한국심리학회지: 산업 및 조직
Korean Journal of Industrial and Organizational Psychology
2003. Vol. 16, No. 2, 205-230

© 한국 산업 및 조직심리학회 2003

# Cognitive Prerequisites for Developing Expertise in Situation Awareness: Role of Pilot Memory*

Young Woo, Sohn  Min Young, Kim†  Jae Wook, Seol  Shin Woo, Kim

Yonsei University

The present research examined the relative roles of capacity-and skill-based cognitive characteristics in performance on a complex aviation task. Specifically, we examined the impact of individual differences in inherent capacity of working memory (WM) and acquired memory skill to encode situational information in an accessible form in long-term memory (LTM) on flight situation awareness (SA). We developed spatial and verbal measures of WM-based capacity and LTM-based skill, and then determined the ability of these measures to predict the differences in performance on SA tasks. While both spatial measures of WM-based capacity and LTM-based skill were important predictors of SA performance, their importance varied as a function of pilot expertise. Furthermore, evidence of a compensatory role of the two memory constructs is indicated. The results have practical implications for aviation professionals interested in pilot training and selection for developing pilot expertise.

*key words : situation awareness, pilot, expertise, working memory, long-term memory*

---

* 이 논문은 2002년도 한국학술진흥재단의 지원에 의하여 연구되었음. (KRF-2002-003-D00446)

† 교신저자 : 김 민 영, 016-396-1778, karis@yonsei.ac.kr

The term "situation awareness (SA)" has been highlighted in the aviation domain because of its prominent role in flight operations. Although it has been considered an essential prerequisite for the safe operation of aircraft, the use of the term is inconsistent overall in the domain of aviation. Its use is most often based on an intuitive understanding and a commonly accepted definition is missing. To fill this gap, aviation psychologists have focused on the cognitive components of SA because of the increasingly cognitive nature of the tasks operators should perform (Durso & Gronlund, 1999; Endsley, 1995a; Sarter & Woods, 1991; Wickens, 1999). Thus, researchers in aviation psychology have sought to determine the cognitive requirements that constitute the SA. The previous studies in the aviation domain provided important support for the centrality of cognitive processes in SA, though they did not address the nature of individual differences in utilizing cognitive processes during SA (see Adams, Tenney, & Pew, 1995; Endsley, 1995a; Wickens, 1999). The objective of this research was to determine the locus of individual differences in pilot ability to accommodate the cognitive demands that building and maintaining SA imposes.

The role of memory is viewed as central to performing tasks that require dynamic and complex processing such as SA (see Durso & Gronlund, 1999; Endsley, 1995a). The content relevant to SA is processed and resides in memory. The accuracy of SA depends also on memory in which incoming information is integrated into coherent interpretation and prediction of aircraft status. Due to the importance of memory components of SA, many SA researchers have focused on the role of memory in a variety of task domains including air traffic control (e.g., Gronlund, Ohrt, Manning, Dougherty, & Perry, 1998), driving (e.g., Gugerty, 1997), and instrument flight (e.g., Doane, Sohn, & Jodlowski, in press).

In order to understand memory processes in SA, it is necessary to understand the construct of SA. The spirit of most definitions of SA can be incorporated into Endsley's (1995a) information processing view (Durso & Gronlund, 1999; Jones & Endsley, 2000; Wickens, 1999). Endsley's (1995a) view defines three levels of SA in terms of component cognitive processes. The first level of SA involves perceiving environmental elements such as other aircraft, terrain, system status and warning lights. The second level involves information integration, a process of activating long-term memory (LTM) knowledge structures in order to organize the perceived situation elements into meaningful and recognizable configurations. The third level includes processes that enable projection of future flight status. This third level of SA uses the goal-relevant activated knowledge structures formed in the second level of SA to predict the status of the aircraft. The accuracy of SA is a function of

activating LTM knowledge structures that facilitate the integration of environmental information and result in a coherent interpretation of the current and future flight status. Recent research suggests that the integration of environmental information takes place in working memory (WM) (see Durso & Gronlund, 1999; Endsley & Garland, 2000). In summary, SA in Endsley's view involves memory processes that dictate whether the resulting level of awareness will facilitate versus detract from pilot performance.

## Research on Cognitive Factors in Situation Awareness

Many SA researchers suggest that understanding component processes is crucial to understanding failures in SA (e.g., Durso & Gronlund, 1999; Endsley, 1995a; Sarter & Woods, 1991). Adams, Tenny, and Pew (1995) suggest that componential analyses will address the practical need for the ability to predict failures in SA. As an example, they describe a list of indicators of incomplete SA that was devised to help pilots detect SA problems.

Recent research suggests that componential analyses of WM and LTM processes (e.g., Ericsson & Kintsch, 1995) are useful in predicting failures in SA (e.g., see Durso & Gronlund, 1999 for a complete review of additional componential research). For example, in a study of U. S. Air Force F-15 pilots, Caretta, Perry, and Ree (1996) showed that the cognitive factors such as verbal working memory, spatial working memory, spatial reasoning, and divided attention were the reliable predictors of SA after controlling for the effects of flight experience. Whereas Caretta et al. (1996) used cognitive factors that were solely based on WM as predictors of SA, Stokes, Kemper, and Kite (1997) used LTM-based knowledge representations as well as WM-based information processing abilities to predict decision-making optimality on a simulated flight situation. Stokes et al. found that both LTM-based knowledge representation measures and WM-based spatial memory measures were predictive of flight decision making.

Although the previous studies pointed to the memory processes as important components of SA, the nature of individual differences in WM and LTM and their relationship in the context of SA have yet to be determined clearly. As an attempt to address the individual differences, the present research examined the locus of individual differences in SA using theoretical analysis based on memory theories developed in a variety of domains (e.g., Groen & Patel, 1988; Just & Carpenter, 1992; Larkin, McDermott, Simon, & Simon, 1980; Sohn & Doane, 1997). In such tasks, successful performance entails processing symbols of sequences and storing their intermediate products over time (Baddeley, 1986; Carpenter & Just, 1989). Reflecting such processing and storage requirements in cognitive processing, individual

differences in complex task performance have been explained in large part by varied ability to satisfy the processing and storage demands that a given task imposes (Engle, Kane, & Tuholski, 1999; Just & Carpenter, 1992; Shah & Miyake, 1996). The focus of the present research was on the locus of the varied ability to satisfy the demands on WM. Testing the validity of possible accounts based on the current cognition theories, this research provides an insight into the cognitive requirements that constitute SA and the effective training methods that optimize acquisition and maintenance of SA.

So far, we have been concerned with how information is processed, but how this information is represented is relevant in the present research because it provides a useful means to examine the nature of memory processes. Of particular interest is the distinction between spatial and verbal representations in memory. Although this separability has long be recognized (see Baddeley, 1986), this distinction has not yet been clearly addressed in the research on SA. Examining whether the present task involves different modality of representations, this research also provides an insight into the modality effect on memory representations of SA.

## Present Research

### Research Objective

The specific research question addressed was whether individual capacity to maintain information in WM or individual skill to retrieve information from LTM or both are the loci of individual differences in SA performance. This objective was achieved by analyzing pilot SA performance in the context of cognitive theories. Of interest was the ability of contrasting theories to explain and predict individual differences in SA performance. The capacity theory of WM proposes that the locus of individual differences in performing a complex task is the fixed capacity to compute and maintain presented information in WM (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992; Shah & Miyake, 1996). Alternatively, the skilled memory theory proposes that the locus of the differences is the acquired skill to encode the presented information efficiently in accessible form in LTM (e.g., Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995).

The capacity and skill theories differ in views about storage location as well as factors influencing individual differences in processing and storage of information. According to the capacity theory, the maintenance of information is solely based on storage in WM, and the locus of individual performance differences is the capacity to maintain the information in WM. This capacity is assumed to be inherently different across individuals. In contrast, according to the skill theory, the maintenance of information is based on storage in LTM in accessible form, and

individual performance differences result from acquired skill to construct and use better retrieval structures. Whereas the former theory proposes that the amount of information that can be maintained in accessible form in WM is limited by a fixed capacity, the latter theory proposes that WM capacity is thought to dynamically change within individuals as a function of their access efficiency during a given task. Thus, our research objective was accomplished by devising measures of WM capacity and LTM-retrieval skill and then comparing their predictive ability of the two sets of measures to predict the differences in performance on SA tasks.

## Overview of Experiment

The present experiment consisted of three tasks. First, the span tasks of Daneman and Carpenter (1980) and Shah and Miyake (1996) were modified to measure individual WM capacity. Many researchers hypothesize separate WM resources for different modalities of cognitive processes (e.g., Daneman & Tardif, 1987; Shah & Miyake, 1996). Based on this hypothesis, the span tasks assessed individual WM capacity for computation and storage of spatial and verbal information.

Second, a situation recall task analogous to the Chase and Simon (1973) chess experiment was devised to measure individual pilots' LTM-retrieval skill. Their experiments showed that chess masters were able to reconstruct more chess pieces on a board than novices if the pieces were originally arranged in a meaningful game position. In contrast, if the pieces were placed in a random configuration, chess master performance fell to novice levels. In the present experiment, the elements to be reconstructed were cockpit situational indices (e.g., altitude, heading, airspeed). The cockpit situations were represented in either spatial or verbal form to examine the effect of presentation modality on pilot ability to construct an accurate mental representation of cockpit situation. The spatial form included a pictorial snapshot of actual cockpit instruments, while the verbal form was a written description of instrument indications. To measure encoding in and retrieval from LTM, participant memory for a cockpit situation was tested after a 30-s delay filled with an intervening task (e.g., Peterson & Peterson, 1959). The intervening task was considered to clear short-term storage of scanned situational information. To measure construction of efficient retrieval structures, we presented participants with both meaningful and non-meaningful cockpit situations. The meaningfulness was defined as whether the cockpit situation represented a plausible real state of an aircraft in the world.

The difference in delayed-recall accuracy of meaningful and non-meaningful situations is thought to reflect pilot ability to construct and

use efficient retrieval structures in LTM (see e.g., Stokes, Kemper, & Kite, 1997). In the meaningful trials, pilots are expected to access their LTM-retrieval structures to encode the situational information. In contrast, in the non-meaningful trials pilots are not expected to have a matching retrieval structure that can be accessed to aid encoding. The benefit of accessing a LTM-retrieval structure during encoding can be calculated by comparing delayed-recall accuracy in meaningful situations to that for non-meaningful situations. In sum, the differences in delayed-recall accuracy between the meaningful and non-meaningful conditions were calculated as an index of situational information accessible in LTM.

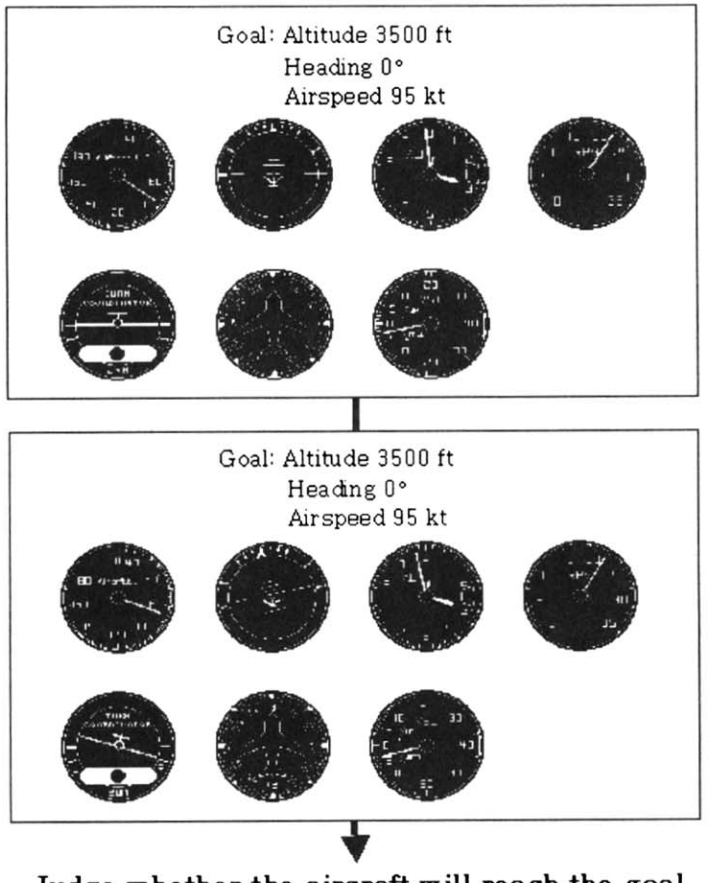



*Figure 1.* An example trial of a SA task.

Finally, the SA task posed to participants in the present research was designed to reflect the three levels of cognitive components defining SA (Dominquez, 1994; Endsley, 1995a). Specifically, participants viewed consecutive screens that showed a goal description and two consecutive cockpits and then judged whether an aircraft depicted by the consecutive cockpit snapshots would reach the specified goal state in the next 5 s (see Figure 1). (The details of task materials and procedures are provided in the methods section.) To successfully perform this task, participants must perceive changes in flight situation elements across various instruments (e.g., status and changes in airspeed, heading, altitude), interpret and understand their meaning with respect to the goal (e.g., I am currently below my desired airspeed, etc.), and predict their future implications given the goal state in mind (e.g., Given my current state and rates and directions of change, I am headed toward my desired flight status). The methods of the three tasks are detailed in the subsequent sections.

## Method

### Participants

Fifty-two pilots from local flight schools participated in this experiment. As a function of pre-experiment questionnaire data obtained from each participant, novice and expert groups contained twenty-five and twenty-seven participants, respectively (For additional questionnaire and classification details, see Doane & Sohn, 2000). Novices on average had total flight time of 86 hours, total flight time of 18 hours in the last 90 days. In contrast, experts on average had total flight time of 1117 hours, total flight time of 31 hours in the last 90 days.

### Materials and Procedure

All participants completed the three tasks (span, situation recall, and SA). Each task is described below.

**Span tasks.** We used two span tasks spatial and verbal requiring participants to process and store different modality of information in WM. In the spatial span task, participants were presented with a set of English capital letters and their mirror images one at a time, each appearing in different orientations (see Figure 2*a*). The task was to remember the orientation of each letter in the correct order while deciding whether each letter is normal or mirror-imaged as quickly and accurately as possible. A single trial consisted of the individual presentation of normal or mirror-imaged letters (F, J, L, P, or R) on a computer screen. Each letter was presented for 2,200 ms in one of seven possible

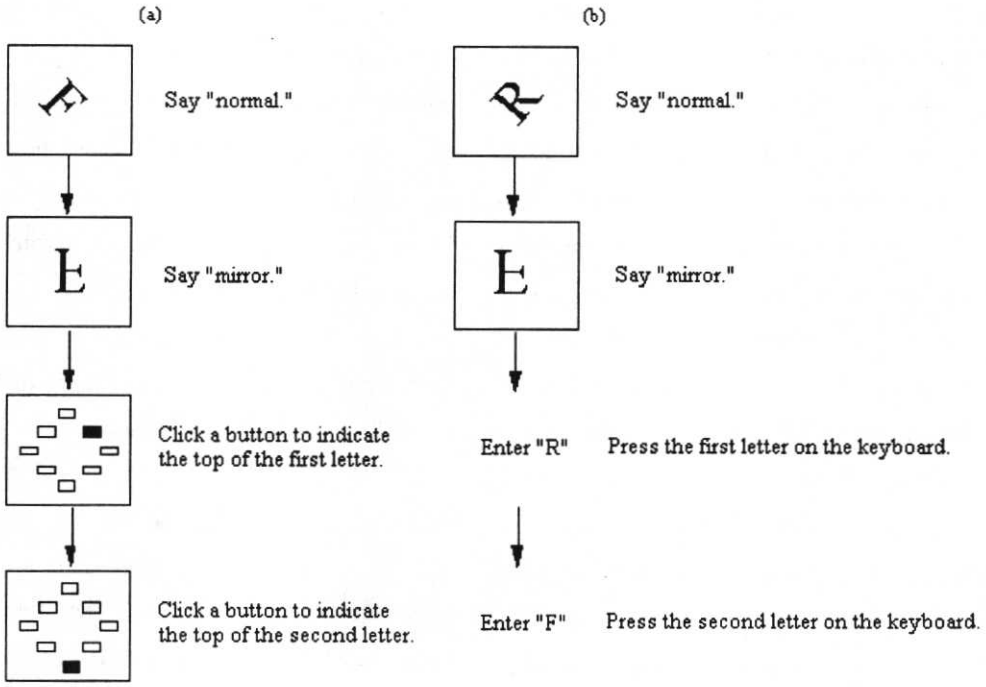



*Figure 2*. Example trials of (*a*) a spatial span task and (*b*) a verbal span task. (Adapted from Shah & Miyake, 1996).

orientations in 45 increments, excluding the upright orientation.

Participants were asked to respond aloud to indicate whether the letter was "normal" or "mirror." After the entire set of letters was presented in a trial, participants were asked to report the correct orientations in serial order by clicking on small buttons that indicated the possible orientations on a diamond-shaped grid (for more details, see Shah and Miyake, 1996).

The structure of the verbal span task was identical to that for the spatial part, except that participants were presented with a set of different letters and asked to recall the letters rather than the orientations (see Figure 2*b*). At each trial, participants were presented with a letter set one at a time, each appearing in different orientations. We added two letters G and Q to the letter set (F, J, L, P, or R) that was used for the spatial span task to increase the variability of the possible set of letters to be recalled. A single trial for this task consisted of the individual presentation of a small set of 7 letters (F, G, J, L, P, R, and Q) presented in random order on a computer screen. Letter presentation was constrained such that the same

letter could appear only once in a given trial. The participant's task was to remember the letters and their order of presentation while deciding whether each letter was normal or mirror-imaged. At the end of a trial, the participant was prompted to type the letters in the order presented using the keyboard.

Each span task included a total of 20 letter sets including 5 sets at each size ranging from two to five letters and participants were presented with increasingly longer sets of letters. After the participants finished 5 trials at a size of letter sets, they were informed of how many letters would appear in the next set of trials and were prompted to click the mouse to begin the next trial. The order of task performance was counter-balanced such that half of the participants started with the spatial task and the other half of the participants started with the verbal task.

**Situation recall tasks.** Participants' task on each trial was to view a sequence of two cockpit situations presented on the screen and recall the situational information. Participants were presented either with pictorial cockpit snapshots as shown in Figure 3*a* (spatial stimuli) or verbal lists that described cockpit situations as shown Figure 3*b* (verbal stimuli). A pair of consecutive cockpit situations (as depicted in Figures 3*a-b*) was shown for 40 s and then they were removed from the screen. Participants then completed an intervening task for a 30-s delay period, where they counted backward by threes as fast as possible for 30 s. For example, a prompt such as "Count backward by 3: 528" was presented on the screen following the disappearance of the cockpit snapshots and participants counted backward by threes from 528: 525, 522, 519, and so on. This intervening task was devised to interrupt maintenance and computation of the display indications in WM.

After the 30-s delay filled with the intervening task, they had to recall the flight situation. For the spatial situation trials, participants reconstructed the situation by manually filling in the indications of display instruments on a blank cockpit frame sheet of paper. For the verbal situation trials, they reported aloud the description of display indications to the experimenter. Participants were asked to recall the value indications that were presented in either the top or bottom situation, and the choice of situation to be recalled was randomly selected. When participants thought they were finished recalling all the indications they could remember for the trial, they pressed the space bar on the computer keyboard to view a pair of cockpit situations for the next trial.

As previously stated, we manipulated the meaningfulness of the cockpit situation. Figures 4*a* and 4*b* depict example non-meaningful situations. The figures show that each snapshot displayed an implausible flight situation that

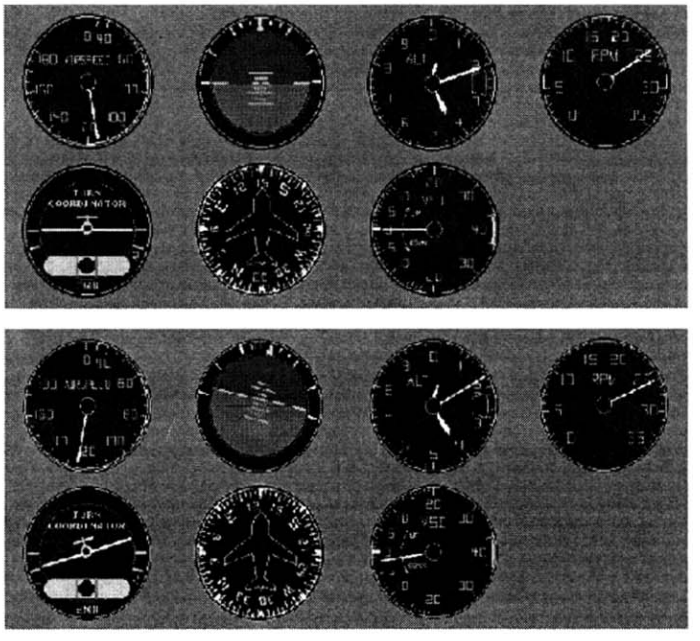


*Figure 3a.* Example spatial cockpit snapshots used for situation recall tasks.

| |
|---|
| Airspeed: 115 kt.<br>Bank: 0°<br>Pitch: level<br>Altitude: 4200 ft.<br>Power: 2500 rpm<br>Rate of turn: 0<br>Heading: 150°<br>Rate of climb: 0 ft./min. |
| Airspeed: 125 kt.<br>Bank: 15° to the left<br>Pitch: 1 dot below<br>Altitude: 4170 ft.<br>Power: 2600 rpm<br>Rate of turn: Standard to the left<br>Heading: 130°<br>Rate of climb: -200 ft./min. |

*Figure 3b.* Example verbal cockpit descriptions used for situation recall tasks.

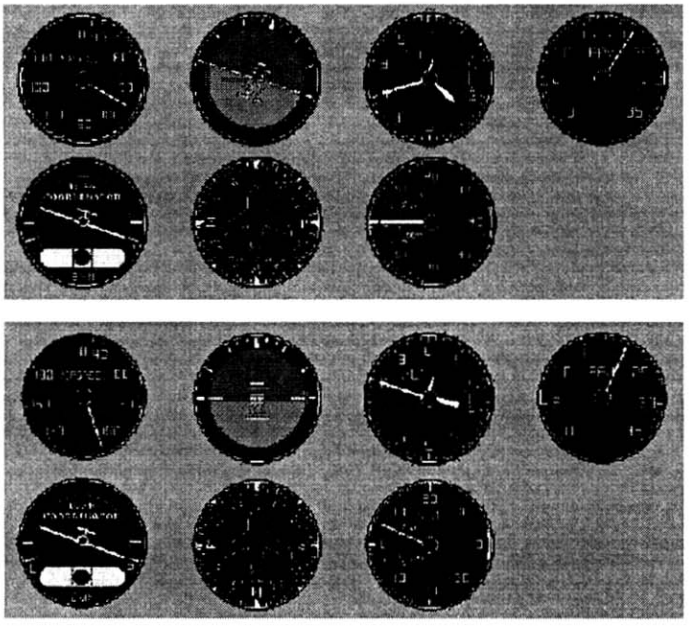


*Figure 4a.* Example non-meaningful snapshots used for situation recall task

| |
|---|
| Airspeed: 90 kt.<br>Bank: 15° to the left<br>Pitch: level<br>Altitude: 3700 ft.<br>Power: 2200 rpm<br>Rate of turn: Standard to the right<br>Heading: 90°<br>Rate of climb: 0 ft./min. |
| Airspeed: 110 kt.<br>Bank: 0°<br>Pitch: level<br>Altitude: 2800 ft.<br>Power: 2100 rpm<br>Rate of turn: Standard to the right<br>Heading: 180°<br>Rate of climb: 500 ft./min. |

*Figure4b.* Example non-meaningful descriptions used for situation recall tasks.

pilots would never encounter, and the pair of snapshots did not depict consecutive flight situations. For example, as depicted by the bottom snapshot in Figure 4*a*, the attitude indicator indicates a level flight while the turn coordinator indicates a turn to the right is in progress. These two events cannot take place at the same time; it is a physical impossibility. Looking at the sequence of two snapshots, the altimeter indicates that the altitude has decreased from 3,700 ft. to 2,800 ft. though the vertical speed indicator indicates that the climbing rate has increased from 0 ft. to 500 ft. per minute. Such a situation cannot occur in the real world.

The situation recall tasks included 9 trials of the spatial task and 9 trials of the verbal task, 6 trials with meaningful situations and 3 trials with non-meaningful situations for each task. The order of task performance was counter-balanced such that half of the participants started with the spatial task and the other half of the participants started with the verbal task. For each task, participants were first presented with a set of trials with meaningful situations and then with a set of trials with non-meaningful situations. This was done to avoid any possible changes in viewing strategy caused by exposure to unexpected non-meaningful situations. We gave participants two practice trials with meaningful situations to familiarize them with the procedure. They received feedback regarding recall accuracy for the practice trials. No accuracy feedback was provided for the experimental trials.

**SA task.** As previously stated, a SA trial was composed of consecutive screens that showed a goal description and two consecutive cockpits (see Figure 1). The goal description indicated the desired state that an aircraft should reach in the near future (i.e., in approximately 5 s) on the three flight performance elements (e.g., "Altitude 3500 ft, Heading 180, Airspeed 90 kt"). The first screen displayed the goal description on top of the screen and right below it the cockpit snapshot that depicted the initial flight situation (at time 1) for 20 s, at which time the first cockpit snapshot disappeared and the second cockpit snapshot appeared. The goal state description remained on top of the screen. The second snapshot depicted the "current" state of the aircraft (at time 2) following changes caused by control movements (not specified to the participants) executed at time 1. This second snapshot depicted the state of the aircraft approximately 5 seconds following the status depicted in the first snapshot. The 5-s flight interval between the two snapshots was sufficient to indicate a noticeable amount of change in cockpit display of aircraft status. The second snapshot remained on the screen until participants pressed a response key, at which time the goal description and the second snapshot disappeared.

Participants were asked to determine if the aircraft depicted in the cockpit snapshots would achieve the goal initially specified in the next 5 seconds without further control movements. That is, they had to determine whether the snapshot sequence indicated a movement toward the goal and whether the second snapshot indicated that movement toward the goal would continue. In essence, they had to mentally predict the state of the aircraft in the next 5 seconds with the constraint that no further control movements would be applied. The participants indicated whether the aircraft was moving in a manner *consistent* with achieving the goal, or in a manner *inconsistent* with achieving the goal by pressing the key marked "C" for consistent or "I" for inconsistent. Inconsistent trials were created by manipulating one of three flight elements (i.e., altitude, heading, airspeed) depicted in the second snapshot to render the "current" (time 2) flight situation inconsistent with the specified goal. The materials included 23 consistent and 27 inconsistent snapshot sequences. Once a consistency judgment was made, the next trial was presented.

Participants were presented with 7 practice trials to familiarize them with the procedure prior to viewing experimental trials. They received accuracy feedback during the practice trials, but feedback was not provided during the experimental trials. The computer recorded consistency judgment accuracy and the time that elapsed between the presentation of the second snapshot and the entry of the consistency judgment.

### General Procedure

The experiment took place in a single session of 2 - 2.5 hrs. After signing an informed consent form, participants completed the flight background questionnaire. Then, the three tasks were administered in the following order: the SA task, the span task, and the situation recall task.

## Scoring and Data Analysis

### WM Capacity Measures

**Scoring for spans.** We used procedures that Daneman and Carpenter (1980) and Shah and Miyake (1996) devised to score a participant's span. A participant's span score for the spatial task was defined as the highest set size for which all of the orientations were recalled in the correct order, for at least *three* of the five sets. If the participant's recall was accurate on two of the five sets at the next set size, half a point was added to the score. If a participant correctly recalled less than three of the five sets at a particular level but was able to recall two or more sets at a higher level (this case was rare, $n$ = 5), the average of the lower and the upper

limits was used as the span score. The possible maximum and minimum scores for the spatial span measures were 5.0 and 1.0 (in cases in which only one or no set was correctly recalled at the two-item level) respectively.

For the verbal task, we used a higher criterion because this task was less difficult than the spatial task. A participant's verbal span score was defined as the highest set size for which all of the letters were recalled in the correct order, for at least *four* of the five sets. If the participant's recall was accurate on three of the five sets at the next set size, half a point was added to the score. If a participant correctly recalled less than four of the five sets at a particular level but was able to recall three or more sets at a higher level (this case was rare, $n$ = 4), the average of the lower and the upper limits was used as the span score. The possible maximum and minimum scores for the verbal span measures were 5.0 and 1.0 (in cases in which two or less sets were correctly recalled at the two-item level) respectively. Although the scoring criteria differed across span tasks due to their varied difficulty, they were applied equitably across participants and resulted in comparable span scores.

We did not take into account participant accuracy in judging whether letters were normal or mirror-imaged. This aspect of span tasks is thought to reflect spatial visualization ability rather than WM capacity (Shah & Miyake, 1996). Regardless, the participants showed 70% or higher percent accuracy on judgment of normal or mirror-imaged orientation. This suggested that the participants made an effort to be accurate and in so doing used the processing component of WM demands imposed by the span tasks.

**Span scores.** A summary of descriptive statistics for the span measures for novices and experts is presented in Table 1. Of interest is whether measures of WM capacity relate to amount of piloting experience. As previously mentioned, the capacity account proposes that individual differences in WM capacity are inherent in the individual and do not vary with expertise. To determine if there was a significant relationship between the measures of WM capacity and piloting expertise,

Table 1. Summary of Descriptive Statistics for Span Measures.

| | Novice | | | | Expert | | | |
|---|---|---|---|---|---|---|---|---|
| Measures | M | SD | Min | Max | M | SD | Min | Max |
| Spatial span | 3.15 | 1.08 | 1.00 | 5.00 | 3.42 | 1.12 | 1.50 | 5.00 |
| Verbal span | 3.59 | 0.92 | 2.00 | 5.00 | 3.63 | 0.98 | 2.00 | 5.00 |

mean span scores for novices and experts were compared. Two-tailed t tests resulted in no significant difference between the two groups in either of the span measures, *ps* > .3. This is consistent with the hypothesis that WM capacity does not change for a given individual as they acquire expertise.

## LTM-Retrieval Skill Measures

**Scoring correct recall responses.** The experimenter scored participant indications for each of the eight flight elements (i.e., airspeed, heading, altitude, pitch, bank, power, rate of climb, and rate of turn) according to explicit rules. Responses for pitch and bank were scored separately although they are indicated by one instrument, the attitude indicator. Responses were scored as correct and full points were given if they matched exactly all corresponding indications. The responses for the pitch, bank, rate of climb, and rate of turn that provide both directional and value indications were given partial credits if only the directional indications were correct. However, no credit was given to the responses if only their value indications were correct because the value information in the wrong direction indicates a more severe loss of memory for the situation than does memory loss for the exact value but retention of the correct direction. For example, a pitch indication of one dot *above* the horizon when the correct indication is two dots *above* the horizon was given half a point because the pitch direction, above the horizon, is correct though the exact value of pitch is not correct. In contrast, a pitch indication of one dot *above* the horizon when the correct pitch indication is one dot *below* for the actual pitch of one dot below the horizon was given zero points because although the values (one dot) match, the direction (above versus below) is incorrect.

**LTM-retrieval skill scores.** LTM-retrieval skill scores were calculated using the differences in recall accuracy between possible and impossible situations. Table 2 shows group mean control and performance LTM-retrieval skill scores for spatial and verbal situation presentations. As shown in the table, each measure was divided into control and performance sub-measures to specify the nature of LTM-retrieval structures constructed and used to facilitate SA performance. These sub-measures were calculated based on the recall accuracy for control and performance flight elements respectively. Control elements indicate input settings of control movements and include pitch, bank, and power. Performance elements indicate the behavior of the aircraft as a result of control movements and include altitude, heading, airspeed, and rates of climb and turn.

Looking at Table 2, experts showed higher LTM-retrieval skill scores overall than novices, $F(1, 50) = 12.78$, $MSE = 0.030$, $p < .01$. As proposed by the skilled memory theory, this

Table 2. Mean LTM-Retrieval Skill Scores for Novices and Experts as a Function of Situation Presentation Modality (Spatial and Verbal) and Flight Element Type (Control and Performance)

| | Novice | | Expert | |
|---|---|---|---|---|
| LTM-Retrieval Measures | M | SD | M | SD |
| Spatial Presentation | | | | |
| Control Element | .18 | .22 | .34 | .19 |
| Performance Element | .20 | .12 | .26 | .13 |
| Verbal Presentation | | | | |
| Control Element | .26 | .17 | .33 | .18 |
| Performance Element | .26 | .11 | .33 | .15 |

result can be attributed to a large body of relevant knowledge and a repertoire of flight states that expert pilots internalize. Particularly, the group difference was significantly pronounced for the control LTM-retrieval skill measure with spatial stimuli, $F$ (1, 50) = 7.22, $MSE$= 0.042, $p < .01$, but not for the other measures, $F$s < 3. This suggests that experts rely more on meaningful configurations of control flight elements than do novices.

## SA Performance Measures

To measure pilot SA performance, we analyzed accuracy data in the context of signal detection theory to determine how sensitive participants were to discriminating between consistent and inconsistent stimuli (Green & Swets, 1966). To determine the observer sensitivity of pilot consistency judgments, correct judgments for consistent stimuli and incorrect judgments for inconsistent stimuli represented hits and false alarms in d' calculations, respectively. Following procedures outlined in Green and Swets, bias scores were also calculated to determine if the group differences are influenced by different criteria novices and experts used for making "consistent" or "inconsistent" judgments. Judgment latency was also measured as another index of SA performance.

Table 3 shows mean d' values, bias scores, and latency for novices and experts. Judgment sensitivity was higher for experts than for novices, $F$ (1, 50) = 8.69, $MSE$ = 0.269, $p < .01$, but there is no group difference in bias or latency, $F$s < 1. The results suggest that the SA task differentiated between novices and experts, although group performance differences were caused by judgment sensitivity, not by judgment criterion or latency differences. Because judgment sensitivity varied as a function of expertise, we used this performance measure as

Table 3. Mean Judgment Sensitivity (d'), Bias Scores, and Correct Judgment Latency for Novices and Experts in SA task.

| | Novice | | Expert | |
|---|---|---|---|---|
| SA Measures | M | SD | M | SD |
| Judgment sensitivity | 1.20 | 0.53 | 1.62 | 0.50 |
| Judgment bias | -0.32 | 0.57 | -0.24 | 0.54 |
| Judgment latency | 8.94 | 2.61 | 8.37 | 2.77 |

the criterion measure to be predicted by the WM capacity and LTM-retrieval skill measures.

## Results and Discussion

The major question of interest in this research was which memory measures are valid indicators of SA performance and what roles WM capacity and LTM-retrieval skill play in predicting SA performance. To address this question, each of two WM capacity measures (spatial and verbal spans) and four LTM-retrieval skill measures (spatial control, spatial performance, verbal control, and verbal performance) was correlated with the SA performance measure. Then, hierarchical regression analyses were conducted using WM capacity and LTM-retrieval skill measures that were highly correlated with SA performance as predictors of SA performance. These analyses allowed us to determine if the effect of one predictor on SA performance was significant when controlling for the other predictors (Cohen & Cohen, 1983).

### Memory Measures Indicative of SA Performance

Table 4 shows the correlation between each of memory measures and SA judgment sensitivity for novices and experts. The roles of specific memory measures varied as a function of expertise. Overall, WM capacity measures

Table 4. Correlation between Memory Measures and SA Judgment Sensitivity for Novices and Experts.

| Memory Measures | Novice | Expert |
|---|---|---|
| WM Capacity | | |
| Spatial Span | .52** | .10 |
| Verbal Span | .30 | .10 |
| LTM-Retrieval Skill | | |
| Spatial Control | .22 | .46** |
| Spatial Performance | .08 | .24 |
| Verbal Control | -.24 | -.08 |
| Verbal Performance | .17 | .21 |

*Note.* $*p < .05$; $**p < .01$.

correlated highly with SA sensitivity for novices, whereas LTM-retrieval skill measures correlated highly with SA sensitivity for experts. One explanation for this difference is that novices rely more on their inherent WM capacity, because they have not acquired reliable LTM-retrieval skill to overcome the WM overload imposed by the SA tasks. Thus, novice pilots with higher WM capacity were able to be more sensitive performers. In contrast, experts with a higher level of LTM-retrieval skill showed greater SA sensitivity.

In each type of memory measure, spatial span was the significant indicator of SA sensitivity for novices, whereas the spatial control LTM-retrieval skill measure was the significant indicator of SA sensitivity for experts. This suggests that SA requires spatial representation more than verbal representation in WM. Given that LTM-retrieval skill measures reflect access efficiency to LTM, we can also infer that efficient LTM access is mediated by configurations of control flight elements.

## Roles of WM Capacity and LTM-Retrieval Skill in Predicting SA Performance

To determine whether WM capacity, LTM-retrieval skill, or their interaction predicted novice and expert SA performance, hierarchical regression analyses were performed. Because spatial WM capacity and spatial control LTM-retrieval skill measures correlated significantly with SA judgment sensitivity, they were used as representative predictors of SA performance.

Hierarchical regression analyses were conducted in three steps. The LTM-retrieval skill measure was entered in the first step; the WM capacity measure was entered in the second step; and the cross-product of these two measures (LTM-retrieval skill x WM capacity) was entered in the third step. This order of entry reflects our assumption that an increase in LTM-retrieval skill would decrease the pilot's reliance on WM capacity as the determinant of SA performance. The cross-product was entered only after the two variables that might be a source of interaction effect on SA performance had been entered. A significant increment in variance accounted for (referred to as $R^2_{change}$) by the variable entered in each step would indicate the unique contribution of that variable to SA performance. The results of the hierarchical regression analyses for LTM-retrieval skill, WM capacity, and their interaction predicting novice and expert SA performance are summarized in Tables 5 and 6, respectively. To help explain them, the relations of WM capacity to SA performance as a function of LTM-retrieval skill for novices and experts are characterized in Figures 4 and 5, respectively. In these figures, high LTM-retrieval skill pilots are pilots with scores equal to their group mean or higher for the LTM-retrieval skill measure, and low LTM-retrieval skill pilots have scores lower than the group mean.

Table 5. Results of Hierarchical Regression Analyses Predicting Novice SA Performance.

| Criterion variable<br>Predictor variable | $R^2_{change}$ | B | $F_{change}$ |
|---|---|---|---|
| SA performance | | | |
| LTM-retrieval skill | .05 | 0.54 | 1.16 |
| WM capacity | .23 | 0.25 | 7.10** |
| LTM-retrieval skill x WM capacity | .00 | 0.11 | 0.06 |

*Note.* $*p < .05$; $**p < .01$.

Table 6. Results of Hierarchical Regression Analyses Predicting Expert SA Performance

| Criterion variable<br>Predictor variable | $R^2_{change}$ | B | $F_{change}$ |
|---|---|---|---|
| SA performance | | | |
| LTM-retrieval skill | .21 | 1.21 | 6.68** |
| WM capacity | .00 | 0.01 | 0.01 |
| LTM-retrieval skill x WM capacity | .18 | -1.14 | 6.89** |

*Note.* $*p < .05$; $**p < .01$.

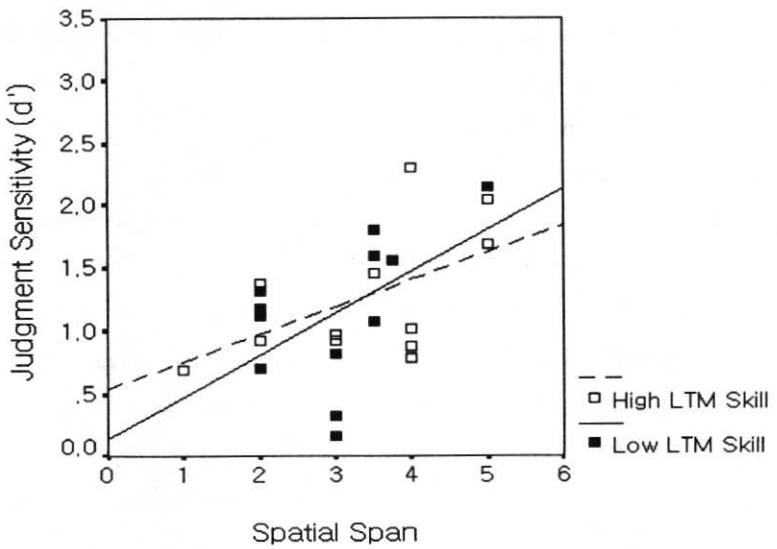



Figure 4. Relation of WM capacity to SA judgment sensitivity as a function of LTM-retrieval skill for novices.

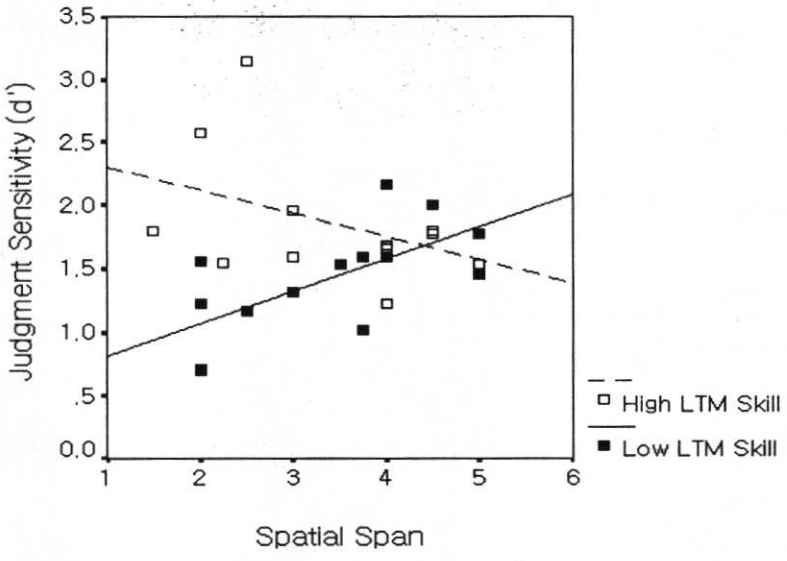



Figure 5. Relation of WM capacity to SA judgment sensitivity as a function of LTM-retrieval skill for experts.

As shown in Table 5 and Figure 4, only WM capacity (spatial measure) predicted novice SA performance, $R^2_{change}$ = .23, $B$ = 0.25, $F_{change}$ (1, 22) = 7.10, $p$ < .01. This suggests that the higher the WM capacity (spatial span), the better the performance on the SA task. In contrast, Table 6 and Figure 5 show that LTM-retrieval skill (spatial control measure) predicted expert SA performance, $R^2_{change}$ = .21, $B$ = 1.21, $F_{change}$ (1, 25) = 6.68, $p$ < .01, which suggests that the higher the LTM-retrieval skill, the better the SA performance. Also, there was a significant LTM-retrieval skill x WM capacity interaction in predicting expert SA performance, $R^2_{change}$ = .18, $B$ = -1.14, $F_{change}$ (1, 23) = 6.89, $p$ < .01. Figure 5 suggests that WM capacity facilitates SA performance for low LTM-retrieval skill experts but not for high LTM-retrieval skill experts. Moreover, for experts with high levels of LTM-retrieval skill, relying on WM capacity decreases SA performance.

## General Discussion

The present research makes both theoretical contributions to understanding the cognitive constituents of SA and practical suggestions for developing pilot expertise and training. The findings from this research suggest that memory processes play a central role in building components of flight situation awareness, and

that the role of WM and LTM change with expertise. These findings address an ongoing discussion among aviation researchers about the relationship between memory processes and SA during flight (e.g., Carretta et al., 1996; Durso & Gronlund, 1999; Endsley, 1995a). In the aviation psychology literature, there are contradictory ideas about the role of WM in SA. For example, Caretta et al. found that cognitive capacity measures based on WM, spatial reasoning, and divided attention predicted SA. In contrast, Durso and Gronlund argued that the correlation between WM and SA is the result not of WM capacity, but of acquired strategies such as LTM skills that allow experienced pilots to compensate for WM overload.

Our results indicate that both WM capacity and LTM-retrieval skill play a role in flight SA, that LTM-retrieval skill compensates for WM load, and that the respective roles of WM and LTM-retrieval skill vary as a function of expertise. WM capacity was critical for less experienced pilots, whereas acquired LTM-retrieval skills were important for more experienced pilots. These findings are consistent with the previous works that emphasized the role of LTM and skilled memory for expert SA (Endsley, 1995a, b; Fracker, 1988, 1991; Hartman & Secrist, 1991). For example, Endsley (1995b) found that expert pilots were able to keep situational information available to answer SA queries for quite a long time after a freeze of system simulation. In contrast, student participants (non-experts) showed poor performance on the similar SA queries (Fracker, 1991).

This research has applied implications for pilot training. Gardner (1983) has argued that a person needs to have a certain predisposition in order to become an expert in a domain, whereas Ericsson and colleagues (Ericsson & Charness, 1994; Ericsson & Lehmann, 1996) contend that a person can learn to perform expertly in a domain through deliberate practice. Our data at least partially support Ericsson's view. Although individual differences in WM capacity had a significant effect on acquisition of a high level of SA for novice pilots, the effect of such individual differences was attenuated for expert pilots with extensive knowledge and skills. Indeed, the reliance on WM capacity should be low for experts who have high levels of LTM-retrieval skill in performing their domain SA tasks.

If WM capacity measures predict novice pilot vulnerability to SA failures, then pilot instructors may be able to use screening tests that assess WM capacity to customize their training according to the cognitive abilities of the student. There are potential applied applications for communications between air-traffic controllers and pilots. For example, during non-nominal flight events, novice pilots will be most vulnerable to failures in SA when their WM load is high. This suggests that interactions with

pilots should be designed to reduce their working memory load.

The findings from this research also have potential implications for the training of expert pilots. Overall, expert SA performance was predicted by LTM-retrieval skill. This suggests a possible future role for LTM-retrieval skill as an assessment tool for pilots. Our results provide an insight into improving pilot LTM-retrieval skill, pointing to pilot knowledge representation varying with expertise. As found in formal domains such as physics (e.g., Chi, Feltovich, & Glaser, 1981), the nature of pilot knowledge representations may vary as a function of expertise. Novices often rely upon and represent surface (or "superficial") features of situations, whereas experts focus on the underlying (or "deep") features. In aviation, operations on the control elements can be cognitively different from operations on the performance elements. The control and performance elements reflect causes and effects of flight situation changes, respectively, and can be seen as corresponding to the deep and surface features of flight situations. As suggested by our data, expert pilots are likely to use configurations of control elements (e.g., bank, pitch, and power) to represent their flight situations rather than performance elements (e.g., heading, altitude, and airspeed).

Thus, one way to improve pilot SA can be to familiarize them with the underlying features of cockpit situations for the event-based training. One of the basic formula taught in ground school training is "attitude + power = performance," that an appropriate input combination of attitude (bank and pitch) and power results in the desirable aircraft performance (see Dogan, 1991). Expanding student pilots' repertoires of meaningful control flight patterns of attitude and power can facilitate their skilled encoding and retrieval of information from LTM. This also suggests that further research is needed to determine if such an explicit training will improve individual LTM-retrieval skill at a given level of expertise. In summary, the present research makes a significant contribution to the goal of determining the cognitive processes that support flight SA. Understanding the cognitive processes that support flight SA will allow us to move beyond the retrospective diagnosis of SA failures. If we can anticipate when cognitive vulnerabilities and flight situations will interact and result in SA failures, we gain the opportunity to intervene and remedy failures before it happens (Durso & Gronlund, 1999)

## References

Adams, M.J., Tenney, Y.J., & Pew, R.W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37(1),* 85-104.

Baddeley, A. D. (1986). *Working memory*. New York:

Oxford University Press.

Carretta, T. A., Perry, D. C., & Ree, M. J. (1996). Prediction of situation awareness in F-15 pilots. *International Journal of Aviation Psychology, 6*, 21-41.

Carpenter, P. A., & Just, M. A. (1989). The role of working memory in language comprehen- sion. In D. Klahr & K. Kotovsky (Eds.), *Advances in the psychology of human intelligence* (Vol. 3, pp. 221-252). Hillsdale, NJ: Erlbaum.

Charness, N. (1991). Expertise in chess: The balance between knowledge and search. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 39-63). Cambridge, England: Cambridge University Press.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associat.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Daneman, M., & Tardif, T. (1987). Working memory and reading skill reexamined. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*, (pp. 491-508). Hillsdale, NJ: Erlbaum.

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

Doane, S. M., Sohn, Y. W., & Jodlowski, M. T. (in press). Pilot ability to anticipate the consequences of flight actions as a function of expertise. *Human Factors*.

Dogan, P. (1991). *The instrument flight training manual*. Seattle, WA: Aviation Book Company.

Dominquez, C. (1994). Can situation awareness be defined? In M. Vidulich, C. Dominquez, E. Vogl, & G. McMillan (Eds.), *Situation awareness: Papers and annotated bibliography*, (pp. 5-15). AL/CF-TR-1994-0085, Armstrong Laboratory.

Durso, F. T., & Gronlund, S. D. (1999). Situation awareness. In F. T. Durso, R. Nickerson, R. Schvaneveldt, S. Dumais, S. Lindsay, & M. Chi (Eds.), *The handbook of applied cognition*, (pp. 283-314). New York: Wiley.

Endsley, M. R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37(1)*, 32-64.

Endsley, M. R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors, 37(1)*, 65-84.

Endsley, M. R., & Garland, D, J. (Eds.). (2000). *Situation awareness analysis and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and the functions of the prefrontal cortex. In

A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. (pp. 102-134). Cambridge University Press.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*, 725-747.

Ericsson, K. A., & Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control.*(pp. 257-297). Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273-305.

Fracker, M. L. (1988). A theory of situation assessment: Implications for measuring situation awareness. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (Vol. 1, pp. 102-106). Santa Monica, CA: Human Factors Society.

Fracker, M. L. (1991). *Measures of situation awareness: An experimental evaluation* (AL-TR-1991-0127). Wright-Patterson AFB, OH: Armstrong Laboratory.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Green, D. L., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Groen, G. J., & Patel, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In M. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 287-310). Hillsdale, NJ: Erlbaum.

Gronlund, S. D., Ohrt, D. D., Manning, A. C., Dougherty, M. R. P., & Perry, J. L. (1998). Role of memory in air traffic control. *Journal of Experimental Psychology: Applied, 4*, 263-280.

Gugerty, L. (1997). Situation awareness during driving: Explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied, 3*, 42-66.

Hartman, B. O., & Secrist, G. E. (1991). Situational awareness is more than exceptional vision. *Aviation, Space and Environment Medicine, 62*, 1084-9.

Jones, D. G., & Endsley, M. R. (2000). Overcoming representational errors in complex environments. *Human Factors, 42*, 367-378.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208*, 1335-1342.

Peterson, L. R., & Peterson, M. (1959). Short-term retention of individual items. *Journal of Experimental Psychology, 58*, 193-198.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology, 1*, 45-57.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125,* 4-27.

Sohn, Y. W., & Doane, S. M. (1997). Cognitive constraints on computer problem-solving skills. *Journal of Experimental Psychology: Applied, 3,* 288-312.

Stokes, A. F., Kemper, K., & Kite, K. (1997). Aeronautical decision making, cue recognition, and expertise under time pressure. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 183-196). Mahwah, NJ: Lawrence Erlbaum.

Wickens, C. D. (1999). Cognitive factors in aviation. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay and M. T. H. Chi (Eds.), *Handbook of applied cognition*. (pp. 247-282). John Wiley & Sons Ltd.

1 차원고접수 : 2003. 5. 28
수정원고접수 : 2003. 8. 13
최종게재결정 : 2003. 8. 25

Footnote

1. We confirmed this fact with flight instructors serving as consultants to the project.

한국심리학회지: 산업 및 조직
Korean Journal of Industrial and Organizational Psychology
2003. Vol. 16, No. 2, 205-230

© Korean Society of Industrial and Organizational Psychology 2003

# 상황인식에 있어서의 전문성 계발을 위한 인지적 요건: 조종사 기억의 역할을 중심으로

손 영 우 김 민 영 설 재 욱 김 신 우

**연세대학교 심리학과**

이 연구는 복잡한 비행 과제수행에 있어서 용량기반 인지와 기술기반 인지 특성간의 관계적 역할을 조사하였다. 특별히 비행 상황인식에 있어 선천적 작업기억 용량과 후천적 장기기억 사용기술에 대한 개인차를 조사하였다. 우리는 이 두 가지 기억의 측정을 위해 공간적, 언어적 측정도구를 개발하였고, 이를 통해 상황인식 과제에 있어서 수행 예측력을 검증해 보았다. 그 결과, 작업기억 용량과 장기기억 사용기술이 모두 상황인식 과제 수행의 중요한 예측 변인이었으나 조종전문성에 따라 그 중요성이 달라졌다. 또한, 이 두 가지 기억구조의 역할이 보상적이라는 증거가 나타났는데, 이러한 결과는 조종사의 전문성 계발 훈련 및 선발을 위한 실용적인 함의를 제공해 준다.

주요어 : 상황인식, 조종사, 전문성, 작업기억, 장기기억