Effects of Unmanned Aircraft Voice Communication Delay on En Route Air Traffic Management Operations

Miwa Hayashi[1] and Jillian Keeler[2]
NASA Ames Research Center, Moffett Field, CA, 94035, USA

Cynthia Wolter[3]
San Jose State University Research Foundation, Moffett Field, CA, 94035, USA

Wayne W. Bridges[4]
Flight Research Associates, Moffett Field, CA, 94035, USA

The present study investigated the effects of remotely piloted unmanned aircraft (UA) voice communication delay on air traffic management operations with different background sector traffic-volume levels. Three one-way UA voice delay lengths of 400, 900, and 2,000 milliseconds (ms) were tested, representing the currently estimated transmission delays for terrestrial, satellite communication (SATCOM), and long SATCOM Command and Control Link System, respectively. All delay values exceeded 390 ms, the FAA's current requirement for the maximum communication latency in the National Airspace System. Eight retired en route air traffic controllers (ATCs) and eight remote pilots (RPs) participated in a human-in-the-loop simulation study, where a simulated UA flight transitioned through an Oakland en route low-altitude sector to a local non-towered airport. The results showed that, when the 2,000-ms voice delay was present, radio transmission step-ons and detect-and-avoid (DAA) alert level elevations increased. In addition, ATC workload, ATC acceptability for the UA, and RP acceptability for the DAA were negatively affected only when the UA had a 2,000-ms voice delay under the high traffic-volume condition, but not in the other conditions.

I. Introduction

A. Motivation

In an effort to integrate unmanned aircraft (UA) into the National Airspace System (NAS), it helps if the behavior differences between the UA and the manned flights are minimized as it would help reduce the overall impacts to the NAS operations. However, some differences are inevitable due to the intrinsic nature of the UA, i.e., the pilot is not onboard and is located distance away from the vehicle. One of such differences is UA voice delay in air traffic control radio communication. Even if the remote pilot (RP) is as responsive as a pilot onboard, due to the law of physics that it takes nonzero time for signals to travel through network paths, there will be always extra delay added to the UA voice communication. The delay can be reduced by selecting a shorter signal path, but the preferred shorter path options may not be always available or economically feasible. The present study aims to investigate issues of the UA voice communication delay primarily from the viewpoints of the NAS air traffic management (ATM) operations.

---

[1] Aerospace Engineer, Aviation Systems Division, AIAA Senior Member.
[2] Research Associate, Human Systems Integration Division, NASA Ames Research Center.
[3] Research Associate, Human Systems Integration Division, NASA Ames Research Center.
[4] Senior Air Traffic Control Associate.

B. Definitions

In the present paper, unless otherwise specified, all voice delays are *one-way* delay, which is the interval between the time when the speech sound is uttered by the speaker and the time when the same sound is heard by the receiver. The *round-trip* delay is the sum of the one-way delays of both directions. It can be simply double the one-way delay if the delays are symmetric. If asymmetric, the one-way delay of each direction is noted. Round-trip delay is important in dialog situations where a speaker is waiting for a verbal response from the other party, as it is the total *extra* delay that the speaker experiences due to the network delay before receiving the anticipated verbal response. In this paper, *delay*, *latency*, and *lag* are used interchangeably. *Transmission* sends the uttered sound, whereas *reception* receives the sound. *Communication* includes both the transmission and reception processes.

C. UA Voice Communication Delay

The current assumption of the UA voice communication is that the RP talks with the air traffic controller (ATC) using the Federal Aviation Administration's (FAA's) existing Very High Frequency analog radio and the onboard voice transmitter and receiver [1]. The connection between the UA and the RP is, then, established by the Command and Control (C2) Link System comprised of a terrestrial or satellite communication (SATCOM) links. Packets of digitized voice and other data are sent via the C2 Link System. The Radio Technical Commission for Aeronautics (RTCA) estimated the achievable one-way latency of the UA C2 Link System to be 226 milliseconds (ms) for terrestrial link and 720 ms for geostationary-satellite-based SATCOM link (Appendix D in [1]). With the expected delays for the NAS voice equipment added, the estimated one-way UA voice delay could be approximately 400 and 900 ms, respectively. Note that these are estimates for technologically achievable delays. In practice, UA can experience longer delays due to extended network path distance especially when on a SATCOM link.

Such a small delay may at first seem insignificant, as a human pilot already has a natural verbal response delay on the order of seconds. For instance, Cardosi reported an average of 3.3-second lag time from the end of an ATC's turn instructions to the beginning of a pilot's responses in en route airspace [2]. However, there is an important difference between the natural delay in the human's verbal response and network delay. In the former case, all parties hear the same sounds (or lack of sounds) at the same time. In the latter case, some hear only delayed sounds (or lack of them) and their speeches always arrive to the other party with a delay. In short, with a voice delay, the audio experiences of the speakers are asynchronous.

In addition to perceived slower responses, the presence of a voice delay is known to increase occurrences of *step-ons*, i.e., accidental simultaneous transmissions by two or more speakers. When two people simultaneously press the push-to-talk button, one or both transmissions may be blocked. Sometimes this also results in a piercing heterodyne noise heard on the frequency. The stepped-on messages must be retransmitted, increasing ATC and pilot workload, as well as potentially delaying aircraft movement at a critical moment and/or forcing the pilot to turn the aircraft without a prior ATC coordination. During a tabletop discussion session with ATCs, RPs, and a pilot, our researchers identified the UA voice delay as one of the major ATM concerns for introducing UA flights into the NAS [3].

The FAA's current requirements for the voice communication latency between the NAS users (e.g., pilot, air carrier) and the specialist (i.e., the ATC) are less than or equal to 250 ms on average, less than or equal to 300 ms for the 99th percentile, and maximum 350 ms [4]. The maximum value of 350 ms was derived from the FAA's simulation study findings [5]. There was additional known delay of 40 ms in the air propagation and onboard avionics. Thus, 390 ms is often referred to as the maximum acceptable voice communication latency within the NAS. The estimated C2 Link System delays (400 and 900 ms) exceed the FAA's maximum acceptable delay value.

D. Past Work on Voice Delay

In the 1960s, Riesz and Klemmer at Bell Telephone Laboratories investigated how transmission delays of satellite telephoning affected the telephone users' perception and the temporal pattern of their conversations [6]. They reported that, when transmission delay circuits were inserted into their administrative staff's phone lines, the phone users were mostly unaware of the presence of the voice delays. The effects of voice delay may not be obvious enough to the users, unlike more explicit degradations such as noise or echo. When the users sensed degradation of the call quality, they tended to blame the other speaker for being inattentive, slow to respond, or excessively interruptive.

In party-line remote conversations like the ATC radio communication, turn-taking is a major challenge as the speakers cannot obtain some of the non-verbal cues from the other speakers. The voice delay amplifies this challenge. Sacks, Schegloff, and Jefferson proposed the following basic set of rules for turn-taking [7]:

1) Current speaker selects the next speaker by verbal or non-verbal cues (tone of voice, gaze, a question with naturally expected respondent, etc.).
2) If no particular person was selected, the first person who started talking gets the next turn.
3) If none takes the turn, the original speaker may resume.

These rules provide a simple yet insightful framework. For instance, when the ATC specifically addresses the RP, it is an exercise of Rule 1, and a long pause, not step-ons, is likely experienced by all before the RP responds. Step-ons are typical results of Rule 2 (self-selection), and Rule 3 can result in a long pause, step-ons, or both.

In the 1990s and early 2000s, FAA-funded studies were conducted to examine the effects of the long voice communication delays associated with digital and/or satellite-based ATC radio communication systems [5, 8]. Note that they differ from the present study by subjecting *all* flights in the airspace to a long and homogeneous voice delay.

Nadler et al. [8] conducted a human-in-the-loop (HITL) simulation study that tested the following four delay conditions, each indicating the ground-to-air and air-to-ground voice delays in milliseconds divided by a slash: 225/0, 169/70, 485/260, and 429/330. The last two of the four delay conditions represented the longer delays associated with the satellite-mediated communication system. The following three traffic levels were tested: Medium (70% of the sector's peak value), High (90%), and Very High (110%). The researchers found that step-ons increased with the satellite delays, but only when communications workload was Very High. They concluded that areas with very high communications workload should continue relying on terrestrial communications.

Sollenberger et al. [5] evaluated the effects of Very High Frequency Digital Link Mode 3 (VDL3) radio equipment that was planned to replace the aging analog ATC radio. They evaluated the following three voice delay conditions in a HITL simulation study: 250, 350, and 750 ms. They observed that, in the 750-ms delay condition, the controller override, a new VDL3 feature, was used more frequently than in the other delay conditions. This finding implied that, without this feature, more step-ons occurred or the critical ATC instructions had to be delayed. The researchers recommended that the VDL3 be implemented with a 350-ms delay.

Rantanen et al. examined ATC's voice delay (Audio Delay, or AD) and the pilot's response and execution delay (Pilot Delay, or PD) in their HITL simulation study [9]. The following conditions were tested: 150-, 250-, 350-, and 1,000-ms for AD; and Zero delay and Realistic (random) delay for PD. Note that AD was inserted only in the ATC transmission, but not in the reception. Thus, it was equivalent to simulating a round-trip delay. (This may not be intuitively obvious, but a round-trip delay can be distributed in any way along the round-trip paths to give the same net effect.) Therefore, these AD values were equivalent to 75-, 125-, 175-, and 500-ms one-way delay, respectively. In their simulations of the Terminal Radar Approach Control (TRACON) airspace ATM tasks, they found that the higher values of AD, as well as the presence of PD, tended to reduce the lateral traffic separation, but the effects were small. They speculated that the large variance in the rates and timings of the ATCs' speeches may have partially masked the effects of the AD. They concluded that an AD less than 1,000 ms (or 500 ms one-way) was negligible.

The next two studies assessed the effects of the voice communication delays applied only to UA flights [10, 11]. Vu et al. investigated the effects of verbal-response and command-execution delays of UA on the ATC acceptability ratings [10]. In their en route airspace HITL simulation study, the following delays were tested: 1,500- or 5,000-ms UA verbal response delay and 1,500- or 5,000-ms UA command execution delay. The verbal response delay was inserted only in the RP's transmission, but not in the reception. Thus, these were equivalent to 750- and 2,500-ms one-way delay, respectively. The researchers reported that the 5,000-ms verbal delay (or 2,500-ms one-way delay) condition increased the number of step-ons, as well as lengthening the manned flights' communication lags, likely due to the increased step-ons. The ATCs tended to rate the manned flights' responses unacceptable more often when the 5,000-ms verbal delay was present. Nevertheless, the ATCs rated all delay conditions acceptable. The researchers reasoned that having only one UA in the sector made the UA's verbal delays tolerable in this study. The UA's speed may have been also too slow (i.e., 110 knots) for the UA's response delay to be felt objectionable to the ATC.

Comstock et al. conducted a HITL simulation study of 14 encounters between UA and manned aircraft in the presence of moderate background traffic in a TRACON airspace [11]. They evaluated the effects of the UA voice delays on the ATC workload and acceptability ratings. The four voice delays tested were 0, 400, 1,200, and 1,800 ms. The other C2 Link System data, such as vehicle commands and position reporting, were not delayed. Like in the Vu et al [10], no effect of the communication delay on controller workload was found. However, the ATCs commented that longer delays would have been disruptive if traffic density were higher, due to the increased numbers of step-ons and repeats caused by the longer delay. The researchers also observed that the ATCs tended to work the quicker-responding manned aircraft first, then the UA with delayed responses.

So far, there has been no empirical evidence confirming that the UA voice delay exceeding the current FAA's requirement negatively impacts ATM operations. Neither of the UA voice delay studies [10, 11] found any major adverse effect on ATM operations when a UA voice delay as long as 750 or 1,800 ms was inserted. The ATCs comment in Ref. [11] implies that the simulated traffic volume may have been insufficient. Thus, our hypothesis in the present study is that the effects of the UA voice delays on ATM operations are greater at higher traffic-volume levels in the sector, where the radio frequency congestion levels tend to be higher. Note that the current FAA requirements on voice communication delay are the same for any traffic density.

E. Objective, Assumptions, and Scope

The goal of the present study was to assess the interaction effect of the UA voice delay length and the sector traffic-volume level on the ATM operation in the en route airspace. No past study has demonstrated this interaction effect for UA traffic flying among the manned flights where only the UA was subject to a long voice delay. Three levels of one-way voice delay were tested—Short (400 ms), Medium (900 ms), and Long (2000 ms)—under two traffic volume level conditions, Low vs. High. The three voice delay lengths roughly represent the terrestrial, SATCOM, and long SATCOM C2 Link System latencies, respectively. The one-way delay length was applied to both transmission and reception of the RP's voice signals.

The present study assumed that all DAA surveillance signals were perfectly accurate. The study focused on the effect of a single UA's voice delay within a sector in nominal operations. Thus, scope of the evaluations excluded the following effects: delays of the UA execution commands and other C2 Link System data, contingency events such as loss of C2 Link, and presence of more than one UA in the sector. The limited experiment design also excluded from scope the derivation of the absolute safety threshold for the UA voice delay for NAS operations.

II. Method

A. Airspace and UA Ownship Flight

The study simulated the combined 40/41 en route low-altitude sector at the Oakland Air Route Traffic Control Center (ZOA). The sector is primarily an arrival and departure sector for the major airports in the Bay Area (i.e., San Francisco, Oakland, and San Jose). Sector 40 extends from the surface to 8,000 ft Mean Sea Level (MSL), and Sector 41 from 8,000 ft MSL to Flight Level 230. Fig. 1 shows the airspace and the major Standard Terminal Arrival (STAR) routes. The sector also provides approach and departure control services to local airport traffic within its airspace (e.g., Santa Rosa, Napa, and Ukiah). ZOA 40/41 often provides traffic advisories to visual flight rules (VFR) traffic.

The UA ownship was a representative of a regional air cargo carrier and modeled similar to ATR-42, a fixed-wing twin turboprop airplane. The UA had a callsign *NASA01* and was on an instrument flight rules (IFR) flight plan. The UA departed from Metro Oakland International Airport (KOAK), a Class C airport, and arrived at Ukiah Municipal Airport (KUKI), a Class E airport without operating airport tower, located at 96 nautical miles (nm) northwest of KOAK. The route is shown in Fig. 1. The scenario started from just before the point where the RP checked in with the ZOA 40/41 ATC frequency during a departure climb to 10,000 ft MSL. The UA's cruise indicated airspeed was about 190 knots. At around 10 nm from KUKI, the UA was handed off to the Common Traffic Advisory Frequency (CTAF) at KUKI. After descending along the final approach path on the RNAV (GPS)-B approach into KUKI down to the Minimum Descent Altitude (2,540 ft MSL), the UA initiated the published Missed Approach maneuvers, and contacted the ZOA 40/41 frequency again. When able, the ATC gave the UA vectors back for the same final approach course to KUKI, and the scenario ended. Each run lasted about 45 minutes.

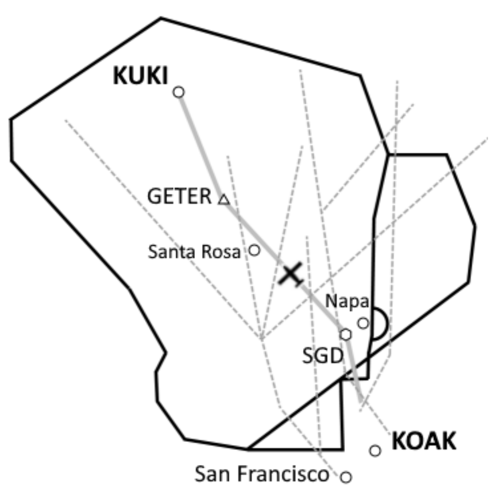


Fig. 1 ZOA 40/41, UA ownship route (solid gray), and major STAR routes (dotted gray).

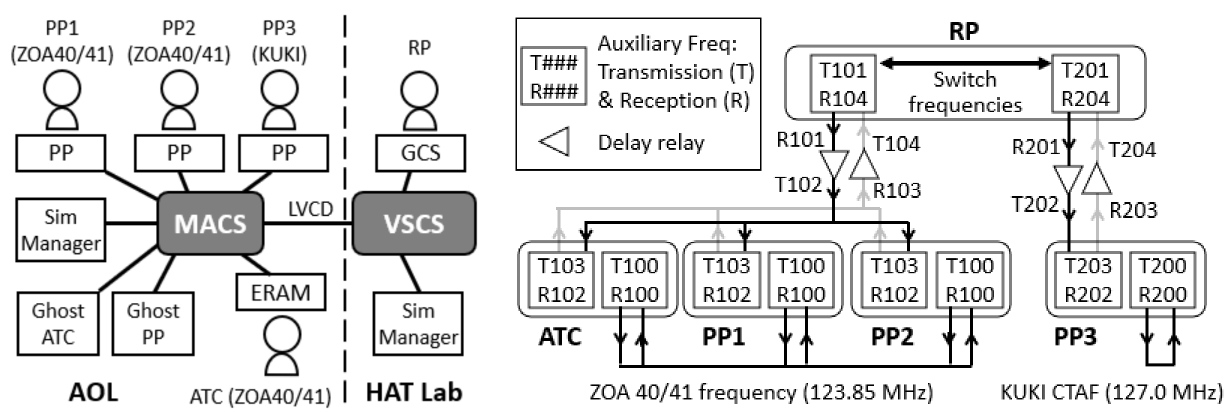
B. Laboratory Setup

The simulation took place in August-September 2021 at the Airspace Operations Laboratory (AOL) and the Human Autonomy Teaming Laboratory (HAT Lab) at NASA Ames Research Center (Fig. 2a). Multi Aircraft Control System (MACS) software [12] in the AOL was used to generate the majority of the ZOA 40/41 background traffic in the study. The MACS-generated traffic was all *cooperative* (e.g., aircraft with a simulated electronic transponder such as Automatic Dependent Surveillance–Broadcast). MACS provided one En Route Automation Modernization (ERAM) radar position workstation for a ZOA 40/41 ATC and three pseudo pilot (PP) workstations. Two of the PPs flew the ZOA 40/41 background traffic, and one flew the CTAF traffic at KUKI. MACS also provided a workstation for MACS simulation manager, as well as ones for ghost ATC and ghost PP, who controlled all the traffic outside the ZOA 40/41.

Within the HAT Lab, the Vigilant Spirit Control Station (VSCS) developed by the Air Force Research laboratory simulated the UA ownship [13]. The VSCS included a simulation manager station with Vigilant Spirit Simulation (VS Sim) and a Ground Control Station (GCS) for the RP. The VS Sim was used to generate the UA flight, as well as inserting one *non-cooperative* flight per run. (See the DAA section for the role of this flight.) The GCS was comprised of two displays, the Tactical Situation Display (TSD) and a status panel showing additional aircraft information. The TSD showed cooperative traffic within a lateral range of 40 nm from the ownship and an elevation of 8,500 ft above and below the ownship. Non-cooperative traffic was displayed only if it was in a lateral detection range of 6.5 nm and an elevation of +/- 15 degrees relative to the ownship. All inputs to VSCS were made via a mouse and keyboard. VSCS and MACS were connected via the Live Virtual Constructive Distributed (LVCD) Environment network so that the traffic generated by each system could fly in the shared simulated airspace environment.

C. Radio Communication

The radio communication among the ATC, RP, and PPs in the two laboratories was provided by the Plexsys Sonomarc system. Two primary frequencies were simulated: 123.85 MHz for the ZOA 40/41 frequency and 127.0 MHz for the KUKI CTAF. Each participant wore a headset with a microphone on a boom and a push-to-talk button. The headset was connected to a radio control panel presented on a tablet. The RP was the only participant who switched between the two frequencies as the UA flight progressed. To simulate voice delays to and from the RP, two Plexsys delay relays, one for each direction, and five auxiliary frequencies were set up per frequency (Fig. 2b). The ATC and PPs transmitted on two auxiliary frequencies simultaneously, of which one was delayed (T103 or T203) and one was not (T100 or T200). The delayed transmissions were sent to the RP, and the ones without delay to the ATC and PPs (as well as to the sim managers for monitoring). The RP transmissions (T101 or T201) were delayed and heard by the ATC and PPs being superimposed with the voices of the others without delay (R100 or R200).



(a) System and Seating in each laboratory. (b) Radio communication connections.

Fig. 2 Laboratory setup.

D. Traffic

Two MACS background traffic scenarios, *A* and *B*, in the ZOA 40/41 were generated based on the traffic data recorded on March 7, 2019. Then, VFR traffic (all cooperative) was added conforming with the typical VFR traffic patterns in the area. Scenario A was designed as a *High* traffic volume scenario, containing traffic level representative of 110% of the Monitor Alert Parameter (MAP) for the combined ZOA 40/41 sectors. Scenario B, a *Low* traffic volume scenario, was a subset of Scenario A, where twenty aircraft were removed to reduce the traffic volume level to 60% of A, or 66% of the MAP. The traffic built up in the first 5 minutes. After that, there were always 16-22 aircraft in the sector in Scenario A (excluding the UA ownship and the non-cooperative traffic), and 8-12 aircraft in Scenario B.

The traffic on the CTAF at KUKI was identical between Scenarios A and B. The airport airspace extended to a 5-nm radius from the airport and from the surface to 3,000 ft above ground level. The types of aircraft and their routes were those typically observed in the area. The CTAF traffic volume was very low as is often expected at airports without operational airport tower.

E. DAA

The VSCS GCS offered DAA alerting and guidance capabilities to help the RP remain DAA Well Clear (DWC) [14]. The DWC thresholds in en route airspace are defined as 450-ft vertical separation, 4,000-ft horizontal separation, and 35-second modified tau (a parameter related to time to Closest Point of Approach [14]) for a cooperative intruder, whereas 450-ft, 2,200-ft, and 0-second for a non-cooperative intruder, respectively. The DAA alerting and guidance were generated by Detect and AvoID Alerting Logic for Unmanned Systems (DAIDALUS) software [15].

In each run, a single non-cooperative aircraft was introduced by VS Sim. This target was designed to generate a DAA Corrective Alert and the associated traffic-avoidance maneuver guidance recommendations to the RP on the TSD. When the time to loss of DWC (LoDWC) became 60 seconds or less, the DAA issued a Corrective Alert. When the time to LoDWC became 30 seconds or less, the alert was elevated to Warning Alert. The non-cooperative aircraft was invisible on the ATC's radar display, so the RP needed to initiate ATC radio communication to request the maneuvers and obtain an ATC approval prior to the execution, if possible. In addition, in Scenario A only, there was one cooperative VFR traffic that intentionally steered toward the UA to trigger a Preventive Alert. Preventive Alerts indicated to the RP that the target was not currently a threat but advised to be monitored.

F. Participants

Eight retired en route ATCs participated in the study. All were retired from the ZOA within the last 16 years (Mean = 5.6). Their experiences at any FAA en route center ranged from 23 to 35 years (Mean = 30.4). Their ZOA experiences ranged from 23 to 34 years (Mean = 28.6).

Eight RP participants were enrolled in the study. Six of them were current military UA pilots, and the other two were civilian pilots. The military UA pilots' approximate flight hours for large UA (i.e., 55 pounds or greater, such as MQ-9) ranged from 1,000 to 3,000 hours (Mean = 2,083). All military UA pilots also had experience in manned flight operations, and their pilot-in-command (PIC) flight hours in manned flights ranged from 50 to 2,000 hours (Mean = 535). The two civilian pilots, including one private pilot and one airline transport pilot, were both current on their IFR rating at the time of the study. Both had over 600 hours of PIC flight time. All participants, including the military and civilian pilots, had experience in civilian airspace operations for either manned or unmanned flight, which was relevant to the simulation environment used in the present study.

Three PPs supported the experiment. All of them had extensive prior experience with the MACS PP workstation user interface. The ghost ATC and the ghost PP were staffed by the research team members. The PPs, the ghost ATC, and the ghost PP were confederates (i.e., non-test participants).

G. Experiment Design

The experiment design was 3×2, including three Voice Delay levels of Short (400 ms), Medium (900 ms), and Long (2,000 ms) and two Traffic Volume levels of A (High) and B (Low). A pair consisting of one ATC and one RP participated together in a single *Session* over the course of two days. Each Session consisted of six runs with all of the 3×2 combinations. The order of the Voice Delay and Traffic Volume conditions were counterbalanced within and between the participant pairs (or Sessions) to reduce potential biases from learning and/or fatigue effect. In total, eight Sessions were conducted in a four-week period, resulting in 48 runs in total. The three PPs stayed at the same position during a single Session, and then rotated positions between Sessions.

Real-time workload ratings and questionnaire responses were collected from both ATC and RP participants as the subjective measurements. In addition to these data, radio communication voices, screen video captures, flight data, and the DAA data were recorded as objective performance measurements.

H. Data Collection Procedure

The ATC and RP participants spent the morning of Day 1 for training, receiving classroom briefing and hands-on training in the corresponding simulator. Then, the six 45-minute data collection runs were conducted in the reminder of Day 1 and Day 2. During each run, ATC real-time workload ratings were recorded at five-minute intervals using a probe presented on the MACS radar display. The RP real-time workload ratings were also collected at the same timing using the VSCS's chat interface. At the end of each run, the post-run questionnaires were administered to both the ATC and the RP to collect their subjective responses. The post-study questionnaire was administered to them immediately following the last post-run questionnaire of Day 2.

The ATC and PPs were informed that the aircraft with the call sign *NASA01* was a remotely piloted UA with a voice delay; however, the length of the delay was concealed from all the participants including the RP. No additional UA indicator was used on the ATC radar screen or the radio phraseology.

I. Statistical Inference

Linear Mixed Models (LMM) repeated-measures regression analysis [16] was applied to analyze the data unless otherwise noted. The effects included in the LMM analysis are listed in Table 1. In the LMM, Participant effect was treated as a random effect, whereas all other effects as fixed effects. Participants could be ATC, RP, or Session (a pair of ATC and RP) depending on the data. All effects were categorical. The reference level (i.e., the baseline) is noted with an asterisk. When there were more than two levels within an effect, a generalized linear hypothesis test was used for contrast testing. Phase effect was only included in the real-time workload rating analysis. The *R* software's *lme4* and *lmerTest* packages were used for the LMM analyses [17-19] and *multcomp* package for the contrast testing [20].

Table 1 Effects included in LMM analysis.

Effect	Levels
Traffic Volume (Traffic)	A (High), B* (Low)
Voice Delay (Delay)	Short*, Medium, Long
Phase	P1* (first quarter hour), P2 (second), P3 (third)
Run Group	RunG1* (the first 2 runs), RunG2 (the second 2 runs), RunG3 (the last 2 runs)
Traffic × Delay	Interaction effect of Traffic and Delay
Participant	{ATC, RP, or Session} 1, 2, 3, 4, 5, 6, 7, 8

* = Reference level

In this paper, Delay, Traffic, and Traffic × Delay interaction effects were the main interests, and only the findings related to these effects are reported. The full results can be found in Ref. [21].

III. Results

A. RP Step-Ons

The waveform voice recordings of the ATC and the PPs' transmissions and the delayed RP transmissions were merged (i.e., T103 and T102 on ZOA 40/41, and T203 and T202 on CTAF in Fig. 2b), and step-ons were counted. This merged audio is a representation of what the ATC and PPs heard on the radio (in contrast to what the RP heard, which would be the merged audio of T104 and T101 on ZOA 40/41, for instance). In the 48 runs, 148 step-ons were observed on both the ZOA 40/41 frequency and the CTAF combined. Among these, 90 involved the RP (i.e., either ATC-RP or PP-RP step-ons). Then, the counts of the RP step-ons per run were analyzed with the LMM. The results showed that the mean RP step-on counts per run increased by 1.75 when the voice delay was Long relative to when the voice delay was Short ($p = 0.033$). The contrast test also found that the step-ons increased in A × Long runs relative to A × Medium runs, as well as in B × Long runs relative to B × Medium runs ($p = 0.026$ and 0.003, respectively). Fig. 3 plots the means and standard errors. The plots confirm the trends that the step-ons increased when the voice delay was Long rather than Short or Medium.

The counts of the non-RP step-ons (i.e., either ATC-PP or PP-PP step-ons) were also subjected to the same LMM analysis, but no significant effect was found. That means that Long voice delay increased only the step-ons that involved the RP, but not the other step-ons.

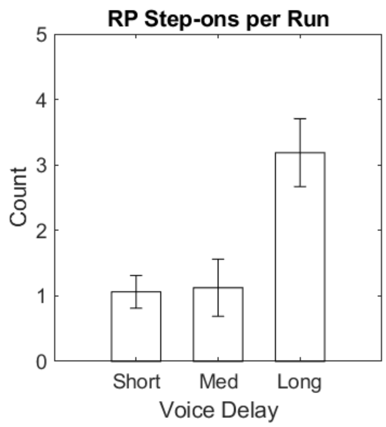


Fig. 3 Means and standard errors of counts of step-ons involving RP by Voice Delay.

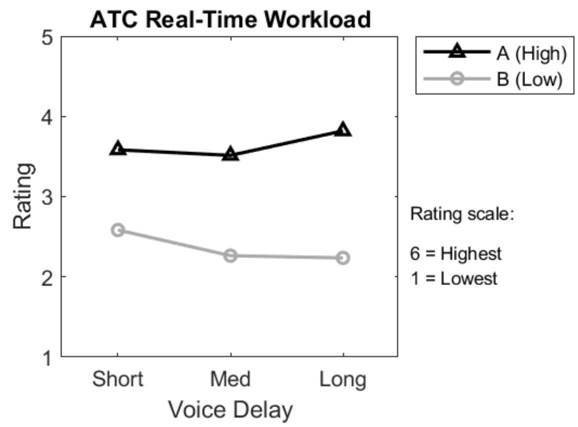


Fig. 4 Means of ATC real-time workload ratings by Traffic Volume × Voice Delay.

B. Elevation of DAA Alert Level

In each of the 48 runs, there was a single scripted non-cooperative traffic conflict for the UA. A Corrective Alert was always issued initially. There were seven runs where the initial alert was elevated to a Warning Alert. Two of them occurred in A × Short runs, two in A × Long runs, and three in B × Long runs.

The runs where the alert was elevated to a Warning Alert were assigned with value 1. The other runs were assigned with value 0. Then, a regular multiple linear regression analysis was applied. The LMM regression analysis was not used as its computation experienced a singularity issue and had a risk of over-fitting. The only difference from LMM was that Participant effect (Table 1) was treated as a fixed effect rather than a random effect. The results found that the non-cooperative encounter in the Long voice delay runs tended to experience more elevations to the DAA Warning Alert level than the Short voice delay runs ($p = 0.042$). The contrast test also found that more alert-level elevations occurred in B × Long runs than in B × Medium runs ($p = 0.043$).

The RPs commented in their questionnaire responses that Long voice delay sometimes caused longer wait times for them to find a radio frequency gap to talk to the ATC, or step-ons and subsequent retransmissions delayed the message to the ATC, leading to a delayed traffic avoidance maneuver. The comments explain why the runs with Long UA voice delay tended to see a DAA Corrective Alert elevated to a Warning Alert more often.

C. ATC Workload

The ATCs' real-time workload ratings were probed at five-minute intervals using the Workload Assessment Keypad presented on the MACS radar display during each run [22]. A six-point scale was used, where 1 corresponded to the lowest workload and 6 the highest. The LMM analysis results showed that the ATC's mean real-time workload ratings increased by about 1 in the A runs compared to the B runs ($p < 0.001$). In A × Long runs, the mean ratings increased by about 1.2 compared to B × Short baseline runs ($p = 0.005$). Fig. 4 illustrated these trends.

Fig. 4 also indicates that, within B runs, the workload ratings slightly reduced in Medium or Long voice delay condition. This was an unexpected trend, and the exact cause was unknown. There is no clear explanation why a longer UA voice delay would consistently result in lower ATC workload than a shorter UA voice delay. Thus, the trend was assumed to be coincidental. The ATCs rated their workload not directly based on the UA voice delay but based on the air traffic management workload level resulting from various factors. The B × Medium and B × Long runs may have been somehow easier than B × Short runs.

The ATCs' post-run NASA Task Load Index (TLX) workload ratings were collected on the post-run questionnaire [23]. The ATCs rated their Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration on a seven-point scale, where 1 corresponded to very low and 7 very high. For Performance rating only, 1 corresponded to perfect performance and 7 failure. Then, the NASA TLX rating was computed as an unweighted average of these six subscale ratings. The LMM analysis results found that the mean ratings increased by 0.94 in A runs than in B runs ($p = 0.002$). Furthermore, the mean ratings increased by 1.8 in A × Long runs relative to B × Short baseline runs ($p = 0.034$). The trends were similar to those in the ATCs' real-time workload ratings.

D. ATC Acceptability for UA Operations

In the post-run questionnaire, the ATCs were asked if they had to adjust their communication strategy for the manned flights to accommodate the UA voice delay, and six out of eight ATCs (75%) responded "Yes" in A × Long runs, a significantly higher percentage of the ATCs than in the other run conditions. The value 1 was assigned to their "Yes" response and 2 to "No." Then, the values were subjected to the LMM analysis. The results showed that the mean value reduced by 0.65 (i.e., more likely to respond "Yes") in A × Long runs compared to B × Short baseline runs ($p = 0.016$). The contrast test also found that the mean value reduced significantly in A × Long runs relative to A × Medium runs ($p = 0.002$). Fig. 5 plots the means of the values and illustrates these trends. According to the ATCs' comments, they made the following strategy adjustments:

- They were prepared for extra delay in the UA response.
- They planned to transmit to the UA only when there was enough time to wait.
- They were prepared to re-transmit the stepped-on communications.
- They ignored the UA when traffic was busy.
- They gave the UA priority and had the other manned flights wait.

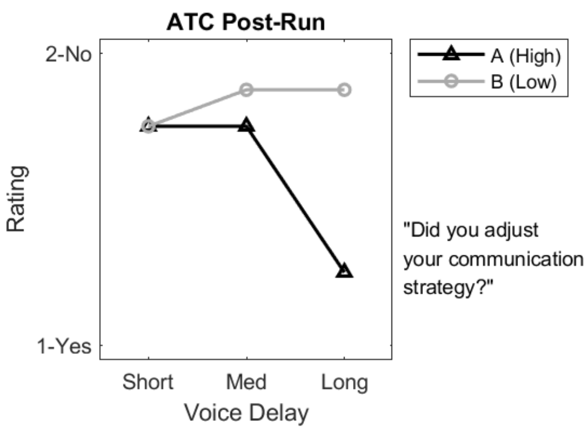


Fig. 5 Means of ATC post-run ratings about adjusting their communication strategy to accommodate the UA voice delay.

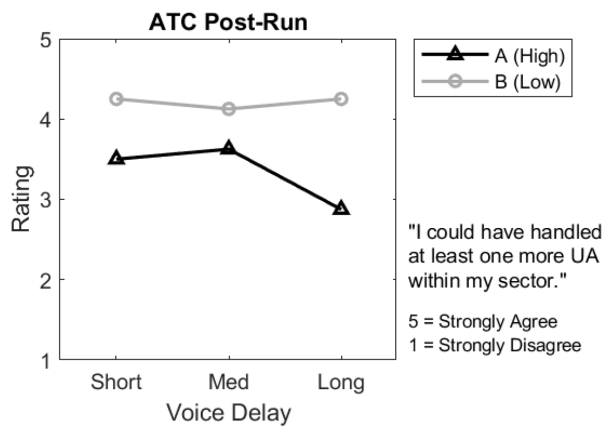


Fig. 6 Means of ATC post-run acceptability ratings for accommodating one more UA.

On the post-run questionnaire, the ATC rated their degree of agreement on the statement, "I could have comfortably handled at least one more UA with the similar voice delay characteristics within my sector" on a five-point rating scale from 1 corresponding to "Strongly Disagree" to 5 "Strongly Agree." The LMM analysis results indicated that the mean ratings reduced by 0.75 (i.e., leaning to disagreement) in A runs than in B runs ($p = 0.021$). The contrast test also showed that the mean rating significantly reduced in A × Long runs relative to A × Medium runs ($p = 0.013$). Fig. 6 plots the means of the ratings to show these trends.

On the same post-run questionnaire, the ATC also rated their degree of agreement on the following statement using the same five-point scale, "Overall, the presence of the UA voice delay did not disrupt the air traffic control operations." The results showed similar trends with the responses to the previous question. The mean ratings reduced by 0.75 (i.e., leaning to disagreement) from A runs to B runs ($p = 0.028$). The contrast test again showed that the rating significantly reduced in A × Long runs relative to A × Medium runs ($p < 0.001$). Fig. 7 shows these trends. The ATCs commented that Long UA voice delays:

- Required more planning for keeping the frequency open.
- Prevented the ATC from moving on to the other traffic.
- Caused too much wait time for the UA responses.

E. RP Acceptability for DAA Alerting and Guidance

The only subjective responses of RPs that resulted in a statistically significant difference were their responses to the post-run question, "The DAA alerting and guidance allowed me to achieve sufficient separation from the traffic conflict(s)." The response scale was from 1 corresponding to "Strongly Disagree" to 5 "Strongly Agree." The LMM analysis did not find any significant effect, but the contrast test found that the ratings significantly decreased (i.e., leaning to disagreement) in A × Long runs compared to A × Medium runs ($p = 0.029$). Fig. 8 shows the trend, although the RP ratings were still very high for all conditions. The high ratings suggest that the RPs were generally confident about these DAA's functions and their ability to maintain separation.

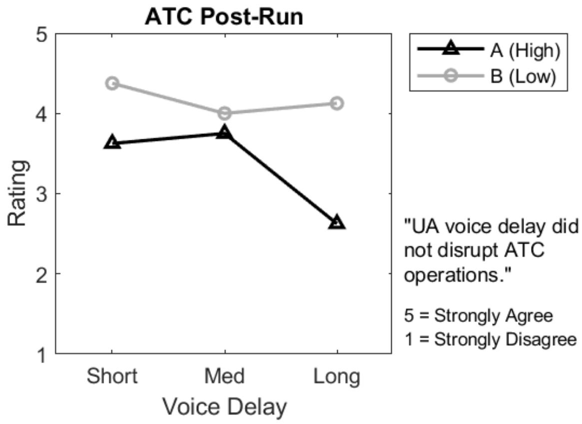


Fig. 7 Means of ATC post-run ratings about the UA voice delay not disrupting the ATM operations.

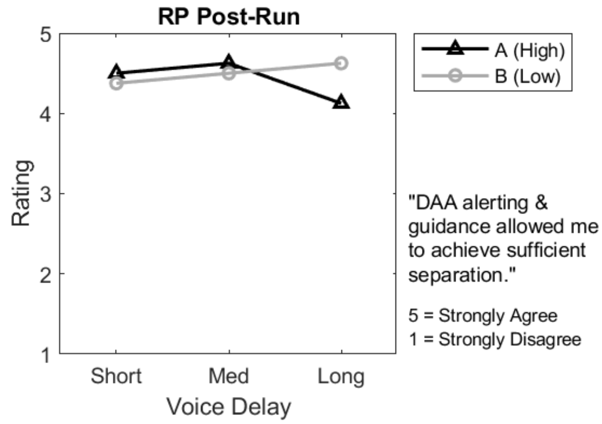


Fig. 8 Means of RP post-run acceptability ratings for DAA alerting and guidance for allowing sufficient traffic separation.

F. UA Indicator

In this simulation, the ATC and PP participants were informed that aircraft with the *NASA01* call sign was a remotely piloted aircraft. Beyond that, there was no information on the ATC radar scope or specific radio communication phraseology that indicated that this aircraft was a remotely piloted aircraft.

One of the questions on the post-study questionnaire asked the ATCs' opinions about the UA indicator. Seven out of eight ATCs responded that they would suggest having an indicator for UA. Four of them suggested to include it in the Flight Data Block because it was an easier location to scan than the aircraft target symbol, as well as historically being a natural place for showing key aircraft characteristics. Three ATCs suggested presenting a different aircraft target symbol for the UA because it would add less clutter on the display and be easier to scan, especially if presented with a different color. One ATC responded that no UA indicator was necessary as manned and unmanned aircraft should be treated identically in the NAS.

No ATC suggested to include a UA indicator in the radio call sign in the post-study questionnaire. However, a few RPs suggested (voluntarily) in their comments that the RP include "unmanned," "remote," etc., with the radio call sign in order to inform other pilots on the frequency of the presence of a remotely piloted UA.

IV. Discussion

A. Effects of UA Voice Communication Delay

The study was able to detect the following negative effects of the Long (2,000 ms) UA voice delay on ATM performance measures regardless of the background traffic volume level:

- The number of transmission step-ons increased.
- The DAA Corrective Alerts for the scripted non-cooperative traffic conflict were more likely to elevate to a Warning Alert level.

The finding that long voice delay increases step-ons is consistent with the previous studies' reporting [5, 8, 10, and 11]. According to some of the RPs' comments, the elevation of a DAA alert level (the second effect above) was often caused by the increased step-ons and subsequent retransmission (the first effect above), or not being able to find a timely gap on the radio to begin speaking.

The study also identified several interaction effects of voice delay and traffic volume, which was the primary interest of this study. The following measures were negatively impacted when the voice delay was 2,000 ms and the traffic volume was high compared to the other conditions:

- The ATCs' workload increased (both real-time and post-run NASA TLX ratings).
- The ATCs adjusted their communication strategy to accommodate the UA voice delay.
- The ATCs' acceptance for accommodating additional UA with the same voice delay characteristics in the sector decreased.
- The ATCs indicated that the UA voice delay was disruptive for the ATM operations.
- The RPs' acceptance for the DAA alerting and guidance for remaining DWC decreased.

The effects of the UA voice delay observed were generally small, which was understandable given that it was only a brief voice delay of 400, 900, or 2,000 ms experienced in only one UA's radio communications among many other manned flights in each scenario. Small effects of voice delay were also consistent with the Rantanen et al. findings [9]. These effects may seem to be too small to be consequential. However, the effects reported above were consistent enough to be statistically significant. In addition, the effects of voice delay were likely underestimated in the study due to the following simulation artifacts:

- The PPs behaved more patiently than pilots in busy airspaces and tended to wait for the RP's turn to respond rather than cutting in.
- The PPs, who remained in the same frequency throughout the run, were less likely to step-on others when checking in a new aircraft to the frequency than pilots in real operations.
- Due to the auxiliary-frequency setup, RP step-ons did not cancel voice transmission. Pressing down the push-to-talk button did prevent the speaker from hearing the other's voice, but the others on the same frequency could still hear the stepped-on transmissions (though often unintelligible).
- UA command execution and other C2 Link System data transmissions were not delayed.
- The ATC and PPs' voices were not delayed.
- The radio voice quality in the lab was clearer than the real-world radio voice qualities.

Furthermore, even in high traffic volume runs, the simulated ATM operations were still simpler than real-world operations, where weather-related issues and additional coordination tasks would be present. If the 2,000-ms voice delay already showed negative impacts in this laboratory study, it would likely be problematic in more complex field operations. If UAs with 2,000-ms voice delays are to be introduced into the NAS en route airspace, additional remedies should be considered, such as increased separation buffers for UAs, restricted background traffic volume levels in areas where UAs operate, limiting the total number of UAs allowed to fly in each sector, Data Link communication, and/or dedicated radio or ground communication channels for the UAs.

No major negative effect was found for the 400- and 900-ms voice delays in the conditions simulated in this study. That does not mean that the operational safety of the UA with these voice delays was verified. The scenarios used in the simulation were limited in scope and were thus unsuitable for such a verification.

B. Turning without Coordination

There was one run in the A × Long condition where the DAA Corrective Alert elevated to a Warning Alert and the RP had to turn the UA without prior coordination with the ATC. The RP's two attempts to talk with the ATC were both blocked by step-ons, and the RP was forced to turn the aircraft immediately and informed the ATC later. The RP commented that, as long as the RP could maneuver the aircraft first, cases like this would not be worrisome. This view appeared to be shared widely among the RP participants, and is in accordance with the federal regulation that states that the PIC is the final authority (14 C.F.R. §§91.3) and the RTCA's assumption that the RP is the PIC in determining actions to remain Well Clear [14]. However, ATCs may have a different view. An ATC (not the one in the same Session with the forementioned RP) commented that any IFR aircraft, including the UA, turning without first asking might be a serious issue. If the UA turns into a busy flow area or the other sector without prior coordination, the ATC would immediately need to coordinate. So, there appeared to be a discrepancy between the ATCs' and RPs' views. The issue is the same for manned IFR flights, but the delayed maneuver in the example above was caused by step-ons due to the long UA voice delay. Thus, the issue becomes more relevant when a UA is present in the airspace.

C. UA Indicator

There have been debates about whether or not the UA should be indicated on the ATC radar screen and/or the radio communications (such as "unmanned," "remote," or "robot" with the aircraft call sign, similar to "heavy," "medevac," etc., identifiers currently used). The opponents claim that these add clutter and require ATC software and procedure modifications. Also, there is an expectation that the UA should be introduced into the NAS operations seamlessly and perform exactly the same way as conventionally piloted aircraft, at least from the ATCs' and the other pilots' viewpoints. If so, no special indicator would be necessary. However, the present study demonstrated that a UA with long voice delay can cause increased step-ons and some other disruptions. Therefore, annunciating that it is a remotely piloted UA may be beneficial. The ATCs suggested to add the indicator to the Flight Data Block or the aircraft target symbol on their radar display. The RPs suggested to add an indicator word to the radio call sign. The radio call sign is helpful in CTAF environments where no ATC is present to oversee the traffic situations.

D. Lack of Awareness of Impacts of Voice Delay

The issue of UA voice communication delay seemed to bring more challenges to the ATC than to the RP. This was evident from the fact that the RPs' workload ratings were generally low, even when the ATCs' ratings increased. The RP was responsible for flying only one aircraft, whereas the ATC was responsible for managing all traffic in the sector. Therefore, the ATCs were the ones whose workflow was most disrupted by the longer wait time and increased step-ons. Simulating only the RP cockpit side would not have revealed the full impacts of the UA voice delay.

Yet, the ATCs themselves may not realize the magnitude of the impacts of the UA voice delay to the NAS operations. Multiple ATC participants told the researchers that they dealt with slow pilot responses regularly, so having a long response delay was not a big problem. Similarly, the RPs also commented that the voice delay, step-ons, and re-transmissions were annoying but not unacceptable. However, there are differences between the natural response delays and voice delays. For instance, the latter increases the likelihood of step-ons. Discounting the negative effects of voice delay may be relevant with the telecommunication research findings that people tend to be unaware of the presence of voice delay and blame the other speaker for being inattentive [6]. So, communicating with ATCs, pilots, RPs, and other stakeholders about the nature of this issue and gaining their supports for mitigations may be a challenge.

V. Conclusion

The present study identified, for the first time to the authors' knowledge, the interaction effects of UA voice delay length and sector traffic-volume level on en route ATM operations in a laboratory HITL simulation. The findings help filling the gap between the current FAA requirements for the NAS voice delay and the past studies' non-findings of negative effect of long UA voice delays.

The UA one-way voice delays of 400-, 900-, and 2,000-ms were evaluated in the study, roughly representing the terrestrial, SATCOM, and long SATCOM C2 Link System transmission delays, respectively. The study found that the ATC workload, the ATC's acceptability for UA operations in the sector, and the RP's acceptability for DAA alerting and guidance for remaining DAA Well Clear were negatively impacted when the UA had 2,000-ms voice delay under the high traffic volume condition, but not with the other conditions. The study also observed that the 2,000-ms UA voice delay increased transmission step-ons and the DAA alert level elevations regardless of the background sector traffic-volume level.

The overall effects of the UA voice delay observed were small, but likely underestimated in this simulation due to several simulation artifacts, such as that our PPs behaved more patiently than pilots in busy airspaces. If the 2,000-ms UA voice delay under high traffic volume condition already exhibited measurable adverse effects in this laboratory simulation, it is likely that the voice delay effect would be even greater in real operations. Therefore, remedies such as increased separation buffers or restricted background traffic density should be considered.

For the 400- or 900-ms UA voice delays, no major adverse effect was found in this study; however, that does not verify that these UA voice delay lengths are safe. These delays still exceed the FAA's current requirement for the maximum voice delay in the NAS (i.e., 390 ms), and require further evaluations. However, the present study's findings may offer grounds for cautious optimism that the background sector traffic level may be used as a control variable to carefully raise the safety threshold for the UA voice delay under low traffic density conditions.

Suggested future research topics are evaluation of the combined effects of voice delays, aircraft command execution delays, and other C2 Link System data transmission delays, as well as the effects of having more than a single UA with voice delays concurrently flying in the airspace.

Acknowledgments

The study was conducted to support the NASA's Pathfinding for Airspace with Autonomous Vehicles (PAAV) subproject within the Air Traffic Management – eXploration (ATM-X) project under NASA Aeronautics Research Mission Directorate. Authors thank the following individuals at NASA Ames Research Center for their assistance: Conrad Gabriel, Kevin Garzon Galindo, Matthew Peel, Annlyle (AJ) Diokno, Conrad Rorie, Casey Smith, Dominic Wong, Terence Tyson, Meghan Chandarana, Matt Gregory, Scott Kalush, Jonathan Luk, Dan Wilkins, and Rob Fong.

References

[1] RTCA, Inc., “Minimum Aviation System Performance Standards for C2 Link Systems Supporting Operations of Unmanned Aircraft Systems in U.S. Airspace,” RTCA DO-377A, Washington, DC, September 2021.

[2] Cardosi, K. M., “Time Required for Transmission of Time-Critical Air Traffic Control Messages in an En Route Environment,” *International Journal of Aviation Psychology*, Vol. 3, No. 4, 1993, pp. 303-313. doi: 10.1207/s15327108ijap0304_4.

[3] Edwards, T. E., Wolter, C., Bridges, W., Evans, E., Keeler, J., and Hayashi, M., "Bow Tie Analysis of the Effects of Unmanned Aircraft on Air Traffic Control," *AIAA Aviation 2021 Forum*, 2021. doi: 10.2514/6.2021-2334.

[4] FAA, “National Airspace System Requirements Document,” NAS-RD-2013, 2014.

[5] Sollenberger, R. L., McAnulty, D. M., and Kerns, K., “The Effect of Voice Communications Latency in High Density, Communications-Intensive Airspace,” Federal Aviation Administration, DOT/FAA/CT/TN03/04, 2003. doi: 10.1037/e597512009-001.

[6] Riesz, R. R., and Klemmer, E. T., “Subjective Evaluation of Delay and Echo Suppressors in Telephone Communications,” *Bell System Technical Journal*, Vol. 42, No. 6, 1963, pp. 2919-41. doi: 10.1002/j.1538-7305.1963.tb00992.x.

[7] Sacks, H., Schegloff, E. A., and Jefferson, G., “A Simplest Systematics for the Organization of Turn-Taking for Conversation,” Language, Vol. 50, No. 4, 1974, pp. 696-735.

[8] Nadler, E., Mengert, P., DiSario, R., Sussman, E. D., Grossberg, M., and Spanier, G., “Effects of Satellite- and Voice-Switching-Equipment Transmission Delays on Air Traffic Control Communications,” *the International Journal of Aviation Psychology*, Vol. 3, No. 4, 1993, pp. 315–325. doi: 10.1207/s15327108ijap0304_5.

[9] Rantanen, E. M., McCarley, J. S., and Xu, X., “Time Delays in Air Traffic Control Communication Loop: Effect on Controller Performance and Workload,” *the International Journal of Aviation Psychology*, Vol. 14, No. 4, 2004, pp. 369-94. doi: 10.1207/s15327108ijap1404_3.

[10] Vu, K.-P. L., Chiappe, D., Morales, G., Thomas Strybel, T. Z., Battiste, V., Shively, J., and Buker, T. J., “Impact of UAS Pilot Communication and Execution Latencies on Air Traffic Controllers’ Acceptance of UAS Operations,” *Air Traffic Control Quarterly*, Vol. 22, No. 1, 2014, pp. 49-80. doi: 10.2514/atcq.22.1.49. doi: 10.2514/atcq.22.1.49.

[11] Comstock Jr, J. R., Ghatas, R. W., Consiglio, M. C., Chamberlain, J. P., and Hoffler, K. D., “UAS Air Traffic Controller Acceptability Study 2: Evaluating Detect and Avoid Technology and Communication Delays in Simulation,” National Aeronautics and Space Administration, NASA/TM-2015-218989, 2015.

[12] Prevot, T., “Exploring the Many Perspectives of Distributed Air Traffic Management: The Multi Aircraft Control System MACS,” *Human-Computer Interaction in Aerospace (HCI-Aero)*, pp. 149-154, 2002.

[13] Fern, L., Rorie, R. C., and Shively, R. J., “NASA’s UAS Integration into the NAS: A Report on the Human Systems Integration Phase 1 Simulation Activities,” *Human Factors and Ergonomics Society Annual Meeting*, 2014, pp. 49-53. doi: 10.1177/1541931214581011.

[14] RTCA, Inc. (2021). Minimum operational performance standards (MOPS) for Detect and Avoid (DAA) systems, RTCA DO-365B, Washington, DC.

[15] Muñoz, C., Narkawicz, A., Hagen, G., Upchurch, J., Dutle, A., Consiglio, M., and Chamberlain, J., “DAIDALUS: Detect and Avoid Alerting Logic for Unmanned Systems,” *IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*, pp. 5A1-1-5A1-12, 2015. doi: 10.1109/DASC.2015.7311421.

[16] West, B. T., Welch, K. B., and Galecki, A. T., *Linear Mixed Models: A Practical Guide Using Statistical Software*, 2nd ed., CRC Press, Boca Raton, FL, 2014. doi: 10.1201/b17198.

[17] R Core Team, R: A Language and Environment for Statistical Computing, Ver. 4.1.0, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/. 2021

[18] Bates, D., Maechler, M., Bolker, B., and Walker, S., “Fitting Linear Mixed-Effects Models Using lme4,” *Journal of Statistical Software*, Vol. 67, No. 1, 2015, pp. 1-48. doi: 10.18637/jss.v067.i01.

[19] Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B., “lmerTest Package: Tests in Linear Mixed Effects Model.” *Journal of Statistical Software*, Vol. 82, No. 13, 2017, pp. 1-26. doi: 10.18637/jss.v082.i13.

[20] Hothorn, T., Bretz, F., and Westfall P., “Simultaneous Inference in General Parametric Models,” *Biometrical Journal*, Vol. 50, No. 3, 2008, pp. 346--363. doi: 10.1002/bimj.200810425.

[21] Hayashi, M., Keeler, J., Wolter, C., and Bridges, W. W., “Effects of Remotely Piloted Unmanned Aircraft’s Voice Communication Delay on En Route Air Traffic Management Operations,” NASA Technical Memorandum, Moffett Field, CA, 2022 (to be published).

[22] Stein, E. S., “Air Traffic Controller Workload: An Examination of Workload Probe,” Federal Aviation Administration Technical Center, DOT/FAA/CT-TN84/24, 1985.

[23] Hart, S. G. and Staveland, L. E., “Development of a Multi-Dimensional Workload Scale: Results of Empirical and Theoretical Research,” *Human Mental Workload*, edited by P. A. Hancock and N. Meshkati, North-Holland, Amsterdam, Netherlands, 1988, pp. 139-183.