Evaluation of an Airborne Spacing Concept, On-board Spacing Tool, and Pilot Interface

Kurt Swieringa[1], Dr. Jennifer L. Murdoch[2], and Brian Baxley[3]
NASA Langley Research Center, Hampton, VA, 23681

and

Clay Hubbs[4]
All Aspect Aerospace Innovations LLC, Parker, CO, 80134

The number of commercial aircraft operations is predicted to increase in the next ten years, creating a need for improved operational efficiency. Two areas believed to offer significant increases in efficiency are optimized profile descents and dependent parallel runway operations. It is envisioned that during both of these types of operations, flight crews will precisely space their aircraft behind preceding aircraft at air traffic control assigned intervals to increase runway throughput and maximize the use of existing infrastructure. This paper describes a human-in-the-loop experiment designed to study the performance of an onboard spacing algorithm and pilots' ratings of the usability and acceptability of an airborne spacing concept that supports dependent parallel arrivals. Pilot participants flew arrivals into the Dallas Fort-Worth terminal environment using one of three different simulators located at the National Aeronautics and Space Administration's (NASA) Langley Research Center. Scenarios were flown using Interval Management with Spacing (IM-S) and Required Time of Arrival (RTA) control methods during conditions of no error, error in the forecast wind, and offset (disturbance) to the arrival flow. Results indicate that pilots' delivered their aircraft to the runway threshold within +/- 3.5 seconds of their assigned arrival time and reported that both the IM-S and RTA procedures were associated with low workload levels. In general, pilots found the IM-S concept, procedures, speeds, and interface acceptable; with 92% of pilots rating the procedures as complete and logical, 218 out of 240 responses agreeing that the IM-S speeds were acceptable, and 63% of pilots reporting that the displays were easy to understand and displayed in appropriate locations. The 22 (out of 240) responses, indicating that the commanded speeds were not acceptable and appropriate occurred during scenarios containing wind error and offset error. Concerns cited included the occurrence of multiple speed changes within a short time period, speed changes required within twenty miles of the runway, and an increase in airspeed followed shortly by a decrease in airspeed. Within this paper, appropriate design recommendations are provided, and the need for continued, iterative human-centered design is discussed.

Nomenclature

ADS-B   = Automatic Dependent Surveillance-Broadcast
AGL   = Above Ground Level
ANOVA   = Analysis of Variance
ASTAR   = Airborne Spacing for Terminal Area Routes
ASTOR   = Aircraft Simulation for Traffic Operations Research
ATC   = Air Traffic Control
ATOL   = Air Traffic Operations Laboratory

---

[1] Graduate Research Assistant, National Institute of Aerospace, NASA LaRC, MS 152, SS.
[2] Research Psychologist, Crew Systems and Aviation Operations Branch, NASA LaRC, MS 152.
[3] Research Engineer, Crew Systems and Aviation Operations Branch, NASA LaRC, AIAA Senior Member.
[4] Manager, 8245 N Sunburst Trail, Parker, CO. 80134, AIAA Senior Member.

ATOS	=	Airspace and Traffic Operations simulation
CDA	=	Continuous Descent Arrival
CDTI	=	Cockpit Display of Traffic Information
CPDLC	=	Controller-Pilot Datalink Communication
DFW	=	Dallas Fort-Worth
DTS	=	Development Test Simulator
EICAS	=	Engine Indicating and Crew Alerting System
EFB	=	Electronic Flight Bag
FAA	=	Federal Aviation Administration
FMC	=	Flight Management Computer
FMS	=	Flight Management System
HITL	=	Human-In-The-Loop
IFD	=	Integration Flight Deck
IM-S	=	Interval Management with Spacing
LaRC	=	NASA Langley Research Center
MACS	=	Multi Aircraft Control System
M	=	Mean (arithmetic average)
MCDU	=	Multi-Function Control Display Unit
MCH	=	Modified Cooper Harper
MCP	=	Mode Control Panel
MFD	=	Multi-Function Display
MSL	=	Mean Sea Level
N	=	Sample Size
NASA	=	National Aeronautics and Space Administration
ND	=	Navigation Display
NOAA	=	National Oceanic and Atmospheric Administration
OPD	=	Optimized Descent Profile
P	=	Probability (level of statistical significance)
PF	=	Pilot Flying
PFD	=	Primary Flight Display
PNF	=	Pilot Not Flying
RTA	=	Required time of Arrival
RUC-13	=	Random Update Cycle (13 kilometer resolution)
SD	=	Standard Deviation
STAR	=	Standard Terminal Arrival Route
TMX	=	Traffic Manager
TOD	=	Top of Descent
TRACON	=	Terminal Radar Approach Control
VFR	=	Visual Flight Rules
VNAV	=	Vertical Navigation
VREF	=	Landing Reference Speed

I. Introduction

By the year 2030, it is anticipated that over one billion passengers will travel by way of U.S. commercial air carriers. Within the U.S., the number of cargo aircraft is expected to double, and the number of passenger aircraft is expected to increase by over 50 percent.[1] If the efficiency of current air traffic procedures is not improved, this projected increase in flight operations will lead to greater noise, pollution, delays, fuel costs, and infrastructure requirements. Therefore, methods that improve operational efficiency and fully utilize existing infrastructure must be identified.

Increasing runway throughput is one method that can minimize the need for new infrastructure, and airborne spacing has been proposed as a viable way to effectively increase that runway throughput. The Federal Aviation Administration's (FAA) Surveillance Broadcast Services Office is leading the development of a concept called Interval Management, in which the flight crew is responsible for flying their aircraft at a speed that achieves their assigned spacing interval at the runway threshold, while Air Traffic Control (ATC) remains responsible for ensuring that all aircraft maintain safe separation.[2] Typically, ATC designates a spacing buffer, which is often twenty to thirty

seconds, in addition to the separation requirement to ensure that separation is always maintained.[3] The goal of airborne spacing (the FAA's Interval Management is one example) is to decrease this spacing buffer by decreasing the standard deviation (SD) of arrival time error. The airborne spacing concept uses a ground based tool to calculate an arrival schedule and to select routes for various aircraft. That information is sent by ATC as a clearance to the flight crew who accepts or rejects the clearance. Upon acceptance of the clearance, an onboard spacing tool uses precise position updates of all involved aircraft to calculate the airspeed the flight crew must fly to achieve that spacing interval by a geographical point.

Several computer batch simulations and human-in-the-loop (HITL) investigations have demonstrated the increase in efficiency that airborne spacing can provide. For example, MITRE's HITL data collection efforts have revealed that current day operations resulted in a mean spacing error at the runway threshold of 24.75 seconds, with a SD of 17.02 seconds, while the largest mean spacing error at the runway threshold associated with multiple airborne spacing scenarios was 2.19 seconds with a SD of 1.32 seconds.[4, 5] Similarly, Baxley et al. (2006) reported that precise runway delivery can reduce ATC's spacing buffer by 10 to 15 seconds, thereby leading to a 5 to 10% increase in runway throughput.[6] While such gains may seem small, Credeur's (1977) analysis of terminal operation benefits showed that a 5% increase in throughput can create a 29% decrease in delays for an airport running at 85% capacity.[7]

When introducing new procedures and new forms of automation into the flightdeck environment, it is essential that procedures, automated tools, and related displays be designed with the human operator in mind to ensure that pilot decision making and task execution are properly supported. A review of the literature reveals that a human-centered approach to the design of an airborne spacing concept is underway, in that the impact of introducing airborne spacing to the flight deck is being assessed from the perspective of the flight crew. MITRE conducted a series of tests examining a near term implementation of airborne spacing, with Cockpit Display of Traffic Information (CDTI) displayed on an Electronic Flight Bag (EFB) and found that, in general, pilots provided favorable ratings of the spacing concept and its implementation.[4,8] CoSpace, another spacing concept being developed by EUROCONTROL, examined displays required to perform airborne spacing and used pilot feedback to guide the refinement of their proposed display formats.[9-12] Findings indicated that pilots desired a spacing trend indicator on the Navigation Display (ND) as well as an autopilot managed mode. At the National Aeronautics and Space Administration's (NASA) Langley Research Center (LaRC), a HITL airborne spacing experiment utilizing Continuous Descent Arrivals (CDA) indicated that pilots found the spacing procedures, commanded speeds, and required workload level to be acceptable.[13] Additionally, a usability study of LaRC's airborne spacing displays found that pilots rated textual and graphical representations of commanded speed, indication of lead aircraft on the ND, and a spacing trend indicator as being useful.[14] Lastly, a three-year effort by the FAA and EUROCONTROL to incorporate all the Interval Management related research described here into an international standard was achieved in the summer of 2011 with the publication of RTCA DO-328, the 'Safety, Performance and Interoperability Requirements Document for Airborne Spacing – Flight Deck Interval Management (ASPA-FIM).[15]

The experiment described in this paper extends previous research by evaluating the latest version of LaRC's onboard spacing algorithm - Airborne Spacing for Terminal Area Routes version 10 (ASTAR10) which supports parallel dependent runway operations while maintaining the ability to support Interval Management with Spacing (IM-S) and Required Time of Arrival (RTA) operations - and by continuing to refine the pilot interface associated with the use of the algorithm's speed guidance.[16,17] Research objectives included: 1) examining the algorithm's performance in terms of the precision of aircraft delivery to the runway threshold and the stability of the aircraft arrival streams, 2) observing flight crew interactions with the algorithm, 3) evaluating requirements for ATC-flight crew communication via Controller-Pilot Datalink Communication (CPDLC), and 4) collecting pilots' reported workload levels and ratings of the acceptability of LaRC's IM-S concept, procedures, speed guidance, and pilot interface. This paper describes the experiment's methodology and the findings associated with qualitative measures and a subset of quantitative data that facilitate an evaluation of the proposed concept, procedures, speed guidance, and pilot interface. Data characterizing the algorithm's performance in terms of aircraft arrival stream stability will be made available in a future publication, while results pertaining to requirements for ATC-flight crew communication via CPDLC are described by Baxley (2011).[18]

II. Methodology

A. Pilot Participants
Participants consisted of 24 commercial airline pilots (i.e., twelve captains and twelve first officers) employed by major U.S. air carriers. All pilots were male and ranged in age from 37-61 years. On average, the pilots had twenty years of airline experience and over 11,000 hours of airline flight time. At the time of the study, seventeen of the

participants served as Boeing 777 pilots; four served as 757/767 pilots; two served as 747 pilots; and one served as a 737 pilot. Twelve of participants flew as single pilots while the remaining twelve flew as members of six, two-person crews. To minimize potential effects associated with different airline operating procedures, all two-person crews were paired from the same airline, and for five of the six crews the pilots flew in the position they did operationally (Captain or First Officer).

B. **Test Facilities and Equipment**

This experiment was conducted at NASA LaRC using three types of simulators and an onboard spacing algorithm capable of presenting pilots with speed guidance. The test facilities and equipment used are described below.

1. Air Traffic Operation Laboratory (ATOL)

The ATOL is comprised of the Airspace and Traffic Operations Simulation (ATOS) platform, containing a network of hundreds of real-time, medium-fidelity aircraft simulators that can be used for batch Monte Carlo studies as well as real-time HITL experiments.[19] When the aircraft simulators are equipped with experimental cockpit displays and pilot interfaces, they are referred to as Aircraft Simulation for Traffic Operations Research (ASTOR) stations. ATC controller stations using the Multi Aircraft Control System (MACS),[20] developed at NASA Ames Research Center, were also integrated into the ATOL to enable confederate air traffic controllers to provide a realistic traffic control environment.

Each ASTOR station is a medium fidelity aircraft and avionics simulation with low fidelity single-pilot interfaces.[19] ASTOR components include: a six degrees of freedom aircraft model, Primary Flight Display (PFD), Multi-Function Display (MFD), autopilot and auto-throttle systems, Flight Management Computer (FMC), Multi-function Control Display Unit (MCDU), Mode Control Panel (MCP), voice communication, Controller-Pilot Datalink Communication (CPDLC), and Automatic Dependent Surveillance-Broadcast (ADS-B). The onboard spacing tool (ASTAR10) was added to the ASTOR stations for this experiment.

2. Development Test Simulator (DTS)

The DTS is a full-scale simulator representative of a large commercial transport category aircraft and is driven by an appropriate aircraft dynamics mathematical model. The DTS has a 210° horizontal by 45° vertical out the window field of view and is equipped with eight D-Sized LCD displays, sidestick controls, rudder pedals, two color MCDUs, and additional interface devices derived from a variety of other transport aircraft. The visual scene used for this experiment was the Dallas Fort-Worth (DFW) terminal environment in a daytime setting.

3. Integration Flight Deck (IFD)

The IFD is a full-scale simulator of a large commercial transport category aircraft and is driven by an appropriate aircraft dynamics mathematical model.[21] The cockpit includes standard ship's instruments representative of a line operations aircraft, and the cockpit's visual system is a panorama system that provides 200° horizontal by 40° vertical field-of-view. The visual scene used for this experiment was the DFW terminal environment in a daytime setting.

4. ASTAR Spacing Algorithm

The ASTAR algorithm produces speed guidance by determining time-to-go until an aircraft and its lead aircraft reach an "achieve by" point along a 4-D trajectory.[16,17] The version of the algorithm used in this experiment, ASTAR10, is capable of supporting dependent parallel runway operations.

ASTAR contains a number of filters and heuristics designed to provide it with greater stability and predictability and to minimize the number of commanded speed changes. To keep unnecessary speed changes from occurring when the aircraft is more than 150nm from the runway, ASTAR10 contains a notch filter that subtracts larger error terms from the commanded speed when the aircraft is far from the runway threshold and subtracts progressively smaller error terms as the aircraft approaches the runway. To aid in the predictability and stability of ASTAR, commanded speeds are bounded to within ±10% of the profile speed. Furthermore, a look ahead function that disables speed increases ten seconds prior to a scheduled speed decrease is also implemented. The ASTAR logic also complies with regulatory requirements, such as flying at 250 knots or less when below 10,000' MSL. The ASTAR logic terminates nulling the spacing error prior to the FAA's mandated stabilization height of 1000' AGL, and outputs as the commanded speed the aircraft's FMC calculated $V_{ref30}+5$ speed. Since the delivery accuracy of the aircraft to the runway threshold increases as pilot conformance increases, this implementation of ASTAR displayed two speeds: a commanded speed and a commanded end speed. The commanded speed displayed the

instantaneous speed, incorporating a typical acceleration or deceleration rate. The commanded end speed was the speed that the aircraft stabilized to at the end of an acceleration or deceleration. Lastly, the speed commands ASTAR provided to pilots were granulated into five knot increments.

5. ASTAR Speed Guidance and Lead Aircraft Display

To assist the flight crew conducting spacing operations, changes were made to the PFD, ND, and MCDU. Advisories and cautions related to the spacing operations were displayed on the Engine Indicating and Crew Alerting System (EICAS).

On the PFD, all normal display elements were retained (airspeed, altitude, etc), and three elements were added to support airborne spacing operations (Figure 1). Clockwise from the bottom left of the PFD, a green speed bug displayed the commanded speed from the spacing algorithm (the magenta speed bug corresponded to the FMC commanded speed) and a three digit green number displayed the commanded end speed from the spacing algorithm (the magenta three digit number immediately below it corresponded to the FMC commanded speed). To the right of the commanded end speed, letters identified which of the four modes the ASTAR algorithm was in. "RTA" indicated the algorithm was using the RTA control method at the runway threshold to calculate the commanded end speed, "IM" indicated the algorithm had valid ADS-B data on one or both of the lead aircraft and was calculating a speed to achieve the assigned spacing interval, "FNL" indicated the algorithm was displaying the FMC calculated final approach speed, and "RVT" indicated an algorithmic or operational error.

The ASTAR speed guidance and lead aircraft information was generally consistent across the ASTOR, DTS, and IFD simulators, however slight differences were intentionally made to explore different aircraft equipage levels, which in turn required slightly different flight crew procedures for the spacing operation. The ASTORs and DTS employed an advanced implementation with the spacing algorithm incorporated into the FMC. This meant the flight crews flying the ASTORs and DTS were able to use VNAV PATH (MCP speed window closed), and the auto-throttle system drove to the spacing commanded end speed. The IFD utilized a near-term implementation where the spacing algorithm and software were not connected to the FMC, requiring crew to use VNAV SPD (manually dial in the commanded end speed into the MCP speed window). Other differences on the PFD included the shape of the speed bugs between the ASTORs, DTS, and IFD, and the fact that the IFD did not have a FMC commanded speed.

Since ASTAR10 supports parallel dependent runway operations, it is possible for an aircraft to have two lead aircraft displayed on the ND. Each lead aircraft was denoted by a double chevron (ASTOR only) or diamond (DTS and IFD), and an aircraft call sign was included in each lead aircraft's expanded data tag (Figure 1). During IM operations, the lead aircraft that was currently controlling the spacing had one of its double chevrons/diamonds changed from white to green.
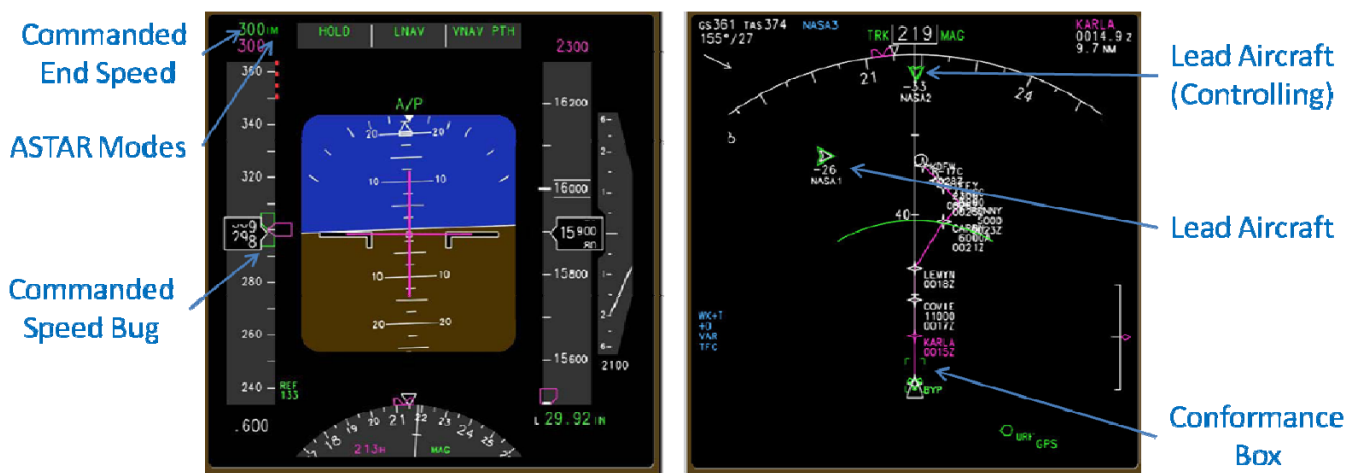

Figure 1. Display elements added to the ND and PFD in support of the spacing operation.

Figure 2 shows three new pages related to the IM operation accessible via the MCDU. The first page displayed summarized information about the lead aircraft, including spacing error, runway, final approach speed, spacing error, and aircraft call sign. The first page was also used to activate the IM-S procedure and, when necessary, to terminate the operation. Two additional IM pages contained supplementary information regarding each lead aircraft, including their IM-S goal, achieve by point, terminate at point, and route.

American Institute of Aeronautics and Astronautics

Figure 2. Three MCDU pages added in support of the spacing operation.

C. Pilot Procedure

The primary responsibility given to the crew was to follow the commanded speed generated by the ASTAR10 spacing algorithm. Once the clearance from ATC was accepted (either RTA to the runway, or RTA to the runway plus relative spacing behind a lead aircraft when able), the pilots were instructed to monitor the commanded speed for changes, monitor the EICAS display for advisories and cautions, and perform all other normal flight duties. In training, pilots were requested to remain within ±5 knots of the commanded speed and within ±200ft of the vertical path.

Within this experiment all IM-S clearances were provided to flight crews through a CPDLC uplink message. After receiving the message, the crew was to load the message into the spacing tool, and then activate the spacing tool via the MCDU. Upon activation, ASTAR calculated the speed required for that aircraft to fly to achieve the assigned RTA or spacing interval. If the crew found the speed acceptable, the procedure had them send ATC an 'Accept' CPDLC downlink message, and then 'Execute' the spacing tool's commanded end speed (similar to 'Execute' of a route modification). This last step caused the commanded speeds and speed bugs to appear on the PDF and lead aircraft information on the ND as discussed earlier. If the clearance was not acceptable, the crew was instructed to respond to ATC with a 'Reject' CPDLC message downlink. If ATC sent an appended clearance, pilots were expected to terminate the current clearance in the spacing software tool, and then follow the process of loading a clearance as discussed in this paragraph.

The CPDLC messages containing the RTA or RTA+IM clearance were intentionally varied in the experiment to include initiating the spacing algorithm with the aircraft within ADS-B range of the lead, and when outside ADS-B range of the lead aircraft. In the later case, the crew flew the speed calculated to achieve the RTA. Once the aircraft was within ADS-B range of the lead aircraft, an EICAS advisory ("IM AC 1 SPACING" or "IM AC 2 SPACING" as appropriate) would appear to notify the flightcrew of the change. When this occurred, the crew was instructed to notify ATC that they were now spacing off a lead aircraft. ASTOR pilots accomplished this by arming a CPDLC report following the acceptance of the original message, while IFD and DTS crews made voice radio calls.

The airborne spacing procedure could be terminated at any time by the crew, ATC, non-conformance of the ownship or lead aircraft to the predicted route, or because of an ASTAR failure. If the procedure was terminated by the crew for operational reasons or due to a failure of the ASTAR spacing tool, the pilots were instructed to contact ATC and continue flying the previous commanded speed until they received further instruction. It should be noted that the pilots were requested not to terminate the IM operation due to crew preference within the context of this experiment.

D. **Experiment and Scenario Design**

1. Experiment Design

This experiment utilized a split-plot design.[22] Each ASTOR pilot, DTS crew, and IFD crew was designated as a *whole plot*, and Simulator Type served as the *whole-plot factor*, since each pilot and crew was assigned to fly one of three types of simulators. Each pilot and crew flew a total of 11 scenarios. Therefore, each whole plot was "split" into 11 *sub-plots*, or treatment conditions. As described below and in the "Scenario Design" section, each flight scenario (i.e., treatment condition) was designed according to a combination of two independent variables (i.e., Error Type and Control Method). Therefore, Error Type and Control Method served as the experiment's *sub-plot factors*.

2. Independent Variables

Flight scenarios were defined using a 2 (Control Method: RTA, and IM-S) x 3 (Error Source: None, Wind, and Offset) matrix (Table 1) to allow an examination of conditions in which the spacing algorithm controlled to an absolute time (RTA) or to a relative time behind a lead aircraft (IM-S), during conditions of no error, error in the forecast wind, and offset (disturbance) to the arrival flow.

Table 1. Experiment Design Matrix.

CONTROL METHOD

		RTA	IM-S
ERROR SOURCE	**None**	Scenario 1 (Replicate 1)	Scenario 2 (Replicate 1)
	Wind	Scenario 3 (Replicate 1) (Replicate 2)	Scenario 4 (Replicate 1) (Replicate 2)
	Offset	Scenario 5 (Replicate 1) (Replicate 2)	Scenario 6 (Replicate 1) (Replicate 2)

Six distinct "nominal" scenarios were created as were second replicates of the wind error and offset error scenarios. One additional exploratory scenario was created to examine off-nominal events. Therefore, each single pilot and two-person crew flew a total of eleven scenarios.

E. **Dependent Measures**

1. Subjective Assessments

Pilots' perceptions regarding LaRC's IM concept and procedures as well as the spacing algorithm's speed guidance and pilot interface were collected through post-scenario questionnaires, post-experiment questionnaires, feedback obtained during post-experiment group debriefing sessions, and workload ratings which were obtained using the Modified Cooper-Harper (MCH) Rating Scale.[26] Use of the MCH scale yields an overall workload rating ranging from "1" (indicating that the instructed task was very easy/highly desirable; operator mental effort was minimal; and desired performance was easily attainable) to "10" (indicating that the instructed task was impossible and could not be accomplished reliably).

2. Human and System Performance Metrics

Quantitative data were collected to characterize human and system performance during Optimized Descent Profiles (OPDs) flown into DFW using IM-S and RTA procedures conducted when no error was introduced, when an error between actual winds and forecast winds existed, and when a change in arrival time was issued to disrupt the arrival stream. Inter-arrival spacing time (also called arrival error), number and location of speed changes, frequency of speed changes, pilot conformance to commanded speeds, aircraft state data, and output from the ASTAR algorithm were recorded.

F. **Scenario Design**

For this experiment, aircraft used flight management system (FMS) guidance to fly the aircraft from its initial position to the runway. Pilots were cleared for the descent prior to the beginning of the scenario and had a route preloaded into the FMS. Before reaching the top of descent (TOD), pilots were issued a CPDLC clearance

containing their spacing instructions. Since all messages provided within this experiment were correct, pilots were expected to load the clearance and accept it. Once the clearance was accepted and the spacing operation was executed, pilots were instructed to follow ASTAR's speed commands to achieve a precise spacing or RTA at the runway threshold. In each of the nominal scenarios, the initial CPDLC clearance was given prior to the TOD; however, scenarios involving the use of RTA procedures during the presence of an offset error included an additional message that was sent when the aircraft reached 9,000ft MSL to create a spacing error. While CPDLC messages were used for the IM-S clearance, voice clearances were used to convey all other communications. Three ATC stations were staffed by confederate controllers who gave landing clearances, provided frequency changes, and handled any unexpected events. In addition to the communications between subject pilots and the confederate air traffic controllers, communication between pilot model ASTORs and ATC were recorded and played to provide background chatter.

Scenarios were designed to simulate a near term NextGen environment. Each scenario contained six piloted aircraft flying different approaches to runway 17C at DFW airport under Instrument Flight Rules (IFR) conditions. Additional traffic was added to the scenarios with pilot model flown ASTOR machines. The routes were created by modifying existing Standard Terminal Arrival Routes (STARs) into DFW, with the piloted aircraft on the MASTY 3, BONHAM 5, and CEDAR CREEK 6 arrivals. Altitude restrictions on the routes were modified to approximate OPDs, simulating near idle continuous descents designed to be fuel efficient and reduce noise.

The winds used in this experiment, derived from data provided by the National Oceanic and Atmospheric Administration (NOAA), were designed to stress the ASTAR algorithm while remaining operationally realistic. Winds varied by altitude, but were kept constant in time and had no vertical component. Wind speed ranged from 97kts at 40,000ft to 12kts at the surface, and a wind shear representative of those observed near DFW was added at an altitude of 5,000ft. The wind shear was positioned so that the wind rapidly changed direction from south to east, causing arrivals from the west to experience a tailwind and arrivals from the east to experience a headwind.

The wind error scenarios contained a discrepancy between the magnitude and direction of the actual winds and forecast winds. The forecast winds were derived from the Rapid Update Cycle weather model with 13 kilometer resolution (RUC-13). The RUC-13 model is able to create weather forecasts ranging from one hour to eighteen hours, with decreasing accuracy as the forecast time increases. This experiment used the RCU-13 error associated with a three hour forecast to simulate an operational event in which the flight crew did not update the wind forecast in their FMS prior to reaching the TOD. The three hour error was then multiplied by a factor of 1.5 to model a case where the RCU-13 model is less accurate due to rapidly changing weather. The forecast wind was compared with measurements provided by NOAA, the truth wind, to obtain the wind error. To add to the wind error, the wind shear was not included in the wind forecast. The resulting wind velocity error was 14kts at an altitude of 40,000ft, and decreased to 4kts at the surface. The wind direction error was 10 degrees everywhere, except the wind shear where the error was 77 degrees. The final wind speed error was approximately one standard deviation away from the mean forecast error at the surface and three standard deviations away at an altitude of 40,000 ft. Therefore, wind errors of this magnitude are not expected to occur very often in during actual flight operations, and the data collected during scenarios involving the wind error should be interpreted with this in mind.

The offset error was implemented differently in the RTA scenarios and the IM scenarios. In the RTA scenarios, an offset was injected by providing each aircraft with a 30 second change to its RTA. This was accomplished by sending an amended CPDLC clearance to each aircraft when it reached an altitude of 9,000ft MSL. An offset error was injected into the IM-S scenarios by giving the aircraft in front of the first aircraft flown by a subject pilot a reduction in its spacing interval, creating a thirty second error that propagated through the stream.

An exploratory scenario was designed to examine various off-nominal events. During the exploratory scenario, the ATC assigned spacing interval was decreased to 75 seconds to simulate IM-S operations in visual flight rules (VFR) conditions. Additionally, a new display element, the conformance box, was added to the PFD during this scenario. The exploratory run contained multiple events; however, the performance of a go-around by the IFD's flight crew as a result of insufficient spacing on the final approach was of primary interest. The DTS was provided with changes to its spacing interval to create adequate space for the IFD to merge back into the traffic flow, so this event was of primary interest as well. Due to the time taken for the IFD to perform its go-around, some of the piloted aircraft were initialized at approximately 10,000ft. As a result, pilots were briefed that this scenario was different from the rest; however, they were not briefed on specific events. Some additional events that pilots experienced during the exploratory scenario included a clearance to space off an aircraft landing on runway 13R and an instance where a lead aircraft on the same route switched to the parallel runway.

G. **Experiment Procedure**

Groups of eight pilots participated in three-day experiment sessions. Three sessions were scheduled over consecutive weeks, each with a different group of pilots, for a total of 24 pilots. Within each group of participants, four flew as single pilots using the ASTOR stations, two flew as a two-person crew in the DTS, and two flew as a two-person crew in the IFD. The six piloted aircraft rotated among six of the fourteen routes into DFW, with each simulator being represented by a distinct, rotating call sign.

All training, data collection, and debriefing for each group was completed over the course of two-and-one-half days at NASA LaRC. The pilots received a half day of training, including several hours of hands-on training in the simulators, supplemented by reading material provided prior to each experiment session. During data runs, all pilots flew their aircraft embedded in the 39 aircraft arriving to DFW. Following each scenario, pilots completed a post-scenario questionnaire. Additionally, all pilots completed a post-experiment questionnaire and group debriefing session after flying the final scenario (i.e., the experiment's exploratory scenario).

III. Results and Discussion

Before evaluating this experiment's implementation of IM-S, it is important to assess the overall viability of the IM-S concept under investigation. Presented below are results associated with the spacing tool's ability to precisely deliver aircraft to the runway, and an evaluation of the spacing operation's successful completion by the pilot participants. Subsequent sections present an evaluation of the IM-S concept and procedures in terms of pilot acceptance and workload levels, characterize pilot interactions with and perceptions of ASTAR10 and its speed guidance, and an describe various aspects of the spacing tool's pilot interface as it was used in this experiment.

All quantitative data were analyzed using General Linear Model Analysis of Variance (ANOVA) tests and, when necessary, Bonferroni post hoc tests. Qualitative data were analyzed using a Wilcoxon signed ranks test. Unless otherwise noted, all results pertain to the experiment's ten nominal scenarios.

H. **Evaluation of IM-S Concept**
3. Runway Delivery Performance

The primary measure of the spacing tool's performance was its ability to precisely deliver aircraft to the runway threshold. Previous experiments have demonstrated that flight deck based spacing can deliver aircraft to the runway with mean errors less than seven seconds and standard deviations less than five seconds.[4,5,13,23-25] In contrast, baseline scenarios of current day operations conducted at MITRE have produced a mean spacing error of 24.75 seconds and a standard deviation of 17.02 seconds.[4,5] The inter-arrival time recorded in this experiment revealed an interaction effect between Control Method and Error Source ($p = 0.002$), and results of post hoc comparisons revealed that scenarios conducted using RTA procedures during the presence of wind error resulted in a greater arrival error when compared with each of the other five Control Method by Error Source combinations ($p < 0.05$). Furthermore, scenarios conducted using IM-S procedures during the presence of wind error resulted in a significantly smaller arrival error when compared with IM-S scenarios involving an offset error and RTA scenarios involving either an offset error or no error ($p < 0.05$). ASTAR's control design was estimated to have a ±4 second error using a five knot granularity in the commanded speed. All of the mean arrival errors in this experiment were below 3.53 seconds, demonstrating the effectiveness of the ASTAR algorithm when large wind and offset errors are present.

Table 2. Arrival Error.

Error Source	RTA		IM-S	
	Mean (s)	SD (s)	Mean (s)	SD (s)
No Error	-3.30	4.87	-1.81	3.87
Wind Error	3.53	3.28	0.90	3.91
Offset Error	-2.30	3.25	-2.16	3.29

4. Pilot Adherence to Speed Commands

Pilots were instructed to keep their speed within ±5 knots (kt) of the commanded speed and their altitude within 200ft of the vertical path. A significant Control Method x Error Source interaction effect was found ($p < 0.0005$), indicating that the effect of error source on pilot speed conformance was dependent upon the control method used. Pilot conformance to speed commands was best during scenarios conducted under conditions of "no error" and was worse when RTA procedures were used during the presence of wind error. The "no error" scenarios were the only scenarios during which the average speed deviance was within the ±5 kt bound. Researcher observations indicated

that many of the segments of non-conformance occurred when commanded speed changes required a large acceleration or deceleration. This was partially due to the fact that the autothrottles used in this experiment had a sensitivity of ±10 knots, forcing the pilots to manually adjust the throttles to maintain a ±5knot interval. Additionally, pilot comments along with researcher observations reveal that there were pilots who preferred allowing the aircraft to slow without using the speed brake to fly a smoother flight for passengers, causing larger speed conformance errors.

Table 3. Pilot Speed Conformance by Error Source (IM-S and RTA data combined).

Error Source	Mean Error (knots)	Standard Deviation (knots)
No Error	3.64	1.20
Wind Error	6.17	2.04
Offset Error	5.84	1.64

1. Pilot Acceptance of the IM-S Concept

Pilots' post-experiment responses were examined to evaluate the acceptability of the spacing concept in terms of additional responsibility, perceived safety, and change in workload as compared with current day operations. When asked whether the added responsibility of meeting a spacing interval was acceptable, 75% of the pilots responded positively. Those who did not find the responsibility acceptable cited concerns regarding the speed behavior and the interface, rather than unacceptability of the airborne spacing concept. Using a scale of 1(Not Safe At All) to 7 (Much More Safe), pilots rated the safety of the spacing operation as being "as safe as" or "slight more safe than" current day operations (M=4.88, SD=1.12, N=24). Using a scale of 1 (Much More) to 7 (Much Less), pilots responded that they would expect the workload of the IM-S spacing operation to be the same or slightly greater than current day operations (M=3.88, SD=1.39, N=24). While concerns were raised regarding the specific implementation of IM-S in this experiment (described below), pilots found the overall IM-S concept acceptable.

2. Pilot Workload

Pilots used the MCH rating scale, ranging from "1" (indicating that the instructed task was very easy/highly desirable; operator mental effort was minimal; and desired performance was easily attainable) to "10" (indicating that the instructed task was impossible and could not be accomplished reliably), to provide assessments of both average and peak levels of workload experienced during each flight scenario.[26] No significant differences were found to exist between workload levels reported for a given control method or among workload levels reported for a given error source (p < 0.05). The pilots rated their mean "average" workload as 1.97 (SD = 0.86, N = 240), and their mean "peak" workload as 2.32 (SD=1.01, N=240), which is comparable to the average workload of 1.87 (SD=0.78, N=207) found in a previous airborne spacing experiment conducted at NASA LaRC.[13] These results indicate that the average workload associated with the task the pilots were asked to perform was easy/desirable, their mental effort was low, and the desired performance was attainable. The pilot's peak workload ratings indicated the task they were asked to perform was mildly difficult; however, the mental effort was acceptable, and an adequate level of performance was attainable. The average workload was not expected to be rated higher than a "3," indicating that the pilots' instructed task had a mild difficulty level and required an acceptable level of mental effort to attain adequate system performance. The data indicate, however, that eight pilots rated their average workload between a "4" and "6." Half of these relatively high workload ratings occurred during the first data collection scenario of a particular group, while the remaining high workload ratings were associated with pilot errors in operating the simulators and undesirable speed change behavior (discussed below).

In addition to providing workload ratings, pilots were given the option of selecting the segment of flight during which their peak workload occurred. Of the 191 flights associated with pilot responses, pilots indicated that their peak workload occurred below 5,000ft 70% of the time, and peak workload occurred when their aircraft was between the altitudes of 5,000ft and 11,000ft 17% of the time. The largest contrast to the aggregate result occurred during the scenarios conducted using RTA procedures in the presence of an offset error. During 38 of these scenarios, pilots reported that their peak workload occurred between 11,000ft and 5,000ft 32% of the time, peak workload occurred below 5,000ft only 50% of the time. It is suggested that the discrepancy between the aggregate results and results for scenarios associated with an offset error and use of the RTA control method is directly related to the issuance of a CPDLC message sent when the aircraft reached 9,000ft, instructing pilots to amend the original spacing clearance.

Table 4. Flight Segment Associated with Peak Workload (IM-S and RTA data combined).

Segment of Flight	Responses (N=191)
>18,000ft (cruise, initial descent, CPDLC)	10%
18,000ft – 11,000ft (descent, approach check)	3%
11,000ft - 5,000ft (TRACON, low altitude merge)	17%
<5,000ft (final approach, configure aircraft)	70%

I. Pilot Acceptance of IM-S Procedures

Pilot acceptance of the IM-S procedures was evaluated using data obtained from the post-scenario questionnaires, post-experiment questionnaires, and from the post-experiment group debriefs sessions. When asked if the IM-S procedures used during this experiment were complete, accurate, and logical, 92% of the pilots answered positively. However, when asked to use a scale of 1 (Very Difficult) to 7 (Very Easy) to rate the ease with which the spacing procedures could be integrated with current day procedures, the pilots' mean response was 4.58 (SD = 1.56, N = 24), indicating that they were somewhat undecided. Many comments also indicated that pilots thought the spacing operation would be associated with a significant learning curve, and that simulator training would be required to adequately learn the airborne spacing procedures. One specific issue that was consistently noted in both questionnaires, debrief sessions, and researcher observation logs was that pilots found it counterintuitive to terminate the existing spacing operation before loading a new spacing clearance. Comments suggested that, upon acceptance of the spacing clearance, the old operation should automatically be overwritten.

Pilots in both the IFD and DTS were asked to characterize how the spacing operation affected the distribution of tasks between the pilot flying (PF) and pilot not flying (PNF), and to indicate whether it changed their crew coordination. Three-quarters of the pilots reported that the distribution of tasks required to complete the spacing operation was desirable and complemented their current distribution of tasks, and 83% of the pilots indicated that the spacing operation did not change their crew coordination. However, the pilots found it somewhat difficult to coordinate the acceptance of CPDLC messages. In this experiment, the PNF was expected to load and accept the CPDLC clearance, while the PF was expected to simultaneously review the information on the IM MCDU page and ensure that ASTAR's commanded speed was acceptable. Half of the pilots commented that the system should allow the PNF to both accept the clearance and execute the spacing operation, and at least one crew ignored the distribution of tasks outlined in the procedures they were asked to follow and completed the operations by having the PNF operate both crewmembers' MCDUs.

J. Evaluation of ASTAR Speeds

3. ASTAR Speed Performance

It is expected that the majority of the workload and disruptions associated with the spacing operation will result from monitoring and implementing speed changes. As described in the section that follows, many of the pilots' critiques concerning the spacing tool were related to undesirable responses driven by large wind and offset errors. Therefore, it is useful to understand the effect of the wind and offset errors on both the number and distribution of speed changes.

The number of speed changes that occur during the course of an arrival has commonly been used as a metric for evaluating the disruptions caused by the spacing operation. Hébraud conducted an experiment using a low fidelity simulator with airline pilots and found an average of one speed change per minute to be acceptable.[10] Other research supports the notion that average speed changes more frequent than one per minute are not acceptable, while speed changes less frequent than one per minute are acceptable.[4,5,8,13] Table 5 shows the number of speed changes experienced by this experiment's pilot participants. The total number of speed changes strongly depended on the error source, with the "no error" condition being associated with the lowest number of speed changes and the wind error condition being associated with the highest number of speed changes. Of the speed changes reported, four correspond to scheduled profile speed changes. Note that these four speed changes do not take into account the additional number of speed changes that would normally be issued by ATC in a traditional spacing operation. Thus, scenarios involving the "no error" condition had an average of 5.5 additional speed changes; scenarios involving an offset error had an average of 7.9 additional speed changes; and scenarios involving a wind error had an average of 12.7 additional speed changes, when ASTAR was activated for an average of 26 minutes per scenario. To better understand the commanded speeds, the total number of speed changes can be examined in conjunction with distributions of speed change frequency and speed changes as a function of distance to go.

Table 5. Number of Speed Changes per Treatment Condition.

	RTA		IM	
Error Source	Mean (s)	SD (s)	Mean (s)	SD (s)
No Error	9.28	1.99	9.72	1.87
Wind Error	17.22	3.99	16.11	3.32
Offset Error	10.44	2.30	13.14	1.81

One comment pilots made throughout the post-scenario and post-experiment questionnaires was related to the algorithm providing pilots with multiple speed commands within a short period of time, thereby generating higher workload and causing some pilots to question the rationale of the algorithm. To better understand the distribution of the time between consecutive speeds changes, a histogram was created (Figure 3). When all scenarios were examined together, it was found that 12% of speed changes occurred fewer than 10 seconds after the previous speed change, 41% of speed changes occurred fewer than 30 seconds after the previous speed change, and 71% of speed changes occurred less than one minute after the previous speed change. Scenarios involving the "no error" condition contained the best distribution, with 9% of speed changes occurring fewer than 10 seconds after the previous speed change, 33% of speed changes occurring fewer than 30 seconds after the previous speed change, and 68% of speed changes occurring less than one minute after the previous speed change. These percentages must be read in the context of the total number of speed changes shown in Table 5. While the distribution of the scenario involving the "no error" condition has a similar distribution to the scenarios involving wind and offset errors, the smaller number of speed changes makes it less likely for speed commands to be closely spaced together in time.
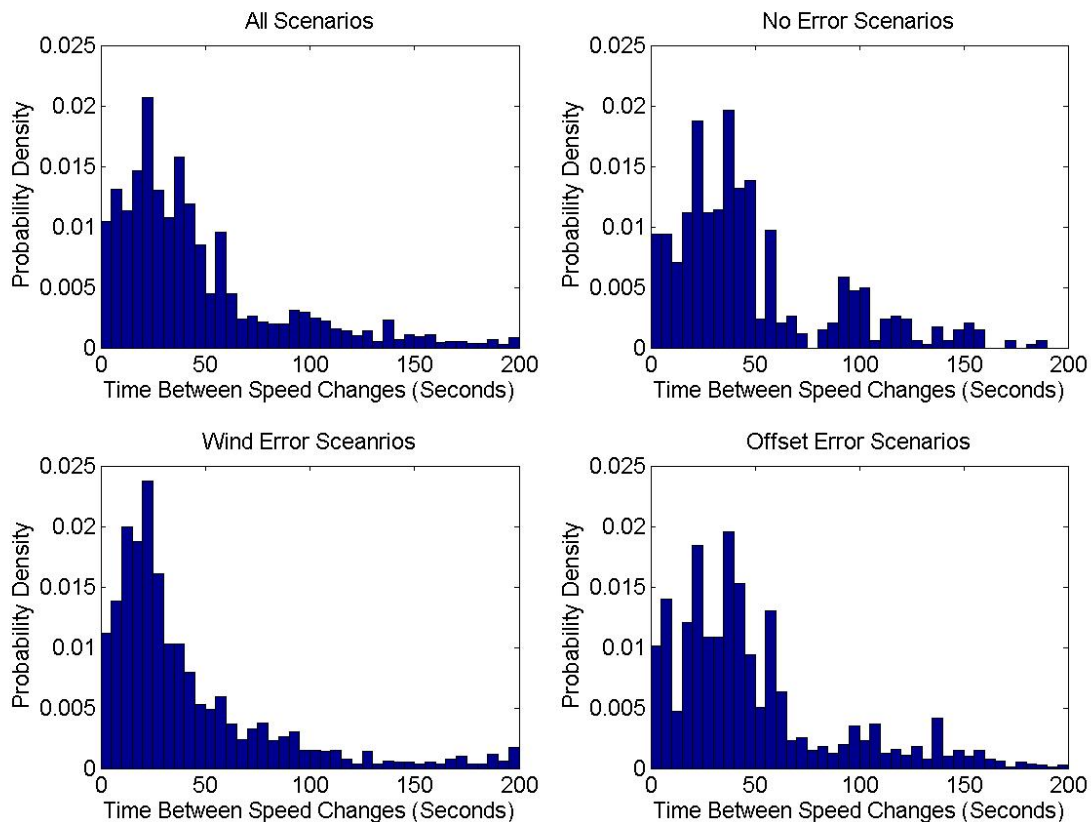

Figure 3. Histogram of the time between consecutive speed changes for each of the three error sources.

Since several pilots reported that the number of speed changes on final approach seemed excessive, the number of speed changes that occurred when the aircraft was on final approach was analyzed, and a histogram of the

American Institute of Aeronautics and Astronautics

location along the arrival was plotted (Figure 4). The shape of the distribution of speed changes as a function of distance-to-go was dictated by the notch filter in ASTAR that ignored some time error when the aircraft was far from the runway, resulting in fewer total speed changes, but resulted in a higher percentage of speed changes as the aircraft approached the runway. When examining all scenarios (i.e., those using the RTA control method and the IM-S control method), the speed changes that occurred on final approach during the scenarios containing wind error were found to be significantly greater than the speed changes on final for the scenarios involving "no error" and the scenarios involving the offset error (p < 0.0005., The scenarios involving "no error" had an average of 2.44 (SD=0.80, N=36) speed changes on final; the scenarios involving the offset error had an average of 2.68 (SD=0.93, N=72) speed changes on final; and the scenarios involving wind error had an average 3.38 (SD=1.10, N=71) speed changes on final. When speed changes were examined as a function of distance to go, 21% of speed changes occurred during the last 10nm of flight, 69% occurred during the last 40 nm, and 89% of speed changes occurred during the last 100 nm. These results are consistent with those indicating that a majority of the pilots' peak workload occurred while below 5,000ft MSL. When asked if the spacing tool required an extraneous expenditure of mental resources that should be used for landing preparation, 67% of pilots rated the expenditure of resources as acceptable, 13% said that the spacing tool freed up mental resources, and 21% stated that the spacing procedure demanded mental resources needed to prepare for landing. Those who stated that spacing procedure's requirement of mental resources was too great cited inadequate cues to notify the flightcrew of a speed change and undesirable speed changes. Thus, even though the workload associated with the spacing task increases closer to the runway, pilot generally found the increase acceptable.
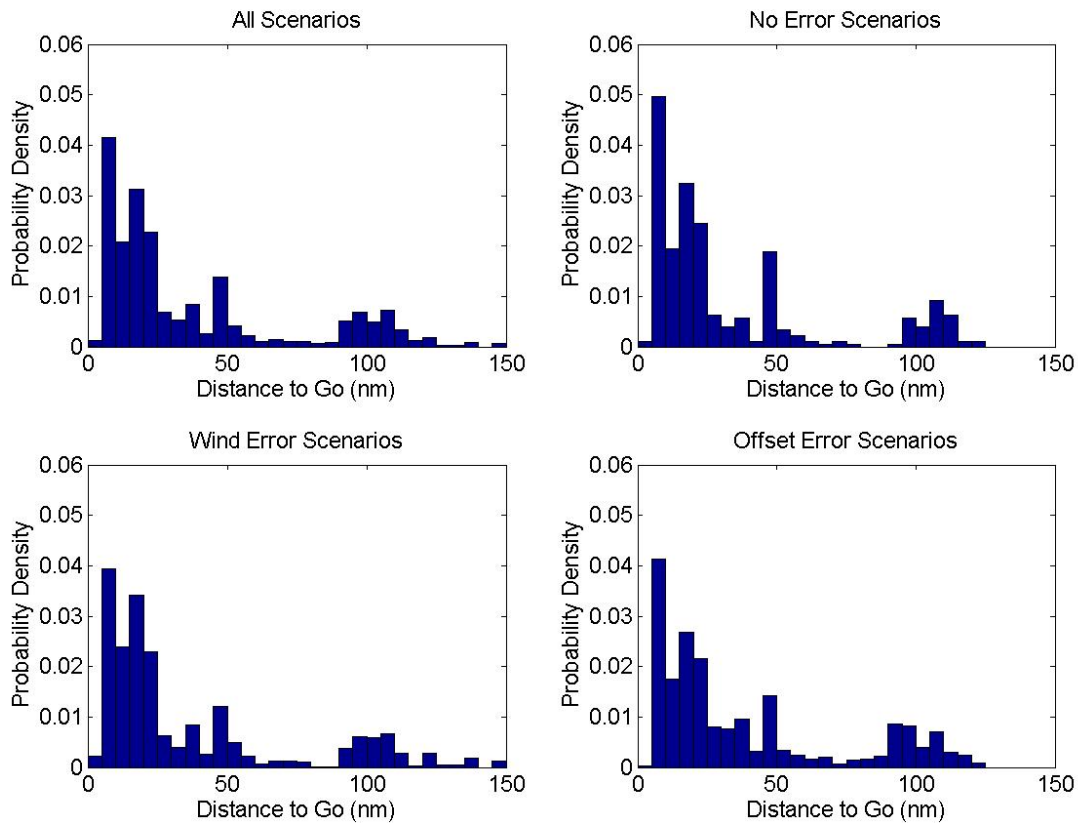

Figure 4. Histogram of speed changes as a function of distance-to-go for each of the three error sources.

A third problem with ASTAR speeds reported by the pilots was related to undesirable speed increases. To attempt to limit the number of speed increases shortly followed by speed decreases, the ASTAR10 algorithm inhibited speed increases 10 seconds prior to a scheduled profile speed decrease. However, some pilots still commented that speed increases were shortly followed by a speed decrease, leading to the perception that the speed increase was unnecessary. Additionally, pilots commented that speed increases were undesirable once they had

begun to configure their aircraft (approximately 20nm from the runway threshold). To better understand the impact of undesirable speed increases, two metrics were examined: the time between a speed increase and a subsequent speed decrease, and the number of speed increases given when the distance-to-go was less than 20nm. The average number of speed increases for all scenarios was 3.58 (SD=2.21, N=180). Overall, aircraft were issued an average of 1.56 (SD=1.40, N=180) commanded speed increases when they were within 20nm of the runway threshold, with aircraft flying in the scenarios involving "no error" receiving a mean number of 0.53 commanded speed increases (SD=0.84, N=180), aircraft flying in the scenarios involving wind error receiving a mean of 1.89 commanded speed increases (SD=1.30, N=180), and aircraft flying the scenarios involving offset error receiving a mean number of 1.74 commanded speed increases (SD=1.49, N=180) within 20nm of the runway threshold. Additionally, the time between a speed increase and subsequent speed decrease was examined, and the results are shown in Table 6. These results indicate that a short time between speed increases and a subsequent speed decrease was a greater problem for the scenarios involving wind error than it was for scenarios involving an offset error or "no error."

Table 6. Distribution of time between a speed increase and a subsequent speed decrease (normalized to number of speeds expected per arrival, with IM-S and RTA data combined).

Time Between Speed Increase and Subsequent Speed Decrease	No Error	Wind Error	Offset Error	All Scenarios
<10 seconds	0.08	0.31	0.08	0.17
<20 seconds	0.19	1.24	0.27	0.65
<30 seconds	0.28	1.82	0.51	1.01
<60 seconds	0.64	2.68	1.09	1.67

As described in more detail below, the pilots' found the commanded speeds acceptable; however, large errors, such as the wind error and offset error used in this experiment, can lead to some instances of less than desirable speed behavior. The ASTAR algorithm was designed with mechanisms to provide pilots with desirable speed behavior. But, based on pilots' comments, it is suggested that the time between consecutive speed changes, the distribution of speed changes as a function of time-to-go, and analysis of speed increases can be used in conjunction with the total number of speed changes as quantitative metrics to help predict pilots' acceptance of speed changes.

4. Pilot Acceptability of Speed Changes
After each scenario, pilots were asked to rate the acceptability, safety, and correctness of ASTAR's commanded speeds and to describe any unexpected behavior exhibited by the algorithm. They were also asked to comment on any discomfort and/or frustration that they experienced as a result of attempting to comply with the commanded speeds. Using a scale of 1 (Completely Disagree) to 7 (Completely Agree), pilots tended to disagree with each of the following statements:
1) there was a time in the scenario when the commanded speed was unsafe (M=1.32, SD=0.81, N=240),
2) there was a time when the commanded speed was incorrect (M=1.46, SD=0.93, N=240),
3) there was a time when they were uncomfortable with the commanded speed (M=1.54, SD=1.20, N=240), and
4) there was a time when information from the spacing tool conflicted with other information provided by the aircraft and/or ATC (M=1.46, SD=0.93, N=240).
Note, however, that the pilots reported that they experienced feelings of frustration and were surprised by the ASTAR's behavior most often during scenarios involving wind error. Using a scale of 1 (Completely Disagree) to 7 (Completely Agree), the pilots tended to agreed that ASTAR's commanded speeds were operationally appropriate (M=6.05, SD=1.30, N=240) and that the frequency of the commanded speed was acceptable at all times throughout each scenario (M=5.7, SD=1.44, N=240). However, the appropriateness of the commanded speeds was found to depend on the error source, with scenarios involving "no error" (M=6.54, SD=0.62, N=240) or offset error (M=6.10, SD=1.22, N=240) being rated as significantly more acceptable than scenarios involving wind error (M=5.75, SD=1.53, N=240) ($p < 0.002$). Of the 240 responses, there were 22 instances where pilots disagreed that the commanded speeds were operationally acceptable and appropriate. Out of the 22 responses, 14 occurred during scenarios containing wind error, and the remaining eight responses occurred during the scenarios containing offset error. A majority of the critiques that pilots provided throughout the experiment can be broken down into three categories: too many speed changes on final, too many speed changes within a short period of time, and undesirable speed increases.

K. Evaluation of the Pilot Interface

When adding new elements to flight deck displays, it is important to ensure that the displays support the philosophy of flight deck design, are developed to complement the flight crew's ability, and are used in the way designers anticipated. Many of the design decisions for the displays used in this experiment were based on previous work.[4,8,9-12,14] On the PFD, the commanded end speed was displayed directly above the FMS commanded speed, and the commanded speed bug was designed to mate with the FMS speed bug. This allows pilots to easily determine if the FMS speed matches the commanded speed, and what the commanded speed is in relation to other parameters displayed on the PFD. Similarly, the modifications to the ND were designed to allow pilots to quickly identify the lead aircraft and maintain situation awareness regarding their progress. Additionally, redundancy was built into the pilot interface. If the pilots failed to implement a speed decrease and were flying too fast, they could see the discrepancy on the PFD, and an "IM DRAG REQRD" message would appear on the EICAS display. If there was an ASTAR error, a caution was displayed on the EICAS display, and the mode on the PFD would change to "RVT." As a result of this human-centered design, pilots found most of the display element's intended to support spacing operations to be useful and intuitive. For example, 63% of the pilots reported that the display elements used in this experiment were easy to understand and positioned in logical locations. However, 29% of the pilots stated that the information conveyed by the displays was adequate and that improvements were needed regarding how and/or where the information was displayed. Most of the pilots recommending that improvements be made focused on the need for a more salient alert associated with changes to the commanded speed.

1. Usefulness of Display Elements

While various display elements did not serve as independent variables within this experiment, qualitative data collected from the pilot participants were used to examine their interaction with the spacing tool in terms of how useful they found each display element and how they monitored the displays. Using a scale of 1 (Detrimental) to 6 (Required), pilots rated all of the display elements on the PFD as "very useful" or "required," with the exception of the box located around the commanded speed (to alert pilots of a speed change) which was rated as moderately useful ($M=4.67$, $SD=1.46$, $N=24$). A majority of the other display elements on the ND, MCDU, and EICAS were rated as moderately useful, with the lowest rated display element, an "IM SPD LIMITED" EICAS message, receiving a rating of slightly to moderately useful ($M=3.38$, $SD=1.31$, $N=24$). In addition to asking pilots about the usefulness of each display element, pilots were asked to rate how frequently they used each display to monitor the spacing operation. Using a scale of 1 (Never) to 5 (All the Time), pilots reported that they monitored the PFD very often ($M=4.26$, $SD=0.75$, $N=24$), monitored the ND moderately to very often ($M=3.86$, $SD=1.11$, $N=24$), and monitored the MCDU slightly to moderately often ($M=2.70$, $SD=0.97$, $N=24$). These data demonstrate that the flight crew used the displays as anticipated and found the displays used in this experiment to be useful.

Since the spacing error was displayed only on the first IM MCDU page, there was concern that pilots would spend more time than desired monitoring that page rather than utilizing the MCDU for normal flight tasks as intended. The metrics collected included the total number of visits to the first IM page and the time spent on the first IM page as a percentage of total flight time. The full crew simulators collected the metric of total visits to either of the MCDUs, and the total time either of the MCDUs showed the first IM page as a percentage of flight time. Results indicate that the pilots visited the first MCDU page an average of 4.44 ($SD=3.13$, $N=180$) times throughout a flight and spent an average of 26% ($SD=30\%$, $N=180$) of the flight on the MCDU. However, median time spent on the first IM page was 11%. Since the IFD and DTS each have two MCDUs, the IM page could be kept on one and not the other. When only the IFD and DTS data were reviewed, the median percentage of time spent monitoring the first IM MCDU page was 17% ($M=22\%$, $SD=21\%$, $N=180$) of the simulation time, and the median number of times the page was visited was 4 ($M=4.58$, $SD=2.7\%$, $N=180$). Since the pilots were more likely to monitor the IM MCDU pages within the context of this experiment, the data suggest that pilots will use the spacing error but will not spend excessive amounts of time viewing the first IM MCDU page.

2. Alerting of the Commanded Speed Changes and Non-Conformance

One of the pilots' major responsibilities with respect to airborne spacing was to monitor the PFD for changes to the commanded speed. Previous experiments have alerted the flight crew to changes in the commanded speed by surrounding the commanded speed with a flashing box,[14] flashing the commanded speed,[8] and changing the commanded speed value field from normal to reverse video.[4] Within this experiment, the alert that was provided to the flight crew was the illumination of a green box around the commanded end speed for ten seconds. Pilot responses reveal that the green box was inadequate to attract the pilots' attention, and it caused the pilots to spend too much time monitoring the PFD for speed changes. Of the 24 pilot participants, 23 thought that a more salient alert, either in place of or in addition to the green box, was needed. Additional comments throughout the

questionnaires as well as researcher observations of flight crews confirm this result. Within the post-experiment questionnaire, the pilots were asked to select whether they desired the currently implemented green box, a flashing box, a chime, or any combination. Sixty-seven percent of pilots indicated that they would like an aural alert to be included as part of a system to notify of them of speed changes; 54% stated that they would like the box around the commanded end speed to flash for ten seconds; and 25% desired both a flashing box around the commanded end speed as well as an aural alert.

While it is useful to gather pilot opinions regarding the type of alert they desire, it is also important to discuss the potential impact from a human factors standpoint. The type of alert used should depend on the importance of the alert. A distinct aural alert will immediately direct a pilot's attention to a new commanded speed. However, it will cause increased disruption if the pilots are in the process of completing other critical tasks. Based on distribution of speed changes as a function of time-to-go and the dependence of the number of speed changes on error source, the disruption will worsen as the aircraft approaches the runway and as conditions worsen. In contrast, a flashing green box would provide a more salient alert than the solid box implemented in this experiment; however, the chance of the pilots missing the speed changes is greater with this implementation than for an aural alert.

In addition to the alert given for a new commanded speed, an EICAS alert, presenting the text "IM DRAG REQRD," was given when the aircraft was greater than 6 knots above the commanded speed and was turned off when the aircraft returned to within 4 knots of the commanded speed. A few pilots claimed that the EICAS message, "IM DRAG REQRD," did not necessarily indicate when they had to apply drag. Data were evaluated to determine how often pilots used the speed brake when the EICAS message appeared. The "IM DRAG REQRD" message was activated an average of 3.53 (SD=1.44, N=180) times per flight, and the flightcrew used the speed brake during 84% (SD=25%, N=180) of these instances, demonstrating that the message resulted in the desired response a majority of the time. The high standard deviation of compliance coupled with researcher observations suggests that there were some pilots who judicially used the speed brake to meet the commanded speed, while there were others who did not desire to use the speed brake to maintain the 5 knot tolerance. It is hypothesized that those who are willing to judicially use the speed brakes will find the "IM DRAG REQRD" message helpful, while those who do not will find it a nuisance.

3. Predictability of Speed Changes

Billings highlighted the importance of predictable automation by using the logic that if pilots are expected to be responsible for the actions of the aircraft, they must be able to monitor automated systems, therefore, the automation must be predictable.[27] The pilots who participated in this experiment tended to agreed with Billings' logic. Using a scale of 1 (Completely Disagree) to 7 (Completely Agree), the pilots moderately agreed that it is important to be able to predict changes to the commanded speed (M=5.46, SD=1.28, N=24). Additionally, pilots moderately agreed that the spacing tool behaved in a predictable manner (M=6.04, SD=0.55, N=24) and slightly agreed that they were able to predict commanded speeds before they occurred (M=5.36, SD=1.37, N=180). Thus, pilots' qualitative ratings indicate both the importance of predictability and that the spacing tool was somewhat predictable. To help drive future design decisions, it is important to discern what information pilots' use for their predictions. Pilots were asked to describe the displays and/or trends that helped them predict changes to the commanded speed, and their comments showed that 46% of them used the first IM page on the MCDU to drive their predictions while 46% used the published speed profile, scheduled speed decreases, or general flight rules to enable their predictions. Lastly, 29% of the pilots stated that they used visual representations of the lead aircraft and TCAS data to predict when speed changes would occur. While pilots will not be able to predict the precise moment when a commanded speed change will be presented, the data indicate that pilots are able to use the displays, profile, and flight regulations to form approximate predictions of the spacing tool's behavior.

4. Conformance Box

Research sponsored by NASA LaRC and by CoSpace indicates that pilots prefer symbology on the ND that graphically displays spacing trends.[10,14] CoSpace implemented a trend indicator, using a small bar with a trend bug located on the side of the ND. Part of a display usability study conducted in 2010 included examining pilot responses to displays: 1) without any trend indicator, 2) with a trend indicator similar to the one developed by CoSpace, and 3) with a green box around the ownship aircraft, called a "conformance box." Results of the usability study show that pilots rated the spacing trend indicator as the least important design element, but that pilots provided higher ratings regarding the understandability and intuitiveness of as well as confidence in the displays that included the presentation of a trend indicator.

Pilots in this experiment were presented with a "conformance box" during the experiment's final, exploratory run. The conformance box is a green box that appears around the depiction of the ownship aircraft on the ND,

indicating how much control authority ASTAR has. If the ownship moves outside the conformance box, the algorithm is predicting that it is no longer possible for the aircraft to meet the spacing goal by a given "achieve by point." The goal of the conformance box was to provide pilots with better predictability regarding the spacing operation. Using a scale of 1(Completely Disagree) to 7 (Completely Agree), pilots slightly to moderately agreed that the conformance box helped them monitor the IM operation (M=5.29, SD=1.78, N=24) and that the conformance box should be part of any display designed to support IM operations (M=5.36, SD=1.62, N=24). However, the pilots were only neutral to slightly in agreement with the statements that the conformance box helped them predict speed changes (M=4.75, SD=1.80, N=24), that it increased the level of safety of IM (M=4.75, SD=1.78, N=24), or that it increased their comfort with IM (M=4.88, SD=1.98, N=24). Additionally, one pilot misinterpreted the conformance box as a separation box, and one pilot would have disobeyed the commanded speed to center his aircraft in the box if his crewmember had not intervened.

Conclusions

This experiment examined the ASTAR10 algorithm's performance along with pilot participants' acceptance of a proposed concept and set of airborne spacing procedures, workload ratings, the spacing tool's commanded speed guidance, and the pilot interface. Although conditions where the algorithm was subjected to large error sources were experienced, pilots generally found the airborne spacing concept and procedures acceptable. For the most part, pilots agreed that the speeds commanded by ASTAR were correct, safe, and comfortable; however, the large wind and offset errors injected into the system resulted in specific algorithm behaviors that pilots found undesirable. These behaviors included too many speed changes within a short period of time, too many speed changes while on final approach, and speed increases that were deemed inappropriate. Workload ratings verify that the spacing tool was easy to use and added relatively little additional workload. Evaluation of the pilot interface used in support of the spacing operation found it was useful and utilized in the way designers anticipated. However, pilots found that the alert designed to notify them of speed command changes was not sufficiently salient, causing them to spend excessive time monitoring the speed command symbology on the PFD.

The results of this HITL experiment demonstrate the viability of the airborne spacing concept and procedures being investigated and suggest ways in which the spacing algorithm and pilot interface may be improved based on empirical data and end user feedback.

Acknowledgments

References

[1]FAA, "FAA Aerospace Forecast Fiscal Years 2010-2030". U.S. Department of Transportation Federal Aviation Administration Aviation Policy and Plans. 2010.

[2]FAA and EUROCONTROL, "Action Plan 1 FAA/EUROCONTROL Cooperative R&D Principles of Operation for the Use of Airborne Separation Assurance Systems". 2001.

[3]Meyn L. A. and Erzberger H., "Airport Arrival Capacity Benefits Due to Improved Scheduling Accuracy," in AIAA Paper AIAA-2005-7376, AIAA 5th Aviation Technology, Integration and Operations (ATIO) Forum, Arlington, Virginia, 2005.

[4]Bone R. S. and Penhallegon W. J., "En-route flight deck-based merging and spacing impact on flight crew operations," in Digital Avionics Systems Conference, 2007. DASC '07. IEEE/AIAA 26th, 2007.

[5]Bone R. S., Penhallegon W. J., and Stassen P., Flight Deck-Based Merging and Spacing during Continuous Descent Arrivals and Approach: Impact on Pilots. MITRE, 2008.

[6]Baxley B., Barmore B., Abbott T., and Capron W., "Operational Concept for Flight Crews to Participate in Merging and Spacing of Aircraft," in 6th AIAA Aviation Technology, Integration and Operations Conference (ATIO), 25-27 Sep. 2006, Wichita, KS, United States, vol. 2006.

[7]Credeur L., "Basic analysis of terminal operation benefits resulting from reduced vortex separation minima," NASA Headquarters, Washington, DC, Tech. Rep, 1977.

[8]Penhallegon W. J. and Bone R. S., "Flight deck-based merging and spacing impact on flight crew operations during continuous descent arrivals and approaches," in Digital Avionics Systems Conference, 2008. DASC 2008. IEEE/AIAA 27th, 2008, p. 3.

[9]Hébraud C., Hoffman E., Papin A., Pene N., Rognin L., and Sheehan C., "CoSpace 2002 light Deck Experiments Assessing The Impact of Spacing Instructions from Cruise to Initial Approach:Volume I," EEC Report No. 388 – Volume I, Feb. 2004.

[10]Hébraud C., Hoffman E., Pène N., Rognin L., and Zeghal K., "Assessing the impact of a new air traffic control instruction on flight crew activity," Human performance, situation awareness and automation: current research and trends: HPSAA II, p. 133, 2004.

[11]Hoffman E., Pene N., Rognin L., and Zeghal K., "Introducing a new spacing instruction. Impact of spacing tolerance on flight crew activity," in Human Factors and Ergonomics Society Annual Meeting Proceedings, 2003, vol. 47, p. 174–178.

[12]Hoffman E., Martin P, Putz T., Trzmiel A., and Zeghal K., "Airborne spacing: Flight deck view of compatibility with continuous descent approach (CDA)," in Proceedings of the 7th   AIAA Aviation Technology, Integration and Operations Conference. September 18-20, Belfast, Northern Ireland., 2007.

[13]Murdoch J., Barmore B., Baxley B., Capron W., and Abbott T., "Evaluation of an Airborne Spacing Concept to Support Continuous Descent Arrival Operations," in 8th USA-Europe Research and Development Seminar - ATM 2009, 29 Jun. - 2 Jul. 2009, Napa, CA, United States.

[14]Volk P. M., Takallu M. A., Hoffler K. D., Weiser J., Turner D. Airborne Precision Spacing for Dependent allel Operations Interface Study. NASA-TM (to be published)

[15]"Safety, Performance and Interoperability Requirements Document for Airborne Spacing-Flight Deck Interval Management (ASPA-FIM)," RTCA DO-328, RTCA, Washington, DC, June 22, 2011

[16]Abbott T. S., "Speed control law for precision terminal area in-trail self spacing," NASA Technical Memorandum, vol. 211742, 2002.

[17]Abbott T., "A Revised Trajectory Algorithm to Support En Route and Terminal Area Self-Spacing Concepts," NASA/CR-2010-216204, Feb. 2010.

[18] Baxley B.T., Hubbs C., Shay R., Karanian J., "Use of Data Comm by Flight Crew to Conduct Interval Management Operations to Parallel Dependent Runways," 11th AIAA   Aviation   Technology,   Integration   and Operations Conference. September 20-22, Virginia Beach, Virginia, 2011

[19]Peters M. E., Ballin M. G., Sakosky J. S., "A Multi-Operator Simulation for Investigation of Distributed Air Traffic Management Concepts," AIAA Modeling and Simulation Technologies conference and Exhibit. August 5-8, Monterey, California, 2002

[20]Prevot T., Smith N. M., and Palmer E. A., "The Airspace Operations Laboratory (AOL) at NASA Ames Research Center," AIAA Modeling and Simulation Technologies Conference and Exhibit, pp. 21-24, 2006.

[21]NASA Langley Research Center: the flight simulation facilities [Online]. URL: http://oim.hq.nasa.gov/oia/scap/docs/SCAP_FLIGHTSIM_112508_508.pdf. [cited 3 August 2011]

[22]Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.

[23]Barmore B. E., Abbot T. S., and Capron W. R., "Evaluation of Airborne Precision Spacing in a Human-in-the-Loop Experiment," in AIAA 5th Aviation, Technology, Integration, and Operations Conference (ATIO), 2005, p. 1–13.

[24]Oseguera-Lohr R. M., Lohr G. W., Abbott T. S., and Eischeid T. M., "Evaluation of operational procedures for using a time-based airborne interarrival spacing tool," in Digital Avionics Systems Conference.

[25]Barmore B., Abbot T. S., Capron W., and Baxley B., "Simulation Results for Airborne Precision Spacing along Continuous Descent Arrivals," in 8th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, 14-19 Sep. 2008, Anchorage, AK, United States.

[26]Wierwille W and Casali J. A valid rating scale for global mental workload measurement. Proceedings of the Human Factors Society 27th Annual Meeting, Norfolk, VA, October 1983

[27]Billings C., "Human-centered aviation automation: Principles and guidelines," NASA-TM-110381, 1996.