Electronic Control System Of Home Appliances Using Speech Command Words

Aye Min Soe, Maung Maung Latt, Hla Myo Tun, Zaw Min Naing

Abstract: The main idea of this paper is to develop a speech recognition system. By using this system, smart home appliances are controlled by spoken words. The spoken words chosen for recognition are “Fan On”, “Fan Off”, “Light On”, “Light Off”, “TV On” and “TV Off”. The input of the system takes speech signals to control home appliances. The proposed system has two main parts: speech recognition and smart home appliances electronic control system. Speech recognition is implemented in MATLAB environment. In this process, it contains two main modules: feature extraction and feature matching. Mel Frequency Cepstral Coefficients (MFCC) is used for feature extraction. Vector Quantization (VQ) approach using clustering algorithm is applied for feature matching. In electrical home appliances control system, RF module is used to carry command signal from PC to microcontroller wirelessly. Microcontroller is connected to a driver circuit for relay and motor. The input commands are recognized very well. The system is a good performance to control home appliances by spoken words.

Index Terms: MATLAB Speech recognition, Feature Extraction, Feature Matching, Mel Frequency Cepstral Coefficient (MFCC), Vector Quantization (VQ), PIC16F887, KST-TX01, KST-RX706

I. INTRODUCTION
Speech Recognition is a technology allowing the computer to identify and understand words spoken by a person using a microphone or telephone. Using a set of pre-programmed commands and instructions, the computer system that understands input speech enables user to have conversations with the computer. User and the computer speaking as commands or in response to events, input, or other feedback would be included in these conversations. Speaking is easier and more sensitive than selecting buttons and menu items. Human speech has changed over many thousands of years to become an efficient method of sharing information and giving instructions. [1] The goal of this paper is to implement smart home appliances controlled system that can be operated by spoken words. The spoken words chosen for recognition are “Fan On”, “Fan Off”, “Light On”, “Light Off”, “TV On” and “TV Off”. The proposed system is composed of two main parts: speech recognition system and home appliances control system. The first and main part of system is personal computer based speech recognition system. Speech recognition system is composed of also two main parts. The first part is speech recognition based on Digital Signal Processing. Second part is interfacing with hardware. The overall block diagram is shown in Fig. 1.

Fig. 1 Overall System Proposed System

II. SPEECH RECOGNITION SYSTEM
A speech recognition roughly consists of two portions. They are speech analysis and pattern recognition.

A. Speech Analysis
The purpose of the speech analysis block is to transform the speech waveform into a parsimonious representation which characterizes the time varying properties of the speech. The speech analysis typically includes two modules, namely data acquisition and feature extraction. The data acquisition module usually contains a microphone and a code from which digitized speech data are generated. The feature extraction is the computation of a sequence of feature vectors which provides a compact representation of the given speech signal. The feature extraction is done on short-time basis. The speech signal is separated into overlapped fixed-length frames. From each frame, a set of frequency-domain or cepstral-domain parameters are derived from each frame, to form the so-called feature vector. There are some basic principles and analysis techniques used in the feature extraction module. They are pre-emphasis, frame blocking and windowing. Discrete Fourier Transform (DFT) computation, spectral magnitudes, Mel-frequency filter bank, logarithm of filter energies, Discrete Cosine Transformation (DCT), Cepstral Weighting, and dynamic featuring.

B. Pattern Recognition
The speech signal is first analyzed and a feature representation is obtained for comparison with either stored reference templates or statistical models in the pattern matching block. A decision scheme determines the word or phonetic class of the
unknown speech based on the matching scores with respect to
the stored reference patterns. There are two types of reference
patterns. The first type, called a nonparametric reference
pattern (or often a template), is a pattern created from one or
more spoken tokens of the sound associated with the pattern.
The second type, called a statistical reference model, is created
as a statistical characterization of the behavior of a collection
of tokens of the sound associated with the pattern. The vector
quantization model [6] is used as the statistical model. There
are three portions in pattern recognition. They are pattern
training, pattern matching and maximum selection.

1. Pattern Training
Pattern training is the method by which representative sound
patterns are converted into reference patterns for use by the
pattern matching algorithm. There are several ways in which
pattern training can be performed, including: Casual training in
which a single sound pattern is used directly to create either a
template or a crude statistical model. Robust training in which
several versions of the sound pattern are used to create a
single merged template or statistical model. Clustering training
in which a large number of versions of the sound pattern is used
to create one or more templates or a reliable statistical model of
the sound pattern.

2. Pattern Matching
Pattern matching refers to the process of assessing the
similarity between two speech patterns, one of which
represents the unknown speech and one of which represents
the reference pattern (derived from the training process) of
each element that can be recognized. When the reference
pattern is a “typical” utterance template, pattern matching
produces a gross similarity (or dissimilarity) score. When the
reference pattern consists of a probabilistic model, the process
of pattern matching is equivalent to using the statistical
knowledge contained in the probabilistic model to assess the
likelihood of the speech (which led to the model) being realized
as the unknown pattern. Pattern matching refers to the process
of assessing the similarity between two speech patterns, one of
which represents the unknown speech and one of which
represents the reference pattern (derived from the training
process) of each element that can be recognized.

3. Maximum Selection
The decision strategy takes all the matching scores (from the
unknown pattern to each of the stored reference patterns) into
account, finds the “closest” match, and decides if the quality of
the match is good enough to make a recognition decision. If
not, the user is asked to provide another token of the speech
(e.g., the word or phrase) for another recognition attempt. This
is necessary because often the user may speak words that are
incorrect in some sense (e.g., hesitation, incorrectly spoken
word, etc.) or simply outside of the vocabulary of the
recognition system.

C. Vector Quantization (VQ)
Vector Quantization (VQ) is a classical quantization technique
from signal processing which allows the modeling of probability
density functions by the distribution of prototype vectors. It was
originally used for data compression. It works by dividing a
large set of points (vectors) into groups having approximately
the same number of points closest to them. Each group is
represented by its centroid point, as in k-means and some other
clustering algorithms. The density matching property of vector
quantization is powerful, especially for identifying the density of
large and high-dimensional data. Since data points are
represented by the index of their closest centroid, commonly
occurring data have low error, and rare data high error. This is
why VQ is suitable for lossy data compression. It can also be
used for lossy data correction and density estimation. Vector
quantization is based on the competitive learning paradigm, so
it is closely related to the self-organizing map model. VQ was
also used in the eighties for speech and speaker recognition.
Recently it has also been used for efficient nearest neighbour
search and on-line signature recognition. In pattern
recognition applications, one codebook is constructed for each
class (each class being a user in biometric applications) using
acoustic vectors of this user. In the testing phase the
quantization distortion of a testing signal is worked out with the
whole set of codebooks obtained in the training phase. The
codebook that provides the smallest vector quantization
distortion indicates the identified user. Vector Quantization (VQ)
approach is applied for training and classification phase. Firstly,
the training set of vectors is used to create the optimal set of
codebook vectors for representing the spectral variability
observed in the training set. And Then distance is measured
between a pair of spectral analysis vectors to able to cluster the
training set vectors as well as to classify spectral vectors into
unique codebook entries. The next step is a centroid
computation procedure. Finally, a classification procedure
selects the codebook vectors that closest to the input vector and
uses the codebook index as the resulting spectral
representation. The classification procedure is essentially a
quantizer. It accepts speech spectral vectors as input and
provides the code index of the code vectors that best matches
the input.

Fig.3 Block Diagram of the VQ Training and Classification
Phase

III. ELECTRONIC CONTROL SYSTEM
The block diagram of home appliances control system for
speech recognition is illustrated in Fig.1. In this diagram,
speech instruction is firstly taken as input to control home
appliances and then a microphone is used to record the person
speech. Secondly, the speech instruction is caught and
transferred the analog signal to digital signal and the recorded
speech is sent to the speech based verification/identification
system. Thirdly, the digital information of speech instruction is
processed and compared by using the MATLAB programming.
Fourthly, the digital information of speech instruction is
outputted through USB port. Finally, PIC receives data from speech recognition block and gives instructions to control home appliances. The control system consists of two sections: transmission section and receiving sections.

**A. Transmission Section**

In the transmission section, there are KS232 module, PIC 16F887 and KST-TX01 (Radio Frequency transmitter module). The KS232 module is used to carry the signal from PC to Microcontroller unit. The signal is retransmitted with baud rate 1200 for RF transmission by KST-TX01 module. This module has four pins: supply pin, data pin, GRN pin, and ANT pin. KST-TX01 technical specific data for wireless transmitter module are (1) transmit power: 1W, (2) operating frequency: 315MHz~433.92MHz, (3) operating temperature: -40℃~80℃, (4) operating voltage: 3V~5V and (5) modulation type: ASK.

**B. Receiver Section**

The receiver section consists of KST-RX706 (RF receiver module), PIC microcontroller, relays, relay drivers and motor driver. In this section, KST-RX706 firstly accept radio signal and then microcontroller read radio signal with baud rate 1200. Microcontroller drives relay and motor driver. The speed of motor is controlled by using Pulse Wide Modulation (PWM) module.

**IV. SOFTWARE IMPLEMENTATION**

This section explains the methods used for speech recognition. These methods are training phase and testing phase.

**A. Training Phase**

Initially, the user must prepare the training files. The speech files are recorded from the microphone and MFCC features are extracted from the input file. Then these features are stored. In this case, the collection of training files is called database. Then, the user must train the system using the files in the database. This is called training phase or pre-processing.

Figure 5 shows the flow chart of the step of training phase. Then, the user must train the system using the files in the database. Initially, the user must prepare the training files. The speech files are recorded from the microphone and MFCC features are extracted from the input file. Then these features are stored. In this case, the collection of training files is called database. Then, the user must train the system using the files in the database. This is called training phase or pre-processing.

**Fig.4 Circuit Diagram of Transmission Section**

**Fig.5 Circuit Diagram of Receiving Section**

**Fig.6 Block Diagram of Home Appliances Control System for Speech Recognition**

**Fig.7 PWM Singal**
B. Testing Phase
In the testing phase, users have to provide the command words as input. In this case, user may use two ways of testing. If user chooses to use the pre-recorded sound file, one of the samples are loaded from test files and read. Then, the modified MFCC features are extracted from the input file. In the next step, the distances between the modified MFCC features and the stored reference models are calculated using Euclidean Distance. Finally, the minimum distance is selected among the distances between the input vectors and codebook vectors. If this minimum distance falls below the local threshold, the system outputs the command word as result. Otherwise, the system determines it is wrong command word. If the user wants to test the system with spoken commands in real time, the sound file to be recognized is recorded from the microphone. To do so, the user must choose time length Typical time length is 2 seconds. In this system, sound files are recorded within this time length. Then the subsequence processes, as above, are carried out and recognition decision.
offline, speech signal is loaded from testing files. And then steps in offline are same steps in real time.

This table includes command words, minimum distance between input vector and code book vector, maximum distance between input vector and code book vector, and execution time.

<table>
<thead>
<tr>
<th>Command words</th>
<th>Distance between input vector and code book vector (Minimum)</th>
<th>Distance between input vector and code book vector (Maximum)</th>
<th>Execution time</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV on</td>
<td>0.0106</td>
<td>1.80</td>
<td>0.44424</td>
</tr>
<tr>
<td>TV off</td>
<td>0.0873</td>
<td>2.5706</td>
<td>0.05861</td>
</tr>
<tr>
<td>Light on</td>
<td>0.0113</td>
<td>2.2317</td>
<td>0.05509</td>
</tr>
<tr>
<td>Light off</td>
<td>0.0960</td>
<td>1.5808</td>
<td>0.070341</td>
</tr>
<tr>
<td>Fan on</td>
<td>0.0144</td>
<td>1.6128</td>
<td>0.063073</td>
</tr>
<tr>
<td>Fan off</td>
<td>0.1764</td>
<td>3.7751</td>
<td>0.061763</td>
</tr>
</tbody>
</table>

V. RESULT AND DISCUSSION

When the user speaks “TV ON” command from microphone, the command is recognized by personal computer using MATLAB software. The recognized command word is input into microcontroller through USB port. In microcontroller, “TV ON” command word is assigned as character “a”. When character “a” is transmitted to one microcontroller, another microcontroller receives the character “a”. And then the signal is driven relay driver to turn on TV. Fig. 9 shows the result of “Light On” command word. “Light On” command word is given as the character “c” in microcontroller and “Fan On” command word as the character “e”. Before testing the real time communication between transmission and receiving section, PIC to PIC serial communication is tested firstly. Figure 6 shows simulation of PIC to PIC wire communication for “Fan on” command. In this figure, it consists of two separate part circuit. One is for reading data from PC and these data would then be stored in a user file. In data communication, it is needed to transmit a start bit for transmitted bits in user file.

When the user speaks a command that does not include in trained commands, the massage box shows “undefined”.

Fig. 8 Speech Command Tests for Electronic Home Appliances
After transmitting start bit, the data in user file will be serially transmitted to another PIC. This PIC receives the transmitted data and the output will go on if the receiving data is equal to the original data. If they are not equal, there is no output condition and the user must check until there is no error in program for the correct result.

**VI. CONCLUSION**

The system has two main parts: speech recognition and smart home appliances electronic control system. Speech recognition is implemented in MATLAB environment. In this process, it contains two main modules: feature extraction and feature matching. Mel Frequency Cepstral Coefficients (MFCC) is used for feature extraction. Vector Quantization (VQ) approach using clustering algorithm is applied for feature matching. In electrical home appliances control system, RF module is used to carry command signal from PC to microcontroller wirelessly. Microcontroller is connected to driver circuit for relay and motor. The input commands are
recognized very well. The system is a good performance to control smart home appliances by spoken words. A wireless based home automation system which can be controlled through spoken commands is presented in this paper. In the proposed system, wireless component is added by Radio Frequency (RF) module. An application for speech command processing is developed.

ACKNOWLEDGEMENTS
I would like to express a great debt of gratitude to Dr. Zaw Min Naing, Deputy Director General, Myanmar Science and Technology Research Department, for his patient guidance, supervision, suggestions and encouragement during a long period of this study. I would like especially thank to my supervisor, Dr. Maung Maung Latt, Rector, Department of Electronic Engineering, Technological University (Taungoo), my co-supervisor, Dr. Hla Myo Tun, Associate Professor and Head, and Dr. Su Su Yi Mon, Associate Professor, Department of Electronic Engineering, Mandalay Technological University, for their useful guidance, patience and giving valuable ideas.

REFERENCES