Human Transient Into-the-Loop Simulation for NGATS

Thomas B. Sheridan[1] and Judith Bürki-Cohen[2]
Volpe National Transportation Systems Center, US Department of Transportation, Cambridge, Massachusetts, 02142, United States

Kevin Corker[3]
NASA, San Jose State University, San Jose, CA, 95192, United States

Serious planning is underway for the United States Next Generation Air Transportation System. While design decisions are far from complete, there is consideration of having aircraft-trajectory control in space and time be determined by some combination of pilots, air-traffic personnel, and computers. Insofar as any control input is continuous in time, because of human perception and decision response times and/or computer multitasking or optimization-cycle times, there are likely to be time delays (e.g., human transients getting "into the control loop") that could threaten control stability. Using the MATLAB® Simulink® dynamic simulation tool this paper examines some hypothetical situations and offers examples of how delays and signal sampling in aircraft control could cause unacceptable oscillations and instability.

Nomenclature

ADS-B = Automatic Dependent Surveillance-Broadcast
ATM = Air Traffic Management
D = transport delay
Fcn = Function
GPS = Global Positioning System
H = hold period
JPDO = Joint Planning and Development Office
K = Gain of controller element (as contrasted to gain of control loop due to multiple dynamic elements)
NGATS = Next Generation Air Transportation System
PIO = Pilot Induced Oscillation
s = Laplace variable
T = aircraft lag-time constant
TCAS = Traffic Collision Avoidance System
Wx = weather

I. Introduction

THERE is a need to accommodate the increase in traffic anticipated for the 2015-2025 period, with hardly any change in airport capacity and a potential reduction in air-traffic-management (ATM) personnel (Fig. 1). An interagency Joint Planning and Development Office (JPDO) for the Next Generation Air Transportation System (NGATS) has envisioned a radically restructured air-traffic-control system from what now exists in the United States. Based on nominal schedules, weather conditions, fuel considerations, global positioning system (GPS) and automatic dependent surveillance-broadcast (ADS-B) capabilities and other constraints, a computer "Evaluator" would determine (near) optimal 4D-trajectories (three in space, one in time) for all enroute phases of flight (climb, cruise, and descent) that would be flown mostly automatically. A surface "Evaluator" would correspondingly determine taxi trajectories. The roles of both flight crew and ATM personnel would change significantly in the direction of their becoming monitors of automation, presumably intervening when anomalies or emergencies occur

---

[1] Senior Transportation Fellow, Human Factors Division, RTV-4G, 55 Broadway
[2] Engineering Psychologist, Human Factors Division, RTV-4G, 55 Broadway, Member
[3] Associate Dean of Engineering, One Washington Square.

that the automation cannot resolve.[1,2] Whether in the air or on the ground, the circumstances under which and the degree to which control would be continuous and be delegated to entities outside the aircraft have yet to be determined. Anticipating at least some continuous remote control of aircraft, however, this paper considers the dangers.

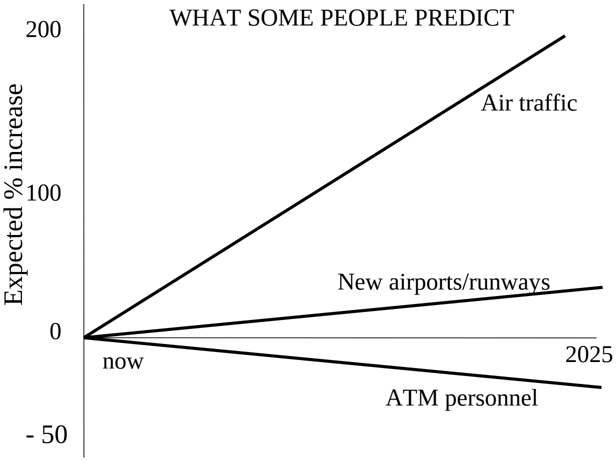


Figure 1. Rough estimate of future demands. *The y-axis shows the estimated percent increase in traffic, the x-axis shows the time from now to 2025.*

II. Human Response Time and Control Loop Instability

Experience in human-automation interaction in aviation, as well as nuclear power and other complex systems where the public is at risk, has demonstrated that humans are poor monitors and slow to comprehend unexpected events, especially when they have not been actively operating within the control loop. Humans need significant time to comprehend the state of the system and get "into-the-loop" to make changes.[3] Probability of time required to achieve a correct response tends to follow a lognormal function of time from event initiation, meaning that delay time for reasonable statistical confidence may be much longer than what is acceptable. Figure 2 illustrates the shape of a typical probability density function for human response time. Figure 3 shows a lognormal model (straight line on plot of x = Gaussian fractiles of cumulative probability in percent, y = log probability of response time in seconds) that makes an excellent fit to experimental data for time taken by nuclear reactor controller teams to recover from a sudden unexpected loss of coolant in a nuclear plant simulator.[4] Each point is a different three-person team undergoing refresher training.

Such delays that are within a control loop can cause instability. For readers not conversant in control theory, a fundamental principle is that whenever feedback control is positive rather than negative *and* the controller-loop gain exceeds unity, then any energy circulating in the loop is continually amplified and the system goes unstable; control error will grow without bound. Such instability can occur in two ways: In the less likely case, there will be positive feedback without oscillations (the roots of the characteristic equation are positive real). In the latter case, the system response will tend smoothly and exponentially to infinity. In the second and more likely case, some components of the system dynamics will cause oscillatory response (there are some complex conjugate roots of the characteristic equation). In this case, feedback will necessarily be positive for any frequency component such that the closed-loop time delay corresponds to one half cycle and the gain of any control loop exceeds unity. Then oscillations will grow without bound.

This instability scenario is familiar for dynamic systems that are controlled by humans. One example occurring for hand-flown aircraft is the much analyzed Pilot Induced Oscillation (PIO).[5] This phenomenon, typically of concern in pitch control of high-performance fighter aircraft, arises from an interaction between the pilot response, the aircraft dynamics, and an oscillation trigger in form of a transient external disturbance, such as atmospheric turbulence. An inexperienced pilot's attempts to minimize time delay or deviations from a pitch reference, or some nonlinear element of the aircraft handling loop, can set up a combination of gain and phase that produces the oscillation (as described above).

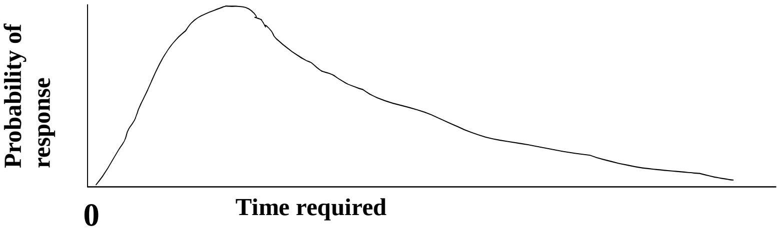


Figure 2. Typical probability density for human response time.

This instability problem has also been experienced in manual control of systems over a communication channel that has a "transport" time delay (e.g., positioning a robot arm in space, where there is significant signal transmission delay due to limited speed of radio waves,[6] or positioning a telemedical diagnostic camera operating over the internet, where the delay is caused by computer signal processing when sending and refreshing high-resolution images). Another example is human-operator control of water level in the steam generator of a nuclear-power plant, where bubble formation causes an effective time delay and distorts the operator's observation of the measured water level.

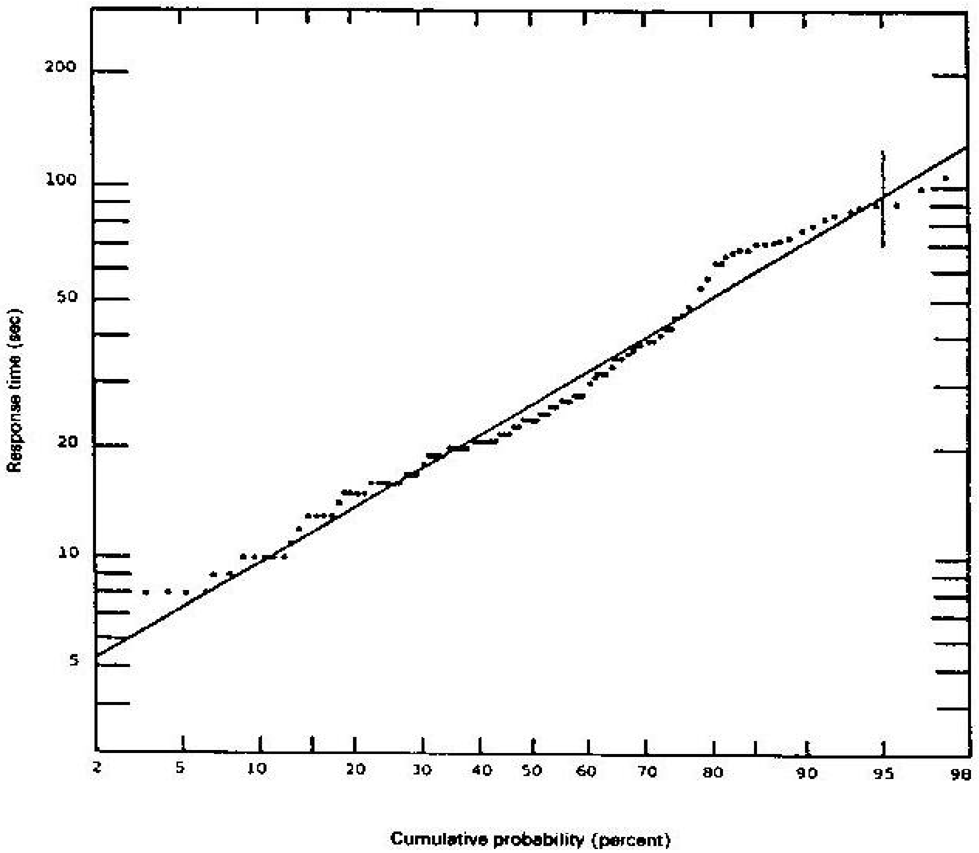


Figure 3. Lognormal fit to nuclear-reactor-team accident recovery times.[4]

III. Where Control-Loop Delays Could Occur in NGATS

Where, specifically, might such delays due to human or computer response time manifest themselves in the NGATS? Here we are not concerned with small time delays and fast, sensitive adjustment of control signals by the pilot in a tight control loop as in a fighter aircraft. Rather, we are concerned with aircraft that have longer time lags, controlled by humans making intermittent control decisions with correspondingly longer time delays while trying to achieve continuous aircraft control. Figure 4 shows the hypothetical scenario of two aircraft (there likely would be many more) being controlled by ATM. For each aircraft, in addition to an inner control loop (heavy lines), there is at least one outer control loop (light solid lines for computer, light dotted lines for human controller) imparting intermittent correction signals. ADS-B is a GPS-dependent aircraft surveillance system. Wx represents weather reporting. Note that in Fig. 4 there is also coupling between the aircraft loops due to the common outer-control agents. This paper will examine only the case of a single outer loop combined with a single inner loop, though the effects of more outer loops can make the problems worse.

Figure 5 shows the topology of air-traffic control (of the current system; the exact nature of NGATS topology is yet to be determined) with several layers of outer loops that potentially could be adding correction signals as shown.

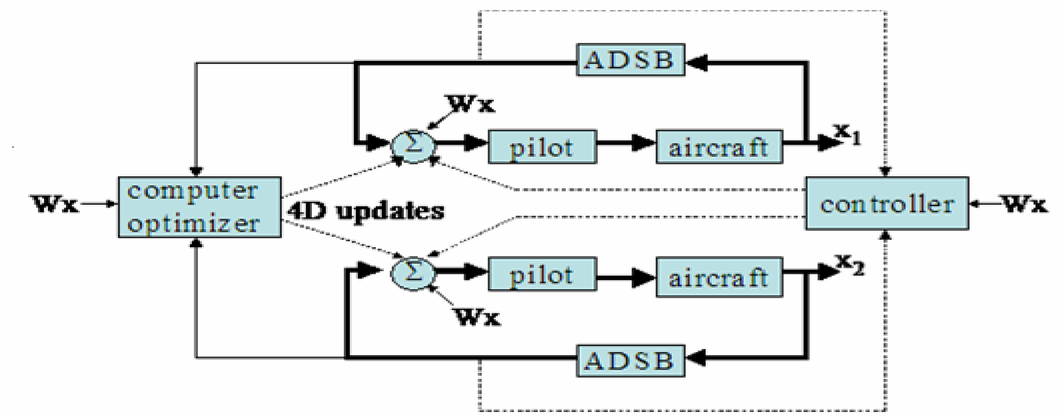


Figure 4. Combined inner (heavy lines) and outer (light lines = computer, dotted lines = human) control loops, and coupling of same, of two aircraft controlled by ATM. *ADS-B is a GPS-dependent aircraft surveillance system. Wx represents weather reporting. Note that there is also coupling between the aircraft loops due to the common outer-control agents*

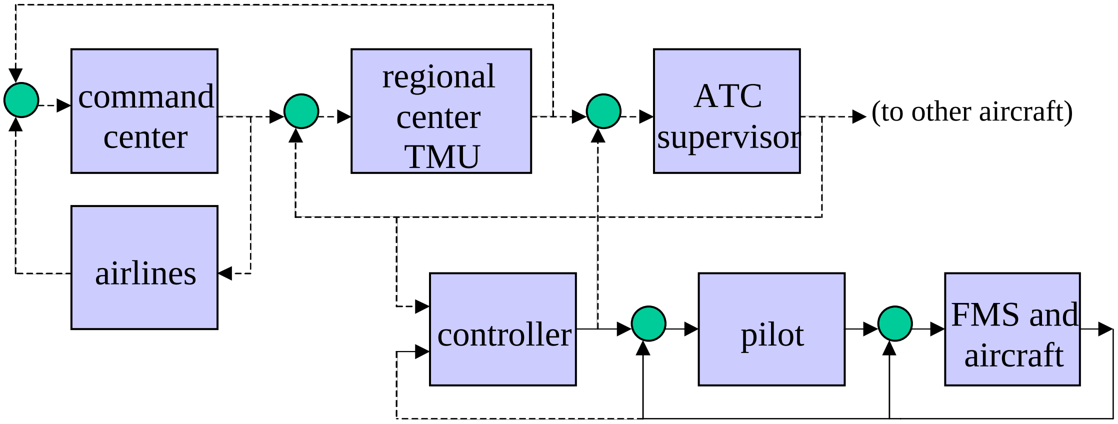


Figure 5. Topology of air traffic control. *Solid lines constitute the inner loop control of the aircraft by the pilot or FMS in roll, pitch, yaw and airspeed. Dotted lines are vectoring corrections that occur on a slower time scale than pilot/FMS control.*

Such air-traffic-control loops could involve data-link communication between either an ATM person or computer on the ground and either a pilot or FMS on-board the aircraft. Figure 6 shows the 2x2 array of such combination possibilities, any of which can pose a delay due to human and/or computer information processing. Hence the potential for response delay in control of any one aircraft.

		Agent on Aircraft	
		Human	**Computer**
Agent on Ground	**Human**	Voice communication with some keyboard entry/computer display	Keyboard entry, computer display
	Computer	Keyboard or graphical data entry, computer display	Datalink

Figure 6. A 2x2 array of communication links within control loops.

Perhaps more likely than continuous time delay is so-called sample-and-hold behavior, wherein the human controller observes the state of the variable being controlled, inputs a constant value (zero-order hold) or constant rate signal (first-order hold) as the control variable, which is maintained until the next sampling operation. In NGATS we are mostly concerned with manifestations of this type of behavior, and will demonstrate below that the sample-hold effect on control is very similar to that of the continuous time delay.

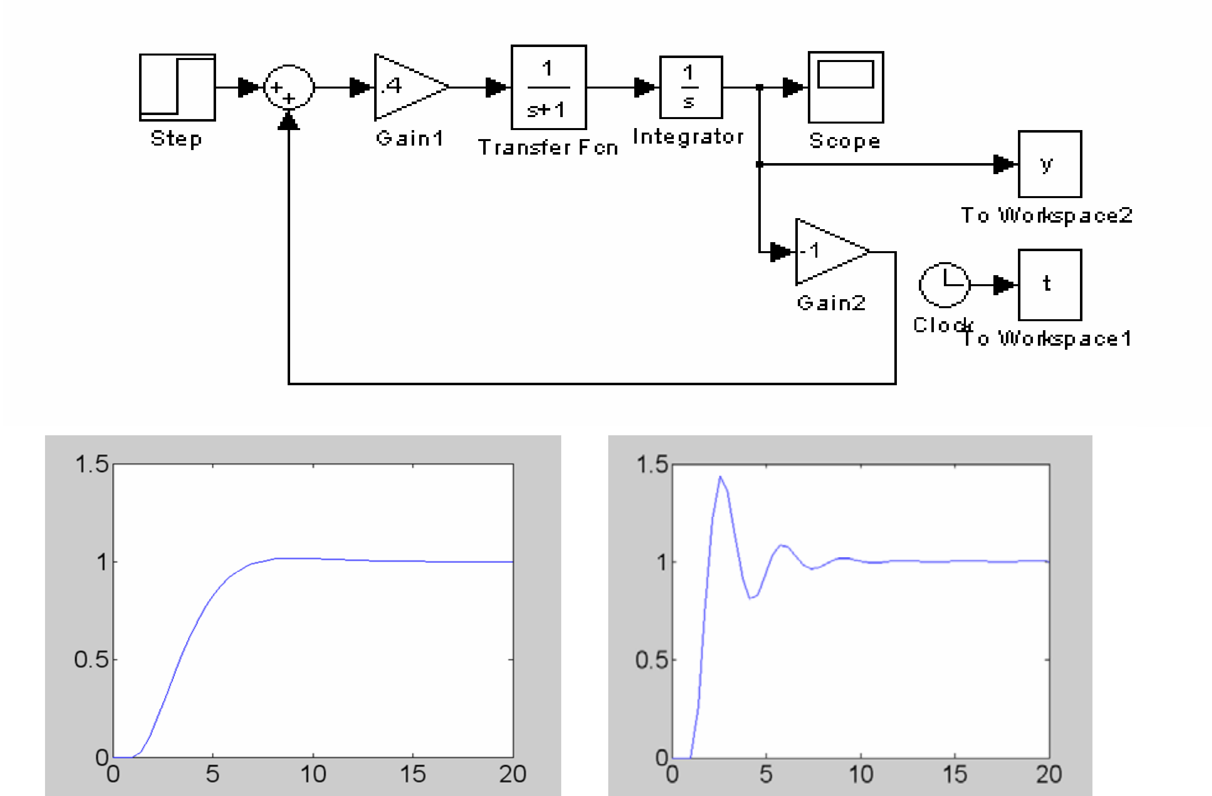
Autonomous computer control of aircraft from the ground might also pose time delays, due to limited bandwidth of datalink systems that must time-share with many up- and downlinked variables as well as sample across a large number of aircraft. One cannot be precise about bandwidth limits at this time. But experience with earth-to-low-space-orbit datalink has been discouraging. Here delays due to multiplexing and other data processing (neglecting speed of light considerations) pose round-trip time delays in the order of several seconds, and therefore preclude continuous control of space robots from the ground.

Aircraft have net lagging dynamics, meaning higher frequency components of circulating signals are damped to below unity gain. This damping is usually sufficient to prevent small response delays (originating from the autopilot or a manually flying pilot) from driving the inner loop unstable. If, however, after receiving communications from the outer ATM-loop (via human or computer elements), pilots repeatedly reset the autopilot settings, this may have the troublesome effect of a loop time delay (see below). We do not yet know what form these outer loops will take, whether they will be continuous or intermittent, and on what basis the outer control loops may be closed. This paper can only illustrate some dangers.

IV. Simulation Results to Show Some Effects of Loop Delays

A. Simple Generic Model

Figure 7 is a generic dynamic-system model of a controlled aircraft responding to a unit step change in set point or reference input. This is a reasonable approximation to a new command to climb, turn, or change speed. Neglecting the time constants for the different degrees of freedom, the two blocks [1/(s + 1)] and (1/s) (in conventional Laplace transform notation) are an approximation to the dynamic transfer function of an aircraft, from control element input to attitude (pitch, roll, or yaw) or longitudinal position output (including the fly-by-wire electronics). The first order lag [1/(1 + s)] approximates the transient change in angular or translational rate in response to control actuation, while the integrator (1/s) is the rate-to-position transfer function. A controller-gain (coefficient) of 0.4 (representing a human pilot in continuous manual flying or an autopilot) nets a smooth closed-



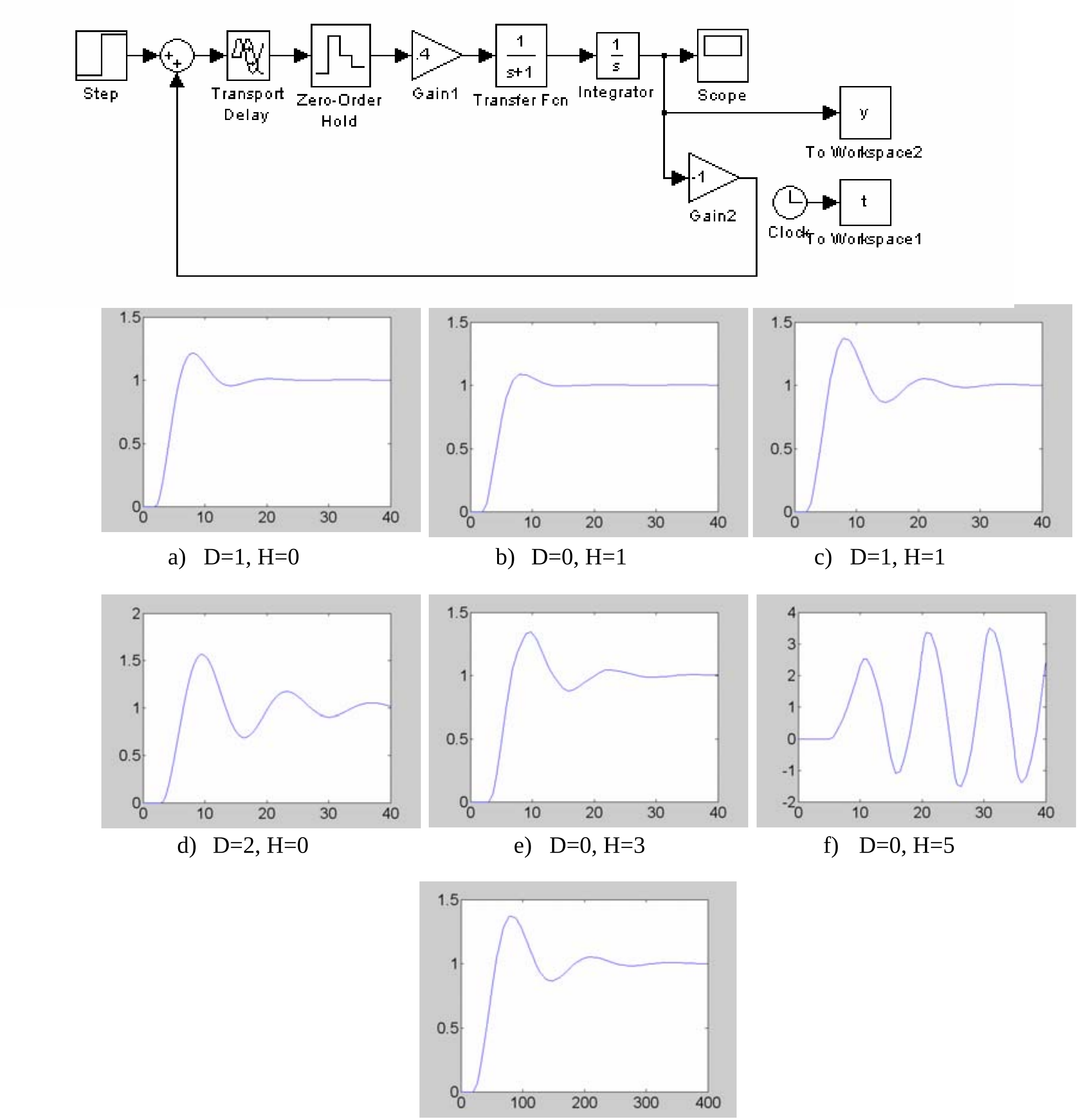
a) Controller gain of 0.4 results in smooth response

b) Controller gain of 4.0 results in unacceptable response

Figure 7. Simulink results for aircraft with coefficient control element in single loop.

loop transient response, as one would expect. This is shown in Fig. 7a using a computer simulation offered by Simulink,® a toolbox of MATLAB.® The time scale (horizontal axis) of the plots is in seconds to fit the dynamics chosen. The vertical axis is scaled relative to the unit step input. The [-1] element provides the essential negative feedback. The “scope,” “y,” “clock,” and “t” blocks are there to make Simulink work but do not enter into the loop dynamics. For the moment we will consider the dynamics of the human pilot or autopilot (other than gain) to be negligible relative to the aircraft dynamics, or to be subsumed by the latter. Figure 7b shows the response when gain is increased by a factor of 10, causing an unacceptable overshoot.

In Fig. 8, we demonstrate the effects of transport delay and sample/hold on the control stability of the reference aircraft dynamics and reference control gain. As seen in Fig. 7, Simulink allows easy change of parameters, so



a) D=1, H=0

b) D=0, H=1

c) D=1, H=1

d) D=2, H=0

e) D=0, H=3

f) D=0, H=5

g) D=10, H=10, T=10, K=0.04

Figure 8. Simulink results for aircraft with transport delay and/or sample and zero-order (constant) hold in single loop. *D = transport delay, H = hold period, K = controller-loop gain, T = aircraft-lag time constant. Note the longer time axis in Fig. 8g, where T=10 and the K=0.04. For Fig. 8a-f, T=1 and K=0.4 as a comparison.*

various combinations have been selected to demonstrate combinations of time parameters of the control elements on both sides of the instability margin. Figures 8a and 8b show that the step response overshoots slightly when delay and hold times are the same as the aircraft-lag time for the given controller gain. The transport delay causes slightly more overshoot. Figures 8c through f show that oscillations increase as delay and/or hold times increase, hold alone again having somewhat less effect than delay alone (compare Fig. 8d and e). When hold lasts long enough (Fig. 8f), the step-response oscillations increase with time, resulting in unbounded instability. This would also be true for delay.

It is important to note that pitch and roll response would be quick, yaw somewhat slower, and longitudinal (speed) changes much slower, due primarily to the different inertias as well as the different effects of the aerodynamic surfaces in these degrees of freedom. For similar reasons, response of a large aircraft will be considerably slower than that of a small aircraft.

B. Stability Analysis

Analysis for different degrees of freedom and different aircraft sizes is made easy because the form and stability of the response are determined by the temporal parameters of the control elements relative to the temporal parameters of the aircraft dynamics. Here we are only concerned with the form and stability of response, not the time constants per se, so we need only explore the effects of time delay in the control elements *relative* to the reference-aircraft model dynamics and control gain.

For example, for open-loop system dynamics of the form K/[(Ts + 1)s], where in the current case controller-gain coefficient K=1 and aircraft-lag time constant T=1, the closed-loop transfer function can be shown by conventional system analysis to be $1/[(T/K)s^2 + (1/K)s + 1]$, the undamped natural frequency is $(K/T)^{0.5}$ and the damping coefficient is $1/[2(KT)^{0.5}]$. If K is set to equal 1/T, then the damping of any closed loop transient response is constant, while the undamped natural frequency is proportional to 1/T.

Happily for our analysis, this relation extends to the gain in relation to *all* the time delays in the loop. This can be seen by comparing Fig. 8g to Fig. 8c, where in Fig. 8g the time constants of not only aircraft lag but also transport delay and sample/hold were increased by a factor of 10, while the controller-gain coefficient was reduced by a factor of 10, relative to the parameters of Fig. 8c. The plots appear identical in form (damping), except that the Fig. 8g transient is ten times longer than that of Fig. 8c. If D, H and T were all equal to 100 and K were 0.004, the same plot would result. *Thus the effects of delay and sample-hold parameters must be evaluated in relation the controller gain K.*

Figure 9 shows how an analysis can be made in terms of the control engineer's Bode plot. This is a dual plot of the open-loop gain and phase shift as a function of frequency ω. Scaling gain and frequency logarithmically as shown has the advantages that for the integrator and the first-order lag (representing aircraft dynamics) gain and phase shift plots can be approximated by straight lines, and effects of these elements plus the controller gain and the delays in the open loop are additive. The critical issue for stability is the value of the resultant open-loop gain when

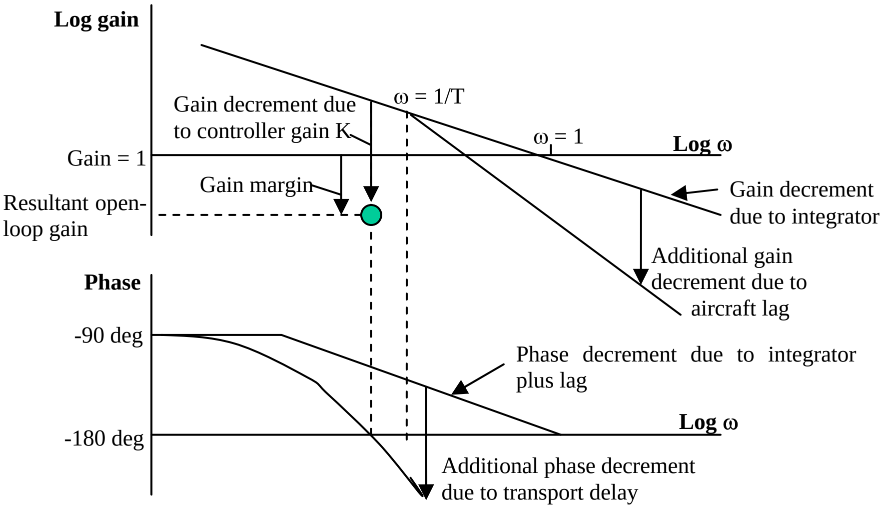
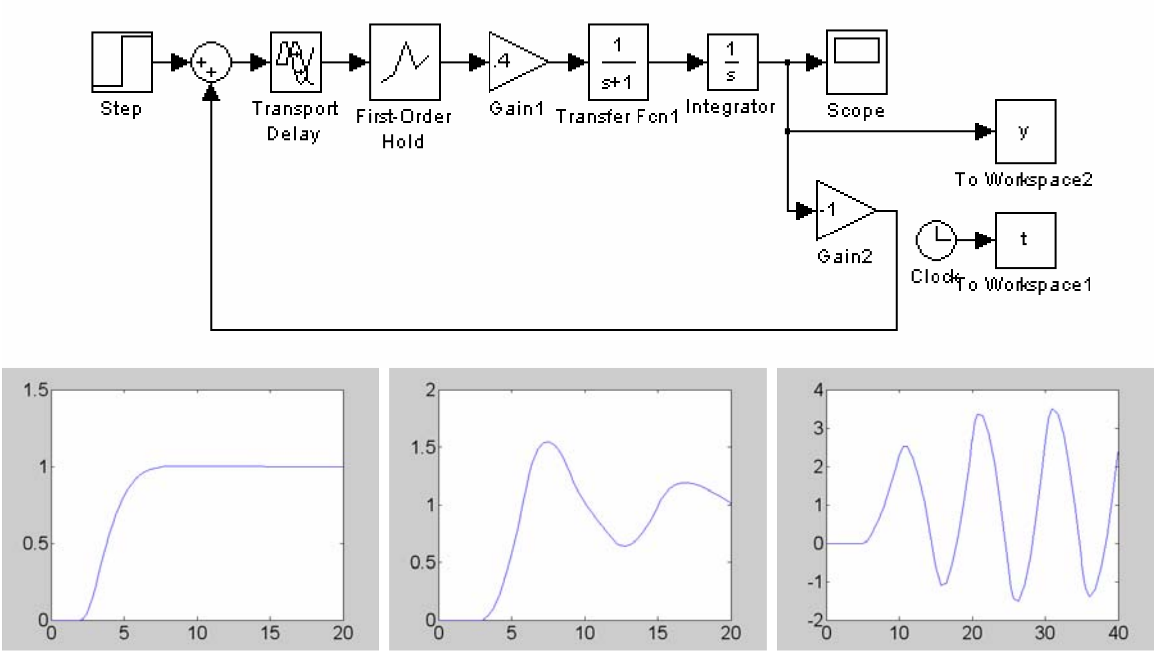


Figure 9. Open-loop gain-phase (Bode) plot showing tradeoff between delay and controller gain.

the open-loop phase is 180 degrees, e.g., whether it exceeds 1 so that energy within the closed loop will regenerate. For the example shown, the resultant open-loop gain (small filled circle) is on the safe side of (less than) one by an amount normally referred to as the *gain margin*. One can observe that as transport delay is increased, the phase crossover of 180 degrees shifts to a lower frequency and the controller gain must be further lowered to maintain a comfortable phase margin.



a) One loop, rate hold, D=0, H=1

b) One loop, rate hold, D=0, H=3

c) One loop, rate hold, D=0, H=5

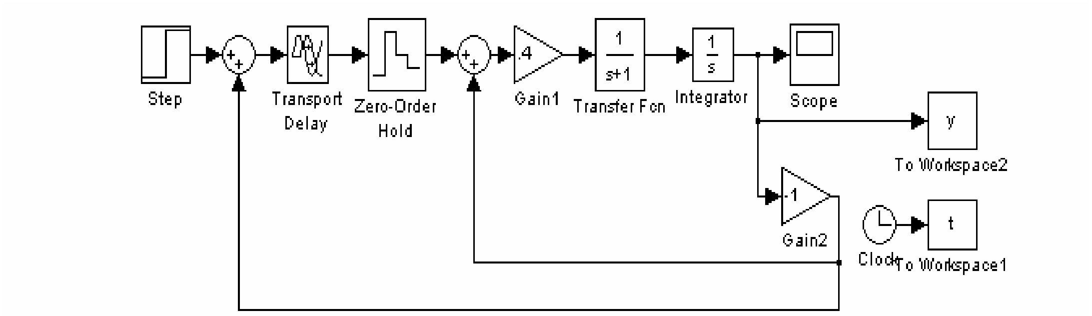
Figure 10. Simulink results for aircraft with transport delay and/or sample and first-order (rate) hold in single loop. *Note the longer time axis in Fig. 9c.*

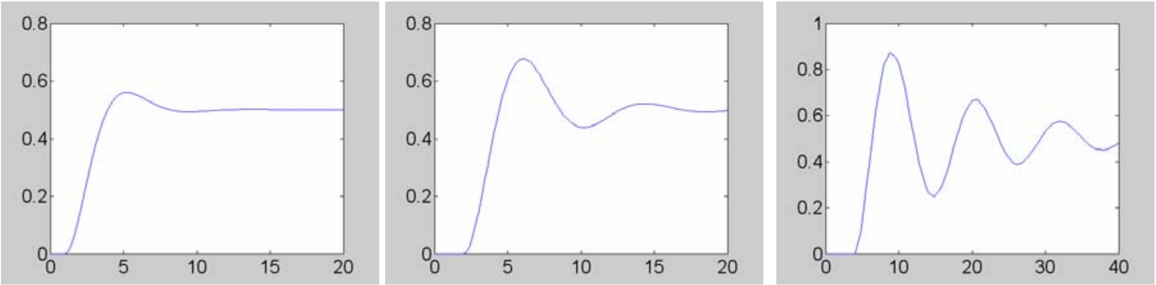
In Fig. 10, we substitute a sample and first-order (rate) hold for the zero-order sample and hold. This results in smoother response for the same hold time than with the zero-order sample/hold, since the rate sample/hold anticipates the rate of change of error. Figure 10a shows a smoother transient than Fig. 10b, where, as the rate-hold time increases, the step response is seen to oscillate, yet still to be somewhat damped. Eventually, with increasing hold time, it becomes positively unstable (Fig. 10c).

C. Two-loop Examples

All three of the model configurations in Fig. 10 were single-loop systems, representing either an aircraft that is manually flown by a pilot (within the aircraft or remotely from the ground) or a possible equivalent where continuous control signals are being fed from an ATM computer. In Fig. 11, we consider the possibility of two-loop control, where the inside loop is continuous and by itself would be well behaved, as the equivalent model of Fig. 7a, but where there is also an outer loop that observes the control error (difference between aircraft state and commanded step reference input) and continuously feeds correction signals to the inner loop. This outer loop might represent a human or computer air-traffic-control agent continuously telling the pilot (or the FMS) what to do. As suggested earlier, in NGATS the interactions between ATM (human or computer), pilot, and FMS have yet to be defined, so these models are purely hypothetical, not modeling anything that will necessarily exist.

However, one example where dual-control loops plus time delay or sample/hold came into play is in TCAS advisories. The mid-air collision in July, 2002, over Ueberlingen in Germany demonstrates this. In that accident, where TACS advised a B757 to descend (which the pilot did) and a TU154 to climb, the controller gave last-minute instructions to the TU154 to descend. The TU154 pilot decided to obey the controller rather than the TCAS advisory (Russian policy is for the pilot to decide which to follow, while the European B757 pilot had been instructed to defer to TCAS). The result was tragic. The outer loop of ATC intervention clearly interfered with the safety alert in the tight inner loop. NGATS may provide many more such opportunities.

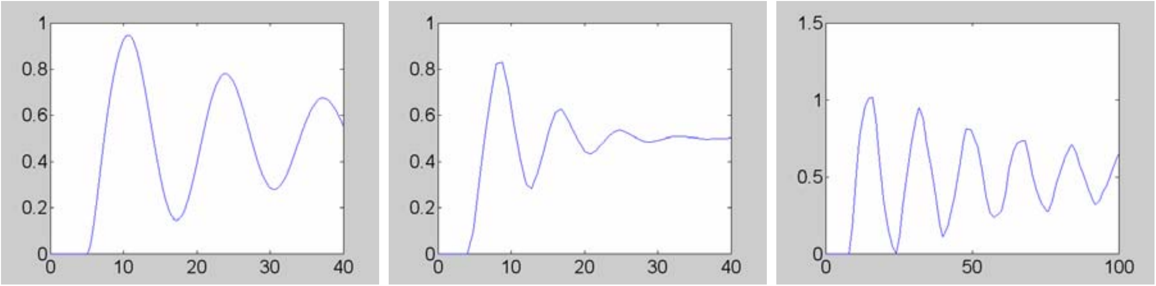




a) D=0, H=0

b) D=1, H=1

c) D=2, H=2



d) D=4, H=0

e) D=0, H=4

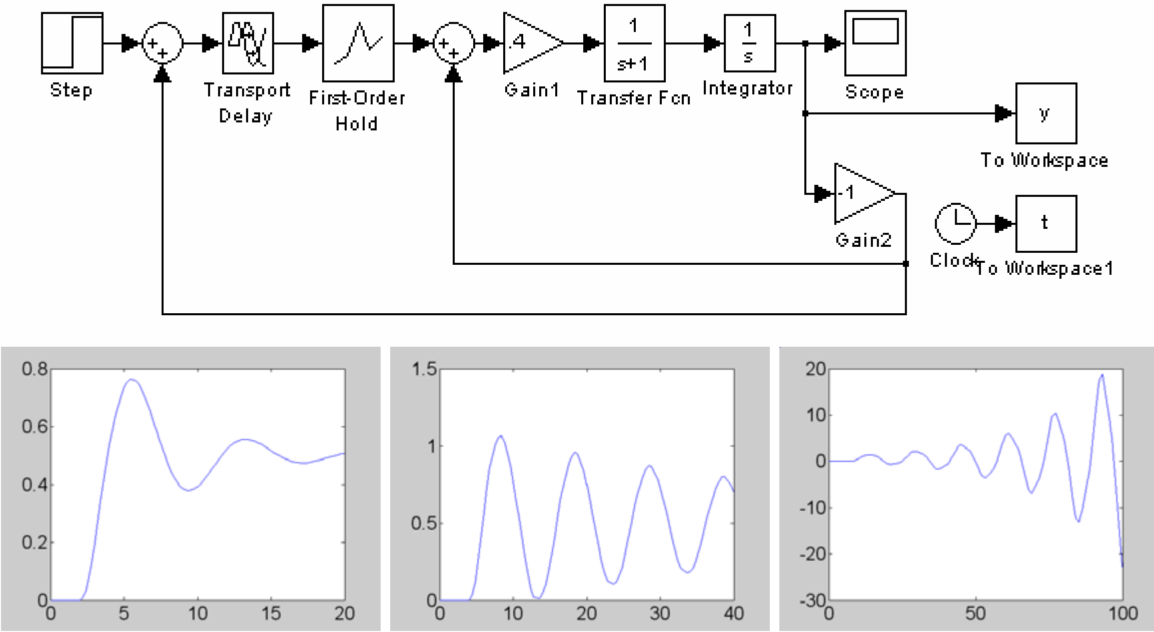
f) D=4, H=4

Figure 11. Simulink results for aircraft with transport delay and/or sample and zero order (constant) hold in secondary (outer) loop. *Note the differing time axes.*

Again we examine step responses under various near-critical combinations of transport delay and sample/hold, remembering that these parameters plus the aircraft-lag time have an effect relative to controller gain K. Stability is achieved if K is small enough, and inversely proportional to these time constants. But smaller gain proportionately slows the transient response, as evident in Fig. 8g.

In the case of the zero-order holds of Fig. 11, the first thing to notice is that the response to a unit-step input asymptotes to a value of 0.5 rather than 1.0. This is obvious in Fig. 11a, which shows the transfer function when the outer-loop-feedback coefficient is unity. In this case when the inner closed-loop transfer function is $1/[(T/K)s^2 + (1/K)s + 1]$, the combined closed-loop transfer function can be shown to be $1/[(T/K)s^2 + (1/K)s + 2]$, i.e., the steady state is 1/2 of the unit step. Again, examining the various near-critical parameter combinations (for the given controller gain) for transport delay D and sample/hold H, one sees that oscillations increase with D and H. D has a somewhat greater effect than H. The transient always converges toward 0.5 (Fig. 11b-f), unless of course D and H are large enough to render the transient unstable.

In Fig. 12, we substitute a sample and first-order (rate) hold and try increasing values of D and H, with similar results as before. The response finally goes unstable in Fig. 12c. We note that for comparable parameters in these two-loop systems, oscillations are larger in the first-order than in the zero-order-hold case. This was not true when there was only one loop (as in Fig. 7-9).



A. D=1, H=1 B. D=2, H=2 C. D=4, H=4

Figure 12. Simulink results for aircraft with transport delay and/or sample and first-order (rate) hold in secondary (outer) loop. *Note the differing time scales.*

Our simulations of step responses in the critical (oscillatory, near unstable) case demonstrate three salient facts:

1) Transport (pure) time delays and/or sample-hold elements (either zero-order or first-order) are tolerable within continuous control loops provided the controller gain is relatively small.

2) But as the delay and hold times increase and controller gain necessarily must decrease to retain the same stability margin (as was observed in Figure 9), the transient response slows more or less proportionately. Delays as can be expected for multitasking demand a small gain that can make transient response unacceptably long.

3) If a second (outer) control loop is established that feeds a continuous reference signal used by the pilot or autopilot, guiding simultaneous control actions by the latter, the delay effects are similar but the steady-state response is 1/2 instead of 1 as it should be.

V. Avoiding the Problem: "Move and Wait"

Why, with the complex hierarchy of control loops in the current air-traffic-control system and acceptably rapid responding aircraft, do instabilities not occur? The answer lies in the fact that the outer loops of the current system operate in a "move-and-wait" mode. "Move and wait" is a term originally applied to human telecontrol of robots on the moon, a situation where the communication round-trip time delay approximates three seconds. In this case it has been amply demonstrated that continuous control requires control gains that make the response unacceptably slow. The only stable operating mode is for the operator to commit to some incremental action without feedback, then wait until execution of that action has settled, and then keep repeating that procedure. If there is an attempt to perform continuous control along with acceptably fast response it is essentially impossible to avoid destabilizing the system.[6]

It is critical to understand that "move and wait" means opening the loop after each sample and control action. It is not the same as "sample and hold." This can be illustrated in the following thought experiment: In a single control loop, when the system output finally matches a fixed reference input (e.g., one), the error is zero and all values stay put. And if an intermediate agent (e.g., air-traffic control), with or without a sample/hold, is interposed and merely communicates the fixed reference signal to the single closed loop, the system will likewise, after a hold delay, stabilize at the communicated fixed reference value with zero error. However if the intermediate agent with a hold is in an outer closed loop, when the system output finally matches the reference input this outer-loop agent samples and holds an error value of zero. This is then communicated as a zero reference input to the inner loop, driving the system output back to zero. The result can be shown to be an intermittent resetting of the system output to one, then zero, and back to one, etc., with a cycle time of twice the hold time.

In the current aviation system the pilot is given traffic-control vectors by controllers verbally. There is no attempt at continuous control by ATM. Currently, the on-board pilot sets the autopilot to perform a specified function and does not attempt to make repeated changes while the response transient is occurring. In NGATS, if there is any consideration of an ATM computer or human making continuous corrections to the control of a given aircraft, it would appear that ATC must operate as an open-loop agent, whose function is to decide what reference value to communicate to the pilot or input to the FMS in the inner loop with the aircraft. ATC observation of zero error in the outer loop (as described in the previous paragraph) is a measure of zero unhappiness. This zero unhappiness, however, should not serve as a basis for resetting the heading (or whatever degree of freedom under control) set point for the inner aircraft control loop to zero.

Within the inner loop a stable, non-oscillatory response can always be achieved by compromising between transient-response speed and any (cognitive and/or computer) processing delays that may exist.

For these reasons, it is imperative that human-(transient into)-the loop simulations be conducted for various phases of the flight profile with off-normal scenarios to determine multitasking pilot and/or computer control delays and to examine the effects of various forms of external control. Such simulations should be conducted as soon as human and computer functions are defined, and not delayed until systems, displays, and controls have been fairly fully developed, as has been the tendency in the past. Much learning can be gained from relatively crude initial simulations that reveal the human cognitive difficulties and response times that may pose problems. These will indicate design needs for procedures, the human-automation interface, and the allocation of authority under different contingencies.

VI. Conclusions

This paper explores hypothetical problems posed by having humans or computers perform continuous control of aircraft from the ground (or even from within the aircraft when subject to time delay), as might be considered in NGATS. Simulations demonstrate how pure transport delays and sample-hold models (representing cognitive and/or computer processing response in getting “into the loop”) produce oscillations or instability unless gains are reduced to the point where response speed may be unacceptable.

Acknowledgments

This paper was prepared as part of a project “Human-Automation Interaction in NGATS” sponsored by the NASA Airspace Systems Program.

References

[1] Sullivan, B., Roberts, D., Hallowell, S., Arbuckle, D., Rhodes, D., Baker, D., Burleson, C., Andrews, M., and Howell, J. “NGATS Concept Brief,” *Tech Hangar* [online database], URL: http://jpdo.digiplaces.com/tech_hangar/ [cited July 2006].

[2] Sheridan, T. B., and Corker, K. M., “Report of a Workshop on Human-Automation Interaction in NGATS,” 2006 (to be published).

[3] Huey, M. B., and Wickens, C. D., *Workload Transition: Implications for Individual and Team Performance*, National Academy of Sciences, Washington, DC, 1993. See also Hancock, P.A., and Desmond, P. A., *Stress, Workload and Fatigue*, Lawrence Erlbaum, Mahwah, NJ, 2001.

[4] Kozinsky, E., Pack, R., Sheridan, T., Vruels, D., and Seminara, J., “Performance Measurement System for Training Simulators,” Electric Power Research Institute Report NP-2719, 1982. See also Sheridan, T. B., *Telerobotics, Automation and Human Supervisory Control*. MIT Press, Cambridge, MA, 1992.

[5] Klyde, D. H., McRuer, D. T., and Myers, T. T., “Unified Pilot Oscillation Theory. Vol 1. PIO Analysis with Linear and Nonlinear Effective Vehicle Characteristics, Including Rate Limiting,” Wright Laboratory, WL_TR 96-3028, Wright Patterson AFB, OH, Dec. 1995.

[6] Ferrell, W. R., “Remote Manipulation with Transmission Delay,” *IEEE Transactions on Human Factors in Electronics*, HFE-6(1), 1965, pp. 24-32.