Using Natural Language to Enable Mission Managers to Control Multiple Heterogeneous UAVs

Anna C. Trujillo[1], Javier Puig-Navarro[2], S. Bilal Mehdi[2], A. Kyle McQuarry[3]

[1]NASA Langley Research Center, MS 492,
Hampton, VA, 23681, USA
anna.c.trujillo@nasa.gov
[2]University of Illinois at Urbana-Champaign
Urbana, IL, 61801, USA
{puignav2, mehdi1}@illinois.edu
[3]Analytical Mechanics Associates, Inc.
Hampton, VA, 23681, USA
andrew.k.mcquarry@nasa.gov

Abstract. The availability of highly capable, yet relatively cheap, unmanned aerial vehicles (UAVs) is opening up new areas of use for hobbyists and for commercial activities. This research is developing methods beyond classical control-stick pilot inputs, to allow operators to manage complex missions without in-depth vehicle expertise. These missions may entail several heterogeneous UAVs flying coordinated patterns or flying multiple trajectories deconflicted in time or space to predefined locations. This paper describes the functionality and preliminary usability measures of an interface that allows an operator to define a mission using speech inputs. With a defined and simple vocabulary, operators can input the vast majority of mission parameters using simple, intuitive voice commands. Although the operator interface is simple, it is based upon autonomous algorithms that allow the mission to proceed with minimal input from the operator. This paper also describes these underlying algorithms that allow an operator to manage several UAVs.

Keywords: Unmanned Aerial Vehicles · Voice Commands · Autonomy · Coordinated Flight · Mission Operator

1   Introduction

Small unmanned aerial vehicles (sUAV) are starting to become ubiquitous because they are relatively cheap and are fairly easy to fly while the potential immediate productivity gain is large for applications such as photography, inspection, and package delivery. As more people find innovative ways to employ sUAVs [1] – such as crop monitoring [2], photography [3], filming [4], package delivery [5], pipeline inspection [6], search and rescue (SAR) [7], and fire monitoring [7] just to name a few – the way humans interact with them will become critical. Current interaction methods typically include manual controllers [8, 9], smartphones [10] and tablets [11], or graphical ground control stations (GCS) [12, 13]. Interacting with all these types of controllers requires the operator

to learn and understand the specifics of the controller and also the sUAV's dynamic behavior rather than having a more natural and higher level teaming relationship with each vehicle. A lack of teaming typically results in increased workload, decreased situation awareness, and trust issues among all active agents. However, with the possibility of communicating with various types of unmanned vehicles (UxVs) by more natural language (NL) methods, such as speech and gestures, the teaming aspect may come to full fruition. NL may also decrease workload [14] and increase situation awareness [15].

Using speech recognition to give commands is becoming pervasive, especially as speech commands move from controlled to natural language. Many people are now comfortable with speech commands beyond the primitive phone tree systems, such as Apple's Siri, Microsoft's speech recognition system, and Google's Now. Great progress has been made in these systems to understand human speech without training [16-19]. However, the word error rate is still rather high for typical conversational speech recognition especially in noisy environments [20]. Various methods to improve on this have been implemented [21] but have only been partially successful.

An area that may benefit from speech recognition is a dispatch scenario with a dispatcher scheduling multiple package deliveries to a defined neighborhood. This application highlights several aspects regarding UAV control by untrained UAV operators. Allowing UAV operators or mission managers with no UAV pilot expertise to control multiple vehicles is critical to fully realize the new missions that UAVs enable; however, more importantly, the inherent ease to control and the stability of many off-the-shelf small UAVs enable casual users to command and control these vehicles without the need for knowledge of aerodynamics, stability and control, weight and balance, etc. Thus, the goal of this research is allowing an inexperienced UAV pilot, an operator, to define and manage a mission. This mission may entail several heterogeneous UAVs flying coordinated patterns or flying multiple trajectories deconflicted in time or space to locations defined by a dispatcher. This mission may be accomplished with a relatively simple interface. For the package dispatcher, this interface allows a dispatcher to easily define locations for packages to be delivered to and then to easily generate and inspect the deconflicted flight paths to ensure on-time delivery.

This paper describes the functionality and preliminary usability measures of an interface that allows an operator, in this case a dispatcher, to define a mission of delivering packages with multiple coordinated UAVs and then to start the mission. This interface includes using natural language, in this case speech, to make inputs beyond the traditional input methods of keyboard, mouse, and touchscreen. With a relatively well-defined, simple vocabulary and using open-source speech-recognition software, the mission manager can input the vast majority of the mission parameters using simple, intuitive voice commands. Furthermore, although the interface is simple, underneath are autonomous algorithms that allow the mission to proceed with minimal operator input. On-going work at NASA Langley Research Center's Autonomy Incubator and the University of Illinois at Urbana-Champaign regarding the underlying algorithms that allow an operator to manage several UAVs is also described.

2 Initial Voice Usability Experiment

An experiment was conducted to begin to measure the efficacy and user acceptance of using voice commands to define a multi-UAV mission and to provide high-level vehicle control commands such as "takeoff" and "land."

2.1 Independent Variables

The primary independent variables were input type (voice or mouse) and order used. Half of the subjects used the mouse input method first and then used voice, while the other half of the subjects first used voice input and then the mouse.

Subject was included as an independent variable because so few were run because this was an exploratory effort. None of the subjects had used a speech recognition system for input before besides traditional phone trees.

2.2 Dependent Variables

The primary dependent variables consisted of the correctness of the mission parameter inputs and the time needed to make all inputs. Other dependent variables were NASA-TLX workload ratings [22] and subjective ratings on a final questionnaire. The NASA-TLX and subjective questionnaire ratings were all normalized to a 100-point scale and all measures were continuous within that scale.

2.3 Procedure

The experiment required each subject to fill in an online form that contained comparable required information that would be needed for a package dispatcher to deliver packages (Fig. 1). The input screen was programmed in Matlab 2015b®[1]. For each run, subjects typed in a simple numeric code for the package code. Then, they defined the initial starting position ("From"), the delivery location ("To"), and the return location ("Return") using either pull-down menus or voice input. Voice input was accomplished using CMU Sphinx4-5prealpha [23] for speech recognition. Next, they inputted the length of the package, which are detailed in Table 1. If another package was to be added, they indicated that by the "Add Another Package" or if done entering packages, the subject indicated "Done." The subject had the system "Calculate Trajectory" and then "Takeoff" once the trajectory was calculated. Later, the subject used "Land" to finish the run. These last commands (*i.e.*, "Add Another Package", etc.) were considered command fields.

After the voice and mouse input experiment runs, all subjects completed a NASA-TLX. At the conclusion of all runs, subjects completed a questionnaire asking them about

---

[1] The use of trademarks or names of manufacturers in the report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.
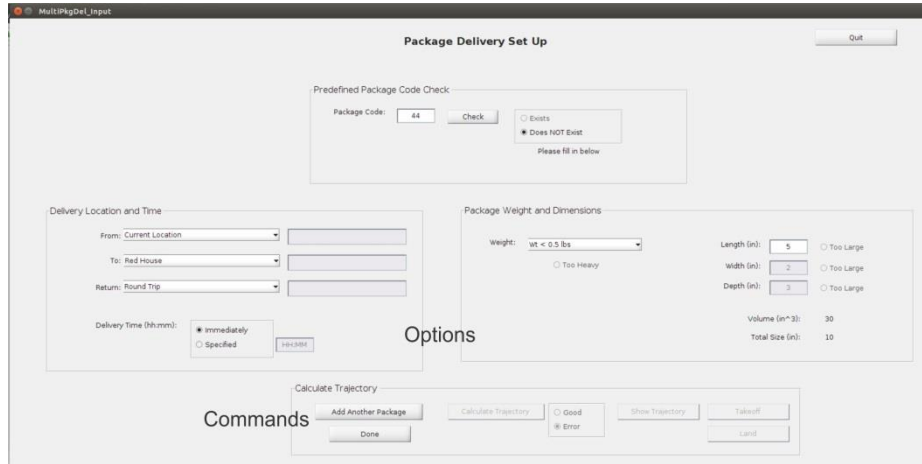
Fig. 1. Package delivery setup screen. The pull down menus and the length input were options. The commands were the pushbuttons such as "Done" and "Takeoff."

their experience in inputting the mission parameters, and starting and stopping the mission using both mouse and voice input.

3 Results

Because this was a preliminary experiment to test the methodology, only four subjects were run. Therefore, only averages and standard errors of the means will be reported with comments on trends.

3.1 Parameter Input Accuracy

All subjects inputted the parameters for the package delivery specification (*i.e.*, "To", "From", "Return", and "Length") with minimal errors. Out of the 288 inputs, there were only five errors (<2% error rate) split between mouse and voice inputs. Therefore, both methods of input appear to be accurate.

Table 1. Option fields accepted inputs.

Option Field	Possible Inputs
Package Code	Integer 1 to 9
From	Current Location
	Forest
	Net
	Red House
	Yellow House
	Green House
To	Red House
	Green House
	Yellow House
Return	Round Trip
	Forest
	Net
	South Lake
	Red House
	Green House
	Yellow House
Length	Integer 1 to 9

However, this initial test was done in a quiet environment with a limited/restricted/controlled vocabulary (see Table 1) and no homophones. Thus, the accuracy of the speech recognition was high [24]. In a noisier environment, possibly with multiple persons talking simultaneously, using push-to-talk headsets or a key word to wake up the system [25] may limit inadvertent voice activation.

3.2 Input Time

Input Specifications. In general, subjects used slightly more time to input package information using voice input (Fig. 2 (a) and (b), and Fig. 3). The required times for voice input may have been influenced by the length of the phrase. Some of the "From" and "Return" phrases were up to three words while the "To" phrases were always two words. The speech recognition system's parsing times for these longer phrases may have increased those times overall.
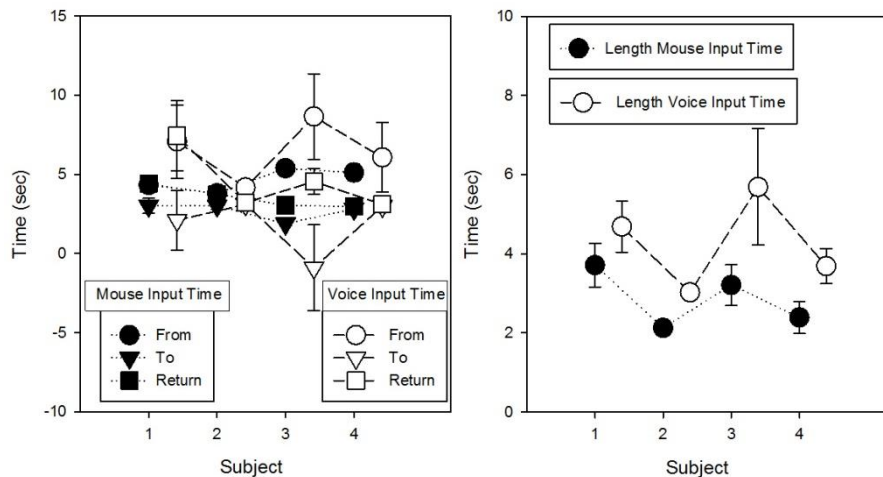

Fig. 2 (a). Time required for each subject to input popup menu information by input type. Popup menu information was To, From, and Return information.

Fig. 2 (b). Time required for each subject to input package length information by input type.

Command Input. The time required by subjects to input commands was generally greater when using voice commands (Fig. 4). Also seen in Fig. 4, the same pattern for each subject when using mouse input was repeated for voice input except it took more time for voice input. Therefore, for commands that are mission critical or safety related (*e.g.*, the requirement to land immediately), it may be necessary to include some type of screen input – such as mouse or touchscreen – in these cases. Lastly, once again the longer phrases (in this case, "Calculate Trajectory") took more time to register when using voice input.

3.3 Workload

The workload in using mouse and voice commands were relatively equal with the voice commands requiring slightly less workload for inputs (Fig. 5). The temporal demand and frustration level between the two input methods were essentially equal. This may have been due to the time it took the voice recognition system to parse the voice input and the timing of the input screen in checking if there was a new voice command. Furthermore, the subjects had no immediate indication on the input screen regarding the voice recognition system and its output until it registered on the screen
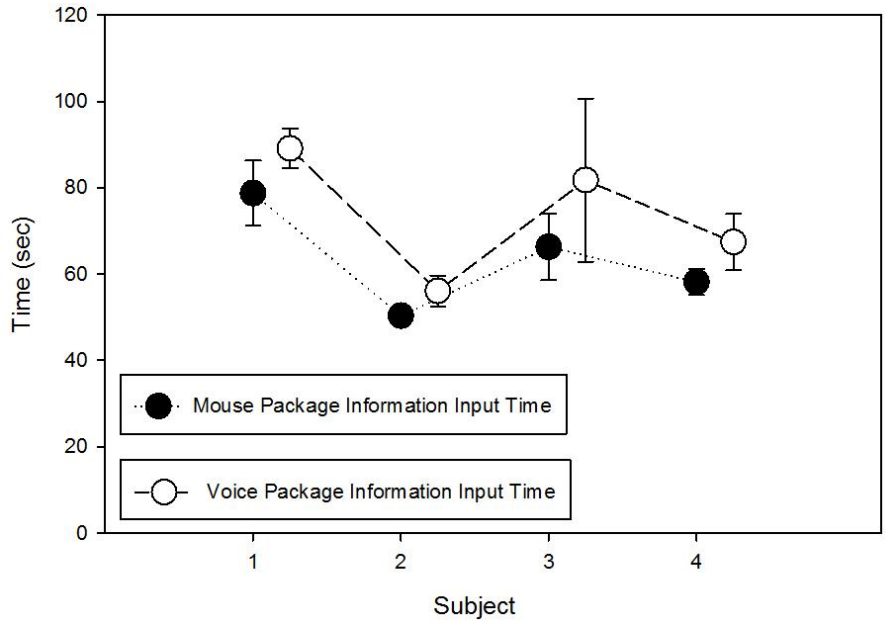
Fig. 3. Total time needed to input all package information for each subject by input type.
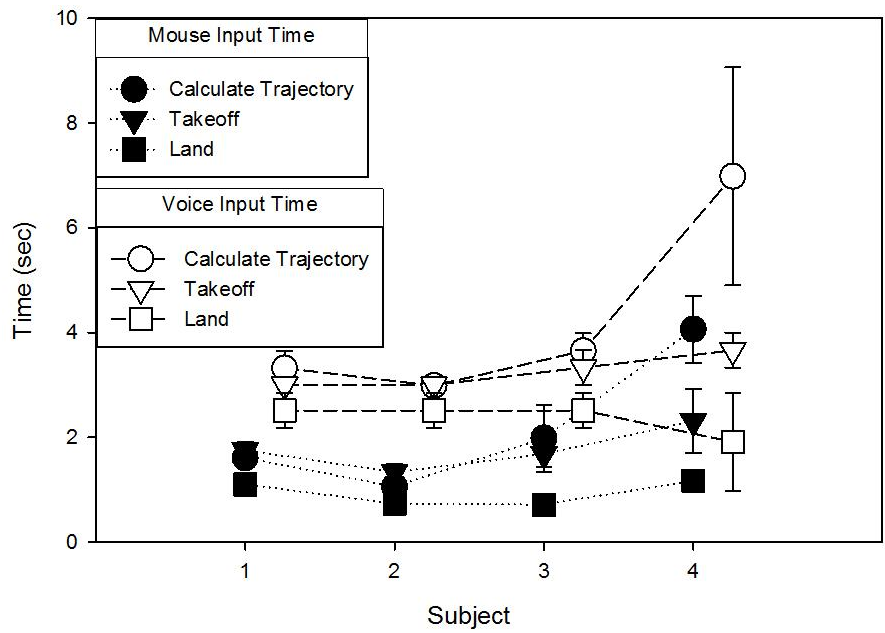

Fig. 4. Input command times for each subject by input type. The input commands consisted of "Calculate Trajectory," "Takeoff," and "Land."
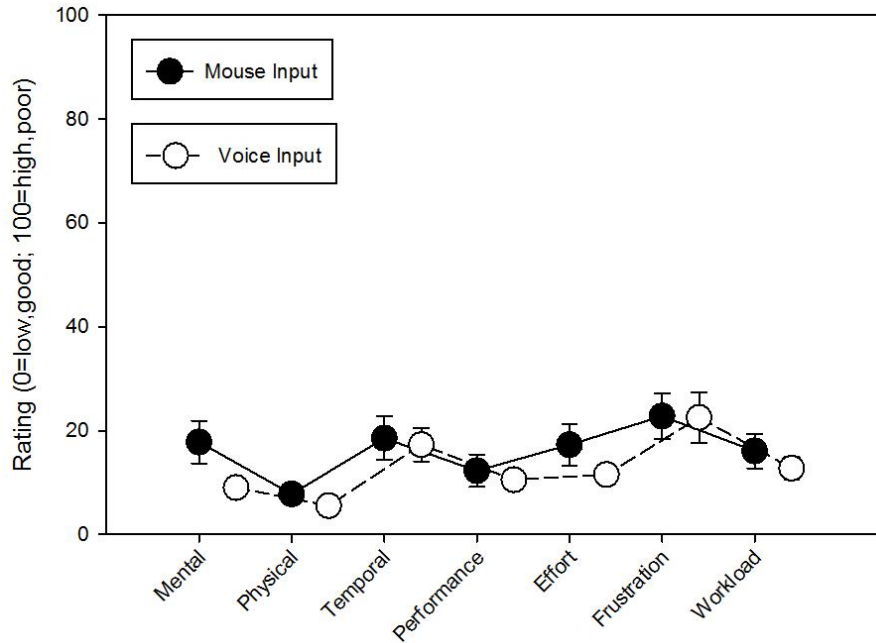
Fig. 5. NASA-TLX ratings across all subjects by input method.

in one of the input fields. Therefore, to decrease the temporal demand and frustration level when using voice input, some type of feedback from the voice recognition system may be required.

3.4 Subjective Preferences

In general, subjects rated using mouse input as slightly easier than voice input when inputting options (*e.g.*, "From") and commands (*e.g.*, "Takeoff"), and in the general ease of use (Fig. 6). For critical input, such as "Land," subjects preferred using mouse input. Again, this may have been due to the responsiveness and time required for the speech recognition system. Surprisingly, subjects indicated that the responsiveness of the mouse was slower than that of voice input. Some subjects indicated that it was tedious to move the mouse around to input the package information. This may have contributed to the mouse input being slower than the voice input. Lastly, subjects had a slight preference for using mouse input. However, their preference of using the mouse for critical input may have swayed this overall preference. Therefore, voice input may be acceptable for non-critical input such as when vehicles are not yet airborne; whereas, direct screen input may be better for critical commands that must be executed immediately.
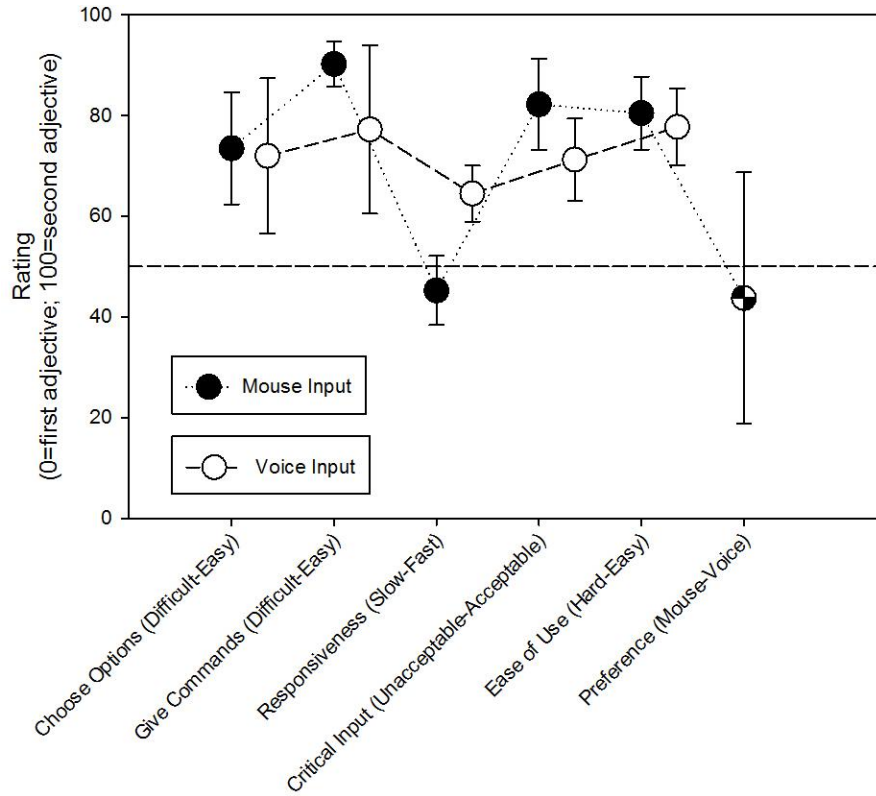
Fig. 6. Questionnaire data.

4   Coordinated Flight Path Generation and Following

Trajectory Generation. Once the packages have all been entered into the system, the system must be able to generate multiple trajectories that consider each vehicle's dynamics and operating characteristics, ensure collision-free maneuvers, and guarantee the desired inter-vehicle coordination for the specific mission [26]. An example algorithm that considers each vehicle's dynamics and coordinates the vehicles in space and time in order to generate each vehicle's trajectory is detailed in [27-30] (Fig. 7). This methodology employs Pythagorian-Hodograph Bézier curves that guarantee satisfaction of boundary conditions, dynamic constraints, and timing schedule of each vehicle. Consequently, the trajectories are provably correct and ensure a safety inter-vehicle distance. Also, the path-following and time-coordination algorithms that complement this autonomous framework have known stability guarantees [31, 32]. These guarantees may engender a higher level of trust in the mission operator that the UAVs will safely arrive at their destinations. This trust will enable a higher functioning system and will facilitate teaming amongst all the autonomous agents whether they are human or machine.

Collision Avoidance. Once the vehicles takeoff, they must have the ability to replan their trajectories to avoid obstacles. Example algorithms that guarantee avoidance along with satisfaction of mission constraints and vehicle dynamic constraints are presented in [33-35] (Fig. 8). In general, these algorithms must first predict a collision and then replan the vehicle's trajectory to avoid the collision. Furthermore, these algorithms are able to avert a possible collision with cooperative or uncooperative obstacles without foreknowledge of the trajectory but with only an online, inaccurate prediction of the obstacle's trajectory. In fact, collision avoidance can be achieved with only the knowledge of the line-of-sight angle only [33].
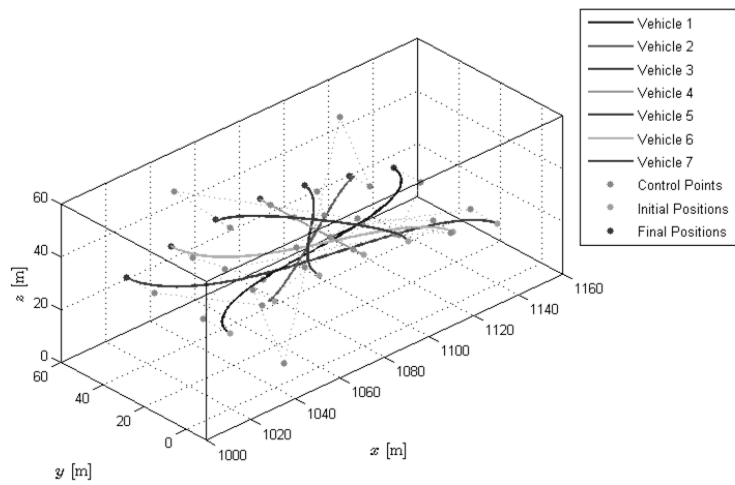

Fig. 7. Three-dimensional temporally deconflicted flight trajectories of 7 mulitrotors.
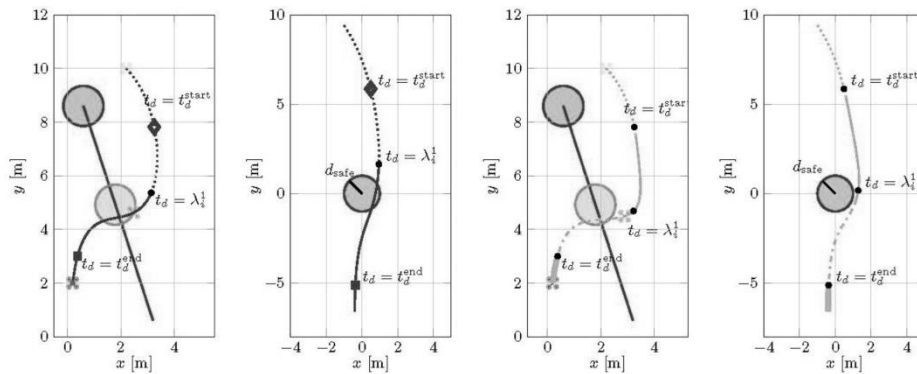

Fig. 8. Illustration of the collision avoidance algorithm: first panel shows the original trajectory (curved line) along with that of the obstacle (straight line). The separation curve, shown in the second panel, overlaps with the obstacle region (circle) indicating a collision. The third panel shows the replanned trajectory (curved line) that ensures collision avoidance. The corresponding separation curve is shown in the fourth panel. Notice that this separation curve remains outside the obstacle region.

Intervehicle Communication. Vehicles will need to coordinate with one another to arrive at a destination at the same time, at prespecified times, or in a time window so as to meet given temporal separation requirements [36] (Fig. 9). In this case, the communications network must ensure adequate communication between the vehicles [32, 37]. In general, the vehicles communicate over a time-varying network, where the quality of service inevitably determines the performance bounds of the coordination algorithm. Once again, performance guarantees may engender a higher level of trust with the mission operator; thus enabling a higher functioning system.

5   Conclusions

The availability of highly capable, yet relatively cheap, unmanned aerial vehicles (UAVs) is opening up new areas of use for hobbyists and for commercial activities. The goal of this research is a relatively inexperienced operator to define and manage a mission using voice commands. This mission may entail several heterogeneous UAVs flying coordinated patterns or flying multiple trajectories deconflicted in time or space to predefined locations.

Fig. 9. (a) Five UAVs arrive at the beginning of glide path within pre- specified arrival windows and separated by approximately 30 sec.

In general, the usability of using voice commands is acceptable. With a relatively well-defined and simple vocabulary, the operator can input the vast majority of the mission parameters using simple, intuitive voice commands. However, voice input may be more applicable to initial mission specification rather than for critical commands such as the need to land immediately due to time and feedback constraints.

Furthermore, although the interface is simple, autonomous algorithms that function transparently to the operator allow the mission to proceed with minimal operator input. This methodology employs algorithms that generate trajectories, coordinate vehicles and avoid collisions with cooperative and uncooperative obstacles using only an online, inaccurate prediction of the trajectory of the obstacles. These algorithms provide rigorous proofs of their performance guarantees. To achieve coordination, vehicle will utilize a wireless communication network, the quality of service of which determines the guaranteed performance bounds of the aforementioned time-coordination algorithm. These guarantees may engender a higher level of trust with the mission operator. This trust will enable a higher functioning system and will facilitate teaming amongst all the autonomous agents whether they are human or machine.

Using voice input for mission specification and using either voice or screen input for commanding the mission combined with guaranteed performance bounds for coordinated flight path generation and following will enable mission operators, rather than UAV pilots, to define and oversee UAV missions. Follow-on research will measure the
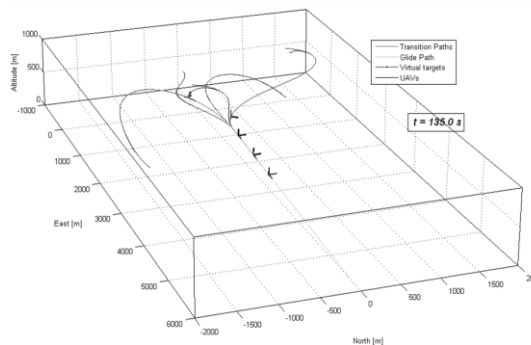
efficiency and acceptability of using voice or screen input for both mission specification and mission control with multiple coordinated sUAVs flying their missions initially in simulation and then in actual flight tests.

References

[1]   D. Jenkins and V. Bijan, "The Economic Impact of Unanned Aircraft Systems Integration in the United States," AUVSI, Arlingtion, VA2013.

[2]   C. Thompson. (2013, 5 April 2013) Unmanned Aircraft to Play Key Role in Future of Agriculture. *Southeast Farm Press*. 2. Available: http://southeastfarmpress.com/equipment/unmanned-aircraft-play-key-role-future-agriculture

[3]   P. Corbett, "Drone Photography Catching on in Real Estate," in *The Arizona Republic*, ed. http://finance-commerce.com/2014/01/drone-photography-catching-on-in-real-estate/: Finance and Commerce, 2014, p. 1.

[4]   T. Johnson, "FAA May Approve Use of Drones for Hollywood Filmmaking," *Variety,* p. 1, 2 June 2014 2014.

[5]   A. Barr, "Amazon Testing Delivery by Drone, CEO Bezos Says," in *USA Today*, ed. http://www.usatoday.com/story/tech/2013/12/01/amazon-bezos-drone-delivery/3799021/#: USA Today, 2013, p. 1.

[6]   B. Jansen, "FAA Approves First Commercial Drone Over Land," in *USA Today*, ed. http://www.usatoday.com/story/money/business/2014/06/10/faa-drones-bp-oil-pipeline-aerovironment-north-shore/10264197/: USA Today, 2014, p. 1.

[7]   M. L. Wald, "Domestic Drones Stir Imaginations, and Concerns," in *The New York Times*, ed, 2013, p. 1.

[8]   "Turnigy 9XR User Manual," Turnigy, Ed., ed.

[9]   "7CAP/7CHP Instruction Manual," vol. 1m23n13606, Futaba, Ed., ed, 2003.

[10] J. Hsu. (2010, 5 April 2010) MIT Researcher Develops iPhone App to Easily Control Swarms of Aerial Drones. *Popular Science*. Available: http://www.popsci.com/technology/article/2010-04/video-using-smartphones-control-aerial-drones

[11] D. Smalley. (2014, 8 February). *Robocopter: New Technology Brings New Capabilities to the Marine Corps*. Available: http://www.onr.navy.mil/Media-Center/Press-Releases/2014/ONR-Demonstrates-Autonomous-Helicopter-Technology.aspx

[12] (2015, 8 February). *Parrot AR.Drone 2.0*. Available: http://ardrone2.parrot.com/

[13] (2016, 8 February). *Prioria Hex Mini*. Available: http://www.prioria.com/hex-mini/

[14] Eurofighter Typhoon. (14 February). *The Cockpit*. Available: https://www.eurofighter.com/the-aircraft

[15] A. C. Trujillo, C. D. Cross, H. H. Fan, L. E. Hempley, M. A. Motter, J. H. Neilan*, et al.*, "Collaborating with Autonomous Agents – 'I'm a doctor Jim, not an engineer' (AIAA 2015-3033)," presented at the Aviation 2015 - 15th AIAA Aviation Technology, Integration, and Operations Conference, Dallas, TX, 2015.

[16] J. Chang. (2011, 10 February). *Speech Recognition Leaps Forward*. Available: http://research.microsoft.com/en-us/news/features/speechrecognition-082911.aspx

[17] J. Koetsier. (2013, 10 February). *Microsoft Doubles Speech Recognition Speed While Improving Accuracy*. Available: http://venturebeat.com/2013/06/17/microsoft-doubles-speech-recognition-speed-while-improving-accuracy/

[18] L. Whitney. (2014, 10 February). *Google Now Tops Siri at Accuracy, Says Analyst*. Available: http://www.cnet.com/news/google-now-tops-siri-at-accuracy-says-analyst/

[19] R. Ritchie. (2015, 10 February). *Siri 'Crushing' Competitors at Language Accuracy*. Available: http://www.imore.com/siri-crushing-competitors-language-accuracy

[20] L. Deng and X. Huang, "Challenges in Adopting Speech Recognition," *Communications of the ACM,* vol. 47, p. 7, January 2004.

[21] Y. McMillian and J. E. Gilbert, "Distributed Listening: A Parallel Processing Approach to Automatic Speech Recognition," in *ACL-08: HLT*, Columbus, OH, 2008, pp. 173-176.

[22] S. G. Hart and L. E. Staveland, "Development of a NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Human Mental Workload*, P. S. Hancock and N. Meshkati, Eds., ed Amsterdam: Elsevier Science Publishers B. V., 1988, pp. 139-183.

[23] Carnegie Mellon University, "CMU Sphinx," http://cmusphinx.sourceforge.net/, Ed., Sphinx4-5prealpha ed, 2015.

[24] National Institute of Standards and Technology (NIST). (2015, 14 February). *Rich Transcription Evaluation*. Available: http://nist.gov/itl/iad/mig/rt.cfm

[25] Amazon.com. (14 February). *Amazon Echo*. Available: http://www.amazon.com/Amazon-SK705DI-Echo/dp/B00X4WHP5E

[26] A. Trujillo, H. H. Fan, C. Cross, L. Hempley, J. P. Navarro, B. S. Mehdi*, et al.*, "Operator Informational Needs for Multiple Autonomous Small Vehicles," *Procedia Manufacturing,* vol. 3, pp. 963-943, 2015.

[27] M. Shanmugavel, A. Tsourdos, R. Zbikowski, B. A. White, C. A. Rabbath, and N. Léchevin, "A Solution to Simultaneous Arrival of Multiple UAVs using Pythagorean Hodograph Curves," presented at the American Control Conference, Minneapolis, MN, 2006.

[28] A. Tsourdos, B. A. White, and M. Shanmugavel, "Cooperative Path Planning of Unmanned Aerial Vehicles," American Institute of Aeronatics and Astronautics, Reston, VA2011.

[29] R. Choe, V. Cichella, E. Xargay, N. Hovakimyan, A. C. Trujillo, and I. Kaminer, "A Trajectory-Generation Framework for Time-Critical Cooperative Missions (AIAA-2013-4582)," presented at the AIAA Infotech@Aerospace Conference, Boston, MA, 2013.

[30] R. Choe, J. Puig-Navarro, V. Cichella, E. Xargay, and N. Hovakimyan, "Trajectory Generation using Spatial Pythagorean Hodograph Bézier Curves (AIAA-2015-0597)," presented at the AIAA SciTech 2015, Kissimmee, FL, 2015.

[31] V. Cichella, I. Kaminer, E. Xargay, V. Dobrokhodov, N. Hovakimyan, A. P. Aguiar*, et al.*, "A Lyapunov-Based Approach for Time-Coordinated 3D Path-Following of Multiple Quadrotors," in *2012 IEEE 51st Annual Conference on Decision and Control (CDC)*, Maui, HI, 2012, pp. 1776 - 1781.

[32] E. Xargay, I. Kaminer, A. Pascoal, N. Hovakimyan, V. Dobrokhodov, V. Cichella*, et al.*, "Time-Critical Cooperative Path Following of Multiple Unmanned Aerial Vehicles over Time-Varying Networks," *Journal of Guidance, Navigation, Control and Dynamics,* vol. 36, pp. 499-516, March-April 2013 2013.

[33] V. Cichella, T. Marinho, D. Stipanović, N. Hovakimyan, I. Kaminar, and A. C. Trujillo, "Collision Avoidance Based on Line-of-Sight Angle," presented at the Conference on Decision and Control (CDC), Osaka, Japan, 2015.

[34] S. B. Mehdi, R. Choe, V. Cichella, and N. Hovakimyan, "Collision Avoidance through Path Replanning using Bézier Curves (AIAA-2015-0598)," presented at the AIAA SciTech 2015, Kissimmee, FL, 2015.

[35] S. B. Mehdi, R. Choe, and N. Hovakimyan, "Multiple Collision Avoidance through Trajectory Replanning Using Piecewise Bézier Curves," presented at the 54th IEEE Conference on Decision and Control (CDC), Osaka, Japan, 2015.

[36] J. Puig-Navarro, E. Xargay, R. Choe, N. Hovakimyan, and I. Kaminer, "Time-Critical Coordination of Multiple UAVs with Absolute Temporal Constraints (AIAA-2015-0595)," presented at the AIAA SciTech 2015, Kissimmee, FL, 2015.

[37] E. Xargay, "Time-Critical Cooperative Path-Following Control of Multiple Unmanned Aerial Vehicles," Ph.D., University of Illinois at Urbana-Champaign, IL, 2013.