"Fly Like This": Natural Language Interfaces for UAV Mission Planning

Meghan Chandarana
Dept. of Mech. Eng.
Carnegie Mellon University
Pittsburgh, PA, USA
mchandar@cmu.edu

Erica L. Meszaros
Dept. of Linguistics
Eastern Michigan Unversity
Ypsilanti, MI, USA
ewicks@emich.edu

Anna Trujillo and Bonnie Danette Allen
Crew Systems & Aviation Operations
NASA Langley Research Center
Hampton, VA, USA
{a.c.trujillo,danette.allen}@nasa.gov

Abstract—With the increasing presence of unmanned aerial vehicles (UAVs) in everyday environments, the user base of these powerful and potentially intelligent machines is expanding beyond exclusively highly trained vehicle operators to include non-expert system users. Scientists seeking to augment costly and often inflexible methods of data collection historically used are turning towards lower cost and reconfigurable UAVs. These new users require more intuitive and natural methods for UAV mission planning. This paper explores two natural language interfaces – gesture and speech – for UAV flight path generation through individual user studies. Subjects who participated in the user studies also used a mouse-based interface for a baseline comparison. Each interface allowed the user to build flight paths from a library of twelve individual trajectory segments. Individual user studies evaluated performance, efficacy, and ease-of-use of each interface using background surveys, subjective questionnaires, and observations on time and correctness. Analysis indicates that natural language interfaces are promising alternatives to traditional interfaces. The user study data collected on the efficacy and potential of each interface will be used to inform future intuitive UAV interface design for non-expert users.

Keywords–natural language; gesture; speech; flight path

I. INTRODUCTION

Many current unmanned aerial vehicle (UAV) enriched applications such as disaster relief [1] and intelligence, surveillance and reconnaissance (ISR) [2], are executed by highly trained operators equipped with a comprehensive knowledge of the vehicle(s) and its control behaviors [3]. Similar to ISR, search and rescue (SAR) missions [4][5] typically employ an intelligent search strategy based on human-defined areas of interest (AOI), and only rely on onboard machine intelligence to locate and identify a target(s) and track to it. This same approach is also employed in suborbital earth and atmospheric science missions that may be collecting data for trend analysis over time across a set of predefined AOIs. In addition to manned flight campaigns, air balloons and satellites are traditionally used to collect data. As new applications emerge, such as atmospheric data collection, the user base shifts from one of experienced operators to one of non-expert users. Therefore, human-robot interaction methods must distance themselves from traditional controllers [5] – whose complexity often makes it arduous for untrained users to navigate – to a more natural and intuitive interface. Systems that work to simulate human-human interaction are found to be more accessible to non-expert users [6].

If available and easily programmable, earth and atmospheric scientists would utilize UAV platforms to collect their data in-situ. UAVs provide a viable method for conducting more comprehensive studies, which may require correlative data to be taken using multiple, coordinated vehicles [3]. Of
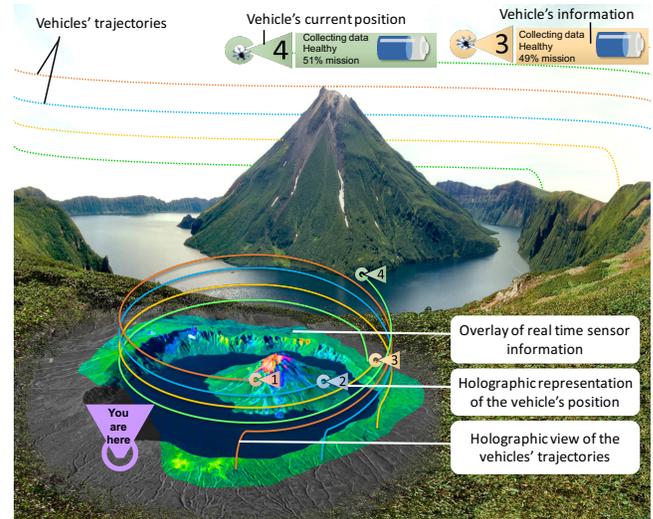

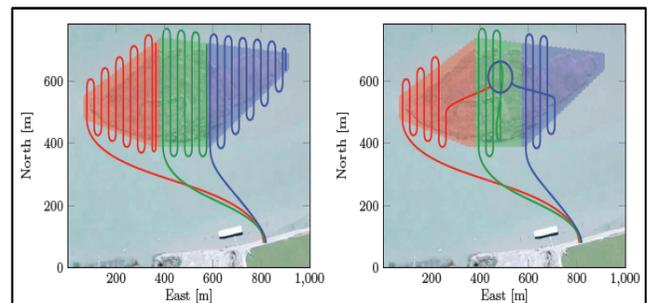Figure 1: Example science mission area of interest (AOI) [7].

Figure 2: UAV search pattern for locating a pollutant [7].

particular interest is their ability to take in-situ sensor measurements in historically hostile or congested environments. Further, data-driven collection based on real-time sampling to point sensors towards, for example, transitions in ozone data or to identify the flow of biomass burning is enabled via real-time replanning for updates of UAV missions and flight paths. Figure 1 illustrates an exemplar science mission AOI and initial search pattern where three UAVs search for the source of a pollutant and then perform a sweeping pattern once within range (Fig. 2) [7]. The UAVs share and fuse maps along with sensor information across platforms during the mission to increase efficiency in locating and tracking the target.

Given current interface and control methods, skilled roboticists and pilots can easily define and program instructions for UAVs. This is made possible by their background knowledge
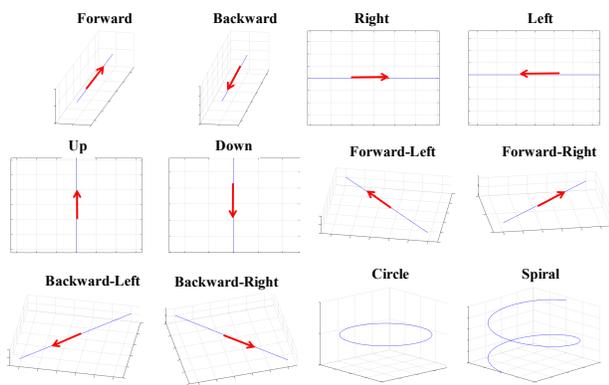
Figure 3: Gesture library of 12 trajectory segments developed by Chandarana et al. [3].

Figure 4: Yes/No message window for the gesture interface.

in the controls architectures required to command complex flight systems. Further, researchers in the area of autonomous aerial missions possess knowledge and insight typical of roboticists and pilots. An understanding of path planning approaches and air vehicle performance is typically required. Airborne (manned) earth science missions are supported by large teams of scientists, engineers, and pilots. Scientists, much like mission commanders, communicate their intent to the engineers and pilots who create a flight profile. This process involves trajectory/route planning of complex, flyable patterns given vehicle and environment. The trajectory/route is generated via negotiation between scientists and engineers such that the desired mission is completed while maintaining safe, executable flight paths. The complex trajectories are often generated/modified in hostile environments (e.g., cargo area of an airplane) where precise, point-and-click interfaces are challenged by factors, such as vibration and dexterity limits (e.g., gloves). The ubiquity and promise of small unmanned aerial systems (sUAS) bring the possibility of reducing dependence on vehicle-specific support, but the gap between science and engineering must be bridged.

Previous researchers looked at several methods for facilitating natural human-UAV interaction. Frequently, these interfaces adopt only a single natural language input. Ng and Sharlin [8] developed a gesture-based library and interface built on a falconry metaphor. Other gesture-based interfaces explore the concept of human-robot teaming where commands like "come here," "stop," or "follow me" communicate intent to the robot or UAV [9] without explicitly defining a flight path [10]. Alternatively, interfaces such as a speech-based interface [11] and a 3D spatial interface [12] have been explored to directly define the flight path of UAV. The work we present here explores the adequacy of common human-human interactions – gesture and speech [13][10] – in the context of an earth science data collection application.

Typically, humans use a combination of gesture and speech for communication. As an initial iteration we explore two distinct natural language interfaces – gesture and speech – for UAV flight path generation. This paper assumes the use of a single autonomous UAV. We compare the performance, efficacy, and ease-of-use of the three interfaces through user studies. Participants use a library of trajectory segments to build several flight paths. The library was developed by gathering information from atmospheric scientists about typ-
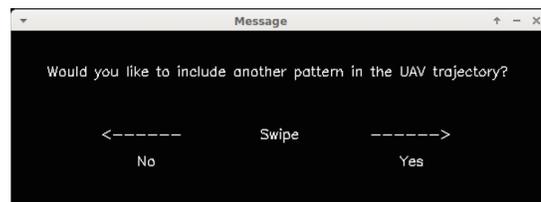
ical desired UAV flight paths to obtain measurements and further breaking them into easily defined primitives [14][15]. Although the given flight paths seen in the remainder of this paper are designed to reflect those of interest to an atmospheric scientist, the same requirement for flight path generation can be seen in variety of other applications such as search and rescue, reconnaissance, etc. This paper evaluates the current instantiation of both natural language interfaces as compared to the mouse baseline. The results will aid in the future development of a multimodal interface that makes use of the strengths from both the gesture and speech interfaces.

The paper is organized as follows. Section 2 describes the three interface frameworks. Section 3 gives an overview of the experimental setup. The results and discussion are given in Sections 4 and 5 respectively. Finally Section 6 provides some concluding remarks and identified future work.

II. INTERFACE FRAMEWORKS

The remainder of this paper will focus on the gesture and speech interfaces, as well as a mouse baseline. The interfaces allow the user to build complex flight paths by defining individual trajectory segments. The subjects are able to use the library of 12 trajectory segments developed by Chandarana et al. [3] to build their desired final flight path (Fig. 3). Using the framework developed by Chandarana et al., each of the natural language interfaces are built with a user flow as follows: (1) define a desired trajectory segment, (2) image of the chosen segment is displayed as confirmation, (3) message asks the user if they would like to define another trajectory segment, if **Yes** (4) repeat step 1, if **No** (5) the user defined flight path is displayed. The framework then combines the segments into one flight path by automatically defining additional parameters [3]. The segments are then automatically combined into a flyable path. All systems make two assumptions about the trajectory library: (1) the *Circle* segment is defined as parallel to the ground and clockwise and (2) the *Spiral* segment is defined as a spiral upward in the clockwise direction.

A. Mouse Interface

The mouse interface consists of a drop-down menu, which includes the 12 trajectory segments in the library (Fig. 3). It assumes that the user will not choose the same trajectory segment two times in a row. A drop-down menu was chosen for this study because it is a selection method familiar to users of a mouse interface and can therefore serve well as a baseline. The user can select a desired trajectory segment by clicking on it in the drop-down menu. As mentioned previously, once a segment is chosen an image of the segment is displayed on the screen to the user as visual confirmation of their choice. For the case of the mouse interface, the user can click on the yes/no window in order to include another segment or finish the flight path.

B. Gesture Interface

For these user studies the gesture interface developed by Chandarana et al., was used [3]. In the gesture interface, a user's gestures are tracked using a commerical-off-the-shelf sensor – a Leap Motion Controller (Leap) SDK v2.2.6 – which has sub-millimeter accuracy. The three infrared cameras provide 8 ft^3 of interactive space [16]. The Leap is placed on the table in front of the user while they sit/stand based on their comfort. The current system assumes that the user is performing the gestures with their right hand.

In contrast to the mouse interface, the gesture interface users perform gesture movements to represent each trajectory segment. The Leap sensor provides more of a natural language interface for the user. This allows them to represent trajectory segments by imitating their shape rather than systems such as the Myo armband, which selects gestures based on discriminability alone [17]. The gesture input is characterized using the linear support vector machine (SVM) model trained by Chandarana et al. For each gesture movement the Leap tracks the palm of the user's hand for three seconds. The eigenvalues and movement direction throughout the gesture are then extracted from the raw data and classified using the trained model [3]. For the yes/no message window, the user must swipe *Right* for **Yes** and *Left* for **No** (Fig. 4).

C. Speech Interface

The speech interface uses a commercial-off-the-shelf headset microphone from Audio-Technica PRO 8HEmW [18] in conjunction with the speech-to-text software CMUSphinx4-5prealpha ("CMU Sphinx"). The CMU Sphinx software was used with the built-in US-English acoustic and language models. This software is a product of Carnegie Mellon University and benefits from more than 20 years of research on speech-recognition. It is ideally suited to this project because it allows for easy customization. The standard version of CMU Sphinx was modified for this application through the creation of a dictionary of allowable words. Four of the formation segments specified in Figure 3 are compound words, e.g., "Forward-left," which consists of both the word "Forward" and the word "left." Therefore, this dictionary contains only eight formation words ("Forward", "Backward", "Right", "Left", "Up", "Down", "Circle", and "Spiral") plus "yes" and "no" for the **Yes** and **No** choices in the message window. In addition, a rule-based grammar was created in order to allow the system to hear the compound formation names.

Similar to the mouse interface, the speech interface presents users with a drop-down selection of the 12 trajectory segments. Rather than selecting the desired segment using the mouse, however, users specify a segment by speaking its name into the microphone. The speech input is then broken down into *phonemes*, or small and distinct units of sound that usually correspond to consonants and vowels, which are in turn compared to the application-specific dictionary of phones and mapped to one of the twelve formations. For the yes/no message window, the system only listens for the words "yes" or "no".

III. EXPERIMENTAL SETUP

Two single input user studies were conducted. Each subject who participated was asked to use two different natural language interfaces: (1) either a gesture or speech natural language interface (Sections 2B and 2C respectively) and (2) a baseline mouse interface (Section 2A). All subjects were allowed to sit or stand in front of the computer screen.
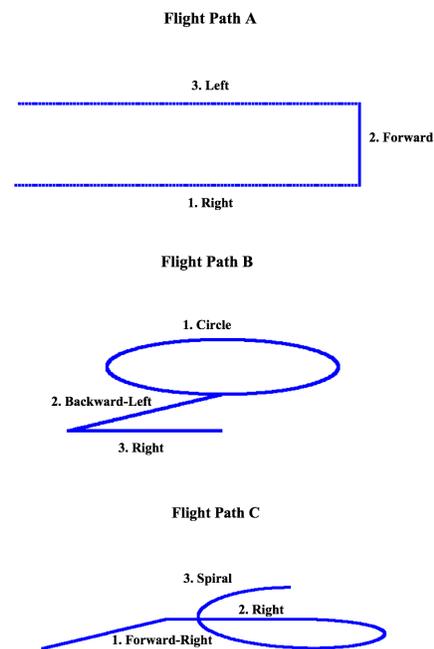

Figure 5: The three flight paths subjects were asked to build in the single input user studies.

The user studies were designed to test the ease-of-use and efficacy of each natural language interface for the purpose of UAV flight path generation. For each trial the subject was asked to define three complete flight paths. Each flight path included three segments. The flight paths ranged in difficulty level and included one common segment — a *Right* — for comparison (Fig. 5). The *Right* segment appeared at different positions in the three flight paths to avoid any bias in segment order. The order of the flight paths was randomized and counterbalanced among the subjects. Each user study was carried out in the following order: (1) subject reads and signs Privacy Act Notice and Informed Consent Form, (2) researcher(s) explains purpose of experiment, (3) subject fills out background questionnaire, (4) researcher trains subject, (5) subject builds given flight paths one at a time (for each interface), and (6) subject fills out subjective questionnaire and NASA TLX (for each interface type) [19][20]. As part of step 2 subjects were told they would be asked to build three flight paths with three segments each.

The subjects were given a printout of the trajectory segment library (Fig. 3) during training and were allowed to keep the printout during testing. Before each trial, the subject was given a printout – with labels – depicting the desired flight path to be built (one of the three shown in Fig. 5). They were allowed to study the flight path for only five seconds before the trail began, but were allowed to keep the printout for reference throughout the entire duration of the run.

In order to correctly define each flight path subjects needed to define the first segment, select *Yes* to add another segment, define the second segment, select *Yes* to add another segment, define the third segment, select *No* to complete the flight path. All errors seen from defining a segment can be attributed to one of six: (1) misinterpreted by system, (2) extra segment, (3) human error – misinterpreted flight path or ended trial too early, (4) combination error – segment misinterpreted
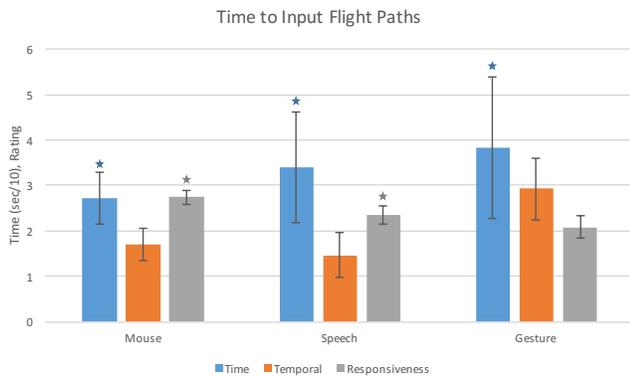
Figure 6: The normalized average time to input flight paths and subject's rating of temporal load and responsiveness of the interfaces.

Figure 7: The average number of errors segments for each input method on a scale from 0 to 3 segments.

by system + human error, (5) combination error – segment misinterpreted by system + extra segment, and (6) combination error – extra segment + human error.

There were 13 subjects who participated in the gesture user study and 14 who participated in the speech user study. All subjects were full time employees at a research center. Subjects who participated in the gesture user study did not participate in the speech user study and vice versa. All participants also used the mouse interface for a baseline comparison. The order of interface use was counterbalanced throughout the subject pool. For both gesture and speech user studies, the same three flight paths were used (Fig. 5). The order in which each subject was asked to build the flight paths was counterbalanced throughout the subject pool, but was kept the same for the mouse interface and the natural language interface runs within the same subject. The subject was asked to fill out a subjective questionnaire and NASA TLX workload assessment survey after using each interface. Researchers also collected time to complete each given flight path and correctness of each flight path defined. The correctness data was collected through observations made by the researcher(s).

IV. RESULTS

The following results were derived from the background questionnaire, NASA TLX(s), and subjective questionnaire. The results will show the time taken to input the given flight paths, the subject's impression of the temporal workload and responsiveness of all 3 interfaces. Input errors will be given for each interface. Mouse interface results are combined as the same interface was used for both sets of user studies. Lastly, we will present the subjective measures of overall impression of how likely subjects are to use the interface method again in the future.

All data was analyzed using an analysis of variance (ANOVA) with IBM SPSS version 24. Tests of Between-Subject effects were run on the independent variables: (1) subject, (2) run, (3) input method, (4) flight path, (5) input x flight path, (6) subject x flight path, and (7) subject x input. A Tukey HSD Post-Hoc test was then run on any non-interaction significant independent variables. The significance values reported assume a $p \leq 0.05$. Error bars are shown for the standard error of the mean in each figure.

The NASA TLX asked each subject to rate their temporal workload on a scale from 0 to 10 – 0 being low temporal load and 10 being high. A separate NASA TLX was used for each

interface used by the subject. In the subjective questionnaire, each subject rated their overall impression (difficulty) of the interface, the responsiveness (speed) of the interface and how likely they were to use the interface again in the future. All subjective questions used a likert scale between 1 and 5. The 1 for the impression rating represented the interface was easy to use and 5 meant it was difficult. In responsiveness, 1 indicated that the interface was too slow, 3 meant it responded at the right speed, and 5 meant the system was too fast. A 1 for likelihood represented that the subject was not likely to use the interface again and 5 that the subject was very likely to use the interface again.

23.08% of Mouse-Gesture user study subjects had previous experience with flying UAVs for an average of 170.67 hours of flight time. 76.92% of subjects said they were right-handed, but all were comfortable using their right hand. Only 7.69% of the subjects had previous experience with a gesture-based interface (other than a cell phone or tablet).

Only 7.12% of Mouse-Speech subjects had previous experience with flying UAVs for an average of 30 hours of flight time. 71.43% of the subjects had previous experience with using a speech-based interface before. This included interfaces such as Siri and Amazon Echo.

A. Time to Input Flight Paths

Figure 6 displays the average time to build a flight path (blue), the average rating of temporal load (orange), and the average rating of responsiveness (gray) for each interface. The average time values given in blue were normalized (divided by 10) to fit on the same graph as the responsiveness and temporal load ratings. The colored stars indicate the input methods that were significantly different from each other.

The time it took for subjects to build a flight path and the subject's temporal load were statistically significant for the input interface method ($F_{(2,58)} = 43.601$, $p \leq 0.01$; $F_{(3,32)} = 3.867$, $p \leq 0.02$ respectively). Responsiveness ratings given by each subject were not significant ($F_{(3,31)} = 2.284$, $p = 0.098$). The time taken to implement flight paths was statistically different as indicated with the blue stars. The mouse method was the fastest input method, however, the responsiveness and temporal load indicated that the different between the mouse, speech and gesture input methods was small. The responsiveness of the mouse interface was statistically different from the speech, but not the gesture (gray stars). Although the time taken to define flight paths with the speech interface was
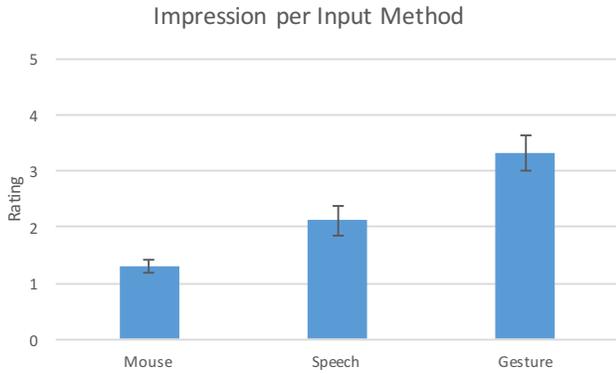
Figure 8: The average impression subjects had about the difficulty of each input method.
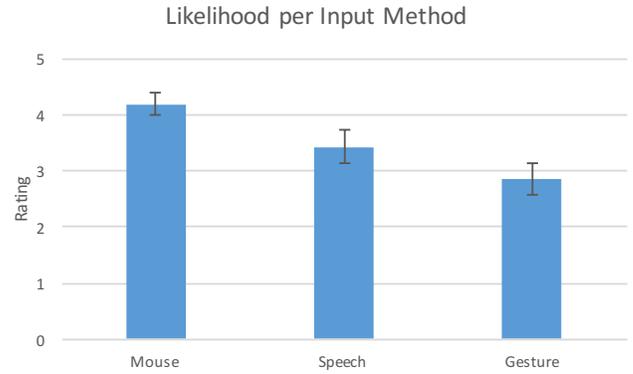

Figure 9: The average likelihood that the subjects would use each interface again.

TABLE I: AVG. % OF FLIGHT SEGMENTS CORRECT

	Flt A % Cor	Flt B % Cor	Flt C % Cor
Mouse	97.62%	100%	98.81%
Speech	95.24%	69.05%	92.86%
Gesture	87.18%	71.79%	64.10%

more than the time taken with the mouse interface, subjects rated their temporal workload lower for the speech interface.

B. Input Errors

The average percentages of correct segments for each flight path are given in Table I. The mouse interface values shown are the average of the values calculated in the all trials combined. For each flight path built, the number of incorrectly defined trajectory segments was counted. The average number of incorrect segments per input method is given in Figure 7. The average number of errors per flight path is statistically significant for the input interface ($F_{(2,58)} = 27.903$, $p \leq 0.01$). All input methods are statistically different from each other.

C. Subjective Preferences

The average impression of each input method given by the subjects was statistically significant ($F_{(3,32)} = 25.458$, $p \leq 0.01$). Similar to the results in the total error per input method, Figure 8 shows that all input methods are statistically different from each other. Figure 9 shows the average likelihood that subjects would use each input method again. Although the ratings are statistically significant ($F_{(3,32)} = 8.618$, $p \leq 0.01$), none of the interfaces are statically different from each other.

V. DISCUSSION

Initial analysis indicates that differences among the input modalities does not seem to drive the total number of errors. The total number of wrong segments was fairly low, with almost no errors using the mouse input method and a low number of errors using the speech interface. This is likely due to familiarity with these types of interface; most subjects use mouse-based interfaces on a daily basis, with 71.43% reporting that they have used speech-to-text systems such as Siri or Amazon Echo previously. The error rate for the speech interface is just above the error rate for the mouse input, except for Flight Path B, potentially indicating an area of focus for improvements to the speech interface system.

Similar to results seen from Trujillo et al. [21], users tended to perform relatively well on each individual flight path segment, though observations indicated that they frequently performed better than they thought they did. With limited contemporaneous feedback and no ability to compare performance to other users or other sessions, users were frequently unaware of their level of success. This often surfaced in their own assessment of their performance on the NASA TLX, as well as, in comments made during experimentation.

Unsurprisingly, the mouse input method proved the fastest method to input flight paths. However, the difference between the mouse, speech, and gesture modalities, as indicated by the temporal and responsiveness responses, was small. The mouse and speech interface temporal results are comparable, while the gestural interface temporal results are only slightly elevated. The responsiveness of all three interfaces is remarkably similar, with mouse and speech both being statistically different.

Users indicated a lower overall impression of difficulty for the mouse interface than for the natural language interfaces. Despite this, users still expressed a likelihood for choosing to use a speech interface again in the future. Users were almost neutral about using the gesture interface again. For both categories, the mouse interface received better scores, which is unsurprising as it is the most familiar. However, the differences were not substantial. Instead, these two subjective categories provide valuable data on user acceptance and willingness to use the natural language interfaces in the future.

Based on observations made throughout training and the user studies, most subjects who participated in the gesture user study seemed to think that using gestures to indicate the shape of a trajectory segment was natural. Most of the errors arose due to a simplification of the interface that required users to perform the gestures at a specific time in relation to feedback shown on the screen. For the most part, using speech to define the trajectory segment shapes did not seem extensible for more complex shapes, which could be more easily defined with gestures. Instead, speech would be better suited to providing information that could augment the gesture input, such as specifying length, radius and height. Such numerical data would otherwise be difficult to intuitively convey with gestures.

While both the speech recognition software and hardware suggest that they work in noisy environments, this initial user study was run with limited background noise conflicting with the speech commands. Because real-life situations will often include at least some degree of background noise, continued research should endeavor to include the effect of noisy envi-

ronments on the accuracy of the speech recognition system. Similarly, while this study used flight paths consisting of three segments, actual science missions may require more complex or lengthy flight paths. Further research should examine whether such changes to flight path length effect the usability of natural language interfaces by leading to fatigue.

Overall, however, analysis of these interfaces has indicated that the natural language interfaces show some promise. Users still successfully used speech and gesture interfaces to define flight paths in only slightly slower times. Continued advancement of their design will enable intuitive, natural language communication between UAVs and human operators, as well as, offer a compelling alternative to traditional interface designs.

Additionally, despite performing faster than other input methods, mouse-based interfaces become a less viable or desirable option outside of the sterile office environment. In the field or on an emergency call, a mouse-based system becomes ill-suited for a trajectory definition application. The results of this study show that alternate natural language interfaces are well-received by users. These alternative interfaces allow for novel ways of defining missions and generating trajectories that lend themselves better to fast-paced field work. Based on these results we can therefore work to improve the next iteration of natural language interfaces so that they are comparable to the results seen by using the mouse-based interface.

VI. Conclusion and Future Work

This paper presented two natural language interfaces for UAV mission planning. User studies were conducted to test the ease-of-use, efficacy and overall acceptance of each interface as compared to a mouse baseline. Overall, the experimental setup proved adequate for gathering data on the efficacy and the potential of individual mouse, speech, and gesture interfaces. This analysis shows that the experimental setup allow for comparison not only of the gesture interface to the mouse interface and the speech interface to the mouse interface, but due to the purposefully similar setup it allows for comparison between gesture and speech interfaces. The analysis indicates that even if users performed better using a mouse interface, they were still able to use the natural language interfaces successfully and were interested in using them in the future. This indicates that natural language interfaces offer an appealing alternative to conventional interfaces, and may provide a more intuitive method of communication between humans and UAVs. Moreover, the data produced in this analysis have indicated areas of each interface that were well-accepted by users, and areas that need to be supported. This is critical information for the design of next generation natural language interfaces.

The focus of this work has been on individual mouse, gesture, and speech interfaces. The data have indicated that while each interface was successfully used to develop UAV flight paths, complementary aspects of each interface were more intuitive and met with greater success. Having identified these strengths, a multimodal interface that combines aspects of the speech and gestural interfaces can be developed to further increase usability and accuracy. Such a combination of both verbal and gestural languages is critical to a truly natural interface [10]. Humans naturally and instinctively use both gestural and verbal modes of communication, indicating that a truly natural language interface should also leverage both [22].

Such a multimodal interface would work to limit any barriers to communication, establishing trust between non-expert users and the system and facilitating improved interaction [13]. More importantly, it would draw on the strengths of the individual interfaces – gesture and speech – and compensate for any limitations in one interface through the use of the other. Future work will examine a next generation multimodal natural language interface used to interact with UAVs.

References

[1] G. Saggiani and B. Teodorani, "Rotary wing uav potential applications: an analytical study through a matrix method," *Aircraft Engineering and Aerospace Technology*, vol. 76, no. 1, pp. 6–14, 2004.

[2] D. Weatherington and U. Deputy, "Unmanned aircraft systems roadmap, 2005-2030," *Deputy, UAV Planning Task Force, OUSD (AT&L)*, 2005.

[3] M. Chandarana, A. Trujillo, K. Shimada, and B. D. Allen, "A natural interaction interface for UAVs using intuitive gesture recognition," in *Advances in Human Factors in Robots and Unmanned Systems*. Springer, 2017, pp. 387–398.

[4] M. L. Wald, "Domestic drones stir imaginations, and concerns," *New York Times*, 2013.

[5] H. Chen, X.-m. Wang, and Y. Li, "A survey of autonomous control for uav," in *Artificial Intelligence and Computational Intelligence, 2009 International Conference on*, vol. 2. IEEE, 2009, pp. 267–271.

[6] D. Perzanowski, A. C. Schultz, W. Adams, E. Marsh, and M. Bugajska, "Building a multimodal human-robot interface," *IEEE intelligent systems*, vol. 16, no. 1, pp. 16–21, 2001.

[7] B. D. Allen and N. Alexandrov, "Serious gaming for test & evaluation of clean-slate (ab initio) national airspace system (nas) designs," 2016.

[8] W. S. Ng and E. Sharlin, "Collocated interaction with flying robots," in *2011 RO-MAN*. IEEE, 2011, pp. 143–149.

[9] T. Naseer, J. Sturm, and D. Cremers, "Followme: Person following and gesture recognition with a quadrocopter," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 624–630.

[10] T. Ende *et al.*, "A human-centered approach to robot gesture based communication within collaborative working processes," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 3367–3374.

[11] M. Quigley, M. A. Goodrich, and R. W. Beard, "Semi-autonomous human-uav interfaces for fixed-wing mini-uavs," in *Intelligent Robots and Systems (IROS). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 3. IEEE, 2004, pp. 2457–2462.

[12] N. Li, S. Cartwright, A. Shekhar Nittala, E. Sharlin, and M. Costa Sousa, "Flying frustum: A spatial interface for enhancing human-uav awareness," in *Proceedings of the 3rd International Conference on Human-Agent Interaction*. ACM, 2015, pp. 27–31.

[13] J. R. Cauchard, J. L. E, K. Y. Zhai, and J. A. Landay, "Drone & me: an exploration into natural human-drone interaction," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 361–365.

[14] A. C. Trujillo *et al.*, "Operator informational needs for multiple autonomous small vehicles," *Procedia Manufacturing*, vol. 3, pp. 936–943, 2015.

[15] A. C. Trujillo *et al.*, "Collaborating with autonomous agents," *15th AIAA Aviation Technology, Integration, and Operations Conference, AIAA AVIATION Forum*, 2015.

[16] D. Bassily, C. Georgoulas, J. Guettler, T. Linner, and T. Bock, "Intuitive and adaptive robotic arm manipulation using the leap motion controller," in *ISR/Robotik 2014; 41st International Symposium on Robotics; Proceedings of.* VDE, 2014, pp. 1–7.

[17] T. Labs, "Myo gesture control armband - wearable technology by thalmic labs," 2015, retrieved: Nov. 2016. [Online]. Available: https://www.myo.com/

[18] Audio-Technica, "Pro 8hemw hypercardioid dynamic headworn microphone," 2016, retrieved: Nov. 2016. [Online]. Available: http://www.audio-technica.com/cms/accessories/b119dcfe66995 ac5/index.html

[19] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," *Advances in psychology*, vol. 52, pp. 139–183, 1988.

[20] J. Byers, A. Bittner, and S. Hill, "Traditional and raw task load index (tlx) correlations: Are paired comparisons necessary?" *Advances in Industrial Engineering and Safety*, pp. 481–485, 1989.

[21] A. C. Trujillo, J. Puig-Navarro, S. B. Mehdi, and A. K. McQuarry, "Using natural language to enable mission managers to control multiple heterogeneous uavs," in *Advances in Human Factors in Robots and Unmanned Systems*. Springer, 2017, pp. 267–280.

[22] "Put-that-there: Voice and gesture at the graphics interface," vol. 14, no. 3. New York, NY, USA: ACM SIGGRAPH Computer Graphics, 1980, pp. 262–270.