

Development of a System for Human Language Commands and Control for a Quadcopter Application

Serdar Gökhan Özkartal

Management Information Systems

Girne American University, Cyprus

E-mail: drnexusgokhan@gmail.com

Received: Nov.3, 2014

Accepted: Nov. 30, 2014

Published: January 1, 2015

doi:10.5296/jmr.v7i1.6538

URL: <http://dx.doi.org/10.5296/jmr.v7i1.6538>

Abstract

This project intends to control and navigate the Quadcopter in an undefined nature using sensors and using only voice recognition. This project aims to improve speech recognition software system and test it. The system processes human voice input and answers the queries of the user.

This system is planned task related with the action of the human voice. In this sense, the application fields can change from speech production, storage, transmission and recognition processes.

Hence, this thesis uses this technology for speech recognition. The main emphasis is given to the Hidden Markov Models, Artificial Neural Network, and Linear Predictive Coding. The objective of this project is to let people have the experience of free flight being independent of a map and without using a mobile phone. This project works using a computer only. My objective was to achieve alternative result in contrast to different methods/systems.

This application is tested with speakers where the test included different tones, emotions and lengths of sample for related methodology. The reports of the system data result are used in the rating and analysis of obtained results. Due to this, test shows us that test the improved speech recognition PC system for mankind language, which was tested on various feelings and lengths is a more accurate system with less error than former researches.

Keywords: Speech recognition, Common Vector Approach, Neural Networks, Hidden Markov Models, Hybrid Systems, Acoustic Modeling, Prediction, Classification, Probability Estimation, Discrimination, Global Optimization, Speech Synthesis Universal Mark-up Language, Quadcopter, OpenCV, EmguCv, UDP Port, Human Detection, Shape detection

1. Introduction

Voice recognition systems in general can be seen as software that converts text to speech data. Development of such system would make it easier and more efficient for the human-machine interaction. However, the development of a voice recognition system has a challenging process. For each language, you have to create different voice recognition software; hence voice recognition software in different languages has to solve the problems caused by specific features in each language. Despite all these difficulties, an effective voice recognition system can provide great benefits in many areas. For example, it will be making easier to do business with call centers, cargo companies, restaurants, military services and search and rescue teams, also with voice recognition system saves money, time and loss of manpower may be prevented. In addition, the voice recognition system can also be used for dictating text. To dictate voice text into any text editor program, it will reduce the time spent and effort. In addition, the transition can be made directly from audio files to text and this will allow people that are having difficulty of hearing to follow their publications. An advanced voice recognition system allows the machine to give voice as a command. Technological developments televisions, cars and other electronic devices can be managed by voice without any intermediary devices.

In this project, four bladed helicopters Quadcopter known by the name of the voice control are developed. In addition Quadcopter control uses C # programming language and the OpenCV library filled with EmguCV using Open Source Shape Detection and identification Human Detection feature was added. I predetermine commands which are "Take Off", "Fly", "Run", "Stop", "System Deactivate", "Security", "Hover", "Land", "Up", "Down", "Go Left", "Go Right", "GoBack", "Move On", "Turn Left", "Turn Right", "Camera2", "Camera1", "Find Human", "Shape Detection", "Find Shape", "Find Circle", "Find Square", "Find Triangle", "Stop Shape Detection", "Wait", "Cancel Detection" and also Quadcopter seamlessly detected by computer commands, and in tests of these commands are being applied smoothly.

1.1 Problem Statement

Core objective of this hypothesis is to improve systems that even in unidentified surroundings the Quadcopter can be controlled and piloted with only onboard sensors, with no use of markers or calibration objects, by means of voice recognition. The most vital sensor to be utilized is the frontal camera as a way to calculate a definite estimate of the Quadcopter pose by employing visual tracking techniques. Effect of this estimate moreover assists us to determine the control commands that are needed to fly and hold a certain position in three-dimensional space and speech recognition.

This method of approach enables a Quadcopter to finish tasks such as:

- retain a flying position despite peripheral obstruction such as wind,
- High-altitude manual control: In spite of straightforwardly piloting the drone, this system allows the pilot to transmit high-level control instructions such as “move by $(\delta x, \delta y, \delta z)^T$ meters” or “fly to position $(x, y, z)^T$ ”,

- going after a specified path made of way points, specified in three-dimensional coordinates in relation to the starting point.

Noteworthy hurdles consist of estimating the unspecified scale of the optical map, making up for the considerable delays that occur in the system. Furthermore another challenge is to deal with the restricted sensor quality presented by merging visual pose estimates with further sensor measurements available. Additionally the system has to be strong in relations to momentary disconnections of visual tracking, missing or corrupted sensor measurements and altering connection quality of the wireless link.

This thesis deals with the responsibility of statistical way in this robust technology as assign to speech recognition and refers to a range of theoretical and practical challenges that are still unresolved in respect of their importance and their consequence on performance for alternative system implementations.

The overall involvement of this thesis research has been to plan and employ an improved voice speech recognition system. The test results have revealed an enhanced voice speech recognition system which is not as sensitive to unknown sounds and consequently has increased accuracy of speech recognition.

2. Algorithms

Voice recognition systems have been used in many areas of our daily lives. We can see voice recognition and voice control applications in the work environment, such as home, car, office, hospitals and factories. Users to interact with people in a voice command systems, which are used in vehicle control and in personal and public accessing data. Voice recognition; isolated, unified and continuous is done in three ways.

Voice recognition systems, can be designed by people independently or person-dependent voice recognition Voice commands in the voice recognition system can provide a model for the whole. This model is professional especially suitable for limited speech recognition systems. It will be recognized when it is more than the number of voice and sound are utilized for modeling phonemes. Audio unit with audio/voice recognition and command recognition method is applied continuously. The voice command recognition system can be used as isolated or continuous speech recognition. In these systems, the commands are variable and if easy integration of these commands to the system is required continuous speech recognition methods should be preferred. Today, in the voice recognition system "state-of-the art" method, hidden Markov models (Hidden Markov Models) HMM is used. Besides voice recognition different methods of HMM can be used. In this study, a subspace classification method for recognizing voice commands which is called CVE (Common Vector Approach) method is also used. CVA results are given close to the performance of HMM. In addition, CVA process has advantages in terms of training and recognition based on HMM. We control the car by giving these commands; Take Off ", " Fly ", " Run ", " Stop ", " System Deactivate ", " Security ", " Hover ", " Land, "Up", "Down", "Go Left", "Go Right ", " Go Back ", " Move On ", " Turn Left ", " Turn Right ", " Camera 2 ", " Camera 1 ", " Finder

Human ", Shape Detection ", Find Shape ", Find Circle ", Square Finder ", Finder Triangle ", Stop Shape detection ", Wait ", Cancel Detection ".

2.1 Common Vector Approach

CVA, voice and image recognition applications are the classification method used in a subspace. With this method, a common vector carrying the constant features of each class is obtained. In both cases CVA sufficient data ($m \geq n$), as well as cases of insufficient data ($m < n$) can be applied. Here the number of voice commands belonging to the class of m vector and n indicates the number of elements in each vector. In this study CVO voice command recognition is performed for the case of insufficient data. CVO 'education belongs to a class with indifference subspace of that class of common vectors and eigenvectors are obtained by stretcher. In case of insufficient data ($m < n$) common vectors and indifference subspace Gram-Schmidt Orthogonalization methods or techniques can be obtained using the subspace. Vector subspaces with technical partners in the relationship between the eigenvectors of the covariance matrix are commonly used.

2.2 Hidden Markov Model

Voice recognition systems, first the music sound waves are converted to an array of vectors or observation. These sequences are handled as a finite-state machine in Markov property. So, the next observation in the observation sequence is not any observation in the past, but depends on the observation now. But in reality, this system has some limitations: models and observations on the state sequence are reserved and it can only be calculated by reference to this case. It is called Hidden Markov Model. Audio signals can be symbolized and processed as finite-state machines. Of course, before rising to the level of formal languages, many calculations are carried out as sound waves. Because, this type of signal processing is entered into the calculations they will not be mentioned. The three major algorithm is being used in the HMM to find the State sequence.

Forward Algorithm: Model is used to find the order of the state and gather all of the possibility situations that may arise.

Viterbi Algorithm: Rather than collecting all the possibilities in Forward Algorithm, Viterbi algorithm matches the best audio vector in the chosen area. So it is a healthy outcome.

Baum-Welch Algorithm: To find more accurate results, Observation sequence calculates from beginning to end and re-calculates thoroughly the probability of the observation. To process audio signals, after separating it into small parts, these parts need to ensure compatibility with sounds and syllables that coincided. To do this, you want to give expressions derive from the grammar, pronunciation dictionary, the word network, to collate many items as required.

2.3 Voice Command Recognition System

General block diagram of a voice command recognition system is given in Figure 2.3. In this system, the endpoints of the first voice command must be specified. In this study energy and zero crossings were used to determine the endpoint measurements. Energy and zero crossing

number of voices are used to determine the threshold values obtained for the end point. The sound feature vector is obtained after determination of the end point. HMM feature vector is determined to find which command it belongs to. Subsequently, this command is sent to the Quadcopter by using numeric of WIFI.

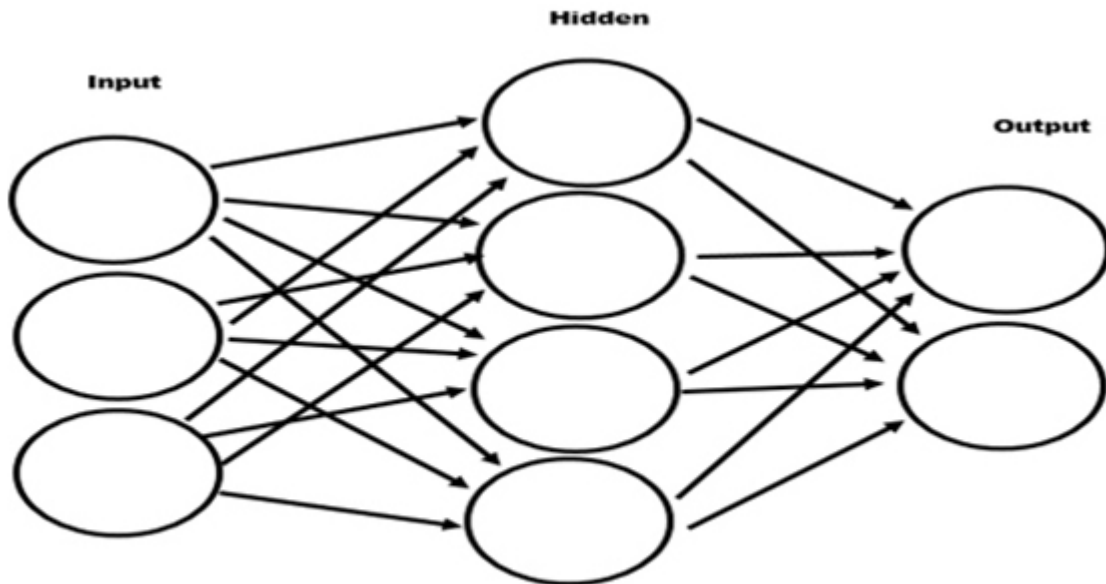


Figure 2.3: A hierarchically structured HMM.

2.4 Speech Recognition Problem

Basically, the speech recognition system task is to get a speech and predict that what is said. So the system is the input speech data and the estimated output sentence (hypothesis sentence). Two main parts (feature extractor and language analyzer) of the system are needed to be able to estimate. Feature extractor, the structure of the sound waves and converts the data into arrays of acoustic attributes that can use language analyzer to create a structure (in the next section gives more detailed information about these tools). Then, the language analyzer takes these attributes that have been removed from the sequence and predicts sentences. Two basic models are used to estimate: the language model and the acoustic models. Detailed information about this model will be given in this section. Basically the following Bayesian estimation process is performed by the equation:

$$W = \text{argmax } P(W|A) = \text{argmax } P(A|W) P(W)$$

Where W is the hypothesis sentences. This sentence in the second side of the equation can be seen through the possible sentences that are most likely coinciding with acoustic series $P(W|A)$ based on the acoustic data (A) the current sentence (W) stands for the probability of acoustic that match with its series. The second part of the equation is obtained by direct application of Bayes' rule. Where $P(W)$, W is the probability of occurrence of a sentence is calculated according to the model and language. $P(A|W)$ according to the sentence given us

is the possibility of the formation of the acoustic index which is calculated using these acoustic modeling. Due to the Bayes rule $P(A)$ is simplified because it is independent from W .

2.5 Attributes Extractor

As mentioned earlier, before voice recognition, voice data must be converted to an array of parameter vector. For this reason, feature extraction audio data divides non-empty sections that intersect with each other in terms of time and remove attributes in each section. Although many different methods for feature extraction are found, it offers similar frequency resolution to detect the human ear therefore; filter-bank analysis is more common. Due to these advantages, we use our voice recognition “sr.LoadGrammar” approach that adapts the filter-bank. Equally spaced frequency Triangular filters and sound waves are generated and multiplied with the Fourier transform of this filter. Thus a weighted sum according to values in the filters is obtained. In this newly created data and the MFCC (Mel-Frequency Cepstral Coefficients) are used to remove a vector attributes.

2.6 Acoustic Model

Hidden Markov Models forms voice recognition system based on the model of the acoustic. Phoneme is the smallest unit of acoustic and phoneme based is usually based on HMM models consists of 3 pieces. HMM the model parameters change state probabilities (state transition probabilities— a_{ij}) state observation densities (state observation density— $b_j(o)$) and the initial state distributions (initial state Distributions - Π). $b_j(o)$ Calculating the probability distribution function is calculated by using Gaussian mixture. Number of Gaussian mixtures which can be used can be increased independently from other sounds around Monophonic use a voice recognition system for providing and for multiple audio that can correspond to a phoneme recognition system however, Monophonic does not provide an effective solution. Stereophonic located in the left and right of each phoneme and phoneme is modeled with stereophonic therefore reduces acoustic variability and provides a more effective recognition.

2.7 Model Language and Grammar

Language models or grammars, while speech recognition, acoustic recognition of spoken words to the context in which they are allowing their identification Grammar rules define manually set sentences, the highest acoustic model sets forth the possibility of a word, that word outside of grammatical rules is not likely to be recognized, the only recognizable words determined by this rule. Therefore grammars are useful only in limited reporter speech recognition systems. BNF is useful in determining this kind hand written grammar. On the other hand, language models, training is trained from text and creates the possibility of sentences in that language. Assuming the sentences are built by words, words consisting of w_i forms W sentences where we can say the equation is as follows:

$$P(W) = P(w_1, w_2, \dots, w_n) = \prod_{i=1}^n P(w_i | w_1, \dots, w_{i-1})$$

Usually N-gram models are used in Language model practices, these models only approaches the right side of the equation by using the previous N-1 words

There for the above equation becomes something like this:

$$P(W) \approx \prod_{i=1}^n P(w_i | w_{i-N+1}, \dots, w_{i-1})$$

N-gram probability is calculated by this formula:

$$P(w_i | w_{i-N+1}, \dots, w_{i-1}) = \frac{C(w_{i-N+1}, \dots, w_{i-1}, w_i)}{C(w_{i-N+1}, \dots, w_{i-1})}$$

Where $C(w_{i-N+1}, \dots, w_{i-1}, w_i)$ N word is specified by showing the training text, $C(w_{i-N+1}, \dots, w_{i-1})$ refers to the number of N-1 passing the text of the specified words. If this is the case $N = 2$ is obtained Bigram language model. For the words that are not recognized in Training data are organized (smoothing) by this method so that it can have probability value assigned to them. Probabilities of word recognition based on statistical distribution models created in this language, it varies like acoustic model.

2.8 Assessment Criteria

There are several measurement methods to determine the performance of speech recognition systems. The most accurate benchmark hypothesis is to evaluate the difference between sentences with reference sentences. For this reason Word Error Rate -WER calculated as follows:

$$WER = \frac{D+S+I}{N} \times 100$$

Where N is the total symbol (label) reference sentences number, D references the deleted symbol hypotheses based on the number of sentences, the number of S is for the changing symbol and I is an extra symbol that represents the number. By knowing the Word Error Rate provides us the percent accuracy. Apart from the base error rate (Stemerror rate - SER) and the letter error rate (Letter error rate - LER) can be defined as the evaluation criteria. SER, after taking out the paragoge words from reference and hypothesis sentences, the base can be found by calculating the Word Error Rate. LER can be calculated by separating the individual letters that determines Word Error Rate between the letters.

3. Problems Encounter in Voice Recognition Systems

Voice recognition system introduced in the number of words and syntax complexity increases, where the recognition percentage decreases. The reasons for this are as follows: the English language has different spellings and meanings of words that have the same pronunciation that are used in the same sentence, the machine can interfere with finding the correct results when analyzing speech. There are multiple types of voice recognition. Isolated word recognition is easier than constant word recognition. Because the first word is relatively more open and articulate. Same as the mutual speech, there is a big difference between reading the speech. Improvisation takes place as in one conversation with constant and variable intonation, in the other, paying attention to intonation in a slower tempo. Speaker dependent and speaker-independent voice recognition systems are also exhibited different performances. Machine recognizing more than one speaker's voice can influence the calculations. During the conversation the noise level of funding is also a major challenge. When it is not possible to stem the growth in technology, voice recognition also increases the rate of development in the same way. Voice recognition is the biggest cause of the difficulty, as in the computer's internal structure in the form of either 1 or 0. As shown on the basis of statistical and mathematical formulas, many used a system based on probabilities and approximations. This is why more and more R & D performed; although voice recognition is an interesting subject, it needs to be dealt with patiently and meticulously.

3.1 Quadcopter

Quadcopters are employed in various situations such as examination and evaluating and furthermore used in researches at universities for studying flight control theory, navigation, real time systems and robotic technology. Nowadays, in numerous universities researches and developing Quadcopters are both operating in a number of angles.

Quadcopters can also execute several tasks for example flip-flap, hovering, and going through a window. Moreover, the use of Quadcopters as a multifunctional examination platform has advantages. Due to Quadcopters straightforwardness of the mechanic composition and the relative cheap price of spare parts beginners can fix and maintain the Quadcopters too. Many academicians are attempting to improve Quadcopters because of their versatile functionality.

Quadcopter projects are more often than not composed of the computer, electrical and mechanical engineers. Seeing that their movement capabilities are very light, Quadcopters are capable of being used in various types of circumstances in which they can fly by themselves with no need of the help of a pilot.

Many research laboratories are working on new developed command techniques and researches. The biggest research laboratories are MIT Aerospace Controls Lab, ETH Flying Machine Arena and Pennsylvania General Robotics, Automation, Sensing and Perception (GRASP).

As pointed out above Quadcopters are employed in various fields. One of those fields is the military. The military uses Quadcopters to track, explore and to rescue. A greatcase for this is the small and quiet UAV, called Aeryon Scout manufactured by Canadian Aeryon Labs.

Pitch and roll x, y and z axis have an effect on the quickness whereby this can be employed without having an effect the yaw axis. Each propeller which rotates in the identical route controls one of the pitch and roll axis. Although amplifying the push power of one propeller, it decreases the others. It maintains the torch balance required for yaw stability. Doing such, offers torch application to roll or pitch axis. Steady propellers allow the movement of Quadcopters to axis. Transaction acceleration is gained either by a non – zero pitches or roll angel.

The intention behind the use of four engines is that it has two fitting rotational symmetries. It is moreun-problematic to generate roll movement with four engines by allocating the push sideways. Complete force on the torch and yaw stays at zero whilst one of the two engines which turn in the equal direction increase in speed whilst the other slows down.

Controlling a conventional helicopter is a difficult job to balance the torch generated by the engine and put on the yaw control. Quadcopters yaw control is in balance. Despite changing the main controller, it remains easy to learn controlling it, it only needs practice. Nowadays, high quality Quadcopters possess gyroscopes which stabilizes the yaw correctly.

Mechanical components, chassis, propellers and engines are required for the manufacturing. If one is to achieve the best performance and uncomplicated control algorithms, engines and propellers ought to be placed at the same distance. These days, Carbon Fiber designs are common due to their weightlessness and sturdiness. Esc module as in RC helicopters control card and batteries are the essential electronic components for a functional Quadcopter. Generally, a simple remote control is used.

4. The Developed System, Its Investigation and Results and Analysis

This chapter examines the development system, its analysis and results. Studying the development system stumbled upon problems along the way. Even though the aim is to reduce errors that were formerly outlined however since the program is custom designed there will be a few restrictions, for example only the needed tasks can be processed, and occasionally more complex tasks cannot be processed.

The surroundings and also sound of the surroundings can have an effect in having clear and high detection of accuracy under stress and noise pertains. For instance, the acoustic noise hindrance can be greater in a closed, indoor environment.

Security is one of the most complex tasks; the program ought not to run on other computers which are not approved to run it. This can be accomplished by making the application validate the computer's hardware information and produce a hardware ID for licensing.

C# is a programming language developed by Microsoft. Even though it is indistinguishable from C++ and Java languages; C++ has distinctions as well. For instance, C# is distinct from C++, C# has 100% object oriented programming. It is not the same as Java in using pointer. In this approach it can operate and compile with the older programming language systems.

In my old system I used Visual Basic. Currently, I have decided to employ .NET C# because of the subsequent reasons; Firstly Microsoft has ceased supporting VB. Secondly,.NET C# has allowed my system to operate better and .Net C# is a simple, modern, object oriented and is a secure programming language.

I have established that .NET C# programming language is simpler to comprehend so I have come to a decision to choose this programming language. Furthermore my programming background persuaded me in deciding for C#. Furthermore, there are numerous open sources for C#. Open sources is a free of charge library which support production and development. The accessibility of the product’s design and application is free of charge

The company of the Quadcopter system that I have created possesses an open source support for C# which is the producer company. As soon as we start to design, and develop our system this open source support, it is viable to call up open source solutions library using .Net C# solution explorer.

The system I have designed is 251 Megabyte. This includes OpenCV and EmguCV library.

4.1 The Application

It is essential to describe the application in this part in order to build up a sound methodology for testing the program. Consequently, in the following pages the author intends to give details concerning the application in addition to its actual functions. This will be accompanied by a clarification of the projected methodology with justification.



Figure 4.1: The Application.

4.2 How Will the Application Be Tested?

The application will be analyzed via various training procedures and sampling as described below. It is important to test this in order to make sure that precision and sensitivity in the

research is examined and achieved frequently. As shown in Figure 4.1, the application will be examined via three major processes, these are the Command list, Navdata and Command and Control button. The application will also incorporate a live camera view.

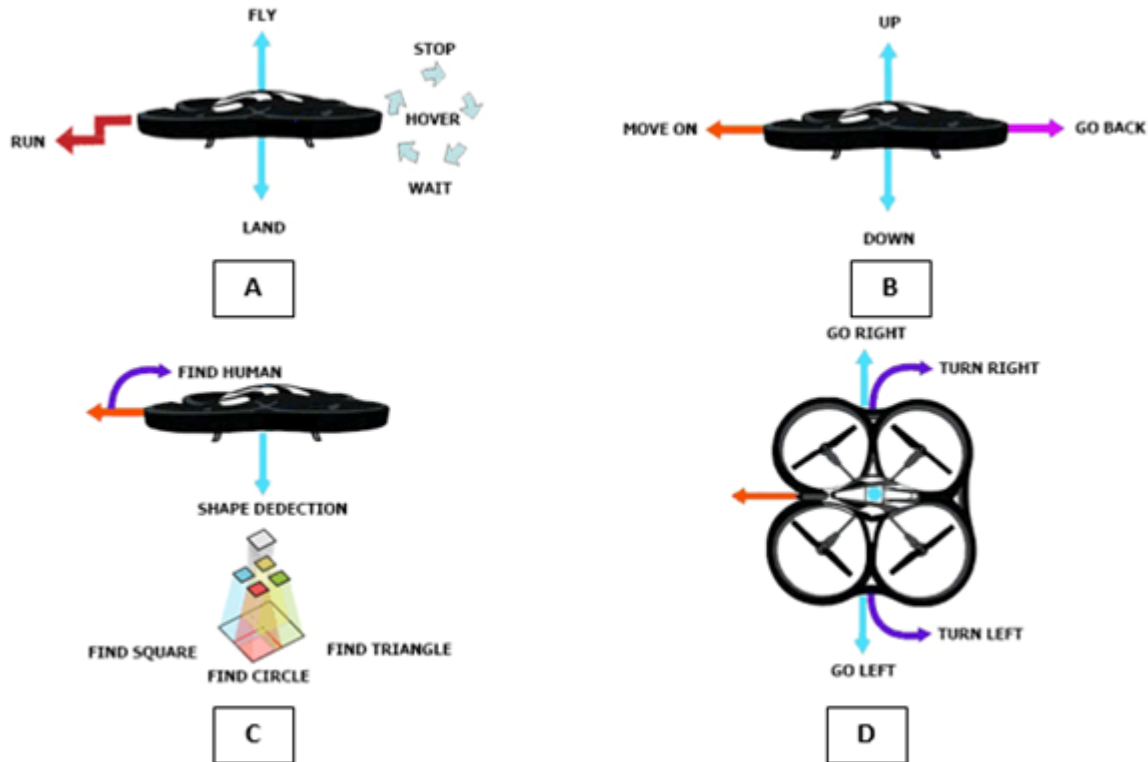


Figure 4.2: Quadcopter Commands.

Figure 4.2 shows how to activate the command list.

As a result the commands that are going be examined are listed below:

- A.** The system is executed with hover together with the wait and stop at the identical level, and also the same with fly and land
- B.** Direction allocation (i.e. move on, go back, up and down)
- C.** Shape/Human detection: when this command is set, the Quadcopter alters the bottom camera view and spots the command shape and so that it can find the human the front camera turns slightly right (i.e. find human, find square, find circle, find triangle).
- D.** Direction command, (i.e. go/ turn left and go/ turn right)

4.3 Flow chart of the system software

The performance of the system was determined by an experiment. The activated command was repeated twenty times in order to estimate the sensitivity; the amount of successful detections was also recorded. The command “RUN and FLY” was applied to estimate the specificity. This sounds similar to “activate”. The Run and FLY command was redone three times but this

time great number of wrong detections were listed. This test was completed for three variation intervals between the microphone and the speaker therefore there will be three various series of volume including: whisper, loud and conversation. Figure 4.3 presents the flowchart for entire process of the sound recognition.

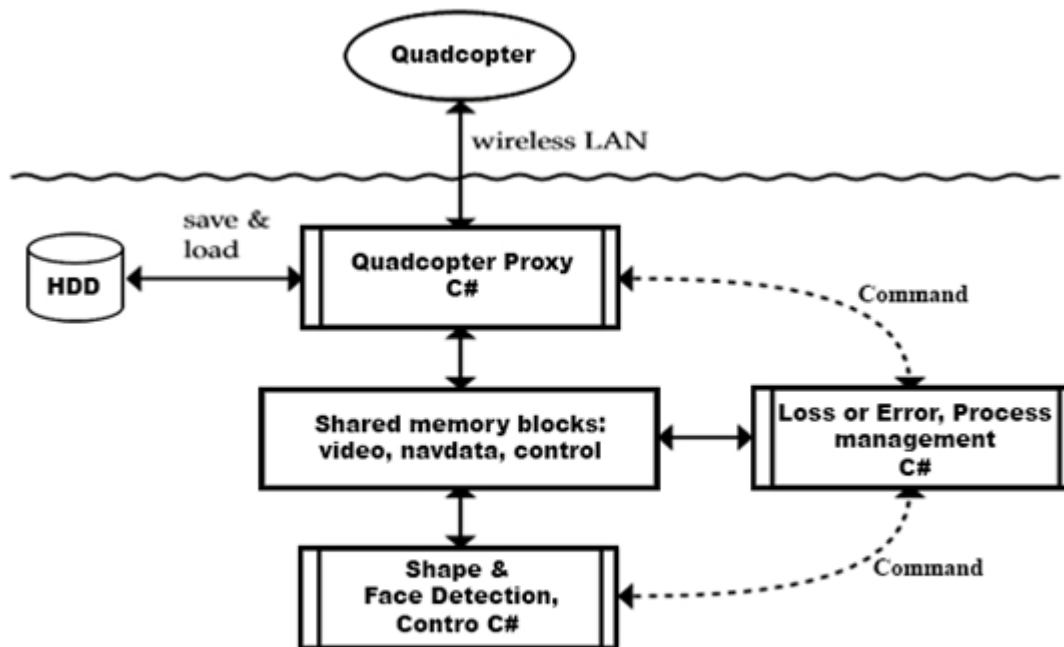


Figure 4.3: Flow chart of the system software.

Quadcopter DHCP as a server create a WIFI that has a default host name IP address 192.168.1.1 which allows one to create a link between a computer and Quadcopter. Once a connection is made among the Quadcopter and the computer we can execute the interface software prepared in C# programming language in order to allow us to process the data that will be sent from the Quadcopter to computer and vice versa. The data is documented on the Hard Disk Drive where it allows us to process. For example, interface software includes control information and identifies algorithms (navdata, the pace of the engine rotary motion, the path of the engine, the remaining battery power, EmguCV and OpenCV library, gas roll, yaw, roll and pitch).

In addition to navdata, Shared Memory Block and Loss or Error too sends data. Shared Memory Block has three functions. These functions are video, navdata and control. Shared Memory Block maintains the function of the computer it requires to itself and creates shared use. It evaluates OpenCV library by means of image processing, navdata and control data. OpenCV library achieves this via video images. It conveys and obtains these data via UDP NavdataPort = 5554, UDP Command Port = 5556 and UDP Video Port = 5555. This enables us to transmit data more rapidly between Quadcopter and the computer.

Loss or Error Process Management calculates approximately the percentages of the commands (RUN, FLY, HOVER, STOP) created at an earlier time by C# system interface

library. The percentages are calculated by making use of Microsoft Speech Recognition program, Hidden Markov Model Algorithms. Hidden Markov Model Algorithm gets hold of the most alike results and transmits the data to Quadcopter by making use of UDP Command port 5556 and the command is applied.

C# OpenCV library has Shape and Face Detection Control and Find Human Algorithm in it. Utilizing these, we depict lines / border around the circle, square, triangle and rectangle. Furthermore, when the Quadcopter recognizes a human being, it halts and creates a visual computer screen of the human being.

4.4 Obtained Result

Speakers correspond to a variety of tones, feelings and lengths in meters. Investigation of Sensitivity and Specificity sheds light on the relationship with other application and its users.

The talkers maintain interactions with the relevant software. Users are represented within the system and exist for specific methods in which the system may be used. The software operates in three modes of connections: tone, feeling and length. Users frequently maintain a documentation of common characteristics. Presuming one identifies a second usage: Similar research laboratories are functioning towards better developed designs and control techniques and researches globally. The leading institutions are MIT Aerospace Controls Lab, ETH Flying Machine Arena and Pennsylvania General Robotics, Automation, Sensing and Perception (GRASP).

My objective in designing this system was to allow people to gain the experience of free flight and not be dependent on a map or a mobile phone but merely utilizing a computer. Furthermore my objective was to achieve a distinctive result when contrasted with the other systems.

The system which I have created is distinct from the current systems. The major distinctions are, not does it only pick up the voice commands rapidly and accurately, but it is free to improvement and is an appealing system for many of individuals and many fields.

The followings are the commands library designed for my Quadcopter control system: ("Take Off", "Fly", "Run", "Stop", "System Deactivate", "Security", "Hover", "Land", "Up", "Down", "Go Left", "Go Right", "Go Back", "Move On", "Turn Left", "Turn Right", "Camera 2", "Camera 1", "Find Human", "Shape Detection", "Find Shape", "Find Circle", "Find Square", "Find Triangle", "Stop Shape detection", "Wait", "Cancel Detection").

The tests I carried out utilizing headsets, suggest that it is accurate to claim that 99% of the commands given by male participants were identified accurately and system acted to 93% of the commands made by female participants.

The results illustrate distinctions as soon as the participants have troubles with their voices (illness), once the participants have unusual pronunciation and as soon as the surroundings are noisy.

The command listed above has been selected particularly for accurate and simple understanding of the system. For instance initially the command to go forward was “Go Forward” however I came to the conclusion that the system fail to differentiate this command with the other commands that have the word “Go” so, as an alternative of “Go Forward” I decided to change the command to “Move On” and consequently the system functioned accurately.

5. Conclusions

In conclusion, numerous researchers are attempting to create Quadcopters due to their multi functionality. Nevertheless, I created a Quadcopter which operates with the use of voice recognition systems, even though voice recognition systems in the broad-spectrum can be viewed as software that converts text to speech data however we utilized this system to command the Quadcopter. Purpose of this paper was to institute asophisticated system that can operate and be commanded even in an unidentified environment using only onboard sensors, with no use of markers or calibration objects, but through voice recognition. My objective in creating this system was to enable people to command the Quadcopter by merely using a computer. The results that I attained based on the tests that I carried out utilizing headsets it is safe to say that 99% of the commands given by male participants were detected accurately and system responded the 93% of the commands which were given by female participants. The result obtained from this test demonstrates distinctions when the participants have difficulties with their voices (illness), once the participants have unusual pronunciation and once the surroundings are noisy. Aside from named limitations, the voice recognition system functions correctly with the given commands.

References

- Abbott, K.R. (2002). *Voice Enabling Web Applications: VoiceXML and Beyond*, Apress, Berkely, USA. <http://dx.doi.org/10.1007/978-1-4302-0850-1>
- Abshire, K.J. and Baron, M.K. (1998). Virtual Maintenance: Real-world Applications within Virtual Environments. *IEEE Proceedings Annual Reliability and Maintainability Symposium*, pp. 132-137. (1988).
- Ackley, D., Hinton, G., & Sejnowski, T. (1985). A Learning Algorithm for Boltzmann Machines. *Cognitive Science*, 9, 147-169. Reprinted in Anderson and Rosenfeld. http://dx.doi.org/10.1207/s15516709cog0901_7
- Allen, J. B. (1994). How do humans process and recognize speech? *IEEE Trans. On Speech and Audio Proc.*, 2(4), 567–577, Oct. <http://dx.doi.org/10.1109/89.326615>
- Amit, Y., Koloydenko, & Niyogi, P. (2005) Robust acoustic object detection. *Journal of the Acoustical Society of America*, 118, 2634–2648. and *Signal Processing*, 1990.
- Austin, S., Zavaliagkos, G., Makhoul, J., & Schwartz, R. (1992). Speech Recognition Using Segmental Neural Nets. In Proc. *IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Bahl, L., Bakis, R., Cohen, P., Cole, A., Jelinek, F., Lewis, B., & Mercer, R. (1981). Speech Recognition of a Natural Text Read as Isolated Words. In Proc. *IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Chen, W., Wang, S., Li, Z., & Feng, C. (2002). Study of Key Techniques Concerning Virtual Assembly System for Equipment Maintenance Training. *Third International Conference on Virtual Reality and its Application in Industry*, 4756, 254-258.

Cole, R. A., Stern R. M., & Lasry. M. J. (1986). *Performing fine phonetic distinctions: templates versus features*. Lawrence Erlbaum Assocs., Hillsdale, N.J

Conference on Pervasive Services 2006 Lyon, Rouillard, J., & Truillet, P. (2005). Enhanced VoiceXML, HCI International, Las Vegas, 2005.

Cooke., M. P. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America*, 119, 1562–1573. <http://dx.doi.org/10.1121/1.2166600>

Loftin, B.R. (1995). Virtual Environments for Aerospace Training, Proceedings of the 1995 IEEE Technical Applications Conference and Workshops, NORTHCON'95, pp. 31-34.

Mabbutt, D. (2013). What is Visual Basic? The "What, Who, When, Where, Why, and How" of VB, http://visualbasic.about.com/od/applications/a/whatisvb_2.htm, (Retrieved: 04/02/2012)

Markov. A (1913). An example of statistical investigation in the text of eugene onyegin, illustrating coupling of tests in chains. Proceedings of the Academy of Sciences of St. Petersburg, Microsoft Basic Logical Expression Evaluation". [Vb.mvps.org](http://vb.mvps.org). <http://vb.mvps.org/tips/Truth.asp>.

Jakob Julian Engel. (2011). Autonomous Camera-Based Navigation of a Quadcopter FAKULTÄT FÜR INFORMATIK DER TECHNISCHEN UNIVERSITÄT MÜNCHEN, Munich, December 15. <http://www.msh-tools.com>

Ozkartal, Gokhan. (2013). The Development of a Speech Recognition Software Application Using HMM, ANN and LPC Models to Minimize Error. *Journal of Management Research*. <http://dx.doi.org/10.5296/jmr.v5i3.3521>

SOAP: Simple Object Access Protocol. (2012). <http://www.w3.org/TR/soap> (Retrieved:02/02/)

SASDK. (2005). <http://www.microsoft.com/presspass/features/2004/mar04/03-24SpeechServer.asp>, (Retrieved, 14.11.2012)

Song, X. (2005). Comparing Microsoft Speech Server 2004 and IBM WebSphere Voice Server V4.2., <http://www.developer.com/voice/article.php/b3381851.html>.5. VoiceXML forum, (Retrieved: 19/03/2012)

Taylor, K. (2005). Configuring Resin as a Development Java Servlet Engine, http://java.about.com/library/weekly/aa_resin_intro2.htm, (Retrieved: 24 March 2011).

Taylor, P., & Isard, A. (1997). SSML: A speech synthesis markup language. *Speech Communication, 21*, 123–133.

Tebelskis, J., & Waibel, A. (1990). Large Vocabulary Recognition using Linked Telispeech 1.2 from Telisma, <http://www.telisma.com/> (Retrieved:08/02/2012)

The Business Value of Speech. Retrieved March 7, 2005, from <http://www.microsoft.com/speech/businessvalue/speech/default.aspx>, (Retrieved: 15/01/2012) UDDI, Universal Description, Discovery and Inte-gration protocol, <http://www.uddi.org>

Voice Technology Goes Mainstream with Release of Microsoft Speech Server. (2004). <http://www.microsoft.com/presspass/features/2004/mar04/03-4SpeechServer.asp>.