**ACKNOWLEDGEMENT**

This project has been prepared under **Department of Computer Engineering**, **Khwopa Engineering College** (***Purbanchal University***). We are very grateful to our department for giving us an enthusiastic support.

It gives us pleasure to write down here the name of **Er. Ishwor Gautam** lecturer, Khwopa Engineering College as our Supervisor for this project. We would like to thank him a lot for guiding properly throughout our project and continuously motivating to tackle the problem faced during this project. He has been always very much supportive and always motivating us to do best than what we are currently and helping to decide correctly.

Besides we are very grateful and thankful to all the respected teachers of Department of Computer, Khwopa Engineering College, for granting us broad knowledge and inspirations within the time period of four years.

Special thanks go to **Er. Amod Ulak, Mr. Jyoti Adhikari, Mr. Arjun Kumar Gaire, Mr. Phadendra Pd. Bhandari** and all colleagues who directly and indirectly extended their hands in making this project work a success.

There may be some errors in our project, so any suggestion regarding the mistake of this project will be always welcomed.

**Amar DUWAL**

**rAJ KAJI SHRESTHA**

**sANJAY RANJIT**

**SHIVA KUMAR SHRESTHA**

**ABSTRACT**

**Voice Command Matching System** (VCMS) is the system that matches the commands that the user speaks on the basis of information extracted from the speech waves. This system makes it possible to use the speaker’s voice to access computer commands such as opening a notepad, closing notepad, opening facebook, shutting down computer, etc.

There are two main states in this system which are recording state and operating state. In recording state, voice is inputted from the user of the system. Five samples of voice are inputted in this process. Feature extraction is done from inputted voice then the features are saved. Next, in operating state, voice is inputted from user and in the same way feature extraction is done as in recording state. The features are then compared with the saved features. If the features are matched with threshold, respective decision taken meanwhile otherwise the system gives chance for retry.

The features or voiceprint is created by extracting its 13 **Mel Frequency Cepstral Coefficient** (MFCC) per frame. For matching purpose, **Euclidean Distance** (ED) & correlation method is used. ED calculates the distance between two vectors which are the command spoken and the stored voiceprints.

**Keywords**: *VCMS*, *MFCC*, *Euclidian distance*

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE No.** | **FIGURE NAME** | **PAGE** |
| Figure 3.1 | Different components of Voice Matching Process & their interactions | 8 |
| Figure 3.1.3.1 | Waveform of recorded audio signal for command “Facebook Kholnuhos Mitra” | 9 |
| Figure 3.1.3.2 | Waveform of trimmed audio signal for command “Facebook Kholnuhos Mitra” | 9 |
| Figure 3.1.4 | Frame Overlapping | 10 |
| Figure 3.1.5 | MFCC Block Diagram | 11 |
| Figure 3.1.5.1 | Audio to spectrum conversion | 12 |
| Figure 3.5.1.3 | Mel-Scale Filter Bank | 13 |
| Figure 3.1.6 | Matching Process | 14 |
| Figure 3.1.7 | Decision Making | 15 |
| Figure 4.2 | Scrum, Agile development Process | 18 |
| Figure A1.1 | Work Breakdown Structure | 30 |
| Figure A2.1 | Use Case Diagram of VCMS | 31 |
| Figure A3.1 | State Diagram of VCMS | 32 |
| Figure A4.1 | VCMS Flowchart | 33 |
| Figure A5.1 | VCMS Application UI | 36 |
| Figure A5.2 | Menu | 36 |
| Figure A5.3 | Training UI | 37 |
| Figure A5.4 | Training UI for Specific Command | 37 |
| Figure A5.5 | Waveform Layout | 38 |
| Figure A5.6 | Audio Recorder | 38 |
| Figure A5.7 | MFCC Calculation Panel | 39 |
| Figure A5.8 | Generate Waveform | 39 |
| Figure A5.9 | Trimming waveform | 40 |
| Figure A5.10 | Memory boost by deleting temporary files | 40 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **TABLE No.** | **TABLE NAME** | **PAGE** |
| Table 5.2.1 | Test Case - Audio Record | 19 |
| Table 5.2.2 | Test Case - Silence Remove | 20 |
| Table 5.2.3 | Test Case - Waveform Generation & Play Back | 20 |
| Table 5.2.4 | Test Case - MFCC Computation | 20 |
| Table 5.2.5 | Test Case - Euclidian Distance & Correlation Computation | 21 |
| Table 5.3.1 | Matching result for command "Facebook KholnuhosMitra" | 21 |
| Table 5.3.2 | Matching result for command "Open Notepad" | 22 |
| Table 5.3.3 | Matching result for command "Shutdown" | 22 |
| Table 5.3.4 | Matching result for command "Open Wordpad" | 22 |
| Table 5.3.5 | Matching result for command "Close Notepad" | 22 |
| Table 5.3.6 | Matching result for command "Light On" | 23 |
| Table 5.4.1 | Summary of Matching Scenario | 23 |
| Table 5.5.1 | Test Environment - Calm Place(Closed Room at Midnight) | 23 |
| Table 5.5.2 | Test Environment - Normal Place(Closed Room at Morning) | 24 |
| Table 5.5.3 | Test Environment - College Premises (At Midday) | 24 |
| Table 5.5.4 | Test Environment - College's Noisy Environment (Operating Generator at Corner of College) | 24 |
| Table 5.5.5 | Test Environment - Noisy Environment (College Expo) | 24 |
| Table 5.6.1 | Test Voice - Male (User - Shiva K. Shrestha) | 25 |
| Table 5.6.2 | Test Voice - Male (User - Raj Kaji Shrestha) | 25 |
| Table 5.6.3 | Test Voice - Female (User - Shanta Shrestha) | 25 |
| Table 5.6.4 | Test Voice - Adult Male (User - Sandesh Danekhu) | 25 |
| Table 5.6.5 | Test Voice - Adult Female (User - Anita Duwal) | 26 |
| Table 5.7 | Summary of Test Voice | 26 |

**LIST OF ABBREVATIONS**

|  |  |
| --- | --- |
| 1-D | One Dimension |
| ANN | Artificial Neural Network |
| ASR | Automatic Speech Recognition |
| AT & T | American Telephone and Telegraph |
| Cmd | Command |
| DCT | Discrete Cosine Transform |
| DFT | Discrete Fourier Transform |
| DTW | Dynamic Time Warping |
| DWT | Discrete Wavelet Transform |
| ED | Euclidian Distance |
| FFT | Fast Fourier Transform |
| GMM | Gaussian Mixture Model |
| GUI | Graphical user interface |
| Hz | Hertz |
| IBM | International Business Machine |
| IDE | [Integrated Development Environment](http://en.wikipedia.org/wiki/Integrated_development_environment) |
| KHz | Kilo Hertz |
| LPC | Linear Predictive Coding |
| MFC | Mel-frequency Cepstral |
| MFCC | Mel-frequency Cepstral Coefficient |
| PC | Personal Computer |
| PCM | Pulse Code Modulation |
| STT | Speech-to-Text |
| TTS | Text-to-Speech |
| UI | User Interface |
| VCMS | Voice Command Matching System |

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CPT** | **TITLES** | **PAGE** |
|  | **BONAFIDE CERTIFICATE** | I |
|  | **ACKNOWLEDGEMENT** | II |
|  | **ABSTRACT** | III |
|  | **LIST OF FIGURES** | IV |
|  | **LIST OF TABLES** | V |
|  | **LIST OF ABBREVIATION** | VI |
|  |  |  |
| **1** | **INTRODUCTION** |  |
|  | 1.1 BACKGROUND | 1 |
|  | 1.1.1 VOICE RECOGNITION | 1 |
|  | 1.1.2 TEXT INDEPENDENT SYSTEM | 2 |
|  | 1.1.3 VOICE COMMAND SYSTEM | 2 |
|  | 1.2 MOTIVATION | 2 |
|  | 1.3 STATEMENT OF PROBLEMS | 3 |
|  | 1.4 OBJECTIVES | 3 |
|  | 1.5 APPLICATION | 3 |
|  | 1.6 SCOPE & LIMITATION | 4 |
|  | 1.7 REPORT STRUCTURE | 4 |
|  |  |  |
| **2** | **LITERATURE REVIEW** | 5 |
|  |  |  |
| **3** | **METHODOLOGY** |  |
|  | 3.1 VOICE COMMAND MATCHING PROCESS | 8 |
|  | 3.1.1 INPUT VOICE | 9 |
|  | 3.1.2 ANALOG TO DIGITAL CONVERSION | 9 |
|  | 3.1.3 VOICE DETECT | 9 |
|  | 3.1.4 FRAMING AND OVERLAPPING | 10 |
|  | 3.1.5 FEATURE EXTRACTION | 10 |
|  | 3.1.5.1 INPUT AUDIO SPECTRUM | 12 |
|  | 3.1.5.2 DISCRETE FOURIER TRANSFORM | 12 |
|  | 3.1.5.3 MEL FILTER BANK PROCESSING | 12 |
|  | 3.1.5.4 DISCRETE COSINE TRANSFORM | 13 |
|  | 3.1.5.5 OUTPUT 13 DCT COEFFICIENT | 14 |
|  | 3.1.6 MATCHING | 14 |
|  | 3.1.6.1 EUCLIDEAN DISTANCE | 15 |
|  | 3.1.6.2 CORRELATION | 15 |
|  | 3.1.7 TAKE DECISION | 15 |
|  | 3.2 TOOLS & PLATFORM | 16 |
|  | 3.3 SYSTEM REQUIREMENTS | 16 |
|  |  |  |
| **4** | **PROJECT DEVELOPMENT LIFE CYCLE** |  |
|  | 4.1 AGILE SOFTWARE DEVELOPMENT | 17 |
|  | 4.2 SCRUM AS AGILE DEVELOPMENT | 17 |
|  |  |  |
| **5** | **TESTING AND RESULT ANALYSIS** |  |
|  | 5.1 OVERVIEW | 19 |
|  | 5.2 TEST CASES | 19 |
|  | 5.3 MATCHING SCENARIO | 21 |
|  | 5.4 SUMMARY OF MATCHING SCENARIO | 23 |
|  | 5.5 TEST ENVIRONMENT | 23 |
|  | 5.6 TEST VOICE | 25 |
|  | 5.7 SUMMARY OF TEST VOICE | 26 |
|  | 5.8 DISCUSSION AND RESULT ANALYSIS | 26 |
|  |  |  |
| **8** | **CONLUSION** |  |
|  | 6.1 CONCLUSION | 28 |
|  | 6.2 FUTURE ENHANCEMENT | 28 |
|  |  |  |
|  | **REFERENCES** | 29 |
|  | **APPENDIX 1: PROJECT MANAGEMENT** | 30 |
|  | **APPENDIX 2: USE CASE DIAGRAM** | 31 |
|  | **APPENDIX 3: STATE DIAGRAM** | 32 |
|  | **APPENDIX 4: FLOWCHART** | 33 |
|  | **APPENDIX 5: NAUDIO FRAMEWORK** | 34 |
|  | **APPENDIX 6: SCREENSHOTS** | 36 |