FLIGHT SAFETY FOUNDATION
Airport Operations
Vol. 24 No. 1 *For Everyone Concerned with the Safety of Flight* January–February 1998

Studies Investigate the Role of Memory In the Interaction Between Pilots And Air Traffic Controllers

One study indicated that the complexity and format of the clearance affect the pilot's ability to remember it. Another study tested whether certain variables affected air traffic controllers' ability to recall specific aircraft flight data.

FSF Editorial Staff

Pilots and air traffic controllers routinely exchange information about aircraft position, destination, heading, altitude, speed, altimeter settings and radio frequencies.

The volume of these communications makes human error inevitable. The problems caused by errors can range from causing an inconvenience to causing an accident.

Two recent studies addressed the role that memory plays in effective air traffic control (ATC). One study, conducted by Judith Bürki-Cohen, experimented with several different structures for spoken ATC clearances in an attempt to determine if one structure is more memorable than others.[1] A second study, by Scott D. Gronlund et al., tested air traffic controllers to determine whether certain variables affected air traffic controllers' ability to recall specific aircraft flight data.[2]

Bürki-Cohen investigated the most effective way for ATC to present clearance information to pilots. Her experiment was intended to answer two questions: What is the best information *format*, to help a pilot remember what was said, and *how much* information should a single transmission contain?

"The first question was motivated by a recent change in [U.S.] ATC communication procedures," Bürki-Cohen said. "Currently, controllers are required to convey all numbers in sequential format; that is, digit-by-digit. For example, a speed of [574 kilometers per hour (kph)] 310 knots has to be conveyed as 'increase speed to three one zero.' Originally, it was assumed that this format is more intelligible in a noisy cockpit than the corresponding grouped format, i.e., 'three hundred and ten.'"

A recent change permitted controllers to utilize a combined format when giving altitudes; that is, to restate an altitude in grouped format (e.g., "thirteen thousand") after first giving the information in sequential format ("one three thousand"). In the Bürki-Cohen report, this is referred to as a restated format. This change was not based on hard evidence, but rather on the controllers' intuition that numbers in grouped format are easier to grasp.

The second question (how much information a single transmission should contain) was motivated by an analysis of audiotapes recorded at air route traffic control centers in 1993.[3] These tapes showed that the greater the complexity of a clearance (i.e., the more information it contained), the more likely it would be to generate an incomplete or erroneous readback or a request for repetition.

The Bürki-Cohen study used airline pilots with a minimum of 3,000 hours of flight time. Tape recordings of ATC clearances simulated a low-sector, en-route environment in U.S. air space. Bürki-Cohen said, "The pilots were asked to assume the role of the nonflying, communicating pilot. They listened to the clearances over headsets and read the clearances back into a microphone. They also set the values on a mockup mode control panel. ... They could do the readback and settings in any order. They were not permitted to use a notepad to aid memory."

Clearances were presented in 36 different combinations of complexity and format. Each clearance contained three, four or five pieces of information; each was presented in sequential, grouped or restated format. Three different message texts were used in each combination, for a total of 108 possible clearances. Two additional clearances were created with unexpected bits of information, such as locating nearby traffic. These raised to 110 the total number of possible clearances that the pilot could be required to respond to.

Each session lasted 45 minutes. The order of presentation of the clearances was balanced across all 24 airline pilots who volunteered for the experiments. Clearances were spoken clearly and not too fast. To avoid the effects of pilot expectations, clearances did not follow a realistic flight scenario. That is, a clearance to descend could be followed by a clearance to maintain altitude but change course and speed. Nevertheless, words such as "reduce," "increase," "descend" and "climb" were used appropriately; clearances were restricted to realistic values only; and U.S. Federal Aviation Administration (FAA) heading/altitude rules were followed.

Bürki-Cohen said, "Another factor that might contaminate the results is the context and order in which the information is presented. Observing the constraints that altimeter readings follow altitude changes and frequency is given last, all possible combinations and orders of information were carefully counterbalanced across the three formats and [three] complexity levels.

"Here is an example of a clearance with five pieces of information in the grouped format:

"'Universal 1642. Reduce speed to two thirty. Fly heading zero four zero. Descend and maintain fifteen thousand. Houston altimeter twenty-nine fifty-two. Contact Houston Center on one thirty-two point twenty-two.'

"As you can see, heading information is always given sequentially.

"Here is another clearance at complexity level five [five bits of information], this time in restated format:

"'Universal 1642. Climb and maintain one six thousand, that's sixteen thousand. Revised Houston altimeter two niner niner niner. Fly heading two four zero. Increase speed to three one zero. Contact Houston Center on one one niner point seven five, that's one nineteen point seventy-five.'

"Based on discussion with controllers, we expanded the recent acceptance of restating altitude to include frequency."

Figure 1 shows the percent of errors for all information in all combinations of format and complexity. (Errors were defined as instances in which the readback or the control panel setting was wrong, or readback and setting were both wrong or both omitted.)

Of Figure 1, Bürki-Cohen said, "Pilots performed remarkably well, especially considering that we were testing unaided recall without a coherent flight scenario. The ... error rate never exceeded 4.2 percent. ...

"Figure [2] shows percent miscommunications summarized over all types of information, by complexity level for each format. Miscommunications include not only errors, but also requests for repeat (regardless of whether the readback or setting [was] correct), and the 42 instances where pilots set the correct number but omitted the readback that was mandatory in our experiment.

"In short, miscommunications include anything that taxes [ATC] resources, be it that controllers have to correct readbacks, repeat information or ask for a readback that they had [earlier] requested."

Figures 3 through 7 (pages 3–5) show experiment results for individual items of numerical information in the clearances:

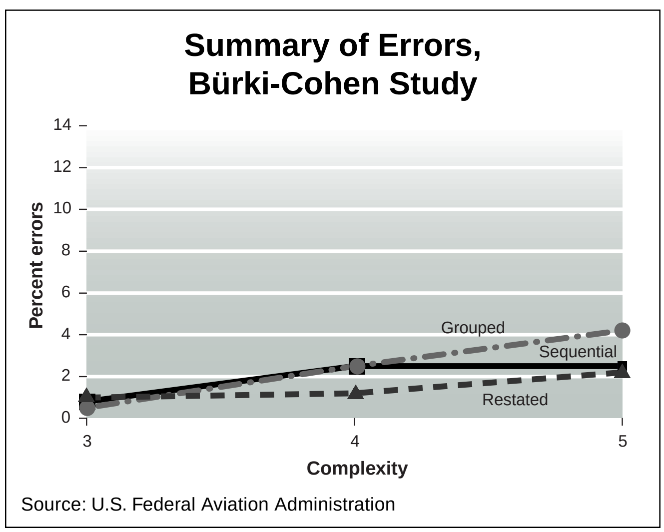


Figure 1

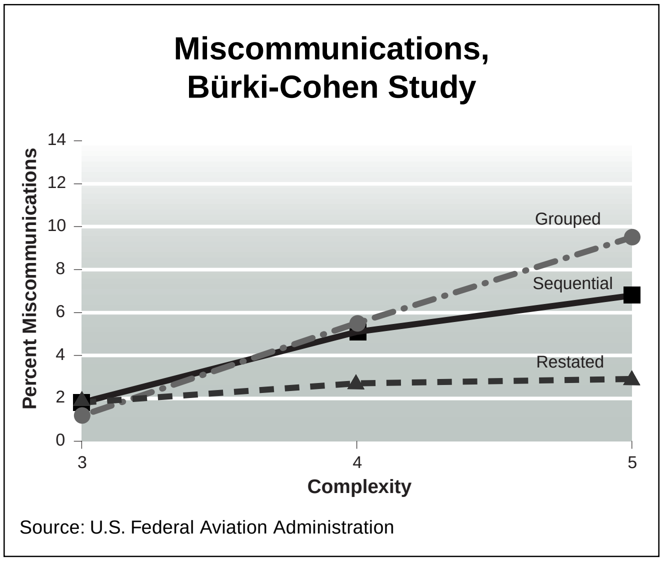


Figure 2

altitude, radio frequency, altimeter setting, heading and airspeed.

Figure 3 shows percent errors and percent miscommunication for altitude data. Bürki-Cohen said, "There were very few errors or miscommunications. The possible reasons for this excellent performance with altitude are threefold. First, altitude is arguably the most important piece of information in any clearance. Second, the numbers used in the low-sector, en-route environment simulated in our experiment cover a relatively small range" — 1,120 meters to 5,185 meters (4,000 feet to 17,000 feet) in 305-meter (1,000-foot) increments. " ... Third, the number itself, with maximum two positions and the 'thousand' remaining constant, do not represent a high memory load."

The results of the experiment for transmission of a radio frequency in a clearance are shown in Figure 4 (page 4). The frequency values ranged from 118.02 megahertz (MHz) to 112.37 MHz and from 123.67 MHz to 135.97 MHz in increments of 0.01 MHz.

In the sequential format, the percent of both errors and miscommunications for radio frequency dropped when complexity increased from four to five. In the grouped format, the percent of errors did not change with the increase in complexity. This phenomenon could be attributed to requests for repeats, which are indicated by the continued increase in percent miscommunications for the grouped format when complexity increased from four to five.

Altimeter settings were not given in clearances with only three pieces of information. In clearances with complexities of four or five, altimeter settings ranged from 982 hectopascals (hPa) to 1050 hPa (29 inches of mercury to 31 inches of mercury) in increments of 34 hPa (0.01 inch of mercury).

Figure 5 (page 4) shows the results of the altimeter experiment. Bürki-Cohen said, "Whereas in grouped format both error and miscommunication rates increase with complexity (the latter as high as 13.59 percent), in sequential format only miscommunications appear affected, reflecting again … that pilots more readily asked for a repeat at higher complexity levels."

Experiment results for heading information in clearances are shown in Figure 6 (page 5). Headings ranged from 10 degrees to 360 degrees in 10-degree increments, and the information was given only sequentially. Bürki-Cohen said, "Again, increased complexity appears to result in a modest decrement of recall."

Figure 7 (page 5) shows the experiment's results for airspeed information. Speeds given ranged from 389 kph to 574 kph (210 knots to 310 knots) in increments of 19 kph (10 knots). Bürki-Cohen said, "Speed is the only type of information that appears to show slightly better recall when it is said in the grouped format, at least with regard to miscommunications. A possible explanation is that grouping speeds helps distinguish

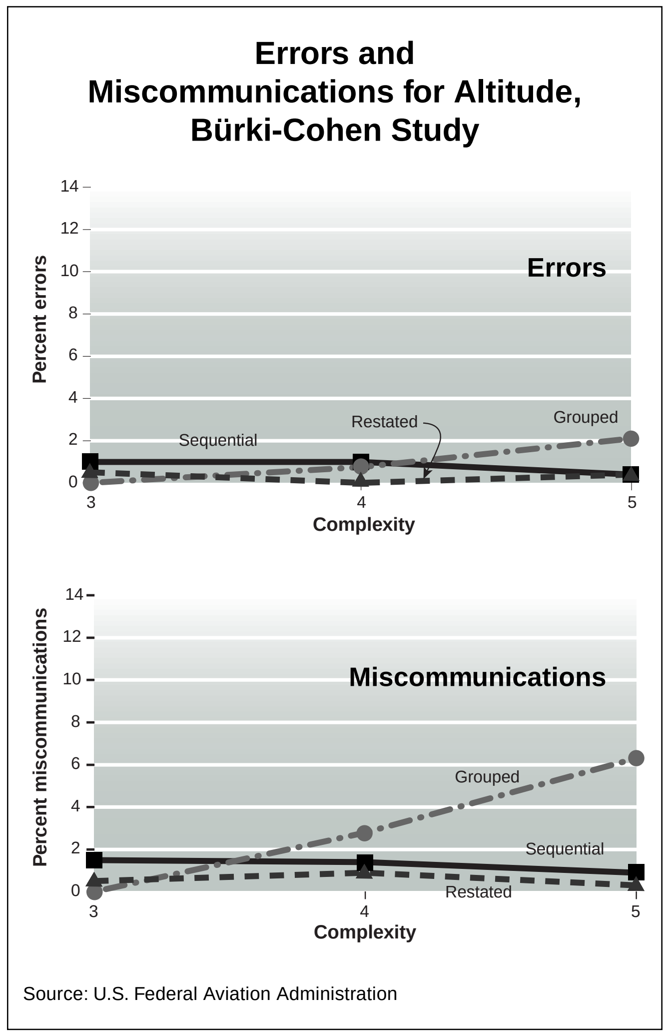


Figure 3

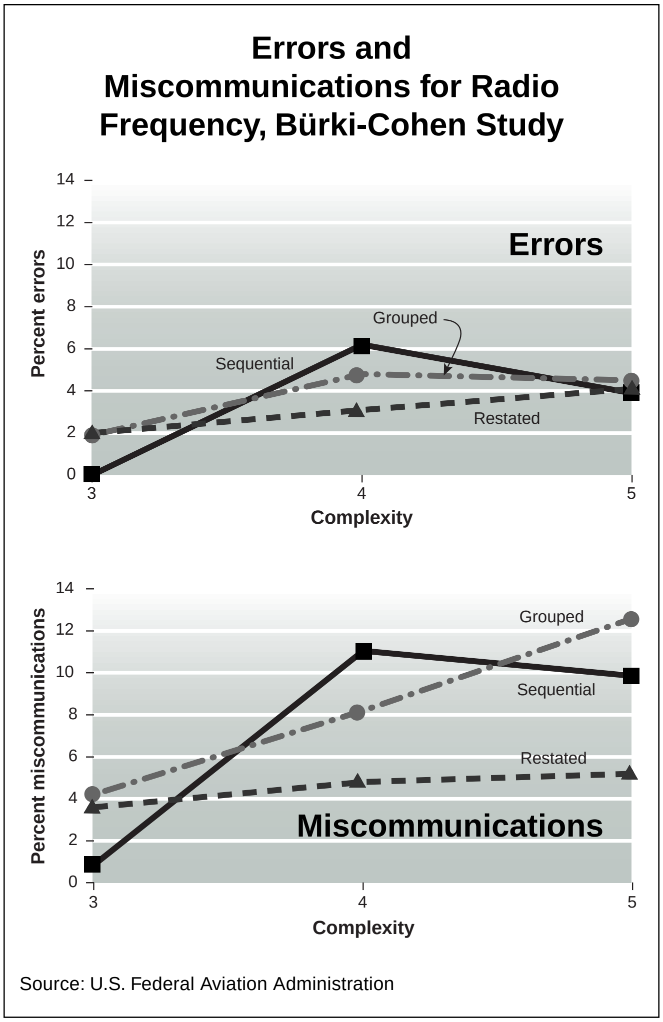


Figure 4

them from the always-sequential headings, with which they overlap in range. In other words, speed and heading are uniquely encoded in the grouped condition."

In conclusion, Bürki-Cohen said, "These results do not support the commonly held opinion that presenting numerical information in grouped format helps. They do support, however, the practice of restating information, possibly regardless of format. Moreover, controllers should be advised that presenting more than three pieces of numerical information in a single clearance may not save time, but lead to errors or ... requests for repeat."

The Gronlund report looked at air traffic control from the viewpoint of the controller. In the Gronlund report, the author said, "In the en route [ATC] environment, the complex dynamic system that confronts the air traffic controller [comprises] a large number of aircraft, coming from a variety of directions, at diverse speeds and altitudes, heading to various destinations. Like most complex, dynamic systems, this one cannot be halted periodically while the controller takes a brief respite. The ability to remain in control of such a ... system requires that the controller maintain situational awareness [SA]. ...

"Means et al.[4] conducted one of the few studies to empirically examine the role of memory in air traffic control. ... Controllers obviously have excellent memory for some information ... and poor memory for other information."

Gronlund was interested in the variables that affect memory for various pieces of aircraft flight data. He said, "One hypothesis was that the probability of recalling information about an aircraft was related to the *amount* of control exercised on the aircraft. ... Means et al. found that twice as much flight data [were] recalled about 'hot' aircraft (aircraft for which controllers exercised a great deal of control) than 'cold' aircraft. ... The second hypothesis was that the *type* of control exercised was related to the information recalled. For example, vectoring an aircraft was found to lead to better retention of its routing information."

Gronlund conducted two experiments to determine whether certain variables affected air traffic controllers' ability to remember flight information. The first experiment tested Means' hypothesis that the controllers' ability to recall flight information about an aircraft was directly proportional to the amount of control exercised over it.

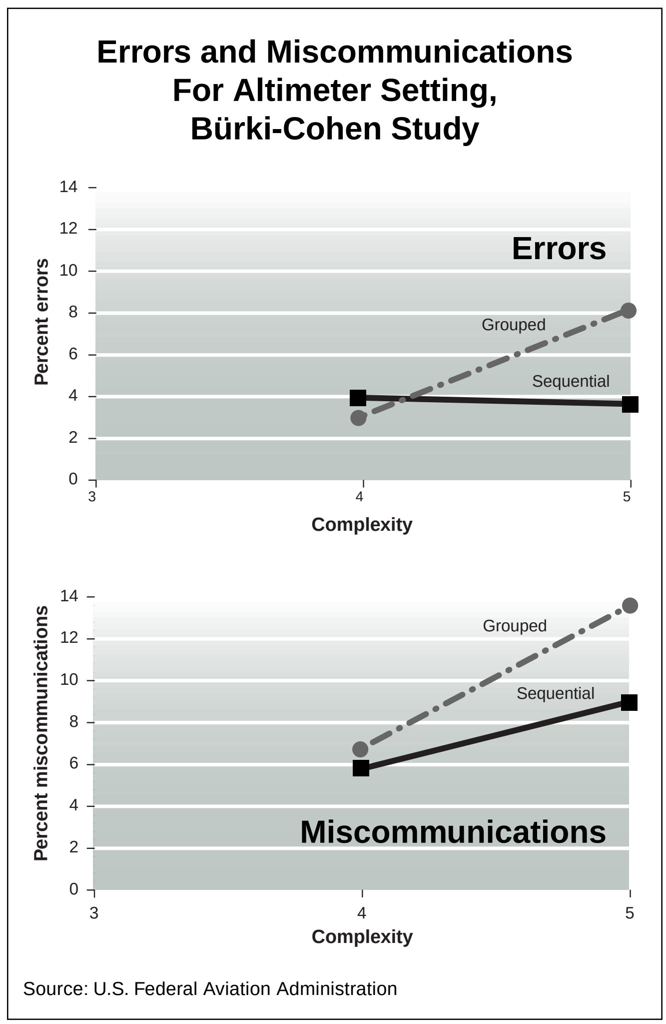
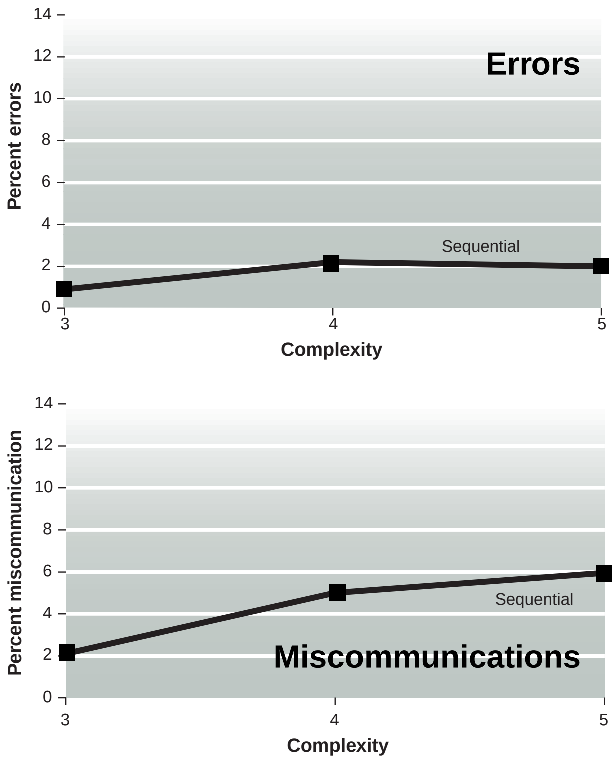


Figure 5

Errors and Miscommunications for Heading, Bürki-Cohen Study



Source: U.S. Federal Aviation Administration

Figure 6

Gronlund said, "We manipulated the number of [controller/pilot] interactions and the number of control actions to produce four experimental conditions, denoted Control 3, Control 1, Interaction 3 and Interaction 1.

"Control 3 aircraft received three control actions, Control 1 aircraft received one control action, Interaction 3 aircraft received three communications, and Interaction 1 aircraft received one communication. ...

"We predicted that controllers would recall more about the Control 3 and Interaction 3 aircraft than about the Control 1 and Interaction 1 aircraft. In addition, performance in the Interaction 3 condition might be better than Control 3 because the same altitude was interacted with three times for the Interaction 3 aircraft, but three different altitudes had been assigned to the Control 3 aircraft. ... In Experiment 1, we focused on altitude information because we knew it was important ... ."

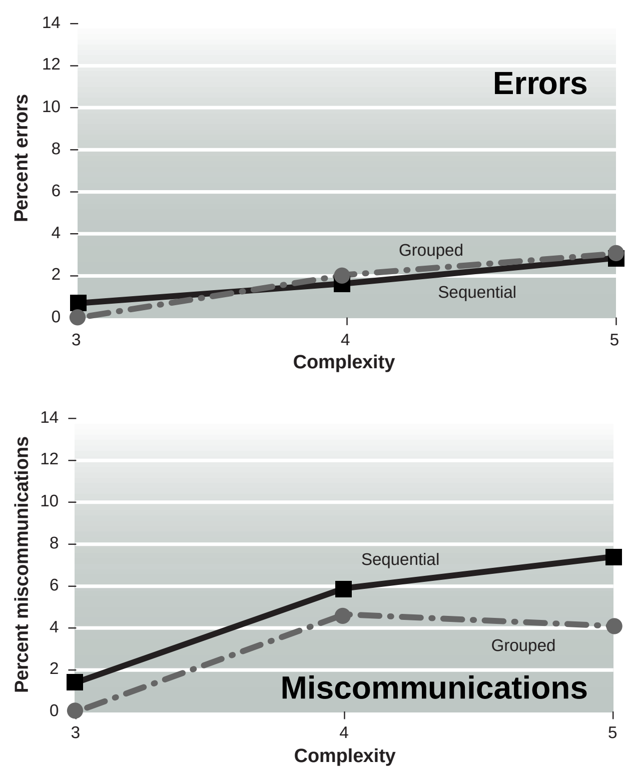
The second experiment amplified the first, attempting to learn just which specific flight information (call sign, altitude, destination, departure point, ground speed, aircraft type) was more readily recalled.

Eighteen full-performance-level (FPL) en route traffic controllers participated in Experiment 1. They had been FPL controllers for an average of 12.4 years, and all were FAA instructors.

Gronlund said, "The experiment was conducted at the Radar Training Facility at the Mike Monroney Aeronautical Center in Oklahoma City, Oklahoma [U.S.]. ... Communications between the controllers and the aircraft take place in the same manner as in the field, although the aircraft were piloted by 'ghost' pilots who controlled simulated aircraft based on the controllers' instructions.

"The equipment consisted of a radar display (the planned view display or PVD), a keyboard and trackball, and a computer readout display. The PVD shows the [two-dimensional] location of the aircraft with an attached data block containing information including the aircraft's call sign, altitude and ground speed."

Errors and Miscommunications for Airspeed, Bürki-Cohen Study



Source: U.S. Federal Aviation Administration

Figure 7

In addition, there was a rectangular paper flight progress strip (FPS) for each aircraft; the strip had 31 fields of information about the aircraft and its mission.

Gronlund said, “Participants worked the R-side, or radar position. Our subject matter expert worked the radar associate’s position and performed all its normal functions ([FPS] marking, communication with other centers, serving as a second pair of eyes to aid the radar controller).”

Three 30-minute flight scenarios were used; they were designed to be as realistic as possible. The scenarios involved a mean of 28.7 aircraft, including nine overflights, 8.7 arrivals and 11 departures. On average, 13 airplanes were displayed simultaneously on the radar screen.

Before the tests began, the participants practiced by answering sample questions. They were told that the scenarios would be stopped from time to time and that they would be asked similar questions about various aircraft. Otherwise, they were told to control traffic normally.

Three times during each scenario, the scenario was stopped and the participants were tested. One test required the participant to duplicate, without looking at the scope, the current radar display on a blank piece of paper.

Gronlund said, “Participants were extraordinarily accurate at their placement of aircraft on the paper sector map. Eighty-four percent of the aircraft recalled were placed within 2.5 centimeters of their actual location within [15 kilometers] eight nautical miles. Overall, the average missed distance was 1.5 centimeters [0.6 inch], or [the onscreen equivalent of] five nautical miles. Ninety percent of all aircraft were recalled.”

A question session followed the map test. Gronlund said, “Three types of questions were asked about a given aircraft, in the following order: (1) Informational — what was American [Airlines Flight] 123’s ... altitude (or ground speed, route, destination, departure point, or aircraft type); (2) metamemorial — rate your confidence in your answer (a range from 0 — absolutely no idea, to 100 — absolutely certain; (3) source — do you *remember* this information (memory was the source) or do you *know* it (answer was based on past experience).” For example, the controller might know that Southwest Airlines Flight 456 was a Boeing 737 because all of Southwest’s aircraft are B-737s.

The most important piece of information was altitude. Gronlund said, “Questions regarding altitude ... made up one-third of all informational questions.”

As shown in Figure 8, altitude was recalled correctly an average of 71 percent of the time, compared with an average of 42 percent recall for other flight data. Gronlund said, “These data do not support the notion of better memory for hot aircraft (Control 3 and Interaction 3) when hot was [defined as] the frequency of interaction or the frequency of control action.”

Gronlund said, “According to Experiment 1, whatever was strengthened by repeated interactions involving the altitude or repeated control actions changing the altitude, it was not the memory for altitude. However, frequent contact might result in increased familiarity of an aircraft’s call sign. Consequently, in Experiment 2, we checked to see if the call signs of aircraft that received more control actions were remembered better. ...

“Experiment 1 showed that what was done with an aircraft did not affect memory for its flight data. In Experiment 2, we [tried] to determine if the role the aircraft played affected memory for its flight data.”

Experiment 2 involved 14 FPL en route air traffic controllers with an average of 11.5 years of experience. Ten scenarios were created. Each scenario was built around a sequencing problem and required greater use of speed and separation control than the scenarios in the first experiment. Each scenario had a mean of 10.6 aircraft, of which 5.9 were overflights, 2.8 were arrivals and 1.9 were departures.

Aircraft were divided into three categories. “Traffic” aircraft were those that the participants were actively separating and monitoring to ensure that separation was maintained. “Not-traffic” aircraft comprised two aircraft that were physically close to each other but not in conflict. “Pretraffic” aircraft were those that might become traffic for one another in the near future. Typically, little might be known about

Experiment 1: Percent Correct for Altitude, by Condition, Gronlund Study

	Traffic	Control 1	Control 3	Interaction 1	Interaction 3	Overall
Altitude	80%	72%	66%	83%	83%	71%

Source: U.S. Federal Aviation Administration

Figure 8

Pretraffic aircraft because they had just entered the airspace.

Gronlund said, "The effect of two aircraft being in conflict in the Traffic condition should be to highlight some piece of flight data, increasing its likelihood of being recalled ... ."

"Twelve aircraft were tested, six that were not on the PVD (called distractors) and six that were (targets). All six of the target aircraft were under the control of the controller. ... There were two targets from each of the three conditions (Pretraffic, Traffic and Not-traffic)."

A scenario was stopped at a predetermined time. The participant turned away from the PVD and completed two tasks. In the first task, the participant was given an aircraft call sign and asked whether that aircraft was on the PVD at the time the scenario was stopped. Performance in call-sign recognition was better for aircraft in the Traffic condition than in either of the other two conditions.

In the second task, the participant was provided with a paper copy of the sector map that showed the location of each aircraft and its call sign. The participant was asked six questions in random order about each aircraft: altitude, ground speed, current altitude status (level, climbing or descending), relationship to the sector (arrival, departure or overflight), direction of flight and destination.

Gronlund said, "[Figure 9] shows the accuracy (percent correct) for each question type for each condition. ... There were no differences among conditions for the altitude question. As in Experiment 1, the greater number of altitude control actions for the Traffic aircraft did not result in better recall for altitude. ...

Experiment 2: Percent Correct for Question Type and Each of Three Conditions, Gronlund Study

	Not-traffic	Pretraffic	Traffic
Altitude	66%	67%	69%
Ground speed	19	25	29
Altitude status	89	94	82
Relationship to sector	83	96	97
Direction of flight	82	82	75
Destination	51	47	93

Source: U.S. Federal Aviation Administration

Figure 9

"For altitude status, we found that performance was best in the Pretraffic condition, next best in the Not-traffic and worst in the Traffic condition; for relationship to sector, Not-traffic was worse than the other two conditions; for ground speed, Not-traffic was worse than Traffic. The only question type for which the Traffic condition was significantly better than both the Not-traffic and Pretraffic conditions was destination.

"Unfortunately, this result was probably an artifact. Performance for the Traffic aircraft was inflated because both Traffic aircraft always had the same destination; that was why [they] had to be sequenced."

Gronlund's conclusions included:

- "The increased number of control actions initiated on Traffic aircraft did affect memory. It improved recognition of the call sign of the aircraft. It did not, however, improve memory for the flight data from that aircraft;
- "Flight data from the Traffic aircraft were not the best remembered. This was contrary to expectations ... ; [and,]
- "The overall low level of performance for ground speed was surprising given that these scenarios were designed to require the use of speed control. However, the poor memory for the exact speed might be caused by the phraseology controllers use. Although controllers instruct pilots to climb or descend to a *specific* flight level, they often tell them to increase or decrease their speed by (for example) 10 knots. Consequently, the controllers remember exact altitude fairly well because they interacted with altitude information, but because they did not deal with exact speed, they do not remember it."

Unsupported were the hypotheses that flight data about hot aircraft would be recalled better, and that the type of control exercised would affect what flight data were recalled.

In both studies, the authors indicated that further research is necessary. Bürki-Cohen is preparing an experiment that will test whether the effects of the format and complexity of a clearance message interact with the rate at which the words are spoken.

Gronlund said, "Why were we unsuccessful in finding variables that affected the recallability of flight data?

"We consider four possibilities. One possibility is that we have yet to discover the correct variables that affect recall. ... A second possibility ... was because memory for the flight data was so vital to task performance that the flight data were not highlighted further by these manipulations. ... A third possibility is that memory is irrelevant to the performance of the controller and, consequently, irrelevant to SA. ... The final possibility we consider is that memory is important to air traffic control and SA, but the wrong measures were used in these studies.”♦

References

1. Bürki-Cohen, J. “How to Say It and How Much: the Effect of Format and Complexity on Pilot Recall of Air Traffic Control Clearances.” *Methods and Metrics of Voice Communications.* Report no. DOT/FAA/AM-96/10. Civil Aeromedical Institute, U.S. Federal Aviation Administration. March 1996.

2. Gronlund, S.D. et al. *The Role of Memory in Air Traffic Control*. Report no. DOT/FAA/AM-97-22. Civil Aeromedical Institute, U.S. Federal Aviation Administration. November 1997.

3. Cardosi, Kim M. *An Analysis of En Route Controller-Pilot Voice Communications.* Report no. DOT/FAA/RD-93/11. U.S. Volpe National Transportation Systems Center, Cambridge, Massachusetts. U.S. Federal Aviation Administration. March 1993.

4. Means, B. et al. *ATC Training Analysis Study: Design of the Next-generation ATC Training System.* Report no. FAA/OPM 342-036. U.S. Federal Aviation Administration. 1988.

Further Reading from FSF Publications

Koenig, R.L. “Excess Words, Partial Readbacks Score High in Analysis of Pilot-ATC Errors.” *Airport Operations* Volume 23 (January–February 1997).

Cushing, S. “Pilot–Air Traffic Control Communications: It’s Not (Only) What You Say, It’s How You Say It.” *Flight Safety Digest* Volume 14 (July 1995).

Visit our World Wide Web site at http://www.flightsafety.org

AIRPORT OPERATIONS

Staff: Roger Rozelle, director of publications; Rick Darby, senior editor; Daniel P. Warsley, senior editor; Joy Smith, editorial assistant; Todd Lofton, editorial consultant; Karen K. Ehrlich, production coordinator; Ann L. Mullikin, assistant production coordinator; and David A. Grzelecki, librarian, Jerry Lederer Aviation Safety Library.

Subscriptions: US$60 (U.S.-Canada-Mexico), US$65 Air Mail (all other countries), six issues yearly. • Include old and new addresses when requesting address change. • Flight Safety Foundation, 601 Madison Street, Suite 300, Alexandria, VA 22314 U.S. • Telephone: (703) 739-6700 • Fax: (703) 739-6708