%%%%%%%%%%%

% Initial conditions for the morel parameters of a
% 3-state HMM consisting of Vertical, Horizontal, Airspeed tracking tasks
A0 = [   0.70     0.15     0.15;
         0.15     0.70     0.15;
         0.15     0.15     0.70];
B0 = [   0.00     0.37     0.15     0.27     0.00     0.16     0.05     0.00;
         0.00     0.32     0.00     0.42     0.26     0.00     0.00     0.00;
         0.39     0.17     0.00     0.00     0.00     0.00     0.05     0.39];
C0 = [   0.34;    0.33;    0.33];

A1 = A0; B1 = B0; C1 = C0;

M = size(A1,1);
log_P_hist = [];

iter = 'y';
while iter == 'y',

         %%%%% Forward & Backward Algorithm (Scaled) %%%%%
         [alpha_scaled, scale, log_P] = forward_scale(Observs, A1, B1, C1);
         [beta_scaled] = backward_scale(Observs, A1, B1, C1, scale);

         display('*** Initial Parameters ***')
         A1
         B1
         log_P

         %%%%% Parameter Estimation (Baum-Welch) %%%%%
         [A1, B1, C1] = ParaEst(A1, B1, alpha_scaled, beta_scaled, Observs);
         [alpha_scaled, scale, log_P] = forward_scale(Observs, A1, B1, C1);
         log_P_hist = [log_P_hist, log_P];

         %%%%% Viterbi Algorithm %%%%%
         [state_seq, P, del, psi] = viterbi(A1, B1, C1, Observs);

         display('*** Iteration Results ***')
         A1
         B1
         log_P

         iter = input('One more iteration? [y] or [n] = ','s');

end

C1
log_P_hist

% Plot
MB = size(B0,2);
figure('Position', [3 377 1020 312])
subplot(311); plot(Observs_time, Observs,'.-');
ylabel('Observation');
v = axis; axis([v(1), v(2), 1, MB]); grid

subplot(613); plot(Observs_time, state_seq,'.-');
ylabel('State');
v = axis; axis([v(1), v(2), 1, M]); grid
xlabel('Sec')
```

forward_scale

```
function [alpha_scaled, scale, log_P] = forward_scale(Observs, A1, B1, C1)

M = size(A1,1);           % # of hidden states
T = length(Observs);      % Max time index

% Initialization
for i=1:M
         alpha_loc(1,i) = C1(i) * B1(i,Observs(1));
end
scale(1) = inv(sum(alpha_loc));
alpha_scaled(1,:) = scale(1) * alpha_loc;

% Induction (scaled)
for t=2:T

         for j=1:M
                  alpha_loc(1,j) = alpha_scaled(t-1,:) * A1(:,j) * B1(j,Observs(t));
         end
         scale(t) = inv(sum(alpha_loc));
         alpha_scaled(t,:) = scale(t) * alpha_loc;

end

log_P = -sum(log(scale));
```

backward_scale

```
function [beta_scaled] = backward_scale(Observs, A1, B1, C1, scale)

M = size(A1,1);           % # of hidden states
T = length(Observs);      % Max time index

% Initialization
beta_scaled = scale(T) * ones(T,M);

% Induction (scaled)
for t=T-1:-1:1

         for i=1:M
                  beta_scaled(t,i) = A1(i,:) * ( B1(:,Observs(t+1)).*beta_scaled(t+1,:)' );
         end
         beta_scaled(t,:) = scale(t) * beta_scaled(t,:);

end
```

ParaEst

```
function [A2, B2, C2] = ParaEst(A1, B1, alpha, beta, Observs)

M = size(A1,1);           % # of hidden states
T = length(Observs);      % Max time index

A_num  = zeros(size(A1));
A_den  = zeros(size(A1,1),1);

B_num  = zeros(size(B1));

for t=1:T-1,
```

```
        A_num_t = zeros(size(A1));
        for i=1:M,
                for j=1:M,
                        A_num_t(i,j) = alpha(t,i)*A1(i,j)*B1(j,Observs(t+1))*beta(t+1,j);
                end
        end

        A_num = A_num + A_num_t;
        A_den = A_den + (sum(A_num_t'))';

        B_num(:,Observs(t)) = B_num(:,Observs(t)) + (sum(A_num_t'))';

        if t == 1,
                C2 = sum(A_num_t') / sum(sum(A_num_t'));
        end

end

for i=1:M,
        A2(i,:) = A_num(i,:) / A_den(i);
        B2(i,:) = B_num(i,:) / A_den(i);
end

C2 = C2';
```

viterbi

```
function [state_seq, P, del, psi] = viterbi(A, B, C, Observs)

M = size(A,1);            % # of hidden states
N = size(B,2);            % # of observations
T = length(Observs);      % Max time index

%%% Preprocessing %%%
% Avoiding log of zero
for i=1:M,
        for j=1:M,
                if A(i,j) == 0,
                        A(i,j) = 1E-3;
                end
        end
        for j=1:N,
                if B(i,j) == 0,
                        B(i,j) = 1E-3;
                end
        end
                if C(i) == 0,
                        C(i) = 1E-3;
                end
end

logA = log(A);
logB = log(B);
logC = log(C);

%%% Initialization %%%
del = logB(:,Observs(1))' + logC';
psi = zeros(1,M);

%%% Rercursion %%%

for t=2:T,

        for j=1:M,
                [y, max_i] = max( del(t-1,:)' + logA(:,j) );
                del_temp(j) = y + logB(j,Observs(t));
                psi_temp(j) = max_i;
```

```
        end
        del = [del; del_temp];
        psi = [psi; psi_temp];

end

%%% Termination %%%

[P, max_i] = max( del(T,:) );

state_seq = zeros(T,1);
state_seq(T) = max_i;

%%% Path Backtracking %%%

for t=T-1:-1:1,
        state_seq(t) = psi(t+1,state_seq(t+1));
end
```

Appendix B:

Participant Briefing Materials

The following materials used during participant briefing are listed:

- Informed Consent Form
- Authorization to Release Protected Health Information
- Briefing materials (used in the experiment in Chapter 3)
- Briefing materials (used in the experiment in Chapter 4)

The same Informed Consent Form and Authorization to Release Protected Health Information were used in both experiments in Chapter 3 and 4 (except that, in the Informed Consent Form, the flight simulator type used was noted Boeing 757, not Boeing 737). Note that the briefing materials of both experiments contain the portions of each experiment that were not included in the analysis of this thesis, i.e., “Display 2” in the experiment in Chapter 3, and the flight segment after passing 1000 ft in the experiment in Chapter 4.

Informed Consent Form

MIT COUHES Protocol 3021 (rev 1/17/2003)

Pilot Instrument Scanning And Attention Pattern Experiments

You have been asked to participate in a study investigating effects of flight instrument display formats on pilots' scanning, visual attention and flight performance during simulated instrument approaches in a fixed base, Boeing 737 research flight simulator. The study is conducted by scientists from MIT and the U.S. Department of Transportation. You are at least 18 years of age. You agree to bring my pilot logbook, and allow the investigators to copy flying background related data from it.

If you agree to participate, you will be asked to fly a series of up to 5 practice approaches and 15 experimental approaches in a given session, which will last up to 3 hours. There are no unusual physical risks involved in this ground based simulator study. Your flight technical error (e.g. airspeed, localizer and glideslope tracking errors) will be monitored during all trials. During the experiment, you will be asked to wear a head band equipped with a magnetic head tracker, infrared video camera, and IR light source to track my head and eye movements. If the headband becomes uncomfortable, you will inform the experimenter, and you can ask for a break at any time. During each approach you will also be asked to verbalize the displayed flight variables you are paying attention to. Your words will be recorded on a video tape along with the image of instrument display and the fixation points, but not with image of your face or any recognizable part of your body. The audio and video records will be used for purpose of post experiment analysis, will be kept locked in the investigator's office and will be destroyed at the end of the project. You will be also asked to answer questions in order to assess your attention mental workload after each approach.

Your name will be kept strictly confidential and not be associated with your data in any publications. You understand you may decline to answer questions, and you may ask questions about the study at any time, and that your participation is strictly voluntary. you are free to withdraw at anytime without penalty. You understand that the time you spent for this study will be compensated by Volpe PPA FA-3E2 A3285.

You understand that you will be given a $100 in gift certificate at Sporty's Shops for participation in the study. (The amount of the certificate will be prorated if you withdraw before completing the experiment.) If you are a Volpe federal employee, you will not be compensated.

In the unlikely event of physical injury resulting from participation in this research, I understand that medical treatment will be available from the M.I.T. Medical Department, including first aid emergency treatment and follow-up care as needed, and that my insurance carrier may be billed for the cost of such treatment. However, no compensation can be provided for medical care apart from the foregoing. I further understand that making such medical treatment available; or providing it, does not imply that such injury is the Investigator's fault. I also understand that by my participation in this study I am not waiving any of my legal rights.

(Page 1)

I understand that by your participation in this study I am not waiving any of your legal rights. (Further information may be obtained by calling the MIT's Insurance and Legal Affairs Office at 253-2822). I understand that I may also contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T. 253-6787, if I feel you have been treated unfairly as a subject.

I have been informed as to the procedures and purpose of this experiment and agree to participate. I have received a signed copy of this statement.

_____

Participant: Name (Print)

_____

Participant: Signature Date

(Page 2)

Authorization to Release Protected Health Information

AUTHORIZATION TO RELEASE PROTECTED HEALTH INFORMATION

The privacy law, Health Insurance Portability & Accountability Act (HIPAA), protects my individually identifiable health information. The privacy law requires me to sign an authorization in order for researchers to be able to use or disclose my protected health information for research purposes in the study entitled "Pilot Instrument Scanning and Attention Pattern Experiments".

I authorize Dr. Charles M. Oman and his research staff to use and disclose my protected health information for the purposes described below.

My protected health information that may be used and disclosed includes:

> Logbook and personal experiences relating to flying background, and eye movement, flight performance and workload data obtained during the experiment.

My protected health information will be used for a study investigating effects of flight instrument display formats on pilots' scanning, visual attention and flight performance during simulated instrument approaches.

The Researchers may use and share my health information with:

- The M.I.T. Committee on the Use of Humans as Experimental Subjects
- Federal, state and local government representatives, when required by law
- M.I.T. representatives if applicable
- Dr. Michael Zuschlag of the DOT Volpe Research Center

The researchers agree to protect my health information by using and disclosing it only as permitted by me in this Authorization and as directed by state and federal law

I do not have to sign this Authorization. If I decide not to sign the Authorization I will not be allowed to participate in the research study.

After signing the Authorization, I can change my mind and not let the researcher disclose or use my protected health information (revoke the Authorization). If I revoke the Authorization, I will send a written letter to Dr. Charles M. Oman, (Address: MIT Room 37-219, Cambridge MA 02139) to inform him of my decision. After I revoke the authorization the researchers may only use and disclose the protected health information already collected for this research study, but my protected health information may still be used and disclosed should I have an adverse event (a bad effect).

If I change my mind and withdraw the authorization, I will not be allowed to continue to participate in the study.

This Authorization does not have an expiration date.

(Page 1)

M.I.T. complies with HIPAA and its privacy requirements and all other laws that protect your privacy. We will protect your information according to these laws. Despite these protections, there is a possibility that your information could be used or disclosed by someone else to whom it is released in a way that it will no longer be protected.

I have read this authorization form and given the opportunity to ask questions. If I have questions later, I understand I can contact Dr. Charles M. Oman. I will be given a signed copy of this form.

I authorize the use of my identifiable health information as described in this form.

_____
Name of Subject

_____
Name of Legal Representative (if applicable)

_____ _____
Signature of Subject or Legal Representative Date

_____
Name of Investigator

_____ _____
Signature of Investigator Date
(must be same as subject's)

_____
Name of witness

_____ _____
Signature of witness (if required by COUHES) **Date**

(Page 2)

Briefing Materials (Experiment in Chapter 3)

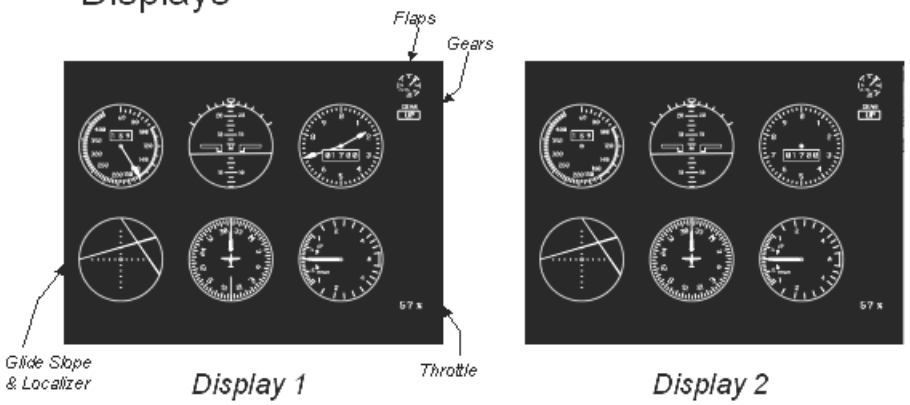
ILS Approach Simulation

Subject Briefing

1

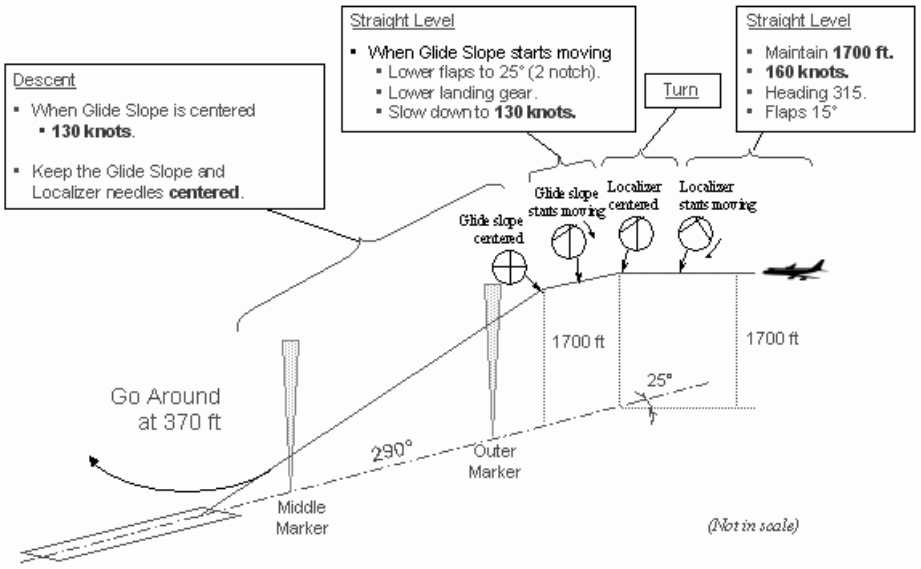
Simulator

- Flight Dynamics
  - B757-200
- Displays



2

Approach Scenario



3

Specific to Simulator Experiment

- Eye-movement tracking device
- "Think aloud"
- Bedford Workload scale

- Informed consent

4

Outline of Simulation

1. Practice
2. Practice the ILS approach of this experiment
3. Practice the approaches with eye-tracker and "Think aloud"
4. 6 data-taking runs

Run #	1	2	3	4	5	6
Display	d2	d1	d2	d1	d2	d1
				^ break?		

5. Comments

5

Briefing Materials (Experiment in Chapter 4)

HUD Simulator Experiment

Subject Briefing

Experimenter : Miwa Hayashi

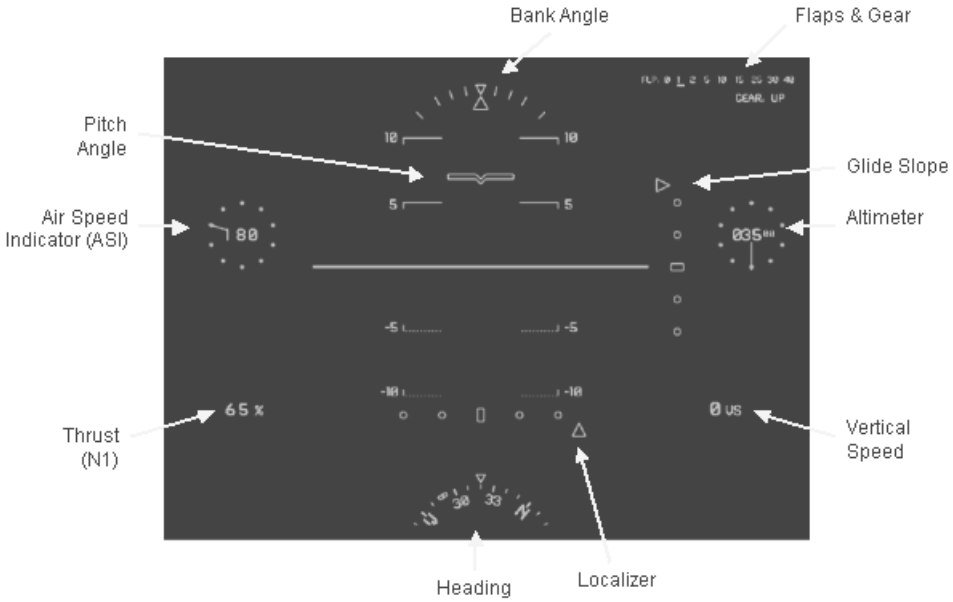
1

Flight Simulator

- Flight Dynamics
  - Boeing 737-400
- Screen Projection
  - Head Up Display symbols
  - Out-the-window view
    - **Mostly showing clouds except the last segment.**
- Pilot Controls
  - Yoke
    - **Trim switch, flap switch, gear button**
  - Throttle levers

2

Head Up Display

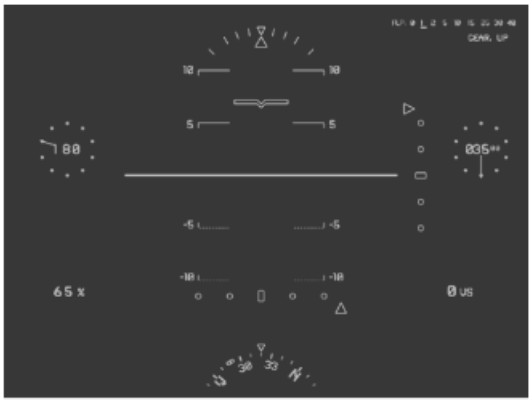


3

ASI & Altimeter Formats

3 ASI & Altimeter formats to be examined. →

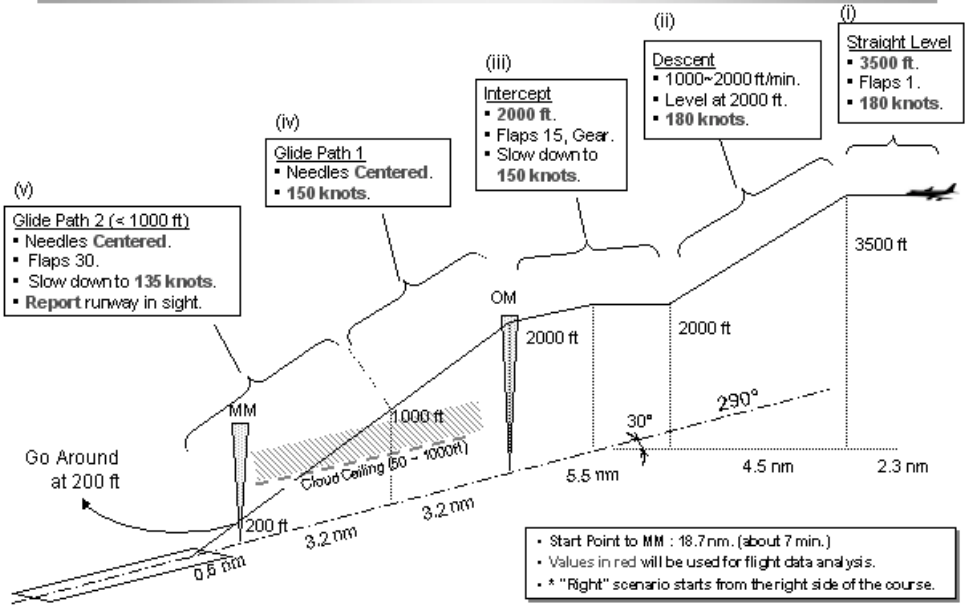
ASI →



← Altimeter

4

ILS Approach Scenario – Left*



5

Data Collection

- Informed consent form
- Subject data sheet (name, e-mail, pilot certificates & ratings, aircrafts flown, total flight time, instrument time, experience of HUD)

- Eye-movement data
- Flight data
- Verbal reports - pitch & bank
- Workload score
- Subjective preference & comments

6

Outline of Simulation

1. Practice the approach
2. Practice the approach with eye-tracker and verbal reporting
3. 9 data-taking runs

Apprch #	1	2	3	4	5	6	7	8	9
Display	D1	D2	D3	D2	D3	D1	D3	D1	D2
Scenario	L	L	R	R	L	L	R	R	L

4. Subjective Preference & Comments

7

Appendix C:

HMM Parameters

The matrices, *A0*, *B0*, and *C0*, are the initial conditions of the HMM model parameters used in the HMM_main program (Appendix A). The matrices, *A*, *B*, and *C*, are the estimation results from the same program. The vector, *log_P_hist*, records all the *logP* values computed during the iteration process. The length of this vector also tells how many iteration loops were performed. For the experiment in Chapter 3, the model parameter matrices for all the pilots and for all the segments are listed here. For the experiment in Chapter 4, due to the large number of the parameters, only those of Pilot 3 (who showed both a 3-state and a 4-state HMM in different segments or with different indicator formats) were listed.

Experiment in Chapter 3

```
*** Pilot 1 ***
Pilot 1, segment (i): 3-state HMM
A0 = [   0.80     0.10     0.10;
         0.15     0.70     0.15;
         0.15     0.15     0.70];
B0 = [   0.00     0.32     0.18     0.27     0.00     0.18     0.05     0.00;
         0.00     0.32     0.00     0.32     0.36     0.00     0.00     0.00;
         0.36     0.22     0.00     0.00     0.00     0.00     0.05     0.37];
C0 = [   0.34;    0.33;    0.33];

A = [    0.781528 0.137924 0.080548
         0.314858 0.629403 0.055739
         0.310023 0.231385 0.458593]
B = [    0.000000 0.483746 0.278080 0.041200 0.000000 0.196973 0.000000 0.000000
         0.000000 0.340820 0.000000 0.152013 0.507167 0.000000 0.000000 0.000000
         0.317890 0.335322 0.000000 0.000000 0.000000 0.000000 0.000000 0.346789]
C = [    0.703458 0.191439 0.105102]

log_P_hist =  -460.9989 -458.4280


Pilot 1, segment (i): Alternative 3-state HMM (Power indicator included in vertical task)
A0 = [   0.80     0.10     0.10;
         0.15     0.70     0.15;
         0.15     0.15     0.70];
B0 = [   0.00     0.32     0.13     0.27     0.00     0.13     0.05     0.10;
         0.00     0.32     0.00     0.42     0.26     0.00     0.00     0.00;
         0.73     0.22     0.00     0.00     0.00     0.00     0.05     0.00];
C0 = [   0.34;    0.33;    0.33];

A = [    0.820673 0.134624 0.044703
         0.316468 0.631928 0.051604
         0.330717 0.279139 0.390145]
B = [    0.000000 0.453447 0.256113 0.044998 0.000000 0.181413 0.000000 0.064028
         0.000000 0.346610 0.000000 0.138500 0.514890 0.000000 0.000000 0.000000
         0.522080 0.477920 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000]
C = [    0.648781 0.189466 0.161753]
```

```
log_P_hist =  -462.8109 -461.1551
```

Pilot 1, segment (ii): 3-state HMM

```
A0 = [   0.80     0.10     0.10;
         0.15     0.70     0.15;
         0.15     0.15     0.70];
B0 = [   0.00     0.32     0.18     0.27     0.00     0.18     0.05     0.00;
         0.00     0.32     0.00     0.32     0.36     0.00     0.00     0.00;
         0.36     0.22     0.00     0.00     0.00     0.00     0.05     0.37];
C0 = [   0.34;    0.33;    0.33];

A = [    0.683760 0.257197 0.059043
         0.123111 0.856892 0.019998
         0.163038 0.391269 0.445692]
B = [    0.000000 0.096173 0.293759 0.177369 0.000000 0.431999 0.000700 0.000000
         0.000000 0.199820 0.000000 0.235270 0.564910 0.000000 0.000000 0.000000
         0.046020 0.313434 0.000000 0.000000 0.000000 0.000000 0.042290 0.598256]
C = [    1.000000 0.000000 0.000000]

log_P_hist =  -602.1820 -598.3334
```

Pilot 1, segment (ii): 2-state HMM

```
A0 = [   0.80     0.20;
         0.30     0.70];
B0 = [   0.05     0.32     0.13     0.27     0.00     0.13     0.05     0.05;
         0.00     0.32     0.00     0.32     0.36     0.00     0.00     0.00];
C0 = [   0.50;    0.50];

A = [    0.714544 0.285456
         0.158672 0.841328]
B = [    0.006914 0.140137 0.235074 0.175384 0.000000 0.345696 0.006914 0.089881
         0.000000 0.196355 0.000000 0.223070 0.580575 0.000000 0.000000 0.000000]
C = [    1.000000 0.000000]

log_P_hist =  -612.9896 -609.7237
```

Pilot 1, segment (iii): 3-state HMM

```
A0 = [   0.80     0.10     0.10;
         0.15     0.70     0.15;
         0.15     0.15     0.70];
B0 = [   0.00     0.32     0.18     0.27     0.00     0.18     0.05     0.00;
         0.00     0.27     0.00     0.27     0.46     0.00     0.00     0.00;
         0.43     0.12     0.00     0.00     0.00     0.00     0.01     0.44];
C0 = [   0.34;    0.33;    0.33];

A = [    0.770032 0.192663 0.037305
         0.229665 0.754401 0.015934
         0.258160 0.291849 0.449992]
B = [    0.000000 0.135085 0.175119 0.344774 0.000000 0.308308 0.036714 0.000000
         0.000000 0.144373 0.000000 0.287444 0.568183 0.000000 0.000000 0.000000
         0.342083 0.181250 0.000000 0.000000 0.000000 0.000000 0.003015 0.473653]
C = [    1.000000 0.000000 0.000000]

log_P_hist =  1.0e+003 *   -1.3215   -1.3182
```

Pilot 1, segment (iii): 2-state HMM

```
A0 = [   0.80     0.20;
         0.30     0.70];
B0 = [   0.05     0.32     0.13     0.27     0.00     0.13     0.05     0.05;
         0.00     0.27     0.00     0.27     0.46     0.00     0.00     0.00];
C0 = [   0.50;    0.50];


A = [    0.790718 0.209282
         0.261125 0.738875]
B = [    0.029029 0.138098 0.158542 0.321520 0.000000 0.279123 0.033495 0.040194
         0.000000 0.145617 0.000000 0.279236 0.575148 0.000000 0.000000 0.000000]
```

```
C = [    1.000000 0.000000]

log_P_hist =  1.0e+003 *   -1.3367   -1.3360
```

***** Pilot 2 *****

Pilot 2, segment (i): 3-state HMM

```
A0 = [   0.80     0.10     0.10;
         0.15     0.70     0.15;
         0.15     0.15     0.70];
B0 = [   0.00     0.32     0.18     0.27     0.00     0.18     0.05     0.00;
         0.00     0.32     0.00     0.32     0.36     0.00     0.00     0.00;
         0.26     0.42     0.00     0.00     0.00     0.00     0.05     0.27];
C0 = [   0.34;    0.33;    0.33];

A = [    0.777005 0.061060 0.161935
         0.237894 0.655147 0.106958
         0.257428 0.092838 0.649734]
B = [    0.000000 0.479609 0.301677 0.001137 0.000000 0.116557 0.101019 0.000000
         0.000000 0.236095 0.000000 0.230961 0.532944 0.000000 0.000000 0.000000
         0.367998 0.367682 0.000000 0.000000 0.000000 0.000000 0.003160 0.261160]
C = [    0.000000 0.000000 1.000000]

log_P_hist =  -460.4763 -454.8730 -452.6376
```

Pilot 2, segment (ii): 3-state HMM

```
A0 = [   0.80     0.10     0.10;
         0.15     0.70     0.15;
         0.15     0.15     0.70];
B0 = [   0.00     0.32     0.18     0.27     0.00     0.18     0.05     0.00;
         0.00     0.40     0.00     0.32     0.28     0.00     0.00     0.00;
         0.36     0.22     0.00     0.00     0.00     0.00     0.05     0.37];
C0 = [   0.34;    0.33;    0.33];

A = [    0.776819 0.130021 0.093160
         0.219141 0.735561 0.045298
         0.179332 0.199551 0.621117]
B = [    0.000000 0.482673 0.262659 0.074167 0.000000 0.163167 0.017335 0.000000
         0.000000 0.407959 0.000000 0.202137 0.389904 0.000000 0.000000 0.000000
         0.476572 0.348955 0.000000 0.000000 0.000000 0.000000 0.055331 0.119143]
C = [    0.074955 0.925045 0.000000]

log_P_hist =  -834.9901 -830.7192
```

Pilot 2, segment (iii): 3-state HMM

```
A0 = [   0.80     0.10     0.10;
         0.15     0.70     0.15;
         0.15     0.15     0.70]
B0 = [   0.00     0.32     0.08     0.47     0.00     0.08     0.05     0.00;
         0.00     0.32     0.00     0.22     0.46     0.00     0.00     0.00;
         0.36     0.22     0.00     0.00     0.00     0.00     0.05     0.37]
C0 = [   0.34;    0.33;    0.33]

A = [    0.753674 0.135034 0.111293
         0.262695 0.668023 0.069282
         0.228979 0.148283 0.622738]
B = [    0.000000 0.348938 0.279873 0.142252 0.000000 0.186582 0.042355 0.000000
         0.000000 0.349048 0.000000 0.193041 0.457911 0.000000 0.000000 0.000000
         0.573988 0.209135 0.000000 0.000000 0.000000 0.000000 0.047624 0.169253]
C = [    0.024434 0.954715 0.020851]

log_P_hist =  1.0e+003 *   -1.2105   -1.2047
```

***** Pilot 3 *****

Pilot 3, segment (i): 3-state HMM

```
A0 = [   0.80     0.10     0.10;
         0.15     0.70     0.15;
         0.15     0.15     0.70];
```

```
B0 = [   0.00     0.32     0.18     0.27     0.00     0.18     0.05     0.00;
         0.00     0.32     0.00     0.32     0.36     0.00     0.00     0.00;
         0.34     0.27     0.00     0.00     0.00     0.00     0.05     0.34];
C0 = [   0.34;    0.33;    0.33];

A = [    0.599194 0.061987 0.338819
         0.338972 0.514764 0.146265
         0.377992 0.163618 0.458390]
B = [    0.000000 0.483149 0.376774 0.032428 0.000000 0.107650 0.000000 0.000000
         0.000000 0.386560 0.000000 0.229809 0.383631 0.000000 0.000000 0.000000
         0.393498 0.446975 0.000000 0.000000 0.000000 0.000000 0.000000 0.159526]
C = [    1.000000 0.000000 0.000000]

log_P_hist =  -434.3645 -428.9493 -426.5116
```

Pilot 3, segment (ii): 3-state HMM

```
A0 = [   0.70     0.15     0.15;
         0.10     0.80     0.10;
         0.15     0.15     0.70]
B0 = [   0.00     0.27     0.18     0.32     0.00     0.18     0.05     0.00;
         0.00     0.32     0.00     0.32     0.36     0.00     0.00     0.00;
         0.37     0.22     0.00     0.00     0.00     0.00     0.05     0.36]
C0 = [   0.34;    0.33;    0.33]

A = [    0.636592 0.121591 0.241817
         0.122486 0.773076 0.104438
         0.294434 0.176802 0.528764]
B = [    0.000000 0.536844 0.315151 0.077615 0.000000 0.055026 0.015364 0.000000
         0.000000 0.516241 0.000000 0.150991 0.332768 0.000000 0.000000 0.000000
         0.386357 0.483141 0.000000 0.000000 0.000000 0.000000 0.006082 0.124420]
C = [    0.223538 0.776462 0.000000]

log_P_hist =  -858.1734 -853.0620
```

Pilot 3, segment (iii): 3-state HMM

```
A0 = [   0.80     0.10     0.10;
         0.15     0.70     0.15;
         0.15     0.15     0.70]
B0 = [   0.00     0.37     0.08     0.42     0.00     0.08     0.05     0.00;
         0.00     0.22     0.00     0.27     0.51     0.00     0.00     0.00;
         0.37     0.22     0.00     0.00     0.00     0.00     0.05     0.36]
C0 = [   0.34;    0.33;    0.33]

A = [    0.682069 0.104936 0.212995
         0.126548 0.716956 0.156496
         0.461650 0.109987 0.428363]
B = [    0.000000 0.523826 0.188699 0.209811 0.000000 0.076499 0.001165 0.000000
         0.000000 0.365526 0.000000 0.201840 0.432634 0.000000 0.000000 0.000000
         0.501621 0.433678 0.000000 0.000000 0.000000 0.000000 0.002595 0.062106]
C = [    0.089034 0.050477 0.860488]

log_P_hist =  1.0e+003 *   -1.2775   -1.2651
```

***** Pilot 4 *****

Pilot 4, segment (i): 3-state HMM

```
A0 = [   0.70     0.15     0.15;
         0.15     0.70     0.15;
         0.15     0.15     0.70]
B0 = [   0.00     0.37     0.15     0.27     0.00     0.16     0.05     0.00;
         0.00     0.32     0.00     0.42     0.26     0.00     0.00     0.00;
         0.39     0.17     0.00     0.00     0.00     0.00     0.05     0.39]
C0 = [   0.34;    0.33;    0.33]

A = [    0.737529 0.167275 0.095196
         0.329661 0.629410 0.040929
         0.236739 0.150682 0.612580]
B = [    0.000000 0.587102 0.344363 0.012313 0.000000 0.056223 0.000000 0.000000
         0.000000 0.517848 0.000000 0.103974 0.378178 0.000000 0.000000 0.000000
```

```
        0.391880 0.239291 0.000000 0.000000 0.000000 0.000000 0.000000 0.368828]
C = [   0.926343 0.059556 0.014101]

log_P_hist =  -372.5250 -370.5004
```

Pilot 4, segment (ii): 3-state HMM

```
A0 = [  0.70     0.15     0.15;
        0.15     0.70     0.15;
        0.15     0.15     0.70]
B0 = [  0.00     0.32     0.15     0.32     0.00     0.16     0.05     0.00;
        0.00     0.32     0.00     0.42     0.26     0.00     0.00     0.00;
        0.39     0.17     0.00     0.00     0.00     0.00     0.05     0.39]
C0 = [  0.34;    0.33;    0.33]

A = [   0.752941 0.168784 0.078276
        0.262141 0.699770 0.038089
        0.307471 0.204662 0.487868]
B = [   00.000000 0.579324 0.304336 0.072424 0.000000 0.023156 0.020760 0.000000
        0.000000 0.417856 0.000000 0.200743 0.381401 0.000000 0.000000 0.000000
        0.387411 0.487903 0.000000 0.000000 0.000000 0.000000 0.027833 0.096853]
C = [   0.787428 0.137969 0.074603]

log_P_hist =  -813.7323 -809.5535
```

Pilot 4, segment (iii): 3-state HMM

```
A0 = [  0.70     0.15     0.15;
        0.15     0.70     0.15;
        0.15     0.15     0.70];
B0 = [  0.00     0.32     0.15     0.32     0.00     0.16     0.05     0.00;
        0.00     0.32     0.00     0.42     0.26     0.00     0.00     0.00;
        0.39     0.17     0.00     0.00     0.00     0.00     0.05     0.39];
C0 = [  0.34;    0.33;    0.33];

A = [   0.624823 0.283601 0.091576
        0.125255 0.832025 0.042720
        0.156970 0.332681 0.510348]
B = [   0.000000 0.673942 0.084073 0.115877 0.000000 0.126109 0.000000 0.000000
        0.000000 0.565080 0.000000 0.143491 0.291429 0.000000 0.000000 0.000000
        0.380780 0.501047 0.000000 0.000000 0.000000 0.000000 0.000000 0.118173]
C = [   0.346500 0.561662 0.091838]

log_P_hist =  -916.6211 -915.2449
```

Pilot 4, segment (iii): 4-state HMM

```
A0 = [  0.70     0.10     0.10   0.10;
        0.10     0.70     0.10   0.10;
        0.10     0.10     0.70   0.10;
        0.05     0.20     0.05   0.70];
B0 = [  0.00     0.22     0.26     0.22     0.00     0.26     0.05     0.00;
        0.00     0.22     0.00     0.22     0.56     0.00     0.00     0.00;
        0.36     0.17     0.00     0.00     0.00     0.00     0.05     0.42;
        0.00     0.75     0.00     0.25     0.00     0.00     0.00     0.00];
C0 = [  0.25;    0.25;    0.25; 0.25];

A = [   0.470391 0.193769 0.116333 0.219507
        0.112559 0.664348 0.032035 0.191058
        0.071911 0.211831 0.519828 0.196431
        0.072517 0.333111 0.041921 0.552451]
B = [   0.000000 0.560946 0.144047 0.078936 0.000000 0.216071 0.000000 0.000000
        0.000000 0.457450 0.000000 0.127977 0.414573 0.000000 0.000000 0.000000
        0.425621 0.442290 0.000000 0.000000 0.000000 0.000000 0.000000 0.132089
        0.000000 0.830499 0.000000 0.169501 0.000000 0.000000 0.000000 0.000000]
C = [   0.011794 0.014489 0.004910 0.968807]

log_P_hist =  -936.6816 -925.6608 -920.5743
```

Experiment in Chapter 4

***** Pilot 3 *****

Pilot 3, segment (i), PGD: 3-state HMM

```
A0 = [   0.700000 0.150000 0.150000
         0.150000 0.700000 0.150000
         0.150000 0.150000 0.700000]
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000]
C0 = [   0.340000 0.330000 0.330000]

A = [    0.662176 0.054339 0.283486
         0.210244 0.523574 0.266182
         0.208935 0.079432 0.711634
B = [    0.000000 0.000000 0.297221 0.000000 0.000000 0.021175 0.352920 0.317628 0.011055
         0.000000 0.000000 0.570759 0.150234 0.279007 0.000000 0.000000 0.000000 0.000000
         0.377604 0.197011 0.423012 0.000000 0.000000 0.000000 0.000000 0.000000 0.002374
C = [    0.053767 0.096677 0.849556

log_P_hist =  -597.056181 -595.072930
```

Pilot 3, segment (i), PD: 3-state HMM

```
A0 = [   0.700000 0.150000 0.150000
         0.150000 0.700000 0.150000
         0.150000 0.150000 0.700000]
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000]
C0 = [   0.340000 0.330000 0.330000]

A = [    0.714776 0.086701 0.198524
         0.152049 0.521497 0.326454
         0.167054 0.085468 0.747478]
B = [    0.000000 0.000000 0.265179 0.000000 0.000000 0.019774 0.369110 0.342745 0.003192
         0.000000 0.000000 0.509315 0.094971 0.395713 0.000000 0.000000 0.000000 0.000000
         0.413576 0.231405 0.352480 0.000000 0.000000 0.000000 0.000000 0.000000 0.002539
C = [    0.000000 0.000000 1.000000

log_P_hist =  -678.263349 -677.087724
```

Pilot 3, segment (i), D: 3-state HMM

```
A0 = [   0.700000 0.150000 0.150000
         0.150000 0.700000 0.150000
         0.150000 0.150000 0.700000
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000
C0 = [   0.340000 0.330000 0.330000

A = [    0.677510 0.052448 0.270042
         0.210884 0.504727 0.284389
         0.207010 0.074065 0.718925]
B = [    0.000000 0.000000 0.300378 0.000000 0.000000 0.019325 0.373616 0.302758 0.003923
         0.000000 0.000000 0.517076 0.131706 0.351217 0.000000 0.000000 0.000000 0.000000
         0.422235 0.257461 0.318291 0.000000 0.000000 0.000000 0.000000 0.000000 0.002013]
C = [    0.573574 0.288147 0.138278]

log_P_hist =  -650.837391 -650.275898
```

Pilot 3, segment (ii), PGD: 3-state HMM

```
A0 = [   0.700000 0.150000 0.150000
         0.150000 0.700000 0.150000
         0.150000 0.150000 0.700000]
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000]
C0 = [   0.340000 0.330000 0.330000]
```

```
A = [    0.615797 0.133423 0.250779
         0.172827 0.483152 0.344021
         0.242091 0.090213 0.667696]
B = [    0.000000 0.000000 0.353465 0.000000 0.000000 0.019396 0.368525 0.258614 0.000000
         0.000000 0.000000 0.499756 0.197465 0.302779 0.000000 0.000000 0.000000 0.000000
         0.476029 0.081897 0.442074 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000]
C = [    0.000000 1.000000 0.000000]

log_P_hist =  -671.951056 -670.176382
```

Pilot 3, segment (ii), PD: 3-state HMM

```
A0 = [   0.700000 0.150000 0.150000
         0.150000 0.700000 0.150000
         0.150000 0.150000 0.700000]
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000]
C0 = [   0.340000 0.330000 0.330000]

A = [    0.592433 0.123754 0.283814
         0.136223 0.573387 0.290390
         0.223795 0.124628 0.651577]
B = [    0.000000 0.000000 0.371364 0.000000 0.000000 0.000000 0.368152 0.250344 0.010140
         0.000000 0.000000 0.525284 0.147689 0.327027 0.000000 0.000000 0.000000 0.000000
         0.402313 0.094047 0.500385 0.000000 0.000000 0.000000 0.000000 0.000000 0.003254]
C = [    0.000000 0.000000 1.000000]

log_P_hist =  -649.980511 -647.560887
```

Pilot 3, segment (ii), D: 3-state HMM

```
A0 = [   0.700000 0.150000 0.150000
         0.150000 0.700000 0.150000
         0.150000 0.150000 0.700000]
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000]
C0 = [   0.340000 0.330000 0.330000]

A = [    0.657681 0.110660 0.231658
         0.337061 0.404718 0.258221
         0.136541 0.099813 0.763645]
B = [    0.000000 0.000000 0.302278 0.000000 0.000000 0.020455 0.436378 0.231826 0.009063
         0.000000 0.000000 0.395736 0.174919 0.429346 0.000000 0.000000 0.000000 0.000000
         0.461281 0.136502 0.394353 0.000000 0.000000 0.000000 0.000000 0.000000 0.007864]
C = [    0.000000 0.000000 1.000000]

log_P_hist =  -691.678567 -689.975851
```

Pilot 3, segment (iv), PGD: 4-state HMM

```
A0 = [   0.700000 0.100000 0.100000 0.100000
         0.100000 0.700000 0.100000 0.100000
         0.100000 0.100000 0.700000 0.100000
         0.100000 0.100000 0.100000 0.700000]
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000
         0.000000 0.000000 0.800000 0.100000 0.050000 0.050000 0.000000 0.000000 0.000000]
C0 = [   0.250000 0.250000 0.250000 0.250000]

A = [    0.564352 0.106754 0.261297 0.067597
         0.109672 0.748947 0.045686 0.095695
         0.216074 0.181302 0.524237 0.078387
         0.123694 0.191315 0.175227 0.509764]
B = [    0.000000 0.000000 0.306913 0.000000 0.000000 0.184670 0.266060 0.235826 0.006531
         0.000000 0.000000 0.257247 0.353705 0.389048 0.000000 0.000000 0.000000 0.000000
         0.455539 0.158448 0.379940 0.000000 0.000000 0.000000 0.000000 0.000000 0.006074
         0.000000 0.000000 0.691591 0.131876 0.069210 0.107324 0.000000 0.000000 0.000000]
```

```
C = [    0.000000 0.932009 0.000000 0.067991]

log_P_hist =  -1155.914774 -1149.768710
```

Pilot 3, segment (iv), PD: 4-state HMM

```
A0 = [   0.700000 0.100000 0.100000 0.100000
         0.100000 0.700000 0.100000 0.100000
         0.100000 0.100000 0.700000 0.100000
         0.100000 0.100000 0.100000 0.700000]
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000
         0.000000 0.000000 0.800000 0.100000 0.050000 0.050000 0.000000 0.000000 0.000000]
C0 = [   0.250000 0.250000 0.250000 0.250000]

A = [    0.627169 0.084804 0.244173 0.043854
         0.098277 0.723418 0.067782 0.110524
         0.231225 0.216509 0.470679 0.081586
         0.196235 0.128596 0.178371 0.496798]
B = [    0.000000 0.000000 0.232106 0.000000 0.000000 0.235103 0.323214 0.208846 0.000731
         0.000000 0.000000 0.199053 0.428088 0.372859 0.000000 0.000000 0.000000 0.000000
         0.477979 0.185881 0.323839 0.000000 0.000000 0.000000 0.000000 0.000000 0.012301
         0.000000 0.000000 0.727588 0.113351 0.052900 0.106161 0.000000 0.000000 0.000000]
C = [    0.000000 0.984731 0.000000 0.015269]
log_P_hist =  -1190.548092 -1183.471128
```

Pilot 3, segment (iv), D: 4-state HMM

```
A0 = [   0.700000 0.100000 0.100000 0.100000
         0.100000 0.700000 0.100000 0.100000
         0.100000 0.100000 0.700000 0.100000
         0.100000 0.100000 0.100000 0.700000]
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000
         0.000000 0.000000 0.800000 0.100000 0.050000 0.050000 0.000000 0.000000 0.000000
C0 = [   0.250000 0.250000 0.250000 0.250000

A = [    0.592578 0.065025 0.290399 0.051998
         0.083451 0.663370 0.138553 0.114626
         0.171128 0.239716 0.516010 0.073147
         0.120204 0.124201 0.206080 0.549515]
B = [    0.000000 0.000000 0.194913 0.000000 0.000000 0.181588 0.346100 0.269189 0.008209
         0.000000 0.000000 0.236828 0.391580 0.371592 0.000000 0.000000 0.000000 0.000000
         0.478245 0.220729 0.283045 0.000000 0.000000 0.000000 0.000000 0.000000 0.017981
         0.000000 0.000000 0.677561 0.144726 0.070145 0.107569 0.000000 0.000000 0.000000]
C = [    0.000000 0.379747 0.000000 0.620253]

log_P_hist =  -1012.491289 -1007.558588
```

Pilot 3, segment (v), PGD: 4-state HMM

```
A0 = [   0.700000 0.100000 0.100000 0.100000
         0.100000 0.700000 0.100000 0.100000
         0.100000 0.100000 0.700000 0.100000
         0.100000 0.100000 0.100000 0.700000]
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000
         0.000000 0.000000 0.800000 0.100000 0.050000 0.050000 0.000000 0.000000 0.000000]
C0 = [   0.250000 0.250000 0.250000 0.250000]

A = [    0.660911 0.076962 0.196454 0.065674
         0.086470 0.698801 0.108051 0.106678
         0.251231 0.188935 0.450050 0.109784
         0.137106 0.185769 0.169499 0.507625]
B = [    0.000000 0.000000 0.266975 0.000000 0.000000 0.325010 0.181517 0.215131 0.011366
         0.000000 0.000000 0.307363 0.353644 0.338993 0.000000 0.000000 0.000000 0.000000
         0.373103 0.177224 0.446788 0.000000 0.000000 0.000000 0.000000 0.000000 0.002885
         0.000000 0.000000 0.770352 0.115221 0.067093 0.047334 0.000000 0.000000 0.000000]
```

```
C = [    0.000000 0.957863 0.000000 0.042137]

log_P_hist =  -837.316038 -831.420022
```

Pilot 3, segment (v), PD: 3-state HMM

```
A0 = [   0.700000 0.150000 0.150000
         0.150000 0.700000 0.150000
         0.150000 0.150000 0.700000]
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000]
C0 =     0.250000 0.250000 0.250000]

A = [    0.666216 0.150177 0.183607
         0.142234 0.663176 0.194590
         0.250523 0.165692 0.583785]
B = [    0.000000 0.000000 0.291266 0.000000 0.000000 0.327290 0.143548 0.229677 0.008218
         0.000000 0.000000 0.306312 0.386864 0.306823 0.000000 0.000000 0.000000 0.000000
         0.408388 0.190581 0.397160 0.000000 0.000000 0.000000 0.000000 0.000000 0.003872]
C = [    0.000000 0.000000 1.000000]

log_P_hist =  -859.011893 -857.940029
```

Pilot 3, segment (v), D: 4-state HMM

```
A0 = [   0.700000 0.100000 0.100000 0.100000
         0.100000 0.700000 0.100000 0.100000
         0.100000 0.100000 0.700000 0.100000
         0.100000 0.100000 0.100000 0.700000]
B0 = [   0.000000 0.000000 0.320000 0.000000 0.000000 0.170000 0.170000 0.170000 0.170000
         0.000000 0.000000 0.320000 0.290000 0.390000 0.000000 0.000000 0.000000 0.000000
         0.230000 0.230000 0.320000 0.000000 0.000000 0.000000 0.000000 0.000000 0.220000
         0.000000 0.000000 0.800000 0.100000 0.050000 0.050000 0.000000 0.000000 0.000000]
C0 = [   0.250000 0.250000 0.250000 0.250000]

A = [    0.676568 0.066415 0.202640 0.054377
         0.188194 0.623023 0.086097 0.102686
         0.160249 0.171168 0.578400 0.090183
         0.266001 0.095176 0.190626 0.448198]
B = [    0.000000 0.000000 0.263084 0.000000 0.000000 0.322280 0.139902 0.220614 0.054120
         0.000000 0.000000 0.212272 0.432522 0.355205 0.000000 0.000000 0.000000 0.000000
         0.454105 0.209587 0.322741 0.000000 0.000000 0.000000 0.000000 0.000000 0.013567
         0.000000 0.000000 0.685009 0.142436 0.042865 0.129689 0.000000 0.000000 0.000000]
C = [    0.000000 0.000000 1.000000 0.000000]

log_P_hist =  -937.377627 -932.952313
```

References

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control, AC-19*(6), 716-723.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. A. Bower (Ed.), *Recent Advances in Learning and Motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Baum, L. E., & Petrie, T. (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics, 37*(6), 1554-1563.

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual Scanning and Pilot Expertise: The Role of Attentional Flexibility and Mental Model Development. *Aviation, Space, and Environmental Medicine, 68*(7), 569-579.

Bishop, C. M. (1996). *Neural Networks for Pattern Recognition*. Oxford, UK: Clarendon Press.

CAB. (1953). *Airplane Airworthiness; Transport Categories Instrument Installations* (Civil Air Regulations, Part 4b). Washington, D.C.: US Civil Aeronautics Board.

Carbonell, J. R. (1966). Queueing model of many-instrument visual sampling. *IEEE Transactions on Human Factors in Electronics, HFE-7*(4), 157-164.

Carbonell, J. R., Ward, J. L., & Senders, J. W. (1968). A Queueing Model of Visual Sampling Experimental Validation. *IEEE Transactions on Man-Machine Systems, MMS-9*(3), 82-87.

Clement, W. F., Jex, H. R., & Graham, D. (1968). Manual control-display theory applied to instrument landings of jet transport. *IEEE Transactions on Man-Machine Systems, MMS-9*(4), 93-110.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B, 39*(1), 1-38.

Doane, S. M., & Sohn, Y. W. (2000). ADAPT: A predictive cognitive model of user visual attention and action planning. *User Modelling and User-Adapted Interaction, 10*(1), 1-45.

Ellis, S. R., & Stark, L. (1986). Statistical Dependency in Visual Scanning. *Human Factors, 28*(4), 421-438.

Ercoline, W. R., & Gillingham, K. K. (1990). *Effects of variations in head-up display airspeed and altitude representations on basic flight performance.* Paper presented at the Proceedings of the Human Factors Society 34th Annual Meeting - Orlando '90, Oct 8-12 1990, Orlando, FL, USA.

FAA. (1970). *Airworthiness Standards: Transport Category Airplanes* (Federal Aviation Regulations, Part 25). Washington, D.C.: US Department of Transportation, Federal Aviation Administration.

FAA. (2001). *Instrument Flying Handbook* (FAA-H-8083-15). Washington, DC: US Department of Transportation, Federal Aviation Administration, Flight Standards Service.

Fitts, P. M., Jones, R. E., & Milton, J. L. (1950). Eye Movements of Aircraft Pilots during Instrument Landing Approaches. *Aeronautical Engineering Review, 9*, 1-5.

Juang, B. H., & Rabiner, L. R. (1985). Mixture autoregressive hidden Markov models for speech signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-33*(6), 1404-1413.

Kaiser, K. (1994). *Improved operational reliability and safety with HUD. The Alaska airlines*

experience. Paper presented at the Proceedings of the 39th Corporate Aviation Safety Seminar (CASS), Apr 13-15 1994, St. Louis, MO.
Kemeny, J. G., & Snell, J. L. (1960). *Finite Markov Chains*. Princeton, NJ: D. Van Nostrand Co., Inc.
Krogh, A., Brown, M., Mian, I. S., Sjolander, K., & Haussler, D. (1994). Hidden Markov models in computational biology: Applications to protein modeling. *Journal of Molecular Biology, 235*(5), 1501-1531.
Levison, W. H., & Elkind, J. I. (1967). Two-Dimensional Manual Control Systems with Separated Displays. *IEEE Transactions on Human Factors in Electronics, HFE-8*(3), 202-209.
Machado, R. (1999). Learning to Fly with Rod Machado, *Microsoft Flight Simulator 2000 Pilot's Handbook* (pp. 41-161): Microsoft Co.
McRuer, D. T., & Jex, H. R. (1967). A Review of Quasi-Linear Pilot Models. *IEEE Transactions on Human Factors in Electronics, HFE-8*(3), 231-249.
Moray, N. (1986). Monitoring Behavior and Supervisory Control. In K. R. Boff & L. Kaufman & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. II). New York: A Wiley-Interscience Publication.
Moray, N., Neil, G., & Brophy, C. (1983). *The behavior and selection of fighter controllers* (Tech. Rep.). London: Ministry of Defense.
Oliver, N., Pentland, A., & Berard, F. (2000). LAFTER: A real-time face and lips tracker with facial expression recognition. *Pattern Recognition, 33*(8), 1369-1382.
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.
Rabiner, L. R. (1989). Tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257-286.
Rabiner, L. R., & Juang, B.-H. (1986). Introduction to Hidden Markov Models. *IEEE ASSP Magazine (Acoustics, Speech, and Signal Processing), 3*(1), 4-16.
Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics, 6*(2), 461-464.
Senders, J. W. (1955). Man's Capacity to use Information from Complex Displays. In H. Quastler (Ed.), *Information Theory in Psychology* (pp. 109-118). Glencoe, Il.: Free Press.
Senders, J. W., Webb, I. B., & Baker, C. A. (1955). The peripheral viewing of dials. *Journal of Applied Psychology, 39*, 433-436.
Starner, T., Weaver, J., & Pentland, A. (1998). Real-time American Sign Language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(12), 1371-1375.
Tole, J. R., Stephens, A. T., Vivaudou, M., Ephrath, A., & Young, L. R. (1983). *Visual Scanning Behavior and Pilot Workload* (NASA CR-3717). Cambridge, MA: Massachusetts Institute of Technology.
USDOD. (1996). *MIL-STD-1787B, Aircraft Display Symbology*. Philadelphia, PA: Defense Automated Printing Service.
Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. *IEEE Transactions on Information Theory, IT-13*, 260-269.
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.

Weir, D. H., & Klein, R. H. (1971). Measurement and analysis of pilot scanning behavior during simulated instrument approaches. *8*(11), 897-904.

Weir, D. H., & McRuer, D. T. (1972). *Pilot Dynamics for Instrument Approach Tasks: Full Panel Multiloop and Flight Director Operations* (NASA CR-2019). Hawthorne, CA: Systems Technology, Inc.

Wickens, C. D. (1980). The Structure of Attentional Resources. In R. S. Nickerson (Ed.), *Attention and Performance VIII* (pp. 239-257). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Young, L. R., & Sheena, D. (1988). Eye Movement Measurement Techniques. In J. Webster (Ed.), *Encyclopedia of Medical Devices and Instrumentation* (pp. 1259-1269): John Wiley & Sons, Inc.

Zuschlag, M., & Hayashi, M. (2003). *Issues and Knowledge Concerning the Use of Head-Up Displays in Air Transports* (DOT-VNTSC-FAA-00-05). Cambridge, MA: US Department of Transportation, John A. Volpe National Transportation Systems Center.