Implementation of Speech Recognition on MCS51 Microcontroller for Controlling Wheelchair

Thiang
Electrical Engineering Department, Petra Christian University
Siwalankerto 121-131, Surabaya, Indonesia
Email: thiang@petra.ac.id, phone: +62-31-2983115

Abstract—This paper describes about implementation of speech recognition on microcontroller. The microcontroller used in this system is ATMEL AT89C51RC microcontroller which is one of the MCS51 family microcontrollers. Speech recognition system is implemented to recognize the word used as the command for controlling movement of a wheelchair.

There are two approaches used to recognize the speech signal. The first approach is Linear Predictive Coding combined with Euclidean Squared Distance. LPC is used as the feature extraction method and Euclidean squared Distance is used as the recognition method. The second approach is Hidden Markov Model, which is used to build reference model of the words and also used as the recognition method. Feature extraction method used in the second approach is a simple segmentation and centroid value. Both approaches work on time domain.

Two DC motors are used as the actuator for driving the wheelchair. Both DC motors are controlled by ATMEL AT89C52 microcontroller and using a simple open loop control system.

Experiments were done to analyze performance of both approaches. Each approach has advantages and disadvantages. The highest average recognition rate that can be achieved using LPC-Euclidean Squared Distance approach was 78.57%. The highest average recognition rate that can be achieved using HMM-Segmentation and Centroid approach was only 32.86%.

Keywords—linear predictive coding, euclidean squared distance, hidden markov model, ATMEL AT89C51RC

I. INTRODUCTION

Automatic speech recognition by machine has been a goal of research for more than four decades. However, in spite of the glamour of designing an intelligent machine that can recognize the spoken word and comprehend its meaning, and in spite of enormous research efforts spent in trying to create such a machine, it is far from achieving the desired goal of a machine that can understand spoken discourse on any subject by all speakers in all environments [1].

The speech recognition system has also been implemented on some particular devices. Some of them are personal computer (PC), digital signal processor, and another kind of single chip integrated circuit. A framework to address the quantization issues which arise in fixed-point isolated word recognition was introduced in [2]. The system was developed using C++ language which is implemented on a PC. Reference [3] describes a speech recognition system using SPHINX-II, an off-the-shelf speech recognition package [4]. In reference [5] and [6], speech recognition system has been tried to be implemented on a FPGA and an ASIC.

This paper introduces the speech recognition which was implemented on a microcontroller. The microcontroller where the speech recognition was implemented on is ATMEL AT89C51RC. This microcontroller is a MCS51 family microcontroller and this microcontroller was chosen because it is popular in Indonesia.

The speech recognition system that is implemented on the microcontroller is used to recognize the word in a speech signal. The words are used as the command for controlling movement of the wheelchair. Therefore, the system was designed to recognize limited number of the words. This is also caused by the limit number of data memory of the microcontroller. There are only seven words used as the command for controlling movement of the wheelchair. They are stop, forward, backward, left, right, up, and down which is used to stop the wheelchair, to move forward the wheelchair, to move backward the wheelchair, to turn left the wheelchair, to turn right the wheelchair, to increase speed of the wheelchair, and to decrease speed of the wheelchair respectively.

Two approaches were implemented to perform the speech recognition. The first approach is Linear Predictive Coding (LPC) and Euclidean Squared Distance (ESD). LPC is used as the feature extraction method and ESD is used as the recognition method. This approach is based on the pattern recognition approach. The second approach applied in this system is Hidden Markov Model (HMM), which is one of the speech recognition approaches that classified as the statistical pattern recognition. HMM is used as the recognition method. As the feature extraction method, a simple segmentation and centroid value is applied.

Section 3 and 4 of this paper describe about this two approaches more detail. The mechanism and hardware design of the wheelchair is explained in the section 2 of this paper. Section 5 presents the experimental results that have been done and the last section is conclusion.

II. MECHANISM AND HARDWARE DESIGN OF WHEELCHAIR

A. Mechanism of Wheelchair

Figure 1 shows the front view and side view of mechanism of wheelchair.
The designed wheelchair has the following specifications:

- Dimension of the wheelchair is 60 cm x 78 cm x 110 cm.
- The wheelchair has four wheels: two pivot front wheels which are move freely in rotation and straight direction, two rear wheels which are actuated by two DC motor with gearbox. One DC motor drives one rear wheel.
- Diameter of front wheel is 10 cm and diameter of rear wheel is 22 cm.
- Specification of the DC motor is 20 V, 2 A, and 200 rpm.
- Maximum linear speed of the wheelchair is about 0.461 m/s or about 1.66 km/hr.

**B. Hardware Design of Wheelchair**

Hardware of the system consists of three main parts. The first part is DC motor control circuit. The circuit consists of controller, diver, and DC motor speed sensor circuit. In this part, an ATMEL AT89C52 microcontroller is used as the controller. The second part is microcontroller minimum system which performs the speech recognition process and microphone interface. In the second part, an ATMEL AT89C51RC microcontroller is used as the speech recognition processor. The third part is interface circuit. This circuit performs the communication between DC motor controller and speech recognition processor. This circuit also read the input command from a set of keypads. An ATMEL AT89C2051 microcontroller is used as the interface in this part. Figure 2 shows block diagram of hardware system.

The speech recognition system was implemented on ATMEL AT89C51RC microcontroller which runs at frequency of 24 MHz and has 32k bytes program memory. With a 24 MHz clock, the fastest time to execute an instruction by the microcontroller is about 0.5 microseconds.

The minimum system of AT89C51RC was designed with 256k bytes external Random Access Memory (RAM). Figure 3 shows the circuit diagram of AT89C51RC minimum system.

ADC0820 was used to convert analog signal of speech to digital signal (see figure 3). This ADC offers a conversion time of 1.5 µs, which is enough for 8 kHz sampling rate. A MD110 microphone from Philips was used as the voice sensor, which converts the voice to electric signal. An amplifier and filter circuit was used to amplify and filter the output signal of microphone, and then, the signal is converted to digital signal by ADC. A 40 dB/Dec high pass filter with cutoff frequency of 20 Hz was implemented to filter the signal.

**III. SPEECH RECOGNITION USING LINEAR PREDICTIVE CODING AND EUCLIDEAN SQUARED DISTANCE**

The first approach of speech recognition implemented on microcontroller is Linear Predictive Coding (LPC), which is combined with Euclidean Squared Distance (ESD) method. LPC is used as the feature extraction method and Euclidean Squared Distance is used as the recognition method. Block
diagram of LPC and ESD training and recognizer system are shown at figure 4 and 5 respectively

![Diagram of Training System Using LPC](image)

**Fig 4. Block Diagram of Training System Using LPC**

In training system, training data are sampled directly from microphone. Then, each training sample is processed using LPC processor algorithm and the result of this process is a set of cepstral coefficients of the speech signal. These cepstral coefficients are used as the reference model. A simple algorithm was implemented to detect the existence of the speech signal. The system reads four consecutive sampling data and then calculates the average of those four data. If the average value is less than a limit value, it means there is no speech signal. If the average value is greater than or equal to that limit value, it means there is a speech signal and then the microcontroller will start to read and record the signal in 0.5 seconds.

The basic steps in the processing of LPC processor include the following[1][7]:

1. **Preemphasis:** The digitized speech signal, \( s(n) \), is put through a low order digital system, to spectrally flatten the signal and to make it less susceptible to finite precision effects later in the signal processing. The output of the preemphasizer network, \( \tilde{s}(n) \), is related to the input to the network, \( s(n) \), by difference equation:

\[
\tilde{s}(n) = s(n) - \hat{a}s(n - 1)
\]

The most common value for \( \hat{a} \) is around 0.95.

2. **Frame Blocking:** The output of preemphasis step, \( \tilde{s}(n) \), is blocked into frames of \( N \) samples, with adjacent frames being separated by \( M \) samples.

If \( x_i(n) \) is the \( i^{th} \) frame of speech, and there are \( L \) frames within entire speech signal, then

\[
x_i(n) = \tilde{x}(ML + n) \quad n = 0, 1, \ldots, N - 1 \quad i = 0, 1, \ldots, L - 1 \tag{2}
\]

3. **Windowing:** After frame blocking, the next step is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. If we define the window as \( w(n) \), \( 0 \leq n \leq N - 1 \), then the result of windowing is the signal:

\[
x_i(n) = x_i(n)w(n) \quad 0 \leq n \leq N - 1 \tag{3}
\]

Typical window is the Hamming window, which has the form

\[
w(n) = 0.54 - 0.46 \cos \left( \frac{2\pi n}{N-1} \right) \quad 0 \leq n \leq N - 1 \tag{4}
\]

4. **Autocorrelation Analysis:** The next step is to auto correlate each frame of windowed signal in order to give

\[
r_i(m) = \sum_{n=0}^{N-1} \tilde{x}_i(n)\tilde{x}_i(n + m) \quad m = 0, 1, \ldots, p \tag{5}
\]

where the highest autocorrelation value, \( p \), is the order of the LPC analysis.

5. **LPC Analysis:** The next processing step is the LPC analysis, which converts each frame of \( p + 1 \) autocorrelations into LPC parameter set by using Durbin’s method. This can formally be given as the following algorithm:

\[
E^{(0)} = r(0)
\]

\[
k_i = \frac{r(i) - \sum_{j=1}^{i-1} a_i^{(j)} r(i-j)}{E^{(i-1)}} \quad 1 \leq i \leq p
\]

\[
a_i^{(i)} = k_i
\]

\[
a_j^{(i)} = a_j^{(i-1)} - k_ia_i^{(i-j)} \quad 1 \leq j \leq i - 1
\]

\[
E^{(i)} = (1 - k_i^2)E^{(i-1)}
\]

By solving the equation 6 to 10 recursively for \( i = 1, 2, \ldots, p \), the LPC coefficient, \( a_m \), is given as

\[
a_m = \alpha_m^{(p)}
\]

6. **LPC Parameter Conversion to Cepstral Coefficients:** LPC cepstral coefficients, is a very important LPC parameter set, which can be derived directly from the LPC coefficient set. The recursion used is

\[
c_m = a_m + \sum_{k=1}^{m-1} \left( \frac{k}{m} \right) c_k \cdot a_{m-k} \quad 1 \leq m \leq p
\]

\[
c_m = \sum_{k=m-p}^{m-1} \left( \frac{k}{m} \right) c_k \cdot a_{m-k} \quad m > p
\]

\[\]
B. Recognizer System Using LPC and ESD

Firstly, an unknown speech signal will be processed by using the LPC processor too. The result of this process is cepstral coefficients of the unknown speech signal. Then, calculation of Euclidean Squared Distance between cepstral coefficients of the unknown speech signal and cepstral coefficients of the reference model is performed. Calculation of Euclidean Squared Distance is done for each reference model by using equation[8][9]:

\[
ESD = \sum_{i=1}^{n} (p_i - q_i)^2
\]

where \( ESD \) is the squared distance between two points, \( P = (p_1, p_2, \ldots, p_n) \) and \( Q = (q_1, q_2, \ldots, q_n) \).

The unknown speech signal will be recognized as the reference model which has the minimum distance to the unknown speech signal.

This approach has been implemented on AT89C51RC microcontroller using assembly language. Some limitations and specifications are applied in the system, i.e.: the maximum time duration of a speech signal is 0.5 second and sampling rate of the signal is 8 kHz, maximum number of reference model that can be stored in the memory is 38 models, maximum LPC order is 16 and maximum dimension of LPC cepstral vector is also 16, and maximum number of data per frame in frame blocking process (part of LPC processor) is 255 and distance between two adjacent frames is less than 255 and greater than or equal to 80.

All the limitations are determined because of the limit memory size that is provided in the AT89C51RC minimum system and also because AT89C51RC is 8-bit microcontroller.

IV. SPEECH RECOGNITION USING HIDDEN MARKOV MODEL

The second approach of speech recognition implemented on microcontroller is Hidden Markov Model (HMM) which is used as the recognition method. In this approach, LPC processor is not used as the feature extraction method because calculation of LPC processor takes much time (about 19 seconds) when it was implemented on AT89C51RC. Therefore, instead of the LPC processor, a simple feature extraction algorithm, segmentation and centroid, was implemented to reduce the calculation time.

A. Design of HMM Structure

Type of HMM that was implemented in this speech recognition system is a left-right model which is also known as Bakis model. This model has the property that, as time increases, the state index increases or stays in the same state. But, in this speech recognition system, the HMM model is designed in such that the state index always increases as the time increases and never stays in the same state. And also, the state index always increases by one to the next state index as the time increases. Figure 6 shows the structure of the HMM that is used in the designed speech recognition system.

By designing the HMM structure like this, the state transition probability distribution is fixed as:

\[
 \alpha_{ij} = \begin{cases} 
 1 & \text{for } i = j = 1 \\
 0 & \text{others} 
\end{cases} 
\]

(16)

The initial state probability distribution is fixed because the initial state is always state 1. The state never starts with other states. So, the initial state probability distribution of this designed HMM structure is:

\[
 \pi_i = \begin{cases} 
 1 & \text{for } i = 1 \\
 0 & \text{others} 
\end{cases} 
\]

(17)

The state of HMM is associated with a segment of the time interval of the speech signal in time domain. The observation symbol of HMM is associated with amplitude of the speech signal.

B. Training System Using HMM

Block diagram of HMM training system is shown at figure 7. The main task of the HMM training system is to create HMM \( \lambda \) model of each word sample. Parameters of HMM \( \lambda \) model consist of number of state (N), number of observation symbol (M), state transition probability distribution (A), observation symbol probability distribution (B), and initial state probability distribution (\( \pi \)). By using HMM structure shown in the figure 6, all the parameters are known except the observation symbol probability distributions (B) which vary depending on the training data set. Therefore, in this particular HMM training system, only observation symbol probability distributions in the state are calculated. The others parameters are pre-defined.

The speech signal is divided into several segments based on the time interval. Each segment will be represented by a centroid value which is determined by calculating the center of area of the speech signal amplitude in the segment. Vector quantization is applied to the centroid value of each segment and result the observation symbol of each state. The sequences of observation symbol of state 1, state 2, up to state N represent the observation sequence of the word.

If number of samples per word that are trained in the HMM speech recognition system is \( N \), and each sample is converted to the observation sequence by the feature extraction and
vector quantization process, the observation symbol probability distribution \( B \) in a state can be calculated by using following equation:

\[
b_j(k) = \frac{n(v_k)}{R}
\]

where \( b_j(k) \) is probability of observation symbol \( v_k \) in the state \( j \), \( n(v_j) \) is number of observation symbol \( v_j \) in the state \( j \), and \( R \) is number of sample.

**C. Recognizer System Using HMM**

Block diagram of HMM recognizer system is shown at figure 8. Same as in the HMM training system, firstly, the speech signal of an unknown word is converted to the observation sequence of that word. This is done by performing the feature extraction and vector quantization process. Then, the observation sequence probability of unknown word for a given model \( \lambda \) \( (P(O|\lambda)) \) is calculated for each HMM word model. The probability of observation sequence calculation is done by using the forward procedure method. By applying the parameter \( A \) and \( \pi \) of the designed HMM (16) and (17) to the forward procedure equations, the probability of the observation sequence, \( O = (o_1, o_2, \ldots, o_N) \), for the given model can be calculated by using the following equation:

\[
P(O|\lambda) = b_1(o_1) \cdot b_2(o_2) \cdot b_3(o_3) \cdots b_N(o_N)
\]

where \( b_N(o_N) \) is probability of N-th observation symbol \( (o_N) \) in state N.

Equation 19 is used to calculate the observation sequence probability of unknown speech signal and the unknown speech signal will be recognized as HMM word model which has the maximum probability of observation sequence \( (P(O|\lambda)) \).

---

**V. EXPERIMENTAL RESULTS**

Because there are two approaches implemented in this system, the experiments are divided into two experiments. The first is experiment for testing the performance of the speech recognition system using LPC and ESD. The second is experiment for testing the performance of the speech recognition system using HMM. In the experiments, either in the training mode or recognizer mode, someone utters the word directly to microphone for giving the command to the wheelchair.

**A. Experimental Results of Speech Recognition Using LPC and ESD**

The experiments were done using 1 sample, 3 samples, and 5 samples of training data per word. The values of the LPC analysis parameters that used in the experiments are:

1. Number of samples in the analysis frame is 240.
2. Number of samples shift between two adjacent frames is 80.
3. LPC analysis order is 10.
4. Dimension of LPC cepstral vector is 12.

Summary of the experimental results are shown at table 1

<table>
<thead>
<tr>
<th>Word</th>
<th>Recognition Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 sample</td>
</tr>
<tr>
<td>Stop</td>
<td>60</td>
</tr>
<tr>
<td>Forward</td>
<td>100</td>
</tr>
<tr>
<td>Backward</td>
<td>20</td>
</tr>
<tr>
<td>Left</td>
<td>0</td>
</tr>
<tr>
<td>Right</td>
<td>0</td>
</tr>
<tr>
<td>Up</td>
<td>0</td>
</tr>
<tr>
<td>Down</td>
<td>80</td>
</tr>
</tbody>
</table>

Average Rec. Rate | 37.14 | 52.86 | 62.86 | 78.57 | 70 | 71.43

As shown at table 1, the highest recognition rate that can be achieved is 78.57%. This recognition rate is resulted from the experiment using 3 training data samples per word. The more number of samples are trained, the higher the average recognition rate is resulted.

From the experiments, we also got that time for training one sample using LPC-Euclidean Squared Distance is about 19 seconds and time for recognizing an unknown word using LPC-Euclidean Squared Distance depends on the number of samples per word that is trained. The more number of samples trained into LPC-Euclidean Squared Distance system, the more time for recognizing a word. And, most of the time is used to perform the LPC analysis calculation.

**B. Experimental Results of Speech Recognition Using HMM**

The experiments were done in various numbers of training data, i.e. experiment using 10 samples, 20 samples, and 30 samples. In this experiment, number of state is 20. The
experiments were also done in various numbers of states in
HMM structure by using 30 states, 50 states, and 80 states. In
this experiment, number of samples is 30.
Table 2 shows summary of the experimental result of
experiment in various numbers of samples. Table 3 shows the
summary of experimental result of experiment in various
numbers of states.

### TABLE 2
SUMMARY OF EXPERIMENTAL RESULTS OF SPEECH RECOGNITION SYSTEM USING
HMM IN VARIOUS NUMBERS OF SAMPLES

<table>
<thead>
<tr>
<th>Word</th>
<th>10 samples</th>
<th>30 samples</th>
<th>50 samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>I</td>
</tr>
<tr>
<td>Stop</td>
<td>10</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Forward</td>
<td>20</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>Backward</td>
<td>30</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Left</td>
<td>30</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Right</td>
<td>0</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Up</td>
<td>10</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>Down</td>
<td>40</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>Average Rec. Rate</td>
<td>20</td>
<td>17.14</td>
<td>22.86</td>
</tr>
</tbody>
</table>

### TABLE 3
SUMMARY OF EXPERIMENTAL RESULTS OF SPEECH RECOGNITION SYSTEM USING
HMM IN VARIOUS NUMBERS OF STATES

<table>
<thead>
<tr>
<th>Word</th>
<th>10 samples</th>
<th>30 samples</th>
<th>50 samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>I</td>
</tr>
<tr>
<td>Stop</td>
<td>0</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>Forward</td>
<td>30</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Backward</td>
<td>50</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Left</td>
<td>60</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Right</td>
<td>30</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Up</td>
<td>20</td>
<td>70</td>
<td>10</td>
</tr>
<tr>
<td>Down</td>
<td>30</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>Average Rec. Rate</td>
<td>31.43</td>
<td>25.71</td>
<td>32.86</td>
</tr>
</tbody>
</table>

As shown at summary table (table 2 and table 3), the highest
recognition rate that can be achieved is 32.86%. This
recognition rate is resulted from experiment of the system
using 30 training data samples per word, 50 states, and 10
observation symbols. But, this recognition rate is not good
enough for a speech recognition system. However, the
experimental results still show that the more number of
samples are trained, the higher the recognition rate is resulted,
although, the recognition rate does not improve too much. It
also shows that the recognition rate does not always increase
when the number of states increases. It is shown when the
number of states is changed from 50 to 80 states.

Comparing with the result from Y.M. Lam, M.W. Mak, and
P.H.W. Leong [2], they achieved recognition rate of 81.33%
using LPC and HMM. In their system, LPC is used as the
feature extraction method and HMM is used as the recognition
method. The difference between their system and this
designed system is on the feature extraction method. In this
system, a simple segmentation and centroid value is used as
the feature extraction method. Thus, it can not be concluded
that HMM is not a good speech recognition system. The bad
result achieved in this system may be caused by the feature
extraction method.

### VI. CONCLUSIONS
Speech recognition has successfully been implemented on
AT89C51RC microcontroller in order to control the
movement of wheelchair. Two approaches of speech
recognition that have been implemented are LPC-ESD and
HMM-Segmentation and Centroid.

From the experimental results, it can be concluded that the
highest recognition rate that can be achieved is 78.57%. This
recognition rate is resulted from the experiment of the system
using LPC-ESD approach with 3 training data samples per
word. But, LPC-ESD approach has disadvantage. LPC-ESD
gives slow response compare with HMM-Segmentation and
Centroid method. Unfortunately, speech recognition using
HMM-Segmentation and Centroid method results a very low
recognition rate.

### REFERENCES