Effects of Communication Modality on Pilot-Controller Coordination during a Simulated m:N Operation

Garrett G. Sadler*, Meghan Chandarana†, Jillian N. Keeler*, Crystal Kirkley*
NASA Ames Research Center, Moffett Field, CA, USA

The last decade or so has seen growing interest in new control paradigms and concepts of operation for uncrewed aircraft systems (UAS) in which multiple aircraft are piloted remotely by a single or relatively small number of people. Referred to as "one-to-many" and "many-to-many" (alternatively, "multi-operator, multi-vehicle")—and frequently expressed as the corresponding ratios, 1:*N* and *m*:*N*—such novel configurations of aircraft and the people who manage them are seen as critical to the path to future operations involving UAS. Examples of industry domains interested in these control paradigms are small package delivery services utilizing small UAS and passenger-carrying, short-range "Urban Air Mobility" (UAM) operations. Stakeholders in such operations have identified communication and coordination of flight activity with air traffic controllers (ATC) as a barrier to operations. In contrast to present-day flight operations, in which a pilot communicates with one ATC on one radio frequency for one aircraft, multi-vehicle operations potentially entail a significant increase in pilot task load for management of comms. New concepts, such as UAS Service Suppliers (USSs) and Providers of Services to UAM (PSUs), have been proposed to address the known bottleneck for Air Traffic Management (ATM) presented by multi-vehicle operations. While progress has been steadily made over years developing USSs and PSUs, it is generally expected that initial UAM operations will rely on traditional voice-over-radio communication with ATC for purposes of ATM. The current study was a human-in-the-loop simulation that had participants, each possessing a Private Pilot License, act as the ground-based pilot-in-command for multiple vehicles in a hypothetical UAM service in the San Francisco Bay Area. The experiment utilized a 2-by-3, within-subjects design in which the pilot's Vehicle Load (4 vs. 12) and Comm System (Voice, Datalink, and a Hybrid) were manipulated. The task given to pilots was to use the Comm System to coordinate flight activity for all aircraft with appropriate controllers, having to obtain departure and arrival clearances at "vertiport" facilities and transition clearances for any intermediate airspaces along the route. Pilots were additionally responsible for compliance with vectoring instructions issued by ATC. Subjective workload questionnaires (NASA-TLX) were administered following each experimental trial. Screen recordings of the pilot's Ground Control Station (GCS) and audio recordings of trials were subsequently coded to obtain performance metrics: response times and error rates. Presented in this paper are results related to pilot responses to vectoring instructions issued by ATC. Workload was found to be significantly higher in the 12-Vehicle condition compared to the 4-Vehicle condition, nearly maxing out the NASA-TLX overall workload scale. There was no significant difference made by the Comm System on workload ratings. Pilots' response times to communications were fastest in the Voice condition, although overall "service time" for compliance was shorter in Datalink and Hybrid conditions in most cases. Errors by pilots were frequent in both Vehicle Load conditions, most perniciously when using the Voice system. The results of this study suggest tradeoffs in advantages and disadvantages of the three comm systems. Recommendations for communication system design are provided taking the tradeoffs into account.

---

* Human Factors Researcher, Human/Machine Systems, Human Systems Integration Division.
† Computer Engineer, Intelligent Systems Division.

I. Nomenclature

$VecRBL$	=	vectoring readback length
$CommRT_{Vec}$	=	vectoring communication response time
$ManRT_{Init}$	=	initial maneuver response time
$ManRT_{Edit}$	=	maneuver edit time
$ManRT_{Serv}$	=	maneuver service time
$RetRBL$	=	return-to-course readback length
$CommRT_{Ret}$	=	return-to-course communication response time
$RetRT_{Init}$	=	initial return-to-course response time
$RetRT_{Edit}$	=	return-to-course edit time
$RetRT_{Serv}$	=	return-to-course service time

II. Introduction

A known barrier to the full-scale implementation of operations involving large numbers of aircraft under the command of one or a small number of pilots is coordination with air traffic control (ATC) for purposes of air traffic management (ATM). It is expected that the volume of uncrewed aircraft systems (UAS) that will be employed in novel categories of aviation will overwhelm the capabilities of present-day ATM infrastructure and its attending ATC workforce [1]. Recognizing this bottleneck to operations, researchers and stakeholders have proposed future concepts that envision highly automated ATM services [2-4]. Examples of such novel solutions include the utilization of UAS Service Suppliers (USSs) for UAS Traffic Management (UTM) [5] and, in the case of Urban Air Mobility (UAM), Providers of Service for UAM (PSUs) [1, 4]. Digital communication systems, such as Controller Pilot Data-Link Communications (CPDLC) and the Aircraft Communications Addressing and Reporting System (ACARS), have been implemented and are regularly used, subject to availability and regulations, in commercial airline operations. For ATM purposes, however, ATC coordination is almost always facilitated through serial communications over radio. There is currently significant investment and research efforts focused on the ATM environment for UAS; nevertheless, in the UAM case, it is expected that the initial UAM ecosystem will utilize existing helicopter infrastructure, including traditional voice-over-radio ATC services [2].

In the late 1990s and early 2000s, researchers performed a series of studies examining performance impacts of various communication modalities for airline operations [6-8]. These studies manipulated communication modality and the time interval between subsequent transmissions. The communication mode variable was comprised by voice over radio, datalink (CPDLC), and a mixed-media "hybrid" mode consisting of both voice and datalink. The time interval variable consisted of long (1 minute) and short (5 second) intervals between transmissions. Researchers found that voice transmission resulted in the shortest overall transaction time for communications (i.e., the total time from the beginning of an ATC transmission to the end of the pilot's response) and that time pressure increased both the transaction time and the number of errors committed by pilots.

The studies described in Refs. [6-8] examined communications with ATC for a (simulated) traditional flight deck consisting of a captain and first officer onboard a (single) commercial airliner. To achieve sufficient economies of scale, common examples of AAM envision the pilot-in-command (PIC) relocated to a position at a ground control station (GCS) and responsible for directing multiple aircraft in a management-by-exception [9] framework. Recently, researchers have begun to study a new control paradigm known colloquially as "m:N" (vocalized as, "em to en") [10-12], which expresses a ratio whereby m is the number of vehicle operators who share responsibilities for N-many vehicles between them. In this context, N is always greater than or equal to m. The m:N paradigm is a broad category of control configurations that may be deployed across a variety of operational contexts, including Unmanned Aircraft System (UAS) Traffic Management (UTM), Advanced Air Mobility (AAM) and Urban Air Mobility (UAM), medium-to-large UAS operations, high altitude platform systems (HAPS), as well as swarms of micro or small UAS (sUAS) [13]. The concept is meant to enable a future state of scalable operations for increasingly autonomous vehicles.

It remains to be determined exactly how communications over the different media (e.g., the voice, datalink, and hybrid modalities of Refs [6-8]) will be characterized in an m:N environment. As of this writing, there is significant interest in some version of m:N from industry [13]. The assumption of an m:N control paradigm has operator situation awareness (SA) impacts for communications. Simultaneous control of multiple aircraft periodically requires operators to switch attention and/or control among vehicles. Research has shown that doing so typically slows response times and increases errors [14, 15]. There is evidence that this cost may be reduced if the participants have a chance to prepare for the switch or receive task-switching cues. Similarly, the design of the operator interfaces for use during m:N operations must support effective information gathering and timely attention-switching [16].

This paper presents the details and results of a human-in-the-loop (HITL) experiment conducted to examine the effects of radio, digital, and radio-digital "hybrid" communication modalities on pilot performance and workload. Participants in this HITL played the role of a multi-vehicle pilot of simulated UAM flights in the San Francisco Bay Area and were responsible for coordinating flight activity for all their aircraft with ATC using the aforementioned communication modalities. This study was conducted in support of NASA's Transformational Tools and Technologies (TTT) Project and in conjunction with two industry partners: Joby Aviation and Wisk Aero. Results of the study will be discussed for generalization to other m:N domains. Finally, recommendations for requirements on communication systems and procedures for m:N will be presented.

III. Method

A. Experimental Design

The present study utilized a two-by-three, within-subjects experimental design. The independent variables consisted of the number of vehicles under participants' management ("Vehicle Load" variable, two levels: "Low" and "High") and the modality utilized to communicate with ATC ("Comms" variable, three levels: "Voice", "Datalink", and "Hybrid"). In the "low vehicle load" condition, each participant managed 4 vehicles. In the "high vehicle load" condition, the participant managed 12 vehicles. In all conditions, participants communicated with a confederate researcher playing the role of ATC. For the "Voice" condition, participates spoke with ATC over serial radio channels resembling current-day ATC in the National Airspace System (NAS). When in the "Datalink" condition, participants communicated with ATC via a chat-based system that involved selecting and sending preformatted messages for clearances. In the "Hybrid" condition, a combination of both digital and voice communications were utilized. Participants managed vehicles using Vigilant Spirit Control Station (VSCS), a software suite originally developed by the Air Force Research Laboratory (AFRL) [17]. Participants were situated at a workstation with three monitors, a keyboard, a mouse, and a push-to-talk (PTT) headset.

Each participant experienced a total of six experimental trials, each 30 minutes in length. Participants were tasked with using the ground control station interface, detailed below, to monitor the location and timelines of aircraft and obtain clearances (departure, transition, and arrival) from ATC as appropriate. Researchers played the role (confederate) of ATC. Controllers additionally reached out to the participant at scripted times with vectoring instructions for aircraft. Vectoring instructions were given to three aircraft (two in quick succession, one individual) in the 4-Vehicle condition. Five aircraft (two in quick succession, and three individual) received vectoring instructions in the 12-Vehicle condition.

B. Participants

Twelve participants (M = 38.75 years of age), each possessing a Private Pilot License (PPL), were recruited for the present study. Pilots had an average of 5,019 hours of crewed (i.e., onboard piloting) flight experience. Ten pilots were instrument flight rules (IFR) rated. Three pilots, each Part 107 certificated, additionally had uncrewed (i.e., remote piloting) experience. These pilots had an average of 28.3 hours of remote flying experience. All remote flying hours were with UAS weighing less than 55 lbs.

C. Training and Experiment Procedures

Participants were trained on how to operate the simulated electric vertical take-off and landing (eVTOL) vehicles for flights between two vertiport (i.e., aerodrome) facilities. Participants additionally received training on how to communicate with ATC across all three comms conditions. During experimental trials, participants acted as the PIC managing multiple vehicles which were at various stages of flight (e.g., departure, enroute, and approach). Flights proceeded between two simulated vertiports in the San Francisco Bay Area: one located at the San Jose International Airport (SJC) and the other located adjacent to the San Francisco Ferry Building (SFF). The route flown between the two vertiports involved passing through various airspaces (Fig. 1) and required that the participant, as the PIC, communicate with ATC control towers at the vertiports and at airports along the San Francisco Peninsula. Participants were responsible for contacting ATC to receive departure and arrival clearances at the two vertiports and to obtain transition clearance through the Class Bravo airspace surrounding San Francisco International Airport (SFO) and Palo Alto Airport (PAO). Periodically throughout the simulation, ATC issued additional instructions and clearances for participants' vehicles (e.g., new altitude assignments, speed changes, heading changes).

Fig. 1 The route between SJC and SFF vertiports. Blue chevrons represent UAM aircraft under control by the PIC and white chevrons represent "background traffic." Airspace boundaries are shown in light blue.

D. Ground Control Station (GCS) Interface

The GCS used by participants was comprised primarily of two monitors (Fig. 2). The main monitor in directly in front of participants included a Tactical Situation Display (TSD) with a map displaying the airspace and vehicles (shown as blue chevrons) being managed by the participant. In addition, overlaid on the map was a sectional view of the area with the four airspaces requiring ATC clearances outlined in light blue. Background traffic in the area was displayed as white chevrons, while checkpoints between the four airspaces along the vehicles' routes were shown as yellow flags. Pilots were able to reference these checkpoints when determining when to place their transition requests for each vehicle.

For the Datalink and Hybrid conditions, an additional chat interface with chatrooms for each of the four towers along the route was provided to the right of the map on the main screen. When a chat message was received from ATC, an aural notification sounded and a "speech bubble" icon appeared beside the target aircraft's chevron on the TSD. In the Datalink condition participants used the chat for all clearances (including arrival, departure, and transition requests) and responding to vectoring instructions issued by ATC, while in the Hybrid condition the chat was only used to request arrival and departure clearances and to respond to vectoring instructions. For all arrival, departure, and transition requests made through the chat, participants were able to use predefined chat messages from a pop-up menu at the bottom of the chat. These messages automatically populated the callsign of the vehicle selected at the time the message was selected.

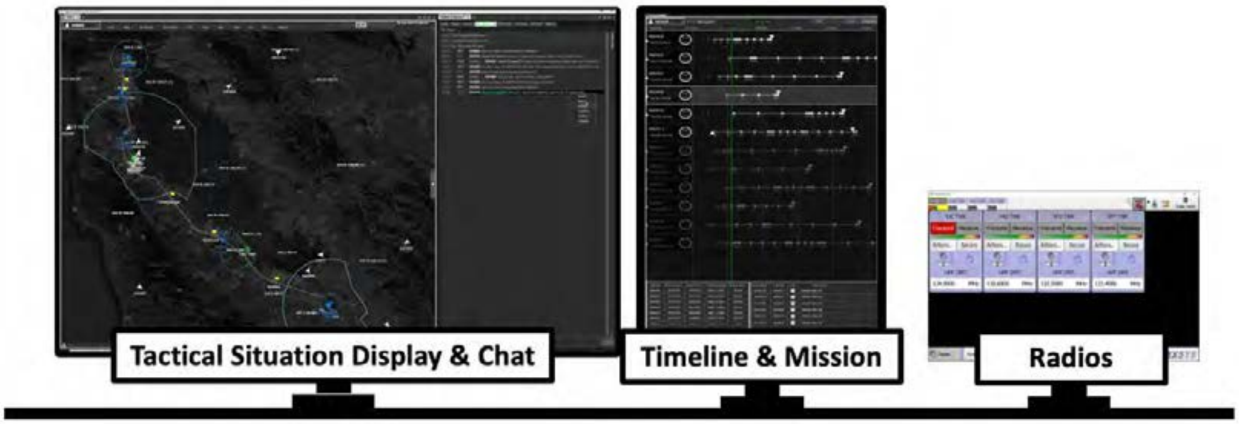

Fig. 2 Diagram of the GCS. The chat panel was not made available in the Voice condition and the radio panel was not made available in the Datalink condition.

On the secondary screen, the GCS provided a timeline view for all vehicles managed by the participant. This included vehicles that were still on the ground waiting for departure clearances. The timeline view also provided information about the remaining battery life of each vehicle. Lastly, the timeline provided information about the next major event for each vehicle including the time remaining until its scheduled departure time if the aircraft was still located at either of the vertiports awaiting takeoff. Under the timeline on the secondary screen, a status and event log was displayed for all managed vehicles.

To the right of the secondary screen a small radio panel was provided. Individual channels for each of the four airspace towers along the route was provided. For the Voice condition, participants used this radio panel for all communications with ATC including arrival, departure, and transition requests, as well as, responding to all ATC calls. The radio panel was also used by participants in the Hybrid condition for transition requests. The radio panel was not used at all by participants in the Datalink condition.

1. Arrival and Departure Requests

Participants were required to monitor both the timeline and the TSD to appropriately send arrival and departure requests. For the Voice condition, all arrival and departure requests were made through the radio. Once approval was granted no future action was required in the case of arrival clearances. In the case of departure clearances, the participants were also required to manually send a takeoff command to the vehicle. This was done by right clicking on the vehicle and choosing the "Launch Takeoff" option (Fig. 3).

Fig. 3 Context menu used to manually send a takeoff command to a vehicle in the Voice condition.

In contrast, for the Datalink and Hybrid conditions all takeoff and landing requests were made through the chat interface. To complete either a takeoff or landing request, participants first selected a vehicle, then switched to the correct tower's chat window (Fig. 4.1). Next, they selected the appropriate arrival or departure clearance message from the pop-up menu and send it to ATC in the chat (Fig. 4.2). ATC would then respond with approval. In the case of departure clearances, the response from ATC also included a clickable link that when pressed, would initiate an automated takeoff sequence onboard the vehicle (Fig. 4.3). For both requests participants were required to acknowledge (i.e., readback) the ATC response via a set of standard responses. This was accomplished by right clicking on the ATC response and choosing one of the standard messages that was provided within the context menu (Fig. 4.4).

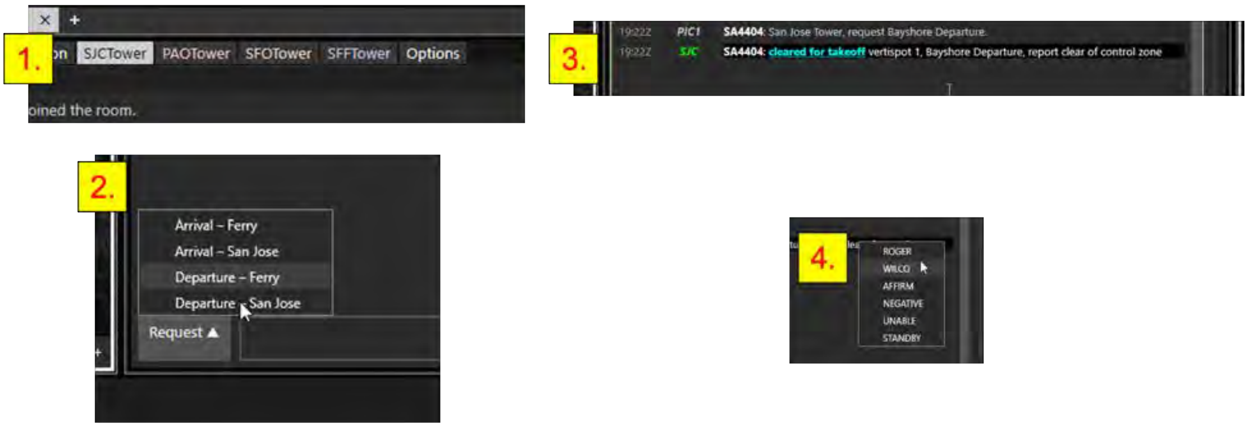

Fig. 4 Sequence of steps required to obtain and respond to arrival and departure clearances from ATC in the Datalink and Hybrid conditions.

2. Transition Requests

Like arrival and departure clearances to make and obtain transition requests, participants in the Voice condition were expected to use the radio panel to call the appropriate ATC tower for each vehicle needing transition clearance. Transition requests into and transition notifications out of airspaces for each vehicle were required. The Hybrid condition also required participants to make transition requests and notifications via the radio. In contrast, the Datalink condition, as for arrival and departure clearances, required participants to send transition calls via the chat interface. Just like for arrivals and departures, the participants first needed to select a vehicle, then select the appropriate tower's chat room. Next, they would use the pop-up menu at the bottom of the chat to select the correct transition message based on the checkpoint that the vehicle was located at (Fig. 5). ATC would then provide approval and the participant would be expected to send an acknowledgement using the same context menu that was used to respond to arrival and departure messages from ATC.

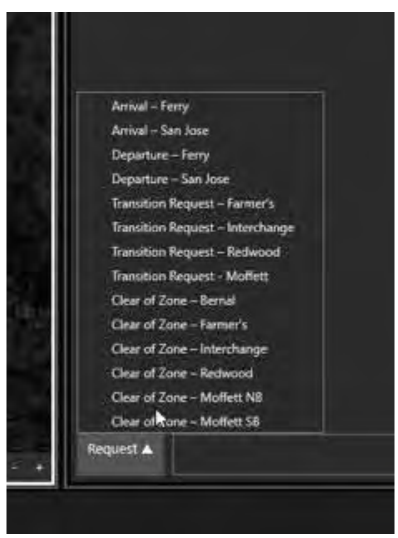

Fig. 5 Pop-up menu in the Datalink conditions chat showing all options for transition, arrivals, and departure clearances.

3. Vectoring

In the Voice condition, as with the transition and arrival/departure requests, all ATC calls were made over the radio. These calls were made by ATC towers on their respective frequencies to specific vehicles. Each call requested that either a speed, altitude, or heading be changed for a particular vehicle. When a call came in over the radio, participants were required to provide a readback of the request on the correct frequency. Then they were expected to double-click on the vehicle to bring up the "Steering Window" (i.e., autopilot menu, Fig. 6.1), Switch to the HOLDS tab (Fig. 6.2), enter the appropriate maneuver parameter (Fig. 6.3) and click send to uplink the maneuver (Fig. 6.4). Approximately one minute later, ATC requested that the vehicle return to course (RTC). Participants again were expected to provide a readback on the correct frequency to ATC. To implement their compliance, participants doubled-clicked on the vehicle again to bring up the Steering Window (Fig. 7.1). This time they chose the NAV tab (i.e., waypoint-to-waypoing menu, Fig. 7.2), selected the waypoint provided by ATC (Fig. 7.3) and clicked send to uplink the maneuver (Fig. 7.4).

Fig. 6 Sequence of actions required to upload a new maneuver to a vehicle in response to an ATC call in the Voice condition.

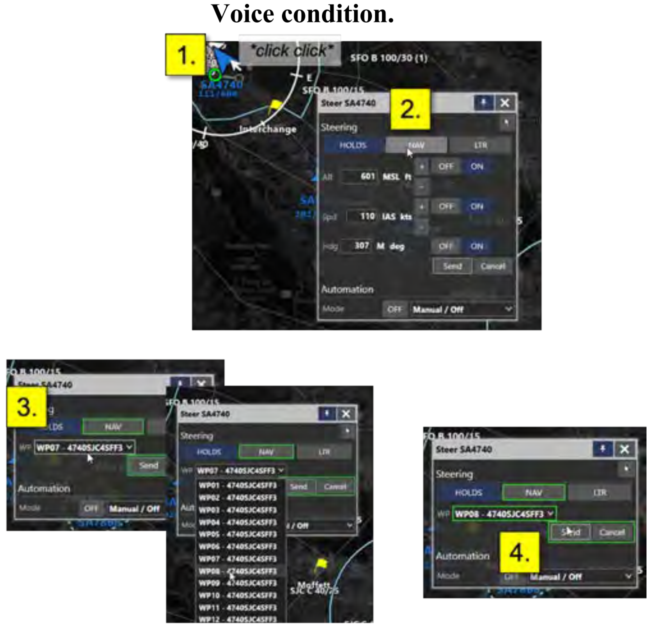

Fig. 7 Sequence of events required to upload a Return to Course maneuver to a vehicle in the Voice condition.

For the Datalink and Hybrid conditions, all ATC communications were received through the chat. For vectoring instructions, each ATC message included a hyperlink (Fig. 8.1) which provided a shortcut to the HOLDS tab of the Steering Window (Fig. 8.2) for the corresponding aircraft. When the window appeared, the instructed value for the maneuver was prepopulated to match the parameter requested by ATC. Participants were required to verify the value and click send to uplink the maneuver. They were then required to perform a readback like the arrival/departure and transition approvals by right-clicking the message (Fig 4.4) and choosing one of the predefined standard readback messages. Importantly—and with consequences related to Section IVB ("Confusion Loops") below—when the pilot sent the readback message to the chat system, the hyperlink for the maneuver was also disabled. As a result, participants were required to uplink the maneuver before performing the readback. Similarly, when a return to course message arrived a minute later, it did so through the chat with a hyperlink shortcut (Fig. 9.1) that when clicked on would automatically open the NAV tab of the Steering Window with the correct waypoint chosen (Fig. 9.2). Participants again would need to confirm the waypoint, send the maneuver to the vehicle, and perform a readback in the chat.

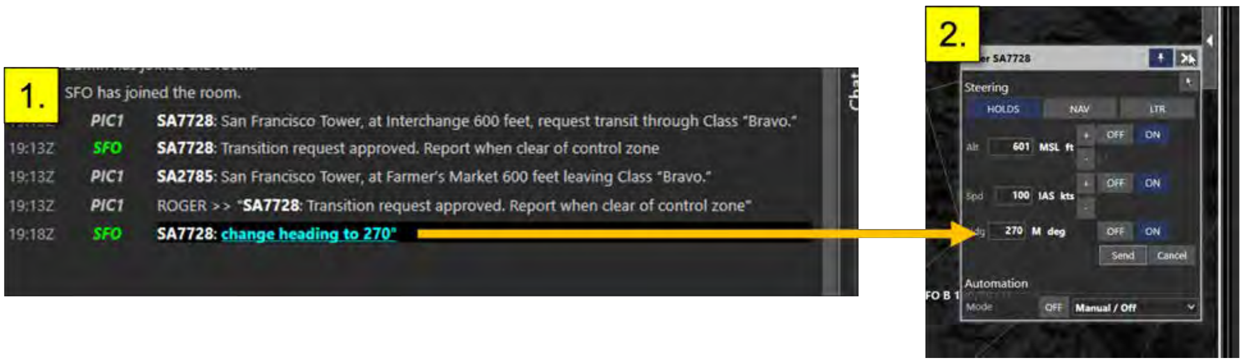

Fig. 8 ATC call for Datalink and Hybrid conditions that is sent through the chat interface. The message contains a hyperlink that automatically brings up the Steering Window and loads the associated heading, speed, or altitude value.

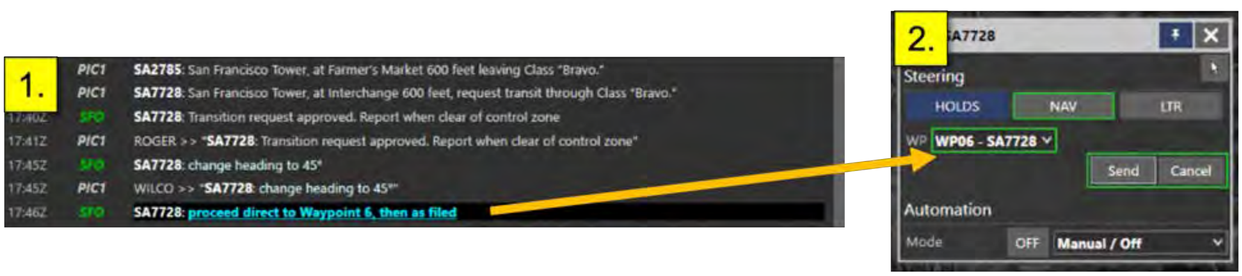

Fig. 9 ATC call for Return to Course for the Datalink and Hybrid conditions is sent through the chat interface. The hyperlink provided automatically brings up the Steering Window with the correct waypoint selected for the maneuver.

IV. Metrics

A. Subjective Workload Ratings

Subjective workload ratings were collected using NASA's Task Load Index (TLX) [18]. The TLX questionnaire was administered following each experimental scenario (i.e., each combination of the Comm and Vehicle Load variables). Weightings for TLX subscales was not employed in this analysis; instead "raw" TLX [19] values are produced here. The Overall Workload score is calculated here as the sum of subscale scores.

B. Performance: Response Time Metrics

Audio and video recordings of trials were coded by researchers in the Human-Autonomy Teaming Laboratory at NASA Ames Research Center to collect timestamps for pilot actions in response to vectoring instructions issued by ATC. The timestamps collected for the calculation of response times (RTs) include: the beginning/end of ATC transmissions, the beginning/end of pilot readback transmissions, times at which the pilot accessed the autopilot and navigation menus for aircraft, and the times at which maneuvers were uplinked to aircraft. See Table 1 for a detailed description of how response times were calculated and for the symbolic designations used for response time metrics. Each category for response times listed below come in pairs: response times related to the vectoring maneuver instructed by ATC and the subsequent response times for the return-to-course (RTC) action.

1. Communication Response Time ($CommRT_{Vec}$ and $CommRT_{Ret}$)

Refers to the length of time between the end of an ATC transmission providing instructions to the participant and the participant's readback of that transmission: $CommRT_{Vec}$ and $CommRT_{Ret}$ denote the RTs to the vectoring and RTC instructions from ATC, respectively.

2. Readback Length (VecRBL and RetRBL)

Refers to the duration of the readback transmission from the pilot in response to vectoring and RTC instructions from ATC. In the Voice condition, the readback length is calculated from the time at which the participant depressed and released the PTT button. In the Datalink and Hybrid conditions, the readback length is calculated starting from the time at which the pilot accessed the readback context menu in the chat system (Fig. 4.4) to the time at which the pilot pressed the "Send" button in the chat panel to transmit the readback to ATC. Readback lengths for the vectoring and RTC instructions are designated by *VecRBL* and *RetRBL* respectively.

3. Maneuver and RTC Initiation Response Time ($ManRT_{Init}$ and $RetRT_{Init}$)

Refers to the length of time between the end of the ATC transmission providing vectoring/RTC instructions and the time at which the participant accessed the Steering Window (Figs. 6.2 and 8.2) for the instructed maneuver, $ManRT_{Init}$, or for returning the vehicle to its previous navigation mode (Fig. 7.2 and 9.2) for RTC, $RetRT_{Init}$.

4. Maneuver and RTC Edit Response Time ($ManRT_{Edit}$ and $RetRT_{Edit}$)

Refers to the length of time starting when the participant accessed the Steering Window to the time the participant uplinked the vectoring maneuver ($ManRT_{Edit}$) or RTC action ($RetRT_{Edit}$) to the vehicle by pressing the "Send" button (Figs. 6.4 and 8.2). If more than a single upload was performed for compliance by the participant, the final upload time was used in the calculation of the edit time.

5. Maneuver and RTC Service Time ($ManRT_{Serv}$ and $RetRT_{Serv}$)

Refers to the length of time starting at the end of the ATC transmission providing vectoring/RTC instructions to the time the participant uplinked the vectoring maneuver ($ManRT_{Serv}$) or RTC action ($RetRT_{Serv}$).

C. Performance: Error Rates Missed Calls, and Clarification Requests

In addition to coding timestamps for the calculation of RT metrics, coders also designated whether transmissions made by the participant contained errors, when calls went without response, and when participants requested clarification from ATC.

1. Frequency Error Rate

Refers to the proportion of readback transmissions made by the participant over an incorrect radio frequency (i.e., transmissions sent to an incorrect ATC authority). Because readback transmissions made using the Datalink and Hybrid Comm systems were sent to the appropriate ATC authority by design, *frequency errors are only recorded for transmissions in the Voice condition*.

2. Callsign Error Rate

Refers to the proportion of readback transmission made by the participant in which an incorrect vehicle callsign was vocalized (e.g., "SimAir7881" instead of "SimAir7081"). Because readback transmission made using the Datalink and Hybrid Comm systems were preformatted with the correct callsign by design, *callsign errors are only recorded for transmissions in the Voice condition.*

Table 1. Description, definition, and formulas used to calculate response time metrics.

Event	Example/Description	Event Category Designation	Associated Timestamp Designations	Associated Response Time (RT) Metrics	
Vectoring call from ATC to pilot	"SimAir2481, San Francisco Tower, change heading to one-four-zero."	T_0	T_0^{on} := start of vectoring call T_0^{off} := end of vectoring call	Vectoring Call Length: $VecCL := T_0^{off} - T_0^{on}$	Vectoring Comm RT: $CommRT_{Vec} := T_1^{on} - T_0^{off}$
Readback (RB) of vectoring call by pilot	"San Francisco Tower, changing heading to one-four-zero, SimAir2481."	T_1	T_1^{on} := start of vectoring RB T_1^{off} := end of vectoring RB	Vectoring RB Length: $VecRBL := T_1^{off} - T_1^{on}$	Initial Maneuver RT: $ManRT_{Init} := T_2^{i} - T_0^{off}$
Vectoring compliance by pilot	Pilot accesses autopilot (AP) menu to edit and uplink the vectoring maneuver (man.) instructed by ATC.	T_2	T_2^{i} := opening of AP menu T_2^{up1} := first man. upload T_2^{upf} := final man. upload	Maneuver Edit Time: $ManRT_{Edit} := T_2^{upf} - T_2^{i}$	Maneuver Service Time: $ManRT_{Serv} = T_2^{upf} - T_0^{off}$
Return-to-course (RTC) call from ATC to pilot	"SimAir2481, San Francisco Tower, proceed direct to [waypoint], then as filed."	T_3	T_3^{on} := start of RTC call T_3^{off} := end of RTC call	RTC Call Length: $RetCL := T_3^{off} - T_3^{on}$	RTC Comm RT: $CommRT_{Ret} := T_4^{on} - T_3^{off}$
Readback of RTC call by pilot	"San Francisco Tower, direct to [waypoint], then as filed, SimAir2481."	T_4	T_4^{on} := start of RTC RB T_4^{off} := end of RTC RB	RTC RB Length: $RetRBL := T_4^{off} - T_4^{on}$	Initial RTC RT: $RetRT_{Init} := T_5^{i} - T_3^{off}$
RTC compliance by pilot	Pilot accesses navigation menu to uplink RTC and resume original route	T_5	T_5^{i} := opening of NAV menu T_5^{up1} := first RTC upload T_5^{upf} := final RTC upload	RTC Edit Time: $RetRT_{Edit} := T_5^{upf} - T_5^{i}$	RTC Service Time: $RetRT_{Serv} := T_5^{upf} - T_3^{off}$

3. Maneuver Error Rate

Refers to the proportion of maneuvers that were uplinked with incorrect parameters: either the maneuver was made in an incorrect dimension (e.g., a speed change to 140kts when a heading change to 140° was instructed) or an incorrect value was entered into the autopilot (e.g., an altitude change to 1200ft when 1100ft was instructed). In the Datalink and Hybrid Comm systems, the autopilot was automatically filled with the parameters of the maneuver instructed by ATC; consequently, *maneuver errors are only recorded for those manually executed in the Voice condition.*

4. Total Error Rate

Refers to the proportion of readback transmissions and/or uplinked maneuvers that contained one of the above three error types. The Total Error Rate was calculated by tallying the number of frequency, callsign, and/or maneuver errors divided by the total number of opportunities for those errors to occur. Because each of the above error types are only recorded for the Voice condition, *the Total Error Rate is only calculated for the Voice condition.*

5. Clarification Requests

Refers to the proportion of calls in which the participant asked ATC for clarification of either the callsign of an aircraft or the parameters of the instructed action. Because callsigns and maneuver parameters are automatically filled into the autopilot in the Datalink and Hybrid conditions, *clarification requests are only recorded in the Voice condition.*

6. Missed Calls

Refers to the proportion of calls that were either ignored (i.e., went entirely unacknowledged) by the participant after ATC provided vectors (Ignored Calls) or calls that were scripted into scenarios but had to be removed because the participant failed to instruct the recipient aircraft to take off (Never Took Off).

D. "Confusion Loops"

An emergent behavior was observed in some rare cases, which we have given the name "confusion loops." This behavior was only possible in the Datalink and Hybrid conditions. In the case of a confusion loop, a pilot would click on the hyperlink issued by ATC for a vectoring maneuver (Fig. 8.1) and uplink the maneuver to the aircraft. After uplinking the maneuver, *the pilot would fail to perform the readback* (Fig 4.4), leaving the hyperlink for the maneuver active and setting up the conditions for the loop. One minute after the vectoring maneuver (and its hyperlink) came through the chat, the RTC message would be sent by ATC. The pilot would then perform the RTC maneuver. In some cases, pilots would additionally fail to perform the readback for the RTC message. In either cases, the pilot would go on to other tasks and at some later time notice that there was an unacknowledged message and hyperlink for an aircraft in a chatroom. The pilot would then *re-perform the maneuver and/or RTC* by clicking on the hyperlink(s), presumably because they believe they accidentally missed the message previously. This loop could go through several repetitions if it happened that the pilot subsequently forgot to perform a readback again, leaving the hyperlink active and waiting in the chat for another go-around of the loop.

V. Results

Two-way repeated measures ANOVAs were performed to analyze the effects of the Comm System and Vehicle Load variables on subjective workload ratings (NASA-TLX) and response time metrics with an α-level of 0.05. Descriptive statistics are reported for error rate, missed calls, clarification requests, and "confusion loop" metrics. Outlier response times were common in the Hybrid and Datalink conditions (Fig. 13), due to incidents of leaving vectoring and/or RTC messages unacknowledged in the chat system for extended periods of time.

A. Subjective Workload Ratings

Unsurprisingly, workload ratings from participant were substantially higher in the 12-Vehicle condition than in the 4-Vehicle condition. Unexpectedly, however, the Comm System did not play a role in driving differences in workload. Overall workload was, on average, 1.97-2.36 times higher in the 12-Vehicle condition compared to the 4-Vehicle condition (Fig 10). A significant main effect of Vehicle Load on Overall workload was observed ($F(1, 11) = 198.70$, $p < 0.001$, $\eta^2 = 0.948$). There was no observed statistically significant main effect of Comm System ($F(2, 11) = 0.052$, $p = 0.949$, $\eta^2 = 0.005$); neither was there a significant interaction effect of the Vehicle Load/Comm System pair ($F(2, 11) = 1.901$, $p = 0.173$, $\eta^2 = 0.147$). Overall workload ratings in the 12-Vehicle condition were so high that they nearly reach the theoretical (raw TLX) maximum score of 42.

This marked difference in overall workload is similarly observed for the TLX subscales: there is little substantive difference in the subscales as a function of the Comm System, whereas the subscales meaningfully differ between the Vehicle Load conditions. Workload ratings for TLX subscales are provided in Figure 11. We found statistically significant main effects of Vehicle Load on all six TLX workload subscales: Mental ($F(1, 11) = 48.59$, $p < 0.001$, $\eta^2 = 0.815$), Physical ($F(1, 11) = 20.25$, $p < 0.001$, $\eta^2 = 0.648$), Time Pressure ($F(1, 11) = 55.45$, $p < 0.001$, $\eta^2 = 0.834$), Performance ($F(1, 11) = 34.65$, $p < 0.001$, $\eta^2 = 0.759$), Effort ($F(1, 11) = 75.96$, $p < 0.001$, $\eta^2 = 0.873$), and Frustration ($F(1, 11) = 20.33$, $p < 0.001$, $\eta^2 = 0.649$). There was no statistically significant main effect of the Comm System variable on any of the TLX subscales: Mental ($F(1.328^{‡}, 11) = 0.324$, $p = 0.641$, $\eta^2 = 0.087$), Physical ($F(2, 11) = 1.064$, $p = 0.362$, $\eta^2 = 0.088$), Time Pressure ($F(2, 11) = 0.048$, $p = 0.953$, $\eta^2 = 0.004$), Performance ($F(2, 11) = 0.209$, $p = 0.813$, $\eta^2 = 0.019$), Effort ($F(2, 11) = 0.314$, $p = 0.734$, $\eta^2 = 0.028$), Frustration ($F(1.378^{‡}, 11) = 0.905$, $p = 0.389$, $\eta^2 = 0.076$).

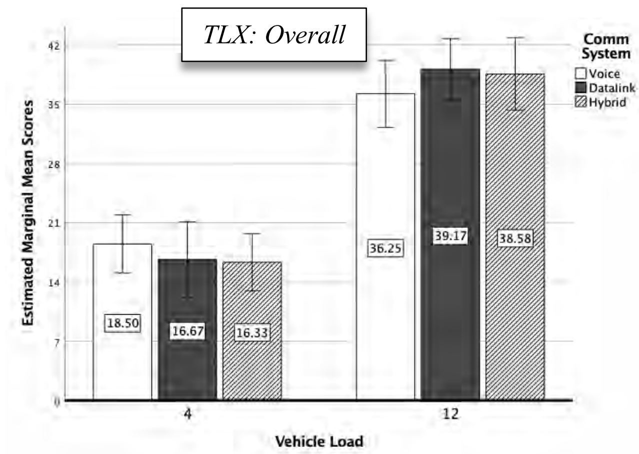

Fig. 10. Estimated marginal means for Overall TLX scores, as computed by the sum of TLX subscale scores for each participant. Error bars represent the 95% confidence interval.

B. Performance: Response Time Metrics

1. Comm Response Time ($CommRT_{Vec}$ and $CommRT_{Ret}$)

After controlling for outliers, statistically significant main effects on $CommRT_{Vec}$ were observed for both the Comm System ($F(2, 9) = 15.14$, $p < 0.001$, $\eta^2 = 0.63$) and Vehicle Load ($F(1, 9) = 19.93$, $p = 0.002$, $\eta^2 = 0.69$) variables. Additionally, a significant interaction of the Comm and Vehicle Load variables was observed ($F(2, 9) = 4.34$, $p = 0.029$, $\eta^2 = 0.33$). Estimated marginal means for $CommRT_{Vec}$ are provided in Table 2 and Fig. 12a.

Table 2. Estimated Marginal Means and Standard Deviations of $CommRT_{Vec}$ (seconds) by condition.

$CommRT_{Vec}$ (seconds)	Comm System					
	Voice		Datalink		Hybrid	
Vehicle Load	*M*	*SD*	*M*	*SD*	*M*	*SD*
4 Vehicles	6.53	6.86	15.28	11.01	14.88	8.05
12 Vehicles	4.03	2.24	21.55	9.02	23.67	10.11

In a similar fashion, significant main effects of both Comm System ($F(2, 10) = 16.46$, $p < 0.001$, $\eta^2 = 0.62$) and Vehicle Load ($F(1, 10) = 48.65$, $p < 0.001$, $\eta^2 = 0.829$) on $CommRT_{Ret}$ were observed. The interaction of the Comm System and Vehicle Load variables on $CommRT_{Ret}$ was also found to be significant ($F(2, 10) = 4.66$, $p = 0.022$, $\eta^2 = 0.318$). Estimated marginal means for $CommRT_{Ret}$ are provided in Table 3 and Fig. 12b.

‡ Mauchly's sphericity test violated; Greenhouse-Geisser *df* used to calculate *F*-ratio.

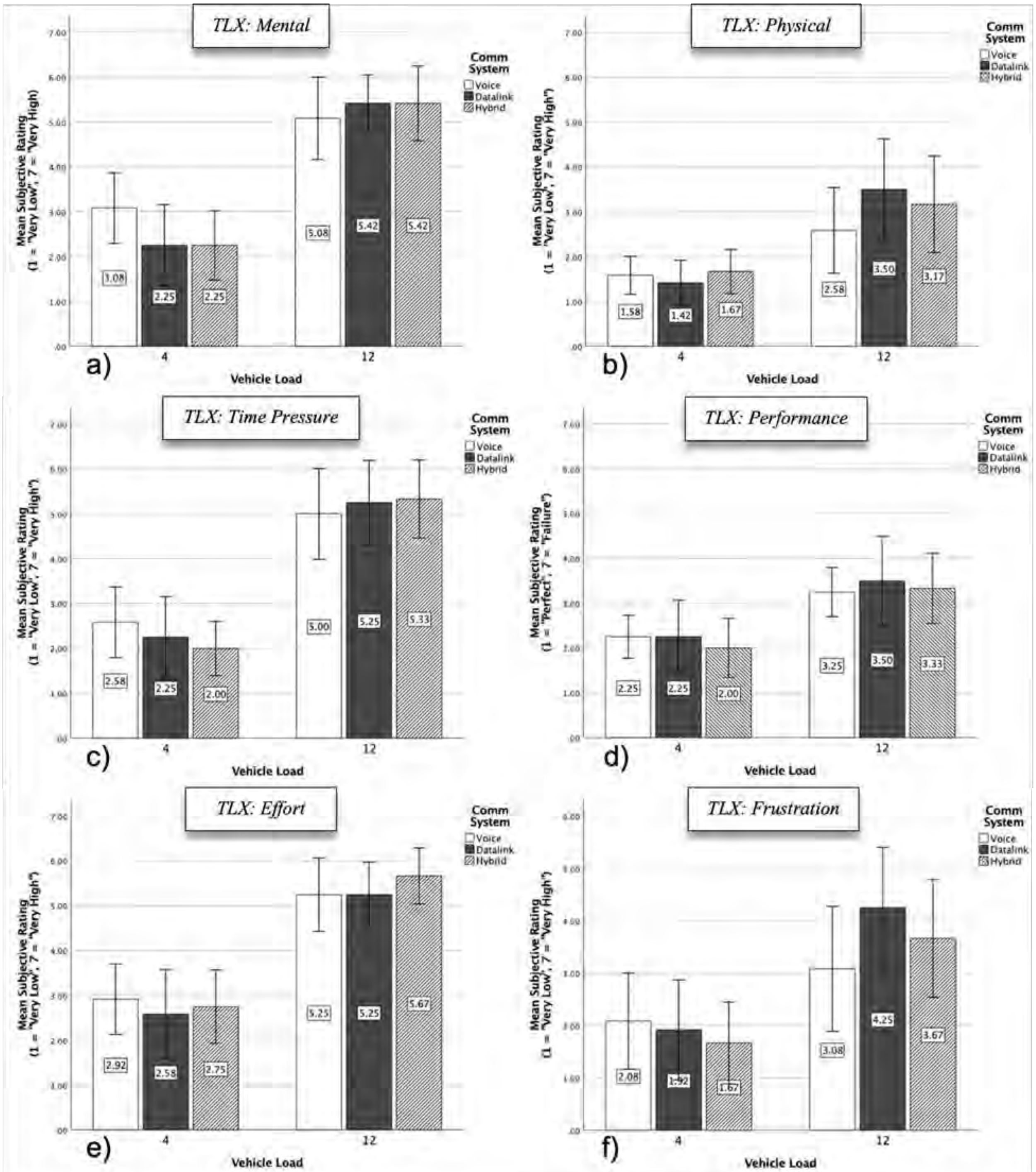

Fig. 11 Pilot workload ratings shown by communication modality and vehicle load.

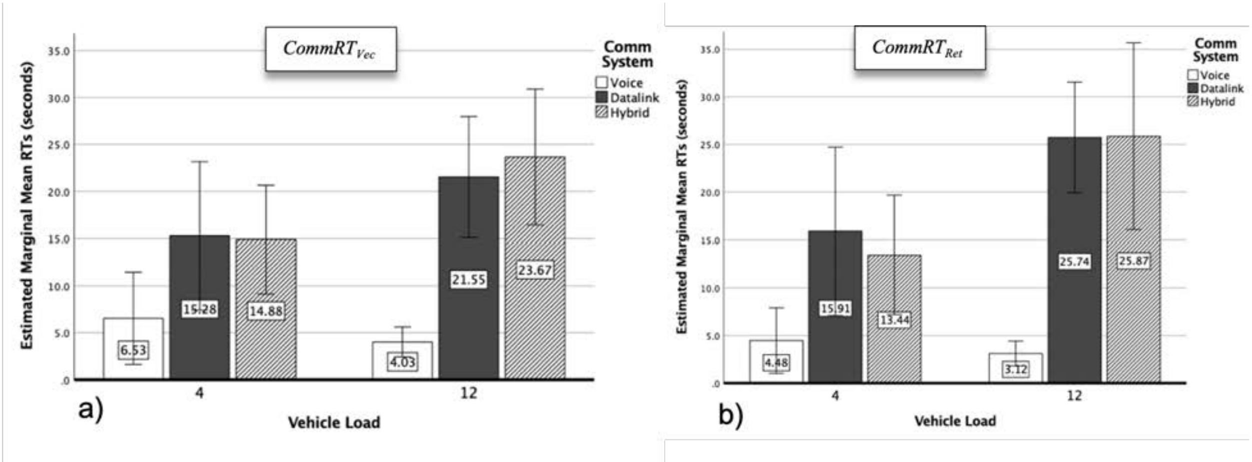

Fig. 12 Estimated marginal means in seconds of $CommRT_{Vec}$ (a) and $CommRT_{Ret}$ (b) by experimental condition. Error bars represent the 95% confidence interval.

Table 3. Estimated marginal means (*M*) and standard deviations (*SD*) of $CommRT_{Ret}$ (seconds) by condition.

$CommRT_{Ret}$ (seconds)	Comm System					
	Voice		Datalink		Hybrid	
Vehicle Load	*M*	*SD*	*M*	*SD*	*M*	*SD*
4 Vehicles	4.48	5.13	15.91	13.11	13.44	9.30
12 Vehicles	3.12	1.94	25.74	8.65	25.87	14.59

2. *Readback Length (VecRBL and RetRBL)*

There was no statistically significant main effect of either Comm System ($F(2, 10) = 0.081$, $p = 0.923$, $\eta^2 = 0.008$) or Vehicle Load ($F(1, 10) = 4.07$, $p = 0.071$, $\eta^2 = 0.29$) on *VecRBL* after controlling for statistical outliers. Furthermore, there was no statistically significant interaction of the two variables ($F(2, 10) = 0.99$, $p = 0.39$, $\eta^2 = 0.09$) on *VecRBL*. Many outliers for *VecRBL* were observed, most commonly in the Datalink and Hybrid Conditions (Fig. 13a). The most egregious outliers occurred in the 12-Vehicle condition. The inordinately long readback lengths were the result of selecting a readback response by right-clicking on the message from ATC and then only actually transmitting (i.e., clicking the "Send" button) after a relatively long period of time had passed—over 6 minutes (376 s) in the worst case. Estimated marginal means for *VecRBL* are provided in Table 4.

Table 4. Estimated marginal means (*M*) and standard deviations (*SD*) of *VecRBL* (seconds) by condition.

VecRBL (seconds)	Comm System					
	Voice		Datalink		Hybrid	
Vehicle Load	*M*	*SD*	*M*	*SD*	*M*	*SD*
4 Vehicles	4.17	1.26	3.45	1.07	3.64	1.67
12 Vehicles	4.21	1.10	4.54	1.10	4.59	2.99

As with *VecRBL* above, there were extreme outliers of *RetRBL* for which controls were exercised. Resultantly, no significant main effect of Comm System (F(1.374[§], 11) = 1.376, $p = 0.27$, $\eta^2 = 0.11$) or Vehicle Load ($F(1, 11) = 0.571$, $p = 0.466$, $\eta^2 = 0.049$) on *RetRBL* was observed. There was no observed significant interaction of Comm System and Vehicle Load on *RetRBL* ($F(2, 11) = 0.152$, $p = 0.86$, $\eta^2 = 0.014$). Several outliers were observed, particularly in the Datalink and Hybrid Comm conditions, from the same cause as elaborated above (Fig 13b). Estimated marginal means for *RetRBL* are provided in Table 5.

[§] Mauchly's sphericity test violated; Greenhouse-Geisser *df* used to calculate *F*-ratio.

Table 5. Estimated marginal means (*M*) and standard deviations (*SD*) of *RetRBL* (seconds) by condition.

RetRBL (seconds)	Comm System					
	Voice		Datalink		Hybrid	
Vehicle Load	*M*	*SD*	*M*	*SD*	*M*	*SD*
4 Vehicles	3.94	0.84	4.10	3.13	3.19	1.57
12 Vehicles	4.29	1.27	4.25	1.58	3.84	1.44

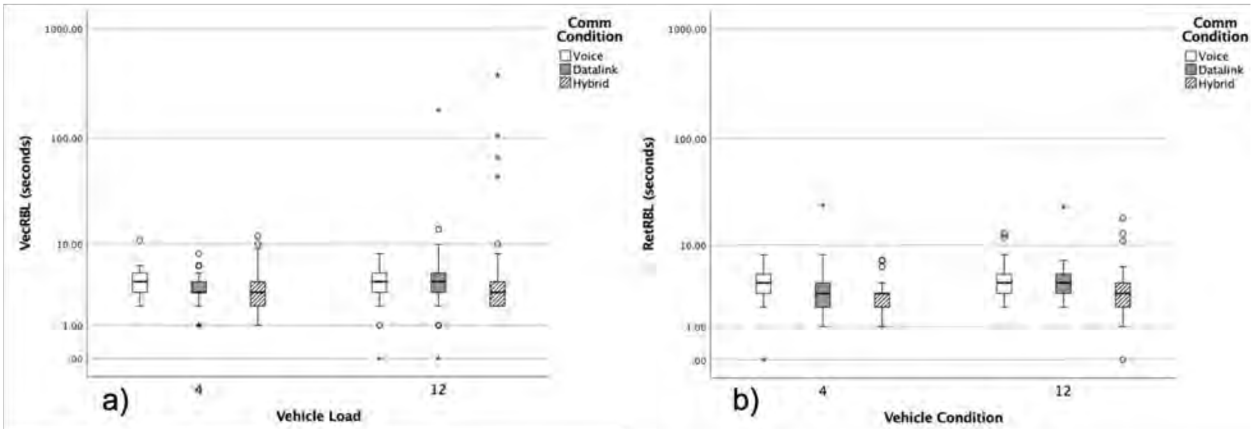

Fig. 13 Boxplots of *VecRBL* and *RetRBL* by condition. Outliers and extreme outliers identified by circles (○) and asterisks (*), respectively. Note that the vertical axes employ logarithmic scales.

3. Maneuver and RTC Initiation Response Time ($ManRT_{Init}$ and $RetRT_{Init}$)

A significant main effect of Vehicle Load on $ManRT_{Init}$ was observed ($F(1, 10) = 5.866$, $p = 0.036$, $\eta^2 = 0.37$). A significant interaction effect of Comm System and Vehicle Load on $ManRT_{Init}$ was found ($F(2, 10) = 5.477$, $p = 0.013$, $\eta^2 = 0.354$) as well. No significant effect of Comm System on $ManRT_{Init}$ was observed ($F(1.328$[**]$, 10) = 3.343$, $p = 0.081$, $\eta^2 = 0.251$). Estimated marginal means of $ManRT_{Init}$ by condition are provided in Table 6 and Fig. 14a.

Table 6. Estimated marginal means (*M*) and standard deviations (*SD*) of $ManRT_{Init}$ (seconds) by condition.

$ManRT_{Init}$ (seconds)	Comm System					
	Voice		Datalink		Hybrid	
Vehicle Load	*M*	*SD*	*M*	*SD*	*M*	*SD*
4 Vehicles	17.55	12.05	5.74	2.12	7.44	5.62
12 Vehicles	13.73	7.55	12.77	8.01	13.58	7.02

As with the maneuver initiation response time above, there was a significant main effect of Vehicle Load on $RetRT_{Init}$ ($F(1, 10) = 13.756$, $p = 0.004$, $\eta^2 = 0.579$). Additionally, a significant interaction effect of Comm System and Vehicle Load on $RetRT_{Init}$ was observed ($F(2, 10) = 5.351$, $p = 0.014$, $\eta^2 = 0.349$). A significant main effect of Comm System on $RetRT_{Init}$ was not found ($F(2, 10) = 0.908$, $p = 0.419$, $\eta^2 = 0.083$). Estimated marginal means of $RetRT_{Init}$ by condition are provided in Table 7 and Fig. 14b.

Table 7. Estimated marginal means (*M*) and standard deviations (*SD*) of $RetRT_{Init}$ (seconds) by condition.

$RetRT_{Init}$ (seconds)	Comm System					
	Voice		Datalink		Hybrid	
Vehicle Load	*M*	*SD*	*M*	*SD*	*M*	*SD*
4 Vehicles	10.61	10.74	5.98	2.53	6.14	4.52
12 Vehicles	9.11	6.82	18.36	9.08	19.93	14.12

[**] Mauchly's sphericity test violated; Greenhouse-Geisser *df* used to calculate *F*-ratio.

4. Maneuver and RTC Edit Response Time ($ManRT_{Edit}$ and $RetRT_{Edit}$)

A significant main effect of Comm System on $ManRT_{Edit}$ was observed ($F(2, 11) = 45.267$, $p < 0.001$, $\eta^2 = 0.805$). There was no significant main effect of Vehicle Load on $ManRT_{Edit}$ ($F(1, 11) = 1.137$, $p = 0.309$, $\eta^2 = 0.094$); neither was there any observed interaction effect of the two variables on $ManRT_{Edit}$ ($F(2, 11) = 1.192$, $p = 0.322$, $\eta^2 = 0.098$). Estimated marginal means of $ManRT_{Edit}$ by condition are provided in Table 8 and Fig. 14c.

Table 8. Estimated marginal means (*M*) and standard deviations (*SD*) of $ManRT_{Edit}$ (seconds) by condition.

$ManRT_{Edit}$ (seconds)	Comm System					
	Voice		Datalink		Hybrid	
Vehicle Load	*M*	*SD*	*M*	*SD*	*M*	*SD*
4 Vehicles	9.78	2.55	2.36	0.88	3.01	1.68
12 Vehicles	9.56	2.46	2.95	1.34	4.34	4.10

As was observed for $ManRT_{Edit}$, there was a significant main effect of Comm System on $RetRT_{Edit}$ ($F(1.248$[††]$, 10) = 60.442$, $p < 0.001$, $\eta^2 = 0.858$). There was no significant main effect of Vehicle Load on $RetRT_{Edit}$ ($F(1, 10) = 0.755$, $p = 0.405$, $\eta^2 = 0.70$). Additionally, no significant interaction effect of the two variables on $RetRT_{Edit}$ was observed ($F(1.256$[††]$, 10) = 0.095$, $p = 0.819$, $\eta^2 = 0.009$). Estimated marginal means of $ManRT_{Edit}$ by condition are provided in Table 9 and Fig. 14d.

Table 9. Estimated marginal means (*M*) and standard deviations (*SD*) of $RetRT_{Edit}$ (seconds) by condition.

$RetRT_{Edit}$ (seconds)	Comm System					
	Voice		Datalink		Hybrid	
Vehicle Load	*M*	*SD*	*M*	*SD*	*M*	*SD*
4 Vehicles	7.00	1.59	2.15	0.72	2.06	0.66
12 Vehicles	7.10	3.06	2.63	0.87	2.40	0.70

5. Service Time ($ManRT_{Serv}$ and $RetRT_{Serv}$)

Results showed a significant main effects of both Comm System ($F(2, 10) = 12.587$, $p < 0.001$, $\eta^2 = 0.557$) and Vehicle Load ($F(1, 10) = 6.415$, $p = 0.03$, $\eta^2 = 0.391$) on $ManRT_{Serv}$. Additionally, there was a significant interaction effect of Comm System and Vehicle Load on $ManRT_{Serv}$ ($F(2, 10) = 8.32$, $p = 0.002$, $\eta^2 = 0.454$). Estimated marginal means of $ManRT_{Serv}$ by condition are provided in Table 10 and Fig. 14e.

Table 10. Estimated marginal means (*M*) and standard deviations (*SD*) of $ManRT_{Serv}$ (seconds) by condition.

$ManRT_{Serv}$ (seconds)	Comm System					
	Voice		Datalink		Hybrid	
Vehicle Load	*M*	*SD*	*M*	*SD*	*M*	*SD*
4 Vehicles	27.24	12.04	8.11	2.33	10.45	6.19
12 Vehicles	23.17	8.28	15.73	7.98	17.93	9.26

A significant main effect of Vehicle Load on $RetRT_{Serv}$ was found ($F(1, 10) = 19.878$, $p < 0.001$, $\eta^2 = 0.665$). There was a significant interaction effect of the two variables on $RetRT_{Serv}$ ($F(2, 10) = 4.272$, $p = 0.029$, $\eta^2 = 0.299$) as well. No main effect of Comm System on $RetRT_{Serv}$ was found. ($F(2, 10) = 0.529$, $p = 0.597$, $\eta^2 = 0.05$). Estimated marginal means of $RetRT_{Serv}$ by condition are provided in Table 11 and Fig. 14f.

[††] Mauchly's sphericity test violated; Greenhouse-Geisser *df* used to calculate *F*-ratio.

Table 11. Estimated marginal means (*M*) and standard deviations (*SD*) of *RetRT*$_{Serv}$ (seconds) by condition.

$RetRT_{Serv}$ (seconds)	Comm System					
	Voice		Datalink		Hybrid	
Vehicle Load	*M*	*SD*	*M*	*SD*	*M*	*SD*
4 Vehicles	17.88	10.54	8.14	2.58	10.06	6.73
12 Vehicles	16.38	6.54	20.99	9.50	24.25	17.63

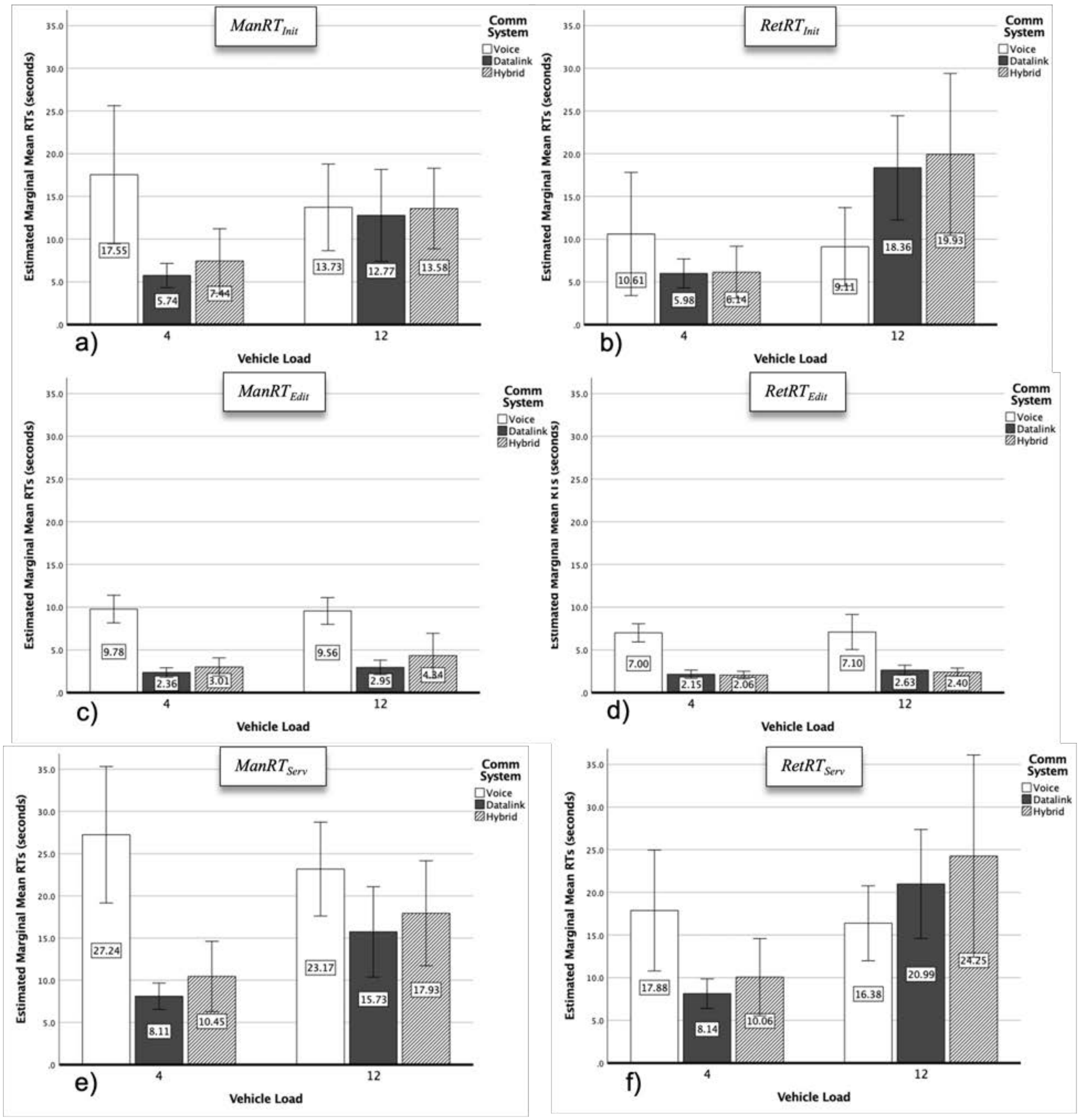

Fig. 14 Maneuver and RTC response times by condition. All values measured in seconds. Error bars represent the 95% confidence interval.

C. Error Rates

1. Frequency Error Rate

When using the Voice Comm System participants were required to use the radio panel (Fig. 2) to manually switch frequencies as appropriate to readback vectoring and RTC instructions issued by ATC. During these readback calls, incidents of transmitting on an incorrect frequency were common, occurring 40.8% of the time overall. Frequency errors were more common in the 12-Vehicle condition (45.0% of calls) than in the 4-Vehicle condition (34.3% of calls). Across Vehicle Load conditions, pilots individually had readback frequency error rates ranging from 0% to 75%, with a median value of 42.9%. This type of error was only possible in the Voice Comm condition. In the Datalink and Hybrid conditions, responses to ATC were automatically addressed into the appropriate chatroom.

2. Callsign Error Rate

Incidents of participants responding to ATC vectoring or RTC instructions using the incorrect callsign in their readback transmissions were less common than frequency errors above, though they occurred notably often. Pilots responded with an incorrect callsign in 8.43% of readbacks overall. Callsign errors were more common in the 12-Vehicle condition (9.26% of calls) than in the 4-Vehicle condition (7.14% of calls). Across Vehicle Load conditions, pilots individually had callsign error rates ranging from 0% to 21.4%, with a median value of 8.12%. This type of error was only possible in the Voice Comm condition. In the Datalink and Hybrid conditions, responses to ATC were automatically preformatted to include the correct callsign.

3. Maneuver Error Rate

Incidents of participants uplinking a maneuver that mistakenly differed from the one instructed by ATC were not common, occurring 4.44% of the time. Maneuver error rates were more common in the 12-Vehicle condition (5.45% of uploads) than in the 4-Vehicle condition (2.86% of uploads). Three quarters of maneuver errors were due to a mis-entered value; the remainder were due to commanding a maneuver in the wrong dimension (e.g., a speed change when heading was instructed). Across Vehicle Load conditions, pilots individually had maneuver error rates ranging from 0% to 16.7%, with a median value of 0%. This type of error was only possible in the Voice Comm condition. In the Datalink and Hybrid conditions, maneuver parameters were automatically populated into the autopilot for the aircraft.

4. Total Error Rate

Here, the total error rate is the overall sum of frequency, callsign, and maneuver errors divided by the total number of opportunities for any of those errors to occur. The total error rate across Vehicle Load conditions was 20.6%. In the 12-Vehicle condition, the rate was 22.8%. The total error rate was 17.1% in the 4-Vehicle condition. Across Vehicle Load conditions, pilots individually had total error rates ranging from 5% to 37.5%, with a median value of 18.8%. As frequency, callsign, and maneuver errors were only possible in the Voice condition, the total error rate is not calculated for the Datalink and Hybrid conditions.

5. Clarification Requests

Pilots requested clarification from ATC regarding a vectoring or RTC instruction 9.60% of the time overall. Clarifications were slightly more common in the 12-Vehicle condition (11.2% of the time) than in the 4-Vehicle condition (7.14% of the time). Across Vehicle Load conditions, pilots individually had clarification request rates ranging from 0% to 28.6%, with a median value of 7.42%. Given that messages in the chatroom are persistent and automatically populated in the Datalink and Hybrid conditions, request clarification rate is only provided for the Voice condition.

6. Missed Calls

For the purposes of this paper, missed calls come in two varieties: ignored calls and deleted calls. Ignored calls refer to cases in which ATC gave a vectoring or RTC instruction to the pilot which, in turn, went entirely unacknowledged from the call's onset through the remainder of the experimental scenario. Deleted calls refer to scripted calls built into the scenario that had to be removed by researchers prior to the call's onset because the participant failed to tell the intended recipient aircraft to depart from its origin. Unlike the errors described in the subsections V.1-V.5, which can only occur in the Voice condition, missed call errors were possible in all Comm System and Vehicle Load conditions. Overall, the rate of ignored vectoring or RTC calls was 2.95%. By Comm Condition, the ignored call rates were 3.29%, 0%, and 5.68% for the Voice, Datalink, and Hybrid conditions, respectively. The ignored call rate was 4.72% for the 4-Vehicle condition and 1.82% for the 12-Vehicle conditions. Individually, pilots had ignored call rates ranging from 0% to 20.8%, where the median value was 0%.

Turning attention to deleted calls, researchers had to remove 4.56% scripted (i.e., planned) vectoring calls for aircraft, due to the pilot's failure to tell the target aircraft to depart (i.e., due to a missed departure). The deleted call rates were 4.17%, 1.08%, and 8.3% for the Voice, Datalink, and Hybrid conditions, respectively. The deleted call rate was 0.93% in the 4-Vehicle condition and 6.74% in the 12-Vehicle condition. Across all condition, pilots individually had deleted call rates ranging from 0% to 20.8%, with a median value of 0%.

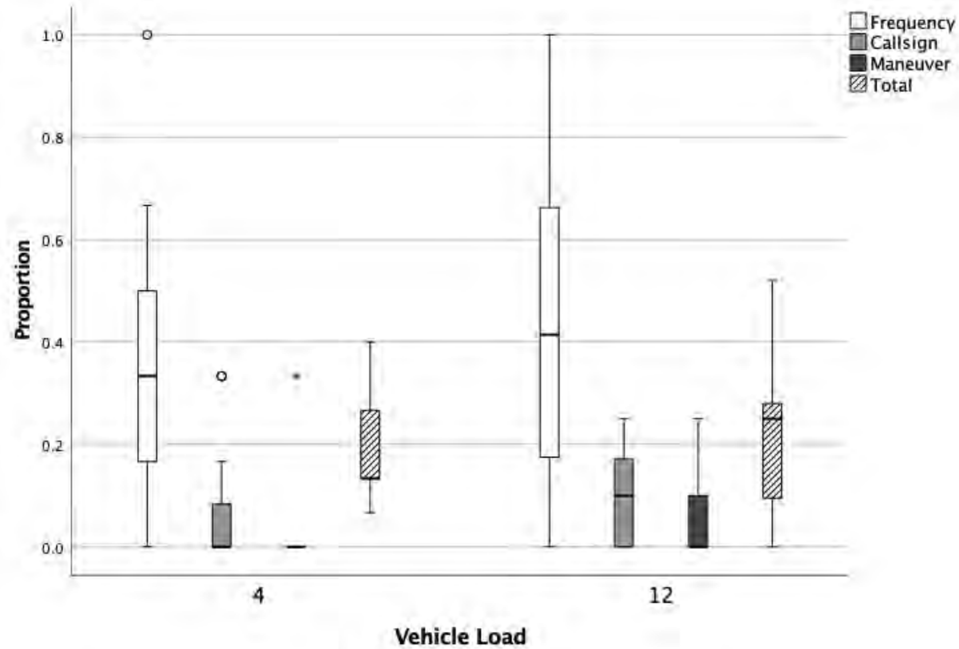

Fig. 15 Boxplots of individual pilot error rates by category and Vehicle Load. Outliers and extreme outliers identified by circles (○) and asterisks (*), respectively.

D. "Confusion Loops"

Because they are the result of repetitively clicking on hyperlinks for vectoring and RTC maneuvers, confusion loops can only occur in the Datalink and Hybrid conditions. When using one of those comm systems, vectoring and RTC calls resulted in confusion loops in a total of 7.34% of cases. Surprisingly, confusion loops were more common in the 4-Vehicle condition than in the 12-Vehicle condition, occurring at rates of 8.82% and 6.42%, respectively. The highest incidence rate of confusion loops was in the Hybrid condition at 9.41%, followed by a rate of 5.43% in the Datalink condition.

VI. Discussion

A. Workload Differences

The sole driver of workload differences was the Vehicle Load variable: overall workload scores more than double on average when working in the 4-Vehicle condition versus the 12-Vehicle condition. The effect of Vehicle Load on workload is by no means surprising and was expected. In a questionnaire administered following the conclusion of all trials, pilots unanimously indicated that the workload was manageable in the 4-Vehicle condition. Responses were split on whether the workload during the 12-Vehicle trials were manageable ("Strongly Agree," $N = 1$; "Somewhat Agree," $N = 3$; "Neither Agree nor Disagree," $N = 4$; "Somewhat Disagree," $N = 3$; "Strongly Disagree," $N = 1$). Of note is the magnitude of overall workload scores in the 12-Vehicle condition, nearly maxing out the scale in all three Comm System conditions.

The lack of any effect of the Comm System on workload was not anticipated. The Hybrid comm system was intended to reduce pilot workload by providing hyperlinks for vectoring and RTC and allowing the pilot to use the radio for transition calls, with the rationale that this partitioning of modalities would shorten overall communication and maneuver/RTC response times. Clearly, the partitioning offered no workload benefit when compared to the Voice and Datalink systems on their own. It is possible that a different schema for partitioning modalities, or that changes to our implementation of the chat system, may have led to a difference in workload.

The authors do not intend to convey that digital versus voice-based systems will never lead to differences in workload. Rather, it is simply the case that the three systems implemented in this study did not produce statistically significant differences in workload ratings. A more thorough exploration of comm systems—including various implementations of the digital comm system and all permutations of modality partitioning for a hybrid system—would be required to make solid conclusions about how communication systems in general affect pilot workload.

B. Response Times

Although the Comm System variable did not affect pilot workload, it did lead to differences in response time metrics for comm response times (both for vectoring and RTC readbacks), maneuver and RTC initiation response times, and for the total service time for vectoring maneuver compliance. Consistent with previous research [6-8], pilots were able to respond to transmissions (vectoring or RTC instructions) from ATC far more quickly in the Voice condition compared to the Datalink or Hybrid conditions, leading to a response time savings of ~10 s in the 4-Vehicle condition and ~20 s in the 12-Vehicle condition.

Whereas the Voice condition allowed pilots to respond quickest to ATC, the Datalink and Hybrid conditions enabled pilots to edit their compliance maneuvers with greater rapidity. In the Datalink and Hybrid condition, pilots edited compliance maneuvers ~7 s and ~5s quicker for vectoring and RTC maneuvers, respectively, compared to the Voice condition. Additionally, the Datalink and Hybrid systems decreased the overall service time for compliance with the vectoring instructions, saving ~20 s and ~6 s on average in the 4-Vehicle and 12-Vehicle conditions, respectively. Although the Comm System variable reduced the total service time for vectoring maneuver compliance, it interestingly did not make a statistically significant difference for RTC service time. This is likely due to the difference in task difficulty of manually uplinking vectoring versus RTC maneuvers. In the former case, the pilot is required to locate the appropriate box for the maneuver dimension (heading, altitude, or speed) in the autopilot (Fig. 6.3) and manually enter the instructed value. For RTC, the pilot only needs to select the appropriate waypoint from a dropdown list (Fig 7.3).

The Vehicle Load variable played a role in response time differences for comm response times, the maneuver and RTC initiation times, as well as the overall service time (for both vectoring and RTC compliance). In all cases in which the Vehicle Load produced a statistically significant main effect in response time, there was additionally significant interaction effects with the Comm System Variable.

When using the Voice system, pilots were quicker to respond to ATC in the 12-Vehicle condition than in the 4-Vehicle condition, though only slightly. In the case of the Datalink or Hybrid systems, pilots took about ~7 s longer to respond to ATC in the 12-Vehicle condition compared to the 4-Vehicle condition. When digital messages came through the chat system, an aural notification sound was rendered; however, this aural notification was identical for all messages. This difference in response times, therefore, is likely due to the difficulty of visually locating and identifying transmissions for aircraft in the Datalink and Hybrid systems, especially when the Vehicle Load is high.

Conversely, the difficulty associated with visually locating vehicles of interest led to longer maneuver and RTC initiation times using the Voice system when compared to the Datalink or Hybrid systems. This is because, in order to initiate the maneuver/RTC, pilots needed to identify and select the target aircraft on the TSD to open the autopilot menu. In the Datalink and Hybrid systems, clicking on the hyperlink would automatically open the autopilot menu for the target aircraft, allowing for quicker maneuver/RTC initiations. This leads to an interesting interaction: whereas the Vehicle Load does not play a significant role in affecting initiation times when using the Voice condition, it does meaningfully affect initiation times in the Datalink and Hybrid conditions. This produces about a ~7 s and ~10 s difference in initiation times for vectoring and RTC, respectively, between the 4-Vehicle and 12-Vehicle conditions.

Given that the service time is the sum of the initiation and edit times—and that there were no significant differences between edit times as a function of our Vehicle Load variable (see Figs. 14c, 14d)—the effects of the Vehicle Load variable on service times for vectoring and RTC are due to the same factors as elaborated for the initiation times.

Finally, an important caveat must be made regarding readback length times. Though we found no statistical difference in readback length times as a function of either the Comm System or Vehicle Load variables, that was only after the removal of outliers, several of them extreme outliers (Fig 13). In the Datalink and Hybrid systems, it was possible for pilots to initiate a readback (Fig 4.4) and then, fatefully, forget to transmit it to the chat room. This mistake led to extremely long readback lengths in six cases, ranging from 68 s to 376 s.

C. Error Rates

1. Frequency, Callsign, and Maneuver Errors (Voice-Condition Only)

When using the Voice system, errors by pilots were commonplace. The error rate was most egregious for frequency errors: pilots were speaking on an incorrect frequency for roughly every four out of ten transmissions! This kind of error, committed this frequently, would likely overwhelm controllers who would have to redirect pilots to the appropriate frequency. The frequency error rate is so high that, unless aided by some other system to ensure transmissions are on the correct channel, using traditional voice-over-radio comms for the *m*:*N* task described in this paper is untenable and would degrade both the efficiency and safety of the ATM framework in the NAS.

Callsign errors have long been an issue in pilot-controller communications [20]. Though far less common than frequency errors, the callsign error rate of 8.43% is not ideal. At the same time, callsign errors are more benign in comparison to the other two categories. Maneuver errors were the rarest error type among the three considered here, occurring 4.4% of the time. Though this category of error occurred at relatively low rate compared to frequency and callsign errors, the consequences errors are not equal. The commission of an incorrect compliance maneuver increases the workload on controllers to catch and correct the error and degrades the overall safety of the local area of the NAS. If ATC is providing vectoring maneuvers for the purposes of separation services, incorrectly executed maneuvers could have catastrophic outcomes.

2. Clarification Requests

Clarification requests came in two types: callsign clarifications and maneuver clarifications. Callsign clarifications refer to instances in which the pilot asked ATC to repeat the callsign for an instruction. Maneuver clarifications refer to cases in which the pilot asked ATC to repeat either the parameters of a maneuver or the waypoint specified for RTC. Overall, clarifications were requested 9.60% of the time. When clarification was requested and provided by ATC, there were no subsequent callsign or maneuver errors. Of note, however, is that initial clarification requests were made on an incorrect frequency in 64.3% of cases.

D. "Confusion Loops"

Instances of confusion loops were most common in the Hybrid condition, resulting in a 9.41% rate. Given that the interaction which sets the stage for a confusion loop is identical in the Datalink condition, where the rate of 5.43% represents a 43% decrease compared to the Hybrid rate, one would expect the two error rates to be similar. At first blush, it is tempting to say that this may be a result of higher mental demand or required effort to use the Hybrid system (it involves managing verbal and textual communications), but this does not accord with the results from pilot TLX ratings. This is further confounded by the fact that confusion loops were more common in the 4-Vehicle condition (8.82%) compared to the 12-Vehicle condition (6.42%). Given the low total number of confusion loops (13 total across participants and conditions), it is possible that these proportions are somewhat spurious in nature: there were 2 loops in the Datalink/4-Vehicle condition, 3 loops in the Datalink/12-Vehicle condition, 4 loops in the Hybrid/4-Vehicle condition, and another 4 loops in the Hybrid/12-Vehicle condition. Breaking the numbers out by run order: 4 loops occurred in the first trial, 3 loops in the second trial, 1 loop in the third trial, 0 loops in the fourth trial, 1 loop in the fifth trial, and 4 loops in the sixth trial. It is possible that an experiment with a larger total number of vectoring calls, resulting in a larger number of confusion loops overall, would produce a different distribution.

VII. Conclusion and Recommendations

Clearly, multiple vehicle control is a demanding task, especially as it relates to the management of communications with the ATM system. Our results show that performing communications for 12 vehicles simultaneously—which may be at different phases of flight, under different ATC authorities, and possessing distinct callsigns—leads to a near saturation in workload, leading to degradations in pilot performance.

Considering the three communication systems employed in this study, there are benefits and drawbacks to the Voice system, on the one hand, and the Datalink/Hybrid system, on the other. The Voice comm system allowed for the shortest response times from pilots and was the most commonly preferred system of our participants. Although Voice was by far the fastest condition for comm response times, it led to the longest overall service times in most cases, due to the time allowance required for manually executing vectoring maneuvers and RTC.

The Voice condition produced an unacceptably high rate of errors in readback transmissions to ATC, most notably in the form of speaking on an incorrect frequency. These transmission mistakes are mitigated by the Datalink and Hybrid systems, although those systems are not totally free from errors. These systems led to the highest number of outlier data points for the readback lengths described above. Additionally, the confusion loop category of error is only possible in the Datalink and Hybrid condition.

This combination of results, where comm responses are fastest and most erroneous by voice and maneuver/RTC response times are best performed by digitally assisted messages (i.e., hyperlinks), suggests that some form of Hybrid system could lead to improved pilot performance for the comm task. Such a system should be designed to avoid the most egregious issues observed in our experimental trials. Our recommendations for such a system are listed below.

- If voice comms are utilized, some form of automated system to ensure the pilot is transmitting on the correct frequency should be required.
- If digital messages are sent to the pilot, there should be an attention-grabbing UI to alert the pilot of an unacknowledged message after some nominal period of time as expired.
- If an approach for pre-loading maneuver parameters into the autopilot similar to ours in the Datalink and Hybrid condition is utilized, the system should be smart enough to detect once the maneuver has been successfully executed to prevent the pilot from re-engaging the maneuver later on (i.e., a confusion loop), whether or not the pilot has communicated the readback for the instruction.

Acknowledgments

The authors would like to thank Igor Dolgov of Joby Aviation and Mark Shikerman of Wisk Aero for their collaboration and input during the design of this study. Additionally, the authors thank Dominic Wong for VSCS software development and support.

References

[1] Federal Aviation Administration, "Urban Air Mobility: Concept of Operations v1.0," 2020.

[2] Federal Aviation Administration, "Urban Air Mobility and Advanced Air Mobility," URL: https://www.faa.gov/uas/advanced_operations/urban_air_mobility [retrieved 8 November 2022].

[3] Hill, B.P., Dwight, D., Metcalfe, M., Griffin, C., Wiggins, S., Metts, C., Bastedo, B., Patterson, M. D., Mendonca, N. L., "UAM Vision Concept of Operations (ConOps) UAM Maturity Level (UML) 4," Publication No. 20205011091, National Aeronautics and Space Administration, 2020.

[4] Price, G. et al., "Urban Air Mobility Operational Concept (OpsCon) Passenger-Carrying Operations," NASA CR-2020-5001587, 2020.

[5] Federal Aviation Administration, "Unmanned Aircraft System (UAS) Traffic Management (UTM): Concept of Operations v2.0," 2020.

[6] Dunbar, M., McGann, A., Mackintosh, M. A., Lozito, S., and Ashford, R., "Re-examination of Mixed Media Communication: The Impact of Voice, Data Link, and Mixed Air Traffic Control Environments on the Flight Deck," NASA TM-2001-210919, 2001.

[7] Martin, L., Verma, S., Jadhav, A., Raghvan, V., and Lozito, S., "Identifying and Refining the Tasks in a Cockpit Data Link Model," *Human-Centered Computing: Cognitive, Social, and Ergonomic Aspects, Volume 3*, edited by Harris, D., Duffy, V., Smith, M., Stephanidis, C., CRC Press, Boca Raton, FL, 2003, pp. 285-289.

[8] Lozito, S., Verma, S., Martin, L., Dunbar, M., and McGann, A., "The impact of voice, data link, and mixed air traffic control environments on flight deck procedures," *Air Traffic Control Quarterly*, Vol. 4, No. 11, 2003, pp. 293-310.

[9] Sheridan, T. B., "Supervisory control," *Handbook of Human Factors*, edited by Salvendy, G., Wiley, New York, NY, pp. 1243-1268, 1987.

[10] Smith, C. L., Sadler, G., Tyson, T., Brandt, S., Rorie, R. C., Keeler, J., Monk, K., Viramontes, J., Dolgov, I., "A Cognitive Walkthrough of Multiple Drone Delivery Operations," AIAA Paper 2021-2330, June 2021.

[11] Sadler,G., Chandarana, M., Tyson, T., Smith, C., Keeler, J., Rorie, R. C., Shyr, M., Wong, D., Scheff, S., and Dolgov, I.,"A Remote, Human-in-the-loop Evaluation of a Multiple-Drone Delivery Operation," AIAA Paper 2022-4002, June 2022.

[12] Chandarana, M., Sadler, G. G., Keeler, J. N., Smith, C. L., Rorie, R. C., Wong, D. G., Scheff, S., Pham, C., Dolgov, I., "Streamlining Tactical Operator Handoffs During Multi-Vehicle Applications," IFAC HMS 2022, September 2022.

[13] NASA, "Multi-Vehicle (m:N) Working Group," [online website], URL: https://nari.arc.nasa.gov/ttt-ram/multi-vehicle [retrieved 8 Nov 2022].

[14] Monsell, S., "Task Switching," *Trends in Cognitive Sciences*, Vol. 7, No. 3, pp. 134-140.

[15] Rubinstein, J. S., Meyer, D. E., and Evans, J. E., "Executive Control of Cognitive Processes in Task Switching," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 27, No. 4, pp. 763-797.

[16] Calhoun, G. L., Ruff, H. A., Behymer, K. J., & Mersch, E. M., "Operator-autonomy teaming interfaces to support multi-unmanned vehicle mission," *Advances in Human Factors in Robots and Unmanned Systems: Advances in Intelligent Systems and Computing*, edited by Savage-Knepshield, P. and Chen, J., Springer, Los Angeles, CA, 2017, pp. 113-126.

[17] Feitshans, G. L., Rowe, A. J., Davis, J. E., Holland, M., & Berger, L. (2008). Vigilant spirit control station (VSCS)—'The face of COUNTER'. In *Proceedings of AIAA Guidance, Navigation and Control Conf. Exhibition*, Honolulu, HI. AIAA Paper (No. 2008-6309).

[18] Hart, S. G., and Staveland, L. E., "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Advances in Psychology*, Vol. 52, 1988, pp. 139–183.

[19] Hart, S. G., "NASA-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergnomonics Society 50th Annual Meeting*, 2006, pp. 904–908.

[20] Cardosi, K., Falzarano, P., & Han, S., "Pilot-Controller Communication Errors: An Analysis of Aviation Safety Reporting System (ASRS) Reports," Department of Transportation Research and Special Programs Administration Report, DOT-VNTSC-FAA-98-4, Springfield, VA, 1998.