A System Architecture for Hands-Free UAV Drone Control Using Intuitive Voice Commands

Megan Landau and Sebastian van Delden
The College of Charleston
66 George Street
Charleston, South Carolina 29424
1-843-953-6905
landaumd@g.cofc.edu, vandeldensa@cofc.edu

ABSTRACT
In this paper, we describe a system architecture that enables an Unmanned Aerial Vehicle (UAV) to be controlled intuitively using simple voice commands. The system has been implemented using the Nuance speech recognition platform along with the DJI Mobile SDK for iOS development in Swift. The control language has been defined using regular expressions and the system has been implemented and tested using a DJI Phantom 4 drone.

CCS Concepts
Human-centered computing → Human computer interaction (HCI) → Interaction paradigms → Natural language interfaces.

Keywords
Voice Control; Natural Language; Robotics;Drones; UAVs

1. INTRODUCTION
Consumers can readily purchase high-end drones online from several vendors. The current top vendors are 3D Robotics, Cheerson Hobby, DJI, Parrot, Walkera, and Yuneec according to a recent Global Consumer Drone Market Report [2016]. These machines are now very easy to fly and do not require an experienced, dexterous pilot in order to perform a variety of missions. A controller with a mixture of buttons, switches, throttles, and rockers/sticks are used to manually pilot the machine from a remote location. Additionally, the controller for the DJI Phantom Series drones, for example, uses an Android or iOS mobile device (purchased separately) to interface with the on-board camera which can be used to autonomously control the drone using built-in computer vision algorithms. Although current high-end drones are relatively easy to fly, they do require that the pilot uses her/his hands to manually control the machine.

This work looks to design an intuitive hands-free approach to piloting a drone in which simple voice commands are used to manually control the drone. Chandarana et al. [2016] recently showed how intuitive gestures can be used for drone control, but that approach still requires the use of hands. Trujillo et al. [2016]'s recently published work is most similar to this work in that it also describes a natural language interface to drone control; however, that paper describes a package delivery system for NASA, uses a different voice recognition platform and drone, and does not focus on the system architecture design. Otherwise, there is relatively little published work in the literature on voice-controlled drones.

The following two sections of this paper describe the system architecture and the implementation of the system which includes a link to a brief video demonstration. The final section of this paper outlines the immediate future direction of this burgeoning work.

HRI '17 Companion, March 06-09, 2017, Vienna, Austria
ACM 978-1-4503-4885-0/17/03.
http://dx.doi.org/10.1145/3029798.3038329.

2. SOFTWARE ARCHITECTURE
The software system architecture alongside its hardware infrastructure is shown in Figure 1.

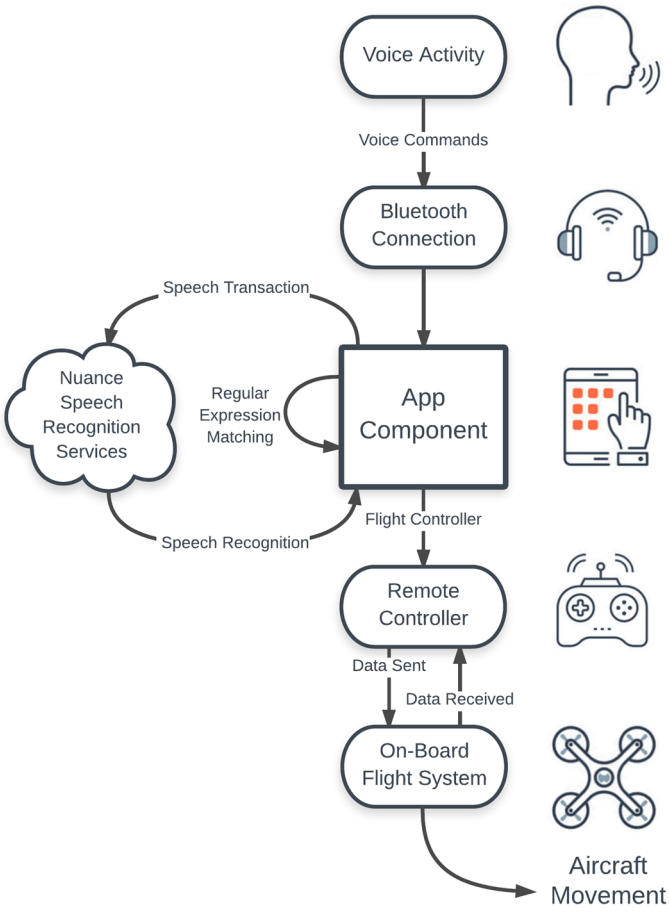

Figure 1. Hardware and Software System Architecture

Before the system can be operational, the user needs to ensure that the remote controller and the drone are powered on, that the app device is attached to the remote controller via the Apple Lightning USB cable, and that the Bluetooth headset is joined with the app device. The user can be talking while the app is running and the Bluetooth device will continuously relay this speech back to the app, which is in constant listening mode. The app is repeatedly sending these transactions to a speech recognition's services, which returns data only if the speech collected was recognizable in the natural language of choice, in our case United States English. Most readily available speech recognition services ensure that irrelevant background noise does not disturb the clarity of commands.

Once a recognizable voice command is issued, such as *"Fly West 6 meters"* or *"Take off"*, the app component will receive the recognized speech and display what was said on the screen. The recognition is interpreted by the app by matching a series of regular expression patterns with the recognized speech. The regular expression patterns are designed to include an intuitive variety of control command options, so that the user is not forced to memorize and repeat a rigid set of instructions. This allows for a more natural interaction with the drone. Regular expressions are used to model the language using a platform-independent mechanism, partially separating it from the implementation in a specific programming language. For example, some of the regular expressions (as Strings) in our Swift-based software implementation, include:

"^\\s*(drone|phantom)?\\s?(go|fly|move|head|come)\\s(east|west|north|south)\\s(to|by|for)?\\s?(to|too)?((?:\\d*\\.)?\\d+)?\\s(feet|foot|meters|meter|m|ft)?$"

"^\\s*(drone|phantom)?\\s?(go|fly|move|head|come)\\s(left|right|up|down|forward|forwards|back|backwards)$"

"^\\s*(drone|phantom)?\\s?(take off|land)$"

A successfully matched command is systematically parsed to collect the necessary information by the app component, and that data is relayed to the On-Board Flight System by way of the app component's physical connection to the DJI Handheld Remote Controller. Successful transmittal of data to the drone results in the execution of the command. The user is also able to cancel missions or instructions while they are being carried out, ensuring flexibility in flight control.

3. SYSTEM DEMONSTRATION

The system has been implemented using the Nuance speech recognition platform along with the DJI Mobile SDK for iOS development in Swift. It has been tested using a DJI Phantom 4 drone. Figure 2 shows an image of the system as it is being used. The user interface on the mobile device is currently rudimentary, displaying the recognized utterance (which can be seen in the lower-left corner of the image), positioning data of the drone, and mission status updates. A video demonstration of the system can be viewed here: https://www.youtube.com/watch?v=roTTViNCnb0.

4. FUTURE WORK

With a working voice-controlled system complete, next steps include expanding the current language capabilities (like adding adverbs to control speed: *"Go up **slowly** to an altitude of 10 meters," "Go left **very quickly**,"* or *"**Quickly** go right 7 meters"*) and merging computer vision algorithms/libraries into the software architecture, enabling a user to issue voice commands that reference visual cues. For example, *"Go North until you see a red truck below."* A comprehensive, domain-neutral language that captures such commands will be defined, implemented, and tested.

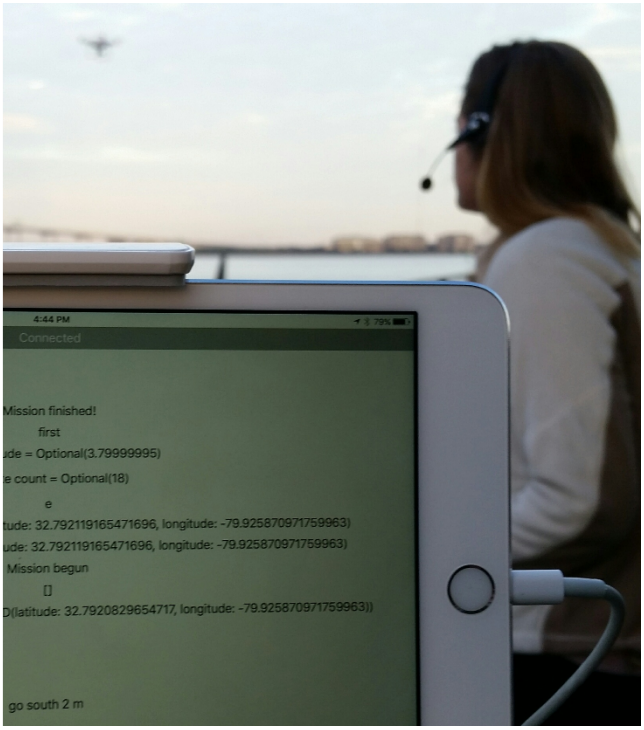

Figure 2. Image of the system in use.

5. REFERENCES

[1] Bell, R. and Perez R., (Session Chairs). 2016. Session on Drones in Geophysics. *The 29th Annual Symposium on the Application of Geophysics to Engineering and Environmental Problems (SAGEEP).* Denver, Colorado.

[2] Chandarana, M., Trujillo, A., Shimada, K., and Allen, B. 2016. A Natural Interaction Interface for UAVs Using Intuitive Gesture Recognition. *Advances in Human Factors in Robots and Unmanned Systems*. Volume 499. Pp 387-398.

[3] Crutsinger, G., Short, J., and Sollenberger, R. 2016. The future of UAVs in ecology: an insider perspective from the Silicon Valley drone industry. *Journal of Unmanned Vehicle Systems.* Volume 4(3). Pp 161-168.

[4] Fisk, K., and Ramos, J. 2016. Preventive Force: Drones, Targeted Killing, and the Transformation of Contemporary Warfare. NYU Press. ISBN 9781479857654.

[5] Global Consumer Drone Market 2016-2020. 2016 http://www.technavio.com/. Report published on October 25, 2016. SKU: IRTNTR10670.

[6] Luppicinia, R., and Sob, A. 2016. A technoethical review of commercial drone use in the context of governance, ethics, and privacy. *Technology in Society*. Vol 46. Pp 109-119.

[7] Sangyoung P., Licong Z., and Samarjit, C. 2016. Design space exploration of drone infrastructure for large-scale delivery services. *In Proc. of the 35th International Conf. on Computer-Aided Design*. New York, NY, USA.

[8] Trujillo, A., Puig-Navarro, J., Mehdi, S., McQuarry, A. 2016. Using Natural Language to Enable Mission Managers to Control Multiple Heterogeneous UAVs. *Advances in Human Factors in Robots and Unmanned Systems.* Volume 499. Pp 267-280.