A Novel Sindhi QoS Aware Framework for Speech Recognition System Based On Designed Dictionary

Fida Hussain Khoso[1,*],Shafique Ahmed Awan[2],Syed Hyder Abbas Musavi[3] , Dhani Bux Talpur[4,5],Ghulam Farooque Solangi[6],Hamza Ahmed[7],Asadullah Shah[8]

[1,*]Department of Computer Science, Dawood University of Engineering and Technology Karachi, Pakistan
[2]Department of Computer Science and IT Benazir Bhutto Shaheed University Lyari Karachi, Pakistan
[3]Department of Electronic Engineering Mehran University of Engineering & Technology, SZAB Campus Khairpur Mir's Sindh, Pakistan.
[4]Department of Information and Computing, University of Sufism and Modern Sciences, Pakistan
[5]Kulliyyah of Information and Communication Technology, International Islamic University Malaysia.
[6]Ph.D Scholar, Department of Computer Science DHA Suffa University Karachi Pakistan
[7]Department of Computer Science, Visitig Faculty at SMIU Karachi Pakistan.
[8]Kulliyyah of Information and Communication Technology, International Islamic University Malaysia.

Abstract— **These days, voice to text recognition system based applications are growing progressively to deal with the daily life task. Many studies have been devised their voice to text systems in different languages. However, voice to text recognition system in the Sindhi language has not been devised yet. With this motivation, in this paper, we propose Sindhi voice to text recognition system based on Hidden Markov Model (HMM). We devise a novel Sindhi Recognition System based on Hidden Markov Model (SRS-HMM) and create a novel Sindhi dictionary to deal with users and their Quality of Service (QoS) requirements. The study formulates the combinatorial problem where minimum error-rate considered as a convex function and maximum accuracy formulated as a concave function. The objective of the study is to improve users input with different accents and enhance their accuracy and minimize the error rate. Simulation results show that, proposed SRS-HMM system gained near-optimal, e.g., 60% of different users with different accents during the performance.**

Keywords— **Hidden Markov Model SRS-HMM Accent Training QoS.**

I. INTRODUCTION

These days, the usage of speech recognition system in different fields has been growing progressively [1]. The applications of speech recognition are widely used in the practice by users [2]. For in1 stance, voice to text transaction, E-Transport, E-Commerce, E-Governance. The acoustic model and phonemes have much importance in speech recognition. Two main issues are discussed in the speech recognition system [3]. (i) How to analyze and model the signal of speech for training or test system. (ii) The new signal needs an adaptive method; it is a challenge on how to adopt a method to recognize new input signal by users. The two preliminary sorts of speech recognition are always considered, such as deterministic (stochastic) and statistical modelling [4]. The deterministic or stochastic use the intrinsic attributes or properties of signal speech, whereas the statistical modelling is linked to attributes of statistical properties. To identify the unit of speech, there are two types which are widely used such as training of acoustic probability of training and recognition unit [5]. The recently proposed methods formulated speech recognition system based on probabilistic technique[6-8]. The techniques widely exploited acoustic-phonetic with decoding component and modelling of language [9]. The preliminary goal of this method is to predict the data analysis of any signal from many parameters such as input, properties and attributes. Whereas module language compares a sequence of elements of different acoustic [10]. It may form a lexical attributes form based on the acoustic model. These attributes to be phonemes, syllable and many syllables, etc. The aforementioned attributes offer the anticipated information about the positioning of the word. These ordering of word in speech signal to be identified via different techniques of modelling for instance grammar based rules and statistics based on the probabilistic method. The Bayesian state widely formulated in the speech recognition system based on the different model [11].

The Hidden Markov Model (HMM) is one of the recognized method in the speech recognition domain and problem. HMM assumes that undiscoverable stochastic process manifests in the all random process. The Markovian always used to solve complex speech recognition system in the real workload practice. The Markov method devoted to the training model which converges on the expression perception test. The state of vector quantization is executed to extricate the acoustic characteristics of a particular unit [12]. Markovian techniques are however the most dependable even with the algorithmic constraints that concern to the fundamental requirements linked to the concentration of the algorithms. Neural Network (NN) have generated interesting issues on speech recognition single character, single word and a single sentence [13]. However, many challenges exist in the current HMM model to solve the speech recognition problem for the Sindhi Language.

(i) The verification of multiple accents input should be validated at the user level. However, the current HMM method cannot handle the verification of input at the local level.

(ii) Due to multi-accent input of different users, the signal and feature extraction along train-data for the Sindhi Language is also a challenging task.

(iii) The HMM-based heuristics or greedy often solved the speech recognition problem. However, due to the complex Quality of Service (QoS) requirements of user applications such as validation, verification, error rate, accuracy, and precision also induce different challenging during the process.

To solve the aforementioned research questions, the research makes the following contribution.

- Proposed a novel system: This study proposed a novel Sindhi Voice to Text as shown in Figure 1. The algorithm goal is to takes multi-accent stochastic inputs from different users, and process them via different components based on QoS requirements of users (e.g., Accuracy and ErrorRate). The proposed algorithm consists of different following components. For instance, MultiAccents User Interface, Identify Input Accent Module, Signal Processing, Feature Extracting, Training Phase & SRS-HMM Algorithm framework as shown in Figure 1. We explain the component in the following points.
- Proposed SRS-HMM Framework:
- Input Processing: The Input processing accepts processed input from users with different accents and process them based on Algorithm 2. The main goal of this algorithm is to process user's requests based on their QoS requirements.
- QoS Processing: This study devises the Quality of Service (QoS) process method which aim to obtain the concave optimization based on equation (1) and convex optimization based on equation (2).
- Fault-Tolerant: This method can handle any kind of failure at user level and system level. We devised the fault-tolerant Algorithm 4 to tackles any kind of failure in the system.
- QoS Parameterizations Module: This module is a monitoring module where all users performance will be checked at the run time of system.
- Sindhi Language Dictionary: This is a novel Sindhi dictionary which is proposed by this study. This dictionary is about 10000 words of Sindhi language which among vegetable name, schools, and healthcare and etc.

The rest of the paper is organized as follows. Section 2 elaborates related work followed by Section 3 which describes the proposed system and its component. Section 4 defined SRS-HMM algorithm in detail. Section 5 discusses the experiments and results while Section 6 concludes the paper.

II. RELATED WORK

These days, voice to text recognition system is widely used in the practice to deal with different tasks. For instance, minimize typing efforts, minimize typing delay, command to machines and etc. Voice to text recognition system proposed in different languages and along with system studies proposed different dictionaries of their languages. Many greed, heuristic, meta-heuristic and dynamic programming based solution suggested to deal with the different voice to text recognition system. However, voice to text recognition system in the Sindhi widely ignored in the research literature as defined in Table 1. Through disc-platelets and series of vibrations the sound is generated by the vocal cards which produces the speech signal by the air from the lungs. As for the conventional transport of information and information processing the speech processing in many specialized workstations in corporate software and telephone making, growing overlap [1].

The studies in [1-4] suggested Arabic workload based system which accepts input in Arabic and convert voice to text for users. The speech processing is a form of natural human communication and it contains one of the best area existing of the processing signal. Understanding of human languages by the computer by following the commands given in human voice has made the speech recognition technology possible. The desire of the people to build mechanical models has been the motivation of communication processing for using research in speech technology and emulating capabilities in the form of verbal communication with human in the computer vision.The development of systems and techniques for input speech to a machine that is our important goal for the area of speech recognition. In this regard, various samples have been collected from area and regions wise in the form of spoken words to simulate basis on environment for comparison the various accents.

The study in [5] suggested Chinese language based recognition system and proposed solution based on heuristic method. The goal was to improve the accuracy of the system user input when it converts from voice to text during process.Many potential applications including command and control, dictation and transcription of recorded speech are covered by automatic continuous speech recognition. It searches audio documents and interactive spoken dialogs. A set of statistical models comprises of the core of all speech recognition systems representing the difference sounds of the language to be recognized. To identify the speech having temporal structure can be encoded as a sequence of the spectral vectors spanning the audio frequency range. The Hidden Markov Model (HMM) provides a natural framework for constructing such models. The Hindi [6] and Russian [7] and Urdu [8] were suggested their system based on their languages. They proposed solutions based on local search, global search and meta-heuristic method to achieve data extraction, improve signal and minimize error.

The study in [9] devised linear integer programming based solution for voice to text recognition system based on French workload. The objective is to minimize error rate of inputs and produce high quality output to the users. The work in [10] considered the English workload as a dataset and proposed solution based on greedy approach to improve lateness of inputs. The lateness determines the less response time during input voice and text process in the system.

The works in [11-13] devised Japanese, German, and Farsi based dynamic programming, meta-heuristic and hidden Markov model based solution to handle voice to text system for users. The objective was to minimize error of the users input and improve the capability of the system during runtime of request [14].

Summary: The existing studies have not support or fulfill the gap because there is no work have done on Sindhi accents and vowels and also on speech recognition system it means there is a big loop hole for research on the said topic. But I have thoroughly studied about other speech recognition system in different languages and I have taken some ideas that how they have worked and which tools and techniques have been used in their research. But my work is very different from others because there is no datasets and database available for Sindhi language. We have collected database in the form of voice from different regions or district wise. We have collected more than 200 samples from male and female through the source of microphone and store in database of Sindhi language. Now we are facing the lot of difficulties and challenges of accent, word error rate and accuracy.

As in Sindhi language there are six unique accents are available so it is the major problem for system understanding. The various accents are listed as Lari, Uttradi. so that we have done some efforts for said system for reduce the error rate and improve the accuracy in this connection. To remove these problems and challenges we have to make model and algorithms for particular problem. In our Sindh region we use different accents and when try to speak in system, it does not recognize accurately accents so it is necessary we should improve error rate and accuracy with help proposed model. In this process we try to connect absolute end to end modeling to utilize HMM model with its approaches for ASR system. Current situation in automatic speech recognition in last decade brought 50% relative improvements in WER by introducing artificial neural networks to all levels of modeling. Traditional state of the art challenged by novel end to end ASR architectures.Enabling factor: generic machine learning tools developed for diverse and complex tasks. ASR has very challenging task, advantages from a method evolution view point. Provides clear performance objective and strong state of art performance to complete against for new approaches.

Table 1: Existing System and Methods

Study	QoS	Workload	Problem	Method	Objective
[1 − 4]	Error	Arabic	Linear	Greedy	Minimize Error
[5]	Error	Chinese	Linear	Heuristic	Max. Accuracy
[6]	Error	Hindi	Linear	Meta-Heuristic	Improve Signal
[7]	Error	Russian	Linear	Local Search	Min. Data Extraction
[8]	Error	Urdu	Linear	Global Search	Min. Error
[9]	Error	French	Linear	Linear Programming	Min. Error
[10]	Error	English	Linear	Greedy	Improve Lateness
[11]	Error	Japanese	Linear	Greedy	Minimize Error
[12]	Error	German	Linear	Dynamic Programming	Minimize Error
[13]	Error	Farsi	Linear	Meta-Heuristic	Minimize Error
Proposed	Error, Accuracy	Sindhi	Combinatorial	Dynamic HMM	Concave, Convex Optimization

III. PROBLEM DESCRIPTION

This study suggests a novel Sindhi Voice Recognition System which aim to take voice with different accents in Sindhi Language and process it in the system. The proposed system consists of different components such as, Multi-Accents User Interface, Identify Input Accent Module, Signal Processing,Feature Extracting, Training Phase and SRS-HMM Algorithm framework as shown in Figure 1. We defined the proposed system with its component in the following way.Multi-Accents User Interface: It is a interactive user interface with system. It takes stochastic random input from different users with different accents and expected their best results from system.

Identify Input Accent Module: This module validated either the input in valid language or not. If it is rather than Sindhi language it will fail, the user input and ask for re-input from scratch. This way, system can validate wrong inputs at user level rather than process level which is time and resource consuming process.Signal Processing: The system exploited Fast Fourier Transform (FFT) approach to process input voice further for the text form. It can help to separate noises, inference from user inputs.Feature Extraction: It process the audio signal into a digital signal form with identified format, such as single character, double character and one full sentence by exploiting the information of a linear predictive model.Training Phase: To fabricate Sindhi speech recognition system dissimilar speech samples having different properties alongwith various formats from dissimilar sources have been collected.

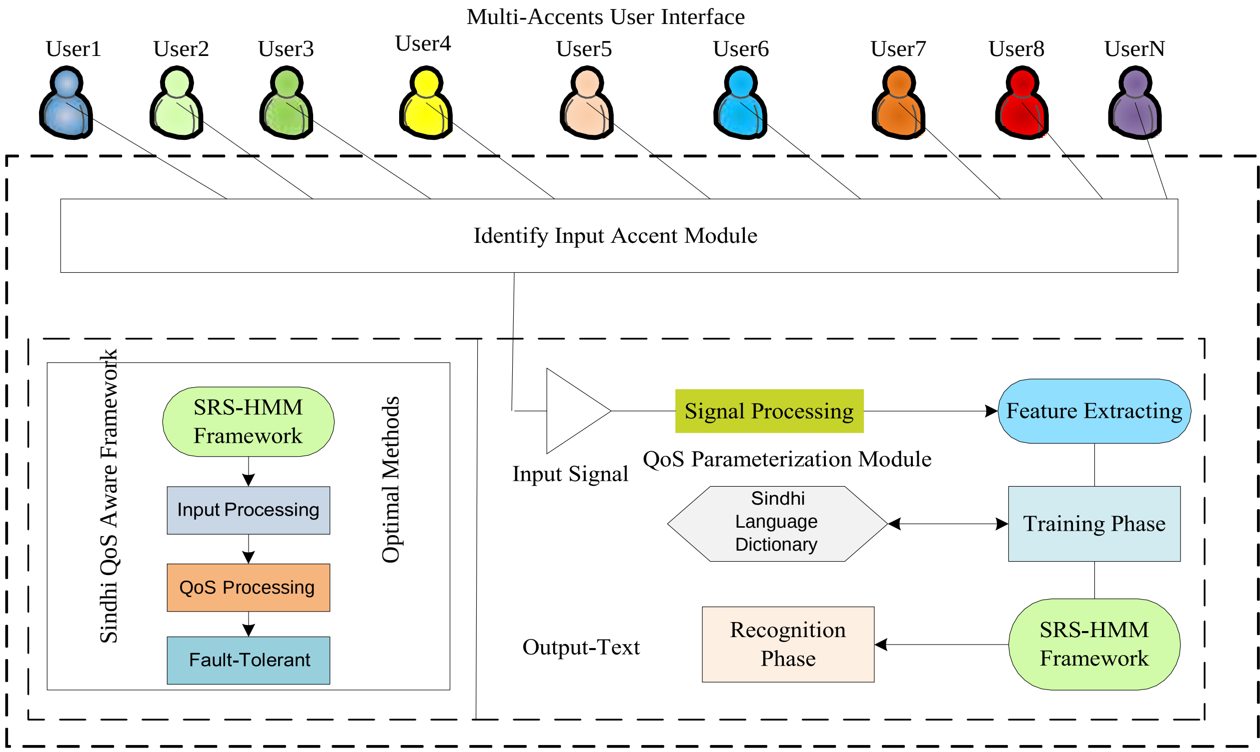

Figure 1: Sindhi Novel System.

Including continuous speech and individual sentences the database contains single word anddouble words. From different cities of Sindh, the spoken words on various subjects was collected. The database was created after collection of large corpus of Sindhi words. In the next sectionwe have given the different properties and comprehensive analysis of data collection. Different words and sentences to speak were given to the speakers. For continuous text as news and lessons the single words double words and sentences were very carefully selected. Male and female names, names of countries and the names of cities were chosen as these words. The names of cities in Sindh, different countries and human names both male and female have been used for selection of these words.

- Input Processing: The Input processing accepts processed input from users with different ac-cents and process them based on Algorithm 2. The main goal of this algorithm is to processusers requests based on their QoS requirements.
- QoS Processing: This study devises the Quality of Service (QoS) process method which aimto obtain the concave optimization based on equation (1) and convex optimization based on equation (2).
- Fault-Tolerant: This method can handle any kind of failure at user level and system level. We devised the fault-tolerant Algorithm 4 to tackles any kind of failure in the system.
- QoS Parameterizations Module: This module is a monitoring module where all users performance will be checked at the run time of system.
- Sindhi Language Dictionary: This is a novel Sindhi dictionary which is proposed by this study.

  This dictionary is about 10000 words of Sindhi language which among vegetable name, schools, and healthcare and etc.

IV. PROBLEM FORMULATION

This study formulated the problem as convex optimization problem. Whereas, accuracy considered as convex optimization problem which convex set, i.e., N number of inputs and with AC number of accents. Whereas, the error-rate Error considered as convex optimization with convex set. i.e., N number of inputs and with AC number of accents. This is bi-objective problem where total performance Tp=W.Accuracy + W.Error of inputs and accent. Tp is further formulated as hidden Markov decision model (HMM) in the environment. We calculate the accuracy of the model in the following way.

$$Accuracy = \frac{D_m \times AC}{N} \times 100\%. \tag{1}$$

The Dm is a dictionary matching process of input N in equation (1). We calculate the error rate of the model in the following way.

$$Error = \frac{U_m \times AC}{N} \times 100\%. \tag{2}$$

Um is un-matched dictionary ratio which is error rate in the input. Equation (2) calculates the error rate of input model. Therefore, we measure the bi-objective in the following way.

$$T_p = 100\% - (W.Accuracy + W.Error). \tag{3}$$

Tp in equation (3) is bi-objective of the considered problem.

We formulate Tp as hidden Markov model (HMM) in the following way.

- an ordered set of states Q;
- a transition probability matrix A, such that aij represents the probability that the machine's next state is the j th state (in the state ordering), given that its current state is the i th state in the ordering; and two special states I $\in$ Q and F $\in$ Q that respectively represent the initial and final states. In other words, we assume that the machine always begins in state I and ends in state F.

Thus, a Markov chain can be represented as a quadruple: (Q, A, I, F). However, we sometimes omit I and F when the initial and final states are not important. A state diagram for a Markov chain (Q, A, I, F) is a directed graph whose vertex set equals Q, and for which each directed edge (i, j) is labeled with aij (note: the edge is usually omitted from the graph if aij = 0).

In this lecture we assume that the states of a Markov chain are occurring in a discrete time sequence. t = 0, 1, . . . , T . We let random variable Qt denote the state that the machine is in at time t.

One important property/assumption that is made with a Markov chain, is that the joint probability distribution for experiencing the sequence of state transitions Q0Q1 · · · QT is computed as

$$T_p = P(Q_0, Q_1, \ldots, Q_T) = P(Q_0) \prod_{i=1}^{T} P(Q_i|Q_{i-1}). \tag{4}$$

1. SRS-HMM Algorithm Framework

The Sindh Recognition System Hidden Markov Model (SRS-HMM) is algorithm framework which consists of different sub-components. These components are Signal Processing, Feature Extracting, Training Phase, Input Processing, QoS processing, and Fault-Tolerant. The study devised a novel SRS-HMM framework and show the process in Algorithm 1.

Algorithm 1: SRS-HMM Algorithm Framework

Input : $\{(Q, A, I, F),\ a_{ij},\ t = 0, 1, \ldots, T,\ N,\ A;$
$\}$
Output: Optimize: T_p;
begin
  foreach *(N ← A as users)* **do**
    Call Signal Processing Based on FFT;
    Call Feature Extracting Based on LPC;
    Call Training Phase Based on Sindhi Language Dictionary;
    Call initial processing Method;
    $(Q, A, I, F) \leftarrow N$;
    a_{ij}=1;
    if $(t = 0\&t <= T)$ **then**
      Call QoS processing method;
      Optimize T_p;
    if *(N ∈ A ← Fail)* **then**
      Call Fault Tolerant method;
    $T_p{}^* \leftarrow T_p$;
  return $T_p{}^*$;
End of Main;

1.1 Signal Processing

The Fast Fourier Transform (FFT) is an algorithm that determines Discrete Fourier Transform of an input significantly faster than computing it directly. In computer science lingo, the FFT reduces the number of computations needed for voice to text problem.

1.2 Feature Extraction

Linear predictive coding (LPC) is a method used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model.

1.3 Call Training Phase Based on Sindhi Language Dictionary

To fabricate Sindhi speech recognition system dissimilar speech samples having different properties along with various formats from dissimilar sources have been collected. Including continuous speech and individual sentences the database contains single word and double words. From different cities of Sindh, the spoken words on various subjects was collected. The database was created after collection of large corpus of Sindhi words. In the next section we have given the different properties and comprehensive analysis of data collection. Different words and sentences to speak were given to the speakers. For continuous text as news and lessons the single words double words and sentences were very carefully selected. Male and female names, names of countries and the names of cities were chosen as these words. These words have been selected form name of cities in Sindh, names of countries and male female names and others.

1.4 Input Processing

The Input processing accepts processed input from users with different accents and process them based on Algorithm 2. The main goal of this algorithm is to process users requests based on their QoS requirements. The process of Algorithm 2 is determined in the following way.

- The algorithm takes N , A, Dm, π parameter as an input.
- Compute Tp$*$ based on equation (3 to solving the linear equation
- Calculate the best policy on the following equation $\pi = (Q, A, I, F) \leftarrow N \leftarrow \pi$
- For each input, users will get best policy in the system based on this condition $N \leftarrow A \leftarrow \pi$
- This process will continue until all inputs to be processed according to QoS requirements.

Algorithm 2: Input Processing

Input : $\{N, A, D_m, \pi\}$;

Output: Optimize: $T_p{}^*$;

begin

  foreach $(N \leftarrow A)$ **do**

    $N \leftrightarrow D_m$ Compute $T_p{}^*$ based on equation (3);

    Solving the linear equation;

    $\pi = (Q, A, I, F) \leftarrow N \leftarrow \pi$;

    if $(\pi^* \leq \pi)$ **then**

      $N \leftarrow A \leftarrow \pi$;

  End of Inner condition;

End of input process;

1.5 QoS Processing

- This study devises the Quality of Service (QoS) process method which aim to obtain the concave optimization based on equation (1) and convex optimization based on equation (2). The study suggests Algorithm 3 to verify the QoS requirements of users input in the model.

Algorithm 3: QoS Processing

Input : $\{N, A, D_m, U_m, \pi\}$;

Output: Optimize: $T_p{}^*$;

begin

  foreach $(N \leftarrow A)$ **do**

    $N \leftrightarrow D_m$ Compute error-rate based on equation (2);

    $N \leftarrow A \leftarrow U_m$;

    Solving the linear equation;

    $\pi = (Q, A, I, F) \leftarrow N \leftarrow \pi$;

    if $(\pi^* \leq \pi)$ **then**

      Calculate Accuracy based on equation (1);

  End of Inner condition;

End of input process;

The process of Algorithm 3 is defined in the following way.

- Algorithm 3 takes an input based on these parameters N , A, Dm, Um, π.
- All inputs must be matched with dictionary N ↔ Dm
- The QoS performance of user inputs to be measured via different states. The initial state, final state and transition state with proper action based on these parameter $\pi = (Q, A, I, F) \leftarrow N \leftarrow \pi$

- For each input, algorithm will find optimal policy i.e., $\pi*$ based on initial policy π and Calculate Accuracy based on equation (1)
- The entire process will repeat until all inputs to be get their results.

1.6 Fault-Tolerant

- In the system, the voice input or process could be trashed during simulation. Therefore, we devise a method which helps users to interact with system without any disruption. In order to avoid interruption of failure, this study devises fault-aware Algorithm 4 which handles any kind of failure in the model.

Algorithm 4: Fault Tolerant

Input : $\{(Q, A, I, F), a_{ij}, t = 0, 1, \ldots, T, N, A;$
}
Output: Optimize: T_p;

begin
    status=[0,1];
    foreach $(N \leftarrow A)$ **do**
        if $(T_p > 100)$ **then**
            $N \leftarrow A \leftarrow =0;$
            failed and status=0;
            Call Input Processing Method;
            Call QoS Processing Method;
        else
            $N \leftarrow A \leftarrow =1;$
            Success and status=1;
End of Main;

The process of Algorithm 4 is defined in the following way.

- This Algorithm takes many parameters from SRS-HMM model i.e., (Q, A, I, F), aij, t = 0, 1, . . . , T , N , A. The main goal is to identify error whether it is on the user level or model level.
- We defined status=[0,1] variable to identify system whether it is in normal mode or fault mod.
- N $\leftarrow$ A $\leftarrow$ =0 failed and status=0, then Call Input Processing Method and QoS Processing Method.
- If it is normal mode it will return Success and status=1.
- This algorithm repeats all process until users get their results with optimal QoS requirements.

1.7 Cases of Improvement

- SRS-HMM is a proposed algorithm framework which improves the input solutions via different as follows. Case-1: In the case-1, SRS-HMM considered the single character input by the users, and generate the output according to their requirements. SRS-HMM assigned weight to input values and compared it with the dictionary based on objective function. The goal is to minimize errors and improve the accuracy of users input from voice to text.

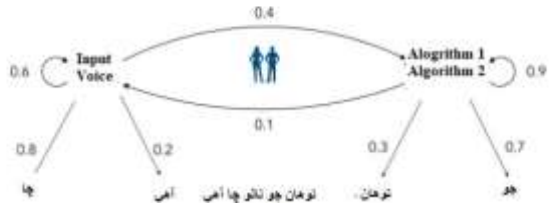

Figure 2: Case-1.

Case-2: In the case-2, SRS-HMM considered the double characters input by the users, and generate the output according to their requirements. SRS-HMM assigned weight to input values and compared it with the dictionary based on objective function. The goal is to minimize errors and improve the accuracy of users input from voice to text.

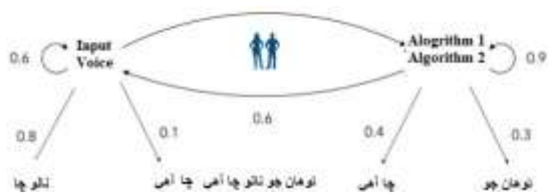

Figure 3: Case-2.

Case-3: In the case-3, SRS-HMM considered the single word input by the users, and generate the output according to their requirements. SRS-HMM assigned weight to input values and compared it with the dictionary based on objective function. The goal is to minimize errors and improve the accuracy of users input from voice to text.

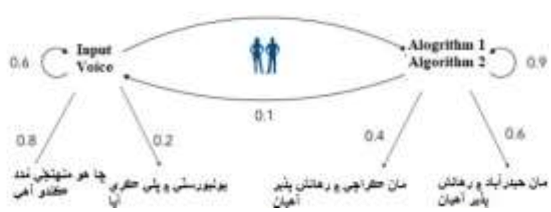

Figure 4: Case-3.

Table 2: Simulation Parameters

Simulation Parameters	Values
Windows OS	Linux Amazon GenyMotion
Centos 7 Runtime	X86-64-bit AMI
Languages	JAVA, XML, Python
Android Phone	Google Nexus 4, 7, and S
Experiment Repetition	160 times
Simulation Duration	12 hours
Simulation Monitoring	Every 1 hour
Evaluation Method	ANOVA Single and Multi-Factor
Sphinx4	System Implementation
Android Operating System	GenyMotion
Application interface	Voice to Text in Sindhi

V. Performance Evaluation

This section measures the performance evaluation of the proposed system based on different inputs with their accents. The system contains the different parameters as shown in Table 2. We built the system based on Sphinx4 framework in visual studio 2020.

5.1 Sample Collection

In research we have collected the samples in various forms and formats as single word names, double words names, Animal names, Fruits names, Birds names, Vegetables names and continuous sentences in Sindhi language. The database was created after collection of large corpus of Sindhi words. The double word names have been selected for collecting the samples of speech from various subjects. The single word, double word names and full sentenced selected are shown in Figure6. The database dictionary contains the speech samples spoken by various subjects the basic dictionary as shown in Figure 5. The subjects were asked to speak these animal names, birds names, vegetable and fruits names as samples were collected.

In this paper, we create the basic dictionary

5.2 Result Discussion

The study evaluates the performance of the system and scheme based on an experiment with different users and with different accents during the simulation. We set the 10,000 native users with 6 different accents during the simulation process. The length of voice 150kb WAV form with the 6 seconds. It means the input should be taken between 0 to 6 seconds, otherwise, the session of input will be finished.

Variable Name	Unicode	Sindhi Letter	Veriable Name	Unicode	Sindhi Letter
sheen	0634	ش	alifMadA	0622	آ
swad	0635	ص	alif	0627	ا
dad	0636	ض	beh	0628	ب
toye	0637	ط	beeh	067B	ٻ
zoye	0638	ظ	peh	067E	پ
aieen	0639	ع	beheh	0680	ڀ
ghain	063A	غ	the	062A	ت
feh	0641	ف	theh	067F	ٿ
peheh	06A6	ڦ	mytheey	067D	ٽ
qaf	0642	ق	tteheh	067A	ٺ
Kaf	06AA	ڪ	ttay	062B	ث
keheh	06A9	ک	jeem	062C	ج
gaf	06AF	گ	dyeh	0684	ڄ
geuh	06B3	ڳ	nyeh	0683	ڃ
ngoeh	06B1	ڱ	cheh	0686	چ
lam	0644	ل	cheheh	0687	ڇ
meem	0645	م	hah	062D	ح
noon	0646	ن	khay	062E	خ
rnoon	06BB	ڻ	dal	062F	د
waw	0648	و	dahal	068C	ڌ
heh	06BE	ھ	dhal	068F	ڏ
hamza	0621	ء	ddal	068A	ڊ
Yeh	064A	ي	ddahal	068D	ڍ
yehSmall	06C1	ہ	zal	0630	ذ
yehHamza	0626	ئ	reh	0631	ر
Min	06FE	۾	rdeh	0699	ڙ
sindhi Ampersand	06FD	۽	zeh	0632	ز
			sceen	0633	س

Sindhi Dictionary Basics

Figure 5: Dictionary of Sindhi Language.

Figure 6: Single Word and Double Word Inputs.

Based on simulation parameters and aforementioned numeric values in the system configuration file, we evaluated the performance of SRS-HMM system based on its components.

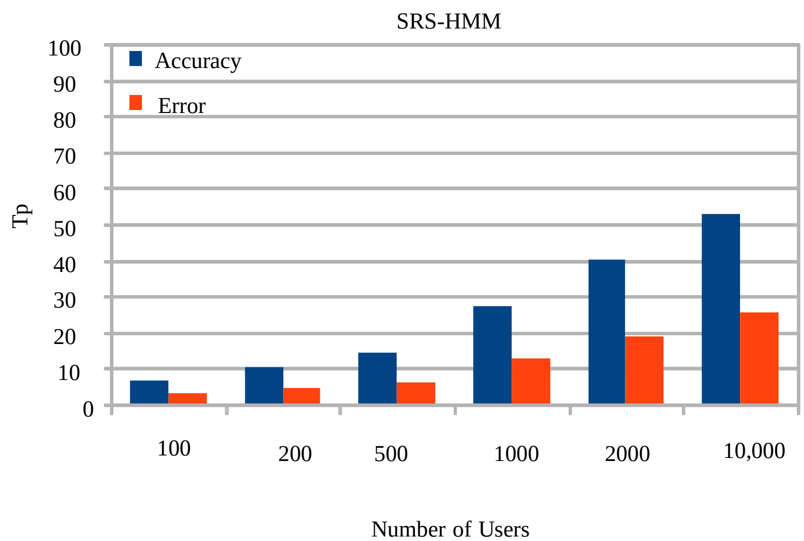

Figure 7: Bi-Objective Optimization of SRS-HMM.

Figure7 shows the performance of objective function Tp in terms of error-rate and accuracy of different 10,000 users with the same accent. With the help SRS-HMM, the system has good accuracy rate and minimum error rate. The SRS-HMM almost gain near-optimal solutions in the system. We called it is obtained optimal solution with is near 50% to 60%. There are a lot of reasons in the system which obtained up to 60% because of the following obstacles. (1) The dictionary has limited vocabulary still many words are required to insert. (2) Informal input also has less accurate results by different age people. (iii) The system can learn after different states as shown in Figure 8. In the end, we can say that we can improve the user's solution via different states.
Figure 8 shows that SRS-HMM can near-optimal solution of 10,000 users with different 6 different accents during simulation with different states. In Figure 8 with state-1, the system has good accuracy ratio, however, still, the error-rate is high in state-1. In Figure 8 with state-2, the system has a good accuracy ratio, improved the error-rate as compared to state-1. In Figure 8 with state-3, the system has a good accuracy ratio, improve the error-rate as compared to state-2. In Figure 8 with state-4, the system has a good accuracy ratio and improved error-rate of all users with different accents as compared to all previous state. It is showed our method is good and adopt dynamic changes and learn by different states and improve the user's solution. The study obtained the near-optimal solution of both concave and convex optimization based on users QoS in the system.

VI. Conclusion

In this paper, we suggest a novel Sindhi voice to text recognition system based on Hidden Markov Model (HMM). We devise a novel Sindh Recognition System Hidden Markov Model (SRS-HMM) based system and create novel Sindhi dictionary to deal with users with their Quality of Service (QoS) requirements. The study formulates the combinatorial problem where minimum error-rate considered.

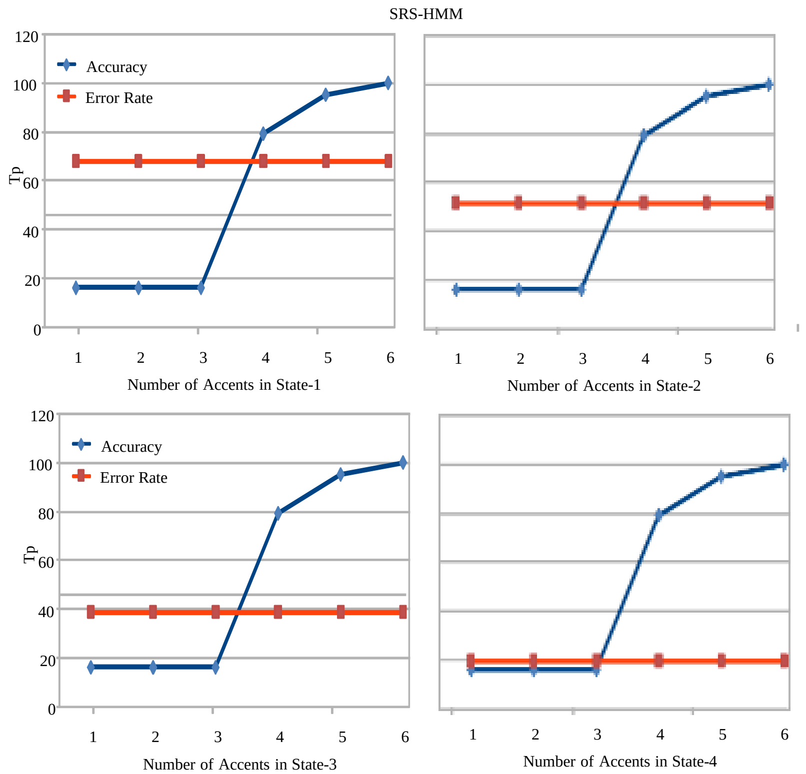

Figure 8: Optimal Results in Different State of SRS-HMM.

as a convex function and maximum accuracy formulated as concave function. The objective of the study to improve user inputs accuracy and minimize the error rate of all given inputs with different accents. Based on simulation results, we showed that, proposed SRS-HMM gained near-optimal, e.g., 60% of different users with different accents during the performance. However, there are many limitations in the system. (i) It accepts less number of accents in the current system. (ii) The Sindhi dictionary has limited vocabulary and cannot support all Sindhi voice words. (iii) The precision, and user input length only support for small length i.e., 6 seconds.

In the future work, we will improve this system with above challenges, and improve the dictionary with the given solution. In the future work, we will propose a novel system for healthcare applications in the Sindhi language with 100 accents and 50,0000 users support.

VII. References

[1] Elharati, Hussien A., Mohamed Alshaari, and Veton Z. Kpuska. "Arabic Speech RecognitionSystem Based on MFCC and HMMs." Journal of Computer and Communications 8, no. 03(2020): 28.

[2] Amich, Heithem, and Mounir Zrigui. "Improvement a Transcription Generated by an Automatic Speech Recognition System for Arabic Using a Collocation Extraction Approach." for the Industry (2019): 85.

[3] Alshayeji, Mohammad, and Sari Sultan. "Diacritics effect on arabic speech recognition." Arabian Journal for Science and Engineering 44, no. 11 (2019): 9043-9056.

[4] Su, Yu-Sheng, Chien-Hsing Chou, Yung-Long Chu, and Zhao-Yu Yang. "A Finger-Worn Device for Exploring Chinese Printed Text With Using CNN Algorithm on a Micro IoT Processor." IEEEAccess 7 (2019): 116529-116541.

[5] Prihandi, Ifan, Indra Ranggadara, and Boy Yuliandi. "Development of Text Recognition Pro- totype with Classification of Neural Networks AND Text-To-Speech in Javanese Scripts Using Incremental Methods." (2019).

[6] Bakken, Jeffrey P., Vladimir L. Uskov, and Nivee Varidireddy. "Text-to-voice and voice-to-text software systems and students with disabilities: a research synthesis." In Smart Education and e-Learning 2019, pp. 511-524. Springer, Singapore, 2019.

[7] @articledharmale2019evaluation, title=Evaluation of Phonetic System for SSSSSSpeech Recognition on Smartphone, author=Dharmale, Gulbakshee J and Patil, Dipti D, journal=International Journal of Innovative Technology and Exploring Engineering (IJITEE), volume=8, number=10, year=2019

[8] Khan, M. Y., and N. A. Siddiqui. "Sign Language Translation in Urdu/Hindi Through Microsoft Kinect." In IOP Conference Series: Materials Science and Engineering, vol. 899, no. 1, p. 012016. IOP Publishing, 2020.

[9] @inproceedingskhan2020sign, title=Sign Language Translation in Urdu/Hindi Through MicrosoftKinect, author=Khan, MY and Siddiqui, NA and others, booktitle=IOP Conference Series: Ma- terials Science and Engineering, volume=899, number=1, pages=012016, year=2020, organiza- tion=IOP Publishing

[10] Jason, C. Andy, and Sandeep Kumar. "An Appraisal on Speech and Emotion Recognition Tech- nologies based on Machine Learning." language 67: 68,(2020).

[11] Sodhar, IrumNaz, Akhtar Hussain Jalbani, and Muhammad Ibrahim Channa. "Identification of Issues and Challenges in Romanized Sindhi Text." Editorial Preface From the Desk of Managing Editor 10, no. 9 (2019).

[12] Meurers, Detmar. "Natural language processing and language learning." The Encyclopedia of Applied Linguistics (2019).

[13] Anusuya, M. A., and Shriniwas K. Katti. "Speech recognition by machine, a review." arXiv preprint arXiv:1001.2267 (2010).

[14] Prasad, Vijay. "Voice recognition system: speech-to-text." Journal of Applied and Fundamental Sciences 1, no. 2 (2015): 191.