Design and Implementaion of Specch Recognition System for Myanmar Language

Ingyin Khaing
Faculty of Information and Communication Technology
University of Technology (Yatanarpon Cyber City)
Pyin Oo Lwin, Myanmar
ingyinkhaing@gmail.com

K Zin Lin
University of Computer Studies, Hardware Department,
Yangon (Bahan Campus), Myanmar
kzinlin78@gmail.com

Abstract *:*

This paper presents speech processing and recognition system for Myanmar continuous language. The speech segmentation methods are based on two simple speech features, namely time domain features and frequency features. To detect the word boundaries, dynamic thresholding method is applied. Then, important features of speech features are extracted by LPC (linear predictive coding) and GTCC (gamma tone cepstral coefficient) approach. K-means is used in feature clustering and HMM is used in recognition process. All the algorithms used in this work are implemented in Matlab. The system obtained the average word error rate of 0.5 with GTCC feature extraction techniques respectively.

Keywords-component; *automatic speech recognition; GTCC; LPC; HMM*

I. Introduction

Automatic Speech Recognition (ASR) is a technology that allows a computer to identify the words that a person speaks into a microphone or telephone. This is a very difficult task, partially because the field is motivated by the promise of human-like performance under realistic conditions. ASR has so far gained some commercial success due to demonstrable increases in productivity by greatly assisting human operators or by replacing the human element altogether. A speech interface in users' own language is a very natural, flexible, efficient, and economical form of communication. ASR has a wide area of applications: command recognition, information inquiry, dictation, personal computer interfaces, automated telephone services, interactive voice response, and special purpose commercial and industrial systems. It can also be used in education for language learning. It has other important applications for handicapped people to help them with their daily life and communication with the rest of the society. It is a technology that makes life easier and very promising.

Speech recognition systems have been developed for isolated words in Myanmar Language. Isolated word recognition, in each word is surrounded by some sort of continuous of pause, is much easier than recognizing continuous speech, in which words run speech into each other and have to be segmented. Continuous speech task is greatly in difficult.

In speech recognition systems, the pre-processing of the speech signal is a key function for extracting and coding efficiently the meaningful information in the signal. In this system, LPC and GTCC feature extraction techniques are used. The basic idea of LPC is to predict the current value of the signal using a linear combination of previous samples, each weighted by a coefficient. Gammatone Cepstral Coefficient (GTCC)

technique is based on the Gammatone filter bank, which attempts to model the human auditory system as a series of overlapping bandpass filters.

LPC is used for both noisy and clean environment and GTCC is better in noisy environment [1]. In a hidden Markov model, the state is not directly visible, but variables influenced by the state are visible. Each state has a probability distribution over the possible output tokens. Therefore the sequence of tokens generated by an HMM gives some information about the sequence of states.

This paper is organized as the follows. Overview of standard Myanmar language is described in Section 1. Hidden Markov Model is detailed in Section 2. Proposed system design is presented in Section 3. Experimental results are showed in Section 4 and the Conclusion is summarized in Section 5.

I. OVERVIEW OF STANDARD MYANMAR LANGUAGE

Myanmar is a tonal language. This means that all syllables in Myanmar have prosodic features that are an integral part of their pronunciation. Standard Myanmar is based on the dialect spoken in the lower valleys of the Irrawaddy and Chindwin rivers. It is spoken in most of the country with slight regional variations. In Myanmar there are 8 main races and 135 sub-races. Myanmar (Burmese) is the official language in Myanmar. A syllable is assigned a tone and each spoken syllable with a different tone will have a different lexical meaning.

A. Tones

The most important feature of the Myanmar language is its use of tone to convey the lexical meaning of the syllables. Myanmar tones can be divided into two groups: static and dynamic. The static group consists of three tones (mid, low, and high) whereas the dynamic group consists of two tones (falling and rising).

B. Stress

The syllable in a word produced with a higher degree of respiratory effort is referred to as "stress." The stressed syllables are usually louder, longer, and higher in pitch than unstressed syllables. The placement of stress on a word in Myanmar is linguistically significant and governed by rules including the monosyllabic word rule and the polysyllabic word rule. For monosyllabic words, all content words are stressed, whereas all grammatical words are unstressed. However, monosyllabic unstressed words when spoken in isolation or emphasized can be stressed as well.

For polysyllabic words, stress placements are determined by the number of syllables as well as the structure of the component syllables in the word. The primary stress falls on the final syllable of a word. The secondary stress is determined by the position of the remaining syllables and whether or not they are linker or non-linker syllables.

C. Vowels and Consonants

The Myanmar alphabet consists of 33 letters and 12 vowels, and is written from left to right. It requires no spaces between words, although modern writing usually contains spaces after each clause to enhance readability. The latest spelling authority, named the *Myanma Salonpaung Thatpon Kyan* (မြန်မာစာလုံးပေါင်းသတ်ပုံကျမ်း), was compiled in 1978 at the request of the Burmese government. Some of the spoken words and their International Phonetic Alphabet (IPA) format are shown in table (1).

TABLE 1 TABLE I: IPA CODE FOR MYANMAR WORDS

IPA	Words in Myanmar	IPA format
ð̲	အညာသား	[ʔəɲàðá]
h̲	ဟုတ်	[**h**oʊʔ]
k̲	ကုန်	[**k**òʊɴ]
k̲ʰ	ခုန်	[**kʰ**òʊɴ]
l̲	လုပ်	[**l**oʊʔ]
l̥̲	လျှပ်	[**l̥**oʊʔ]
ə̲	ခလုတ်	[kʰ**ə**loʊʔ]
̀	ငါ	[ŋà]
̰	င့်	[ŋə̰]

II. HIDDEN MARKOV MODEL (HMM)

HMM is defined as the finite state machine with fix number of states. It is statistical processes to characterize the spectral properties of voice signal. It was two types of probabilities. There should be a set of observation or states and there should be a certain state transitions, which will define that model at the given state in a certain time.

In HMM, the states are not visible directly. They are hidden but the output is visible which is dependent on the states. Output is generated by probability distribution over the states. It gives the information about the sequence of states but the parameters of states are still hidden. HMM can be characterized by following when its observations are discrete.

- N is number of states in given model, these states are hidden in model.
- M is the number of distinct observation symbols correspond to the physical output of certain model.
- A is a state transition probability distribution defined by NxN matrix as shown in equation (1).

  A = {aij}

  $$\text{Aij} = p\{ qt+1 = j|qt = i\}, i \le i, j \le N, \sum_{j=1}^{N} a_{ij} = 1, \quad 1 \le i \le N \tag{1}$$

  Where q_t occupies the current state. Transition probabilities should meet the stochastic limitations

- B is observational symbol probability distribution matrix (2) defined by NxM matrix equation comprises

  $$b_i k = p\{O_t = V_k \mid q_t = j\}, \quad 1 \le j \le N, \quad 1 \le k \le M$$
  $$\sum_{k=1}^{M} b_j(k) = 1, \quad 1 \le j \le N \tag{2}$$

Where V_k represents the K^{th} observation symbol in the alphabet, and O_t the current parameter vector. It must follow the stochastic limitations.

- π is a intinal state distribution matrix (3) defined by Nx1.

$$\pi = \{\pi_i\}$$
$$\pi_i = p\{q_1 = i\}, \quad 1 \le i \le N \tag{3}$$

By defining the N, M, A, B and π, HMM can give the observation sequence for entire model as λ = (A,B, π) which specify the complete parameter set of model [2].

HMM defined forward backward estimation algorithm to train its parameters to find log likelihood of voice sample. Segmental k means algorithm is used to generate the code book of entire features of voice sample. Forward backward algorithm is used to estimate the unidentified parameters of HMM. It is used to compute the maximum likelihoods and posterior mode estimate for the parameters for HMM in training process. It is also known ($P(X_{k|}\, O_{1:t})$ posterior marginal or distribution. For all hidden state variables $X_k \epsilon \{X_1, \ldots, X_t\}$.By given a set of observations as $O_{1:t}$:= $O_1, \ldots, O_t$.

This inference task is commonly known as smoothing [2] [3]. This algorithm uses the concept of dynamic programming to compute the required values for the posterior margins efficiently in two processes first doing the forward estimations and then backward estimation. Segmental K-means algorithm is used to clustering the observations into the k partition. It is the variation of EM (expectation-maximization) algorithm.

III. Proposed System Design

In this system, human speech is taken as input to the system. First human speech is decoded into signals for digital processing. The following figure shows the proposed system design.

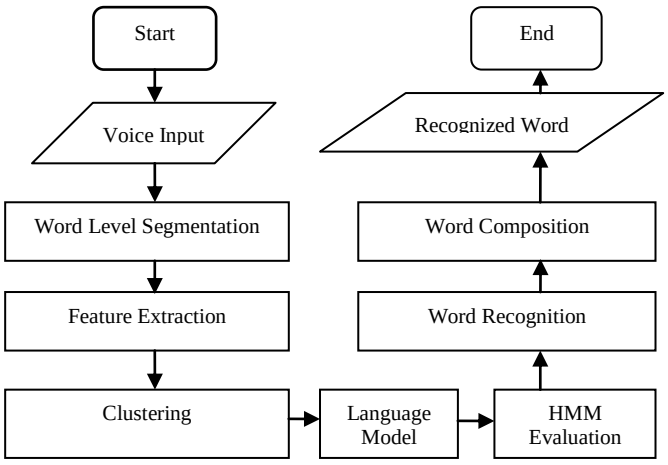

Fig. 1 Proposed System Design

For segmenting continuous Myanmar speech sentences into word/sub-words, time-domain and frequency-domain features extraction is used [4]. To detect the word boundaries, dynamic thresholding criterion is applied [5]. After segmenting the continuous Myanmar words, important features of speech signals are extracted by Linear Predicted Coding (LPC) and Gammatone Cepstral Coefficient (GTCC) approach.

A. Features Extraction

The purpose of feature extraction is to convert the speech waveform to some type of parametric representation (at a considerably lower information rate) for further analysis and processing. This is often referred as the signal-processing front end.

1) Linear Predictive Coding (LPC)

LPC of speech has become the predominant technique for estimating the basic parameters of speech. It provides both an accurate estimate of the speech parameters and it is also an efficient computational model of speech. The basic idea behind LPC is that a speech sample can be approximated as a linear combination of past speech samples. The basic steps of LPC processor include the following:

a) 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 is related to the input to the network, *s(n)* , by difference equation:

$$\tilde{s}(n) = s(n) - a\tilde{s}(n-1) \tag{4}$$

b) Frame Blocking

The output of pre-emphasis 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 l^{th} frame of speech, and there are L frames within entire speech signal, then

$$x_i(n) = \tilde{s}(Ml + n) \tag{5}$$

Where, n = 0,1,…,N – 1 and l = 0,1,…,L – 1

c) 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 ≤ n ≤ N – 1, then the result of windowing is the signal:

$$\tilde{x}_1(n) = x_1(n)w(n) \tag{6}$$

Where, 0 ≤ n ≤ N – 1

Typical window is the Hamming window.

d) Autocorrelation Analysis

The next step is to auto correlate each frame of windowed signal in order to give:

$$r_i(m) = \sum_{n=0}^{N-1-m} \tilde{x}_i(n)\tilde{x}_l(n+m) \qquad m = 0,1,..., p \tag{7}$$

Where the highest autocorrelation value, *p*, is the order of the LPC analysis.

e) 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:

$$
\begin{aligned}
&E^{(0)} = r(0) \\
&k_i = \frac{r(i) - \sum_{j=1}^{i-1} \alpha_j^{i-1} r(|i-j|)}{E_{i-1}} \qquad 1 \le i \le p \\
&\alpha_i^{(i)} = k_i \\
&\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{i-1} \qquad 1 \le j \le i-1 \\
&E^{(i)} = (1-k_i^2) E_{i-1}
\end{aligned}
\tag{8}
$$

For i=1,2,…,p, the LPC coefficient, a_m, is given as

$$a_m = \alpha_m^{(p)}$$

f) 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

$$
\begin{aligned}
&c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) . c_k . a_{m-k} \qquad 1 \le m \le p \\
&c_m = \sum_{k=m-p}^{m-1} \left(\frac{k}{m}\right) . c_k . a_{m-k} \qquad m > p
\end{aligned}
\tag{9}
$$

The LPC cepstral coefficients are the features that are extracted from voice signal and these coefficients are used as the input data of HMM Model. In the HMM model, LPC features are used in building phoneme model.

2) Gamma Tone Cepstral Coefficient (GTCC)

The following steps explain the feature extraction process.

a) Hamming windowing and FFT

The first step of the algorithm is to subdivide a speech sequence into frames. The windowing function is the Hamming window, which aims to reduce the spectral distortion introduced by windowing.

After windowing, Fast Fourier Transform (FFT) is applied to the windowed speech frame. The N-point FFT spectrum, S[k] for 0 ⩽ k ⩽ N-1, of the speech frame is obtained as a result.

b) Gammatone Filterbank

The Gammatone filterbank consists of a series of bandpass filters, which models the frequency selectivity property of the human cochlea. The impulse response of each filter was

$$g(t) = at^{n-1} e^{-2\pi bt} \cos(2\pi f_c t + \phi) \tag{10}$$

Where, a is a constant, usually equals to 1, n is the order of the filter, Φ is the phase shift, f_c and b are the center frequency and the bandwidth of the filter in Hz. The next step is to find out the filter output, X_m,

$$X_m = \sum_{k=0}^{\frac{N}{2}-1} |S[K]|^2 |H_m[K]| \tag{11}$$

c) Equal-loudness

The equal-loudness weight of the m^{th} filter, E_m, can be found by evaluating

$$E_m = \frac{(w^2 + 56.8*10^6)w^4}{(w^2 + 6.3*10^6)^2(w^2 + 0.38*10^9)(w^6 + 9.58*10^{26})} \quad (12)$$

The filter output after equal loudness, $X_{m(e)}$, is simply the product of the filter output and the equal loudness weight

$$X_{m(e)} = E_m X_m \quad (13)$$

d) Logarithmic Compression

The next step of the algorithm is to apply logarithm to each filter output. The aim of this procedure is to simulate the human perceived loudness given certain signal intensity.

Let $X_m(ln+ e)$ be the logarithmically-compressed filter output of the m_{th} Gammatone filter.

e) DCT

To de-correlate the filter outputs, Discrete Cosine Transform (DCT) is applied to the filter outputs. Suppose p is the order of GTCC. The feature vector of one speech frame, which has p number of GTCC coefficients, contains the first *p* DCT coefficients of the filter outputs. In mathematical term, the k_{th} GTCC coefficient of the feature vector is defined as follows.

$$GTCC_k = \sqrt{\frac{2}{M}} \sum_{m=1}^{M} X_{m(\ln+e)} \cos\left(\frac{\pi k(m-0.5)}{M}\right) 1 \le k \le p \quad (14)$$

B. HMM Training

The extracted feature vectors of LPC and GTCC are trained into HMM. The training is done in two steps as

- HMM code book
- HMM training by forward backward re-estimation algorithm

First the code book contains the cluster number specifies to each observation vector, which is obtained by applying the K-means algorithm.

After clustering the training to HMM, parameters begin by applying Forward-backward algorithm. It uses the principle of Maximum likelihood estimation. It returns the state transition matrix A, observation probability matrix B, and the initial state probability vector π. In this phase, the observation vectors being trained in the form of HMM parameters and resulted as the log likelihood of entire speech. This log likelihood is used to store in training database for recognition in real time.

C. Language Model

For continuous speech recognition, a pronunciation dictionary was created that contains the input-output-pronouncing for each word entry where the pronunciation describes the sequence of HMMs that constitute each word. For each word the output is provided as Unicode sequence and the pronunciation is given with the consideration of phoneme as a unit. As in continuous speech recognition, the system recognizes a sequence of words and that's why it is necessary to incorporate a language model. The Regular grammar modelling technique is used as language model which has the properties like finite state model, small vocabulary and restricted grammar.

D. Recognition

HMM recognition recognizes the words on the basic of log-likelihood. It recalculates the log likelihood of speech vector and compares it to the pre stored value of log likelihood. If it matches the entire log value from the database of specified model, then it can recognize the words. Then, words are composed to get the sentences in text. Finally, input speech is transformed into text structure and displayed as output.

IV. EXPERIMENTAL RESULTS

For a continuous speech signal to be recognized, the preprocessing and the feature extraction is done first. Then the signal is recognized using the HMM. This work is based on 558 files gathered from Myanmar native female speaker. The words are taken from the Grade 1 Myanmar Text Book. The speech signals were digitized at a sample rate of 44100 Hz using 16 bits and saved in WAV format.

In this experiment, example of 10 spoken sentences was recognized by the system. The system output was recognized Myanmar words. The performance of speech recognition systems is usually specified in terms of accuracy. Accuracy will be measured in terms of performance accuracy which is usually rated with word error rate (WER).Word error rate can then be computed as:

$$WER = \frac{No.\ of\ \ miss\ recognized\ \ words}{Total\ \ No.\ of\ \ segmented\ \ words} \tag{15}$$

An example of Myanmar continuous sentences are given in Table 2. Table 3 shows the output of the system. Figure 2 illustrates different recognition results between LPC and GTCC feature extraction approach.

TABLE I

EXAMPLE OF MYANMAR CONTINUOUS SENTENCES

Sentences ID	Myanmar Continuous Sentences
S1	ခခရေကုံး မလေးပြိုး
S2	ငါးမဖမ်းရ ပန်းမခူးရ ရေကန် အနီးမဆော့ရ စည်းကမ်းလေးစားပါ
S3	မိုးလရာသီ အဘိုးအိုတို့ လယ်တဲ့ တဲကလေး မိုးယိုနေသလား ကူညီ၍ မိုးပေးပါ
S4	အရှေ့ရွာမှာဘာရှိသလဲလူစည်ကားလှပါသည် ဘုရားပွဲတော်ရှိပါသလား လှည်းစီး၍ လာကြသည် လှေစီး၍ လည်းလာကြသည်
S5	ရွာလူကြီးများကြွလာပြီကြွေပန်းကန်ယူခဲ့ပါ ကျွဲကောသီးထည့်ထားပါ ကျွဲမွေးပြုစုပါရစေ
S6	ကလေးငယ်ငယ် ဘာစားခဲ့သလဲ အမဲသားငါးစားရဲ့လား ကုလားပဲစားပါ
S7	ချွေးတရွဲရွဲ အားကစားပွဲ လွှဲဖယ်၍ မနေပါ ပြေးလွှားကစားကြသည် အားရပါးရ နွှဲပါဦး
S8	စောစော အိပ်ထ ပြုကျင့်က တွင်းပ ကိုယ်ခန္ဓာ ရောဂါခပ်သိမ်း ရှောင်ဖွာတိမ်း ကင်းပြီမ်းလွန်ချမ်းသာ ဥစ္စာ ဓန ပေါကြွယ်ဝ ထွန်းပဥာဏ်ပညာ
S9	လယ်ထဲမှာရေလျှံနေပြီလျှော်စည်းများ မျောပါနေကြသည် လျှောကြမည် သတိထား
S10	နေ့စဉ်မှန်စွာရေချိုးပါ ခေါင်းကိုမှန်စွာ ဖြီးကြပါ ကျန်းမာသန့်ရှင်းရောဂါကင်း

TABLE II
OUTPUT OF THE PROPOSED SYSTEM

Sentences ID with different features	Myanmar Continuous Sentences
S1(LPC) S1(GTCC)	ခ,ခရေ,ကုံးမလေး,ပြိုး; ,ခ,ခရေ,ကုံးမလေး,ပြိုး
S2(LPC) S2(GTCC)	အ,ဖမ်းရ,ပန်းမ,ခူးရ,ရေကန်အနီးမ,ဆော့ရ,စည်း,ကမ်းလေး,လာ,ပါ; ,ဘူး,ဖမ်းရ,ပန်းမ,ခူးရ,ရေကန်အနီးမ,ဆော့ရ,စည်း,ကမ်းလေး,လာ,ပါ
S3(LPC) S3(GTCC)	,အ,အ,ဘိုး,ဘိုး,တို့လဲ,တဲ,တဲကလေး,မိုး,သလား,ကူညီ၍,မိုး,ပေး,ပါ; ,အ,အ,ဘိုး,အို,တို့လဲ,တဲ,တဲကလေး,မိုးယိုနေ,မိုး,ကူညီ၍,မိုး,ပေး,ပါ
S4(LPC) S4(GTCC)	အရှေ့ရွာမှာ,ဘာရှိ,သလဲ,လူ,စည်ကား,လှ,ပါ,သည်,ပါ, ပွဲတော်,ရှိ,ပါ,သလား,လှည်း,စီး၍,လာ,လေ့,သည်,လေ့,စီး၍လည်း,လာ,ကြ,ကြ; အရှေ့ရွာမှာ,ဘာရှိ,သလဲ,လူ,စည်ကား,လှ,ပါ,သည်,ဘုရား,ပွဲတော်,ရှိ,ပါ,သလား,လှည်း,စီး၍ ,လာ,ကြ,သည်,လေ,စီး၍လည်း,လာ,ကြ,ကြ
S5(LPC) S5(GTCC)	ရွာလူကြီးများ,ကြွလာ,ပြီ,ကြွပန်း,ကန်,ယူ,ကော,ပါ,ကျွဲ,ကော,ကျွဲ,ထည့်, ထည့်,ပါ,ကျွေးမွေး,ပြု,စု,ကော,ထည့်; ရွာလူကြီးများ,ကြွလာ,ပြီ,ကြွပန်း,ကန်,ယူ,ခဲ့,ပါ,ကျွဲ,ကော,သီး,ထည့်,ထား,ပါ,ကျွေးမွေး,ပြု, စု,ပါရ,စေ
S6(LPC) S6(GTCC)	ကလေးငယ်ငယ်,ဘာ,စား,ခဲ့,သ,လဲ,အမဲ,သားငါး,စားရဲ့,လား,ကုလား,ပဲ,လား,ပါ; ကလေးငယ်ငယ်,ဘာ,စား,ခဲ့,လဲ,လဲ,အမဲ,သားငါး,စားရဲ့,လား,ကုလား,လား,လား,ပါ
S7(LPC) S7(GTCC)	ချွေးတရွဲရွဲ,အားကစား,ပွဲ,လွှဲ,ဖယ်၍,မနေ,ပါ,ပြေးလွှား,ကစား,ကြ,သည်,အားရ,ပါးရ,နှဲပါ,ဦး; ချွေးတရွဲရွဲ,အားကစား,ပွဲ,လွှဲ,ဖယ်၍,မနေ,ပါ,ပြေးလွှား,ကစား,ကြ,သည်,အားရ,ပါးရ,နှဲပါ,ဦး
S8(LPC) S8(GTCC)	,စောစော,အိပ်,ထ,ပြု,ကျင့်,က,တွင်း,ပ,ကိုယ်,ခန္ဓာ,ရော,ဂါ,ခက်,သိမ်း,ရှောင်,ခွာ,တိမ်း,ကင်း ငြိမ်းလွန်ချမ်းသာ,ဥ,စွာ,ဓန,ပေါ်,ကြွယ်ဝ,ထွန်း,ပဉာဏ်,ပညာ; စောစော,အိပ်,ထ,ပြု,ကျင့်,က,တွင်း,ပ,ကိုယ်,ခန္ဓာ,ရော,ဂါ,ခက်,သိမ်း,ရှောင်,ခွာ,တိမ်း, ကင်းငြိမ်းလွန်ချမ်းသာ,ဥ,စွာ,ဓန,ပေါ်,ကြွယ်ဝ,ထွန်း,ပဉာဏ်,ပညာ
S9(LPC) S9(GTCC)	,လယ်ထဲမှာ,ရေလျှံနေပြီ,မျော,မျော,ပါနေကြသည်,လေ့မျော၍,ဆယ်ယူ,ပါ,လျှောကြမည်, သတိ,ထား; လယ်ထဲမှာ,ရေလျှံနေပြီ,လျှော်စည်းများ,မျော,ပါနေကြသည်,လေ့မျော၍,ဆယ်ယူ,ပါ,လျှော ကမည်,သတိ,ထား
S10(LPC) S10(GTCC)	နေစဉ်,မှန်စွာ,ရေချိုး,ပါ,ခေါင်းကို,မှန်စွာ,ဖြီးကြ,ပါ,ကျန်းမာ,သန့်ရှင်း,ရော,ဂါ,ကင်း; နေစဉ်,မှန်စွာ,ရေချိုး,ပါ,ခေါင်းကို,မှန်စွာ,ဖြီးကြ,ပါ,ကျန်းမာ,သန့်ရှင်း,ရော,ဂါ,ကင်း

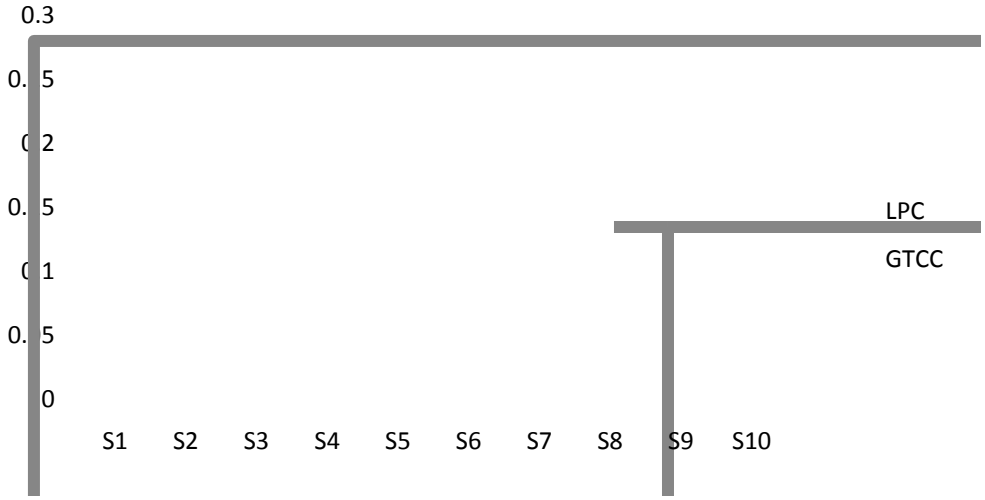

Fig.2 Different recognition results between LPC and GTCC

The system obtained the average word error rate (WER) of 1.18 and 0.5 based on LPC and GTCC features respectively.

V. CONCLUSIONS

This paper presents an automatic speech recognition system for Myanmar language using the appropriate features and recognizer. This paper clearly describes the design and implementation details of the entire development task using the HMM. Speech segmentation was done using time-domain feature and frequency-domain feature and the recognition accuracies obtained using standard vector comparison method like Euclidean distances. The results also demonstrate the superiority of the features using LPC and GTCC features.

ACKNOWLEDGMENT

I am very grateful to Dr. K Zin Lin for fruitful discussion during the preparation of this paper and also specially thank to Rector, Professors and colleagues from Technology University (Yatanarpon Cyber City), Myanmar.

REFERENCES

[1] O. Cheng, W. Abdulla, Z. Salcic, "Performance Evaluation of Front-end Processing for Speech Recognition ", School of Engineering Report No. 621

[2] L. Rabiner, Fellow, IEEE 'A Tutorial On Hidden Markov Model And Selected Applications In Speech Recognition, Proceedings Of The IEEE, Vol. 77, No. 2, February 1989

[3] http://www.cs.brown.edu/research/ai/dynamics/tutorial/Documents/HiddenMarkovModels.html

[4] T. Giannakopoulos, "Study and application of acoustic information for the detection of harmful content and fusion with visual information" Ph.D. dissertation, Dept.of Informatics and Telecommunications, University of Athens, Greece, 2009.

[5] T. Giannakopoulos, A. Pikrakis and S. Theodoridis "A Novel Efficient Approach for Audio Segmentation", Proceedings of the 19th International Conference on Pattern Recognition (ICPR2008), December 8-11 2008, Tampa, Florida, USA.