Towards Creation of Sindhi Speech Database

Fida Hussain Khoso[*,1],Dhani Bux Talpur[2,3],Shafiq Ahmed Awan[4],Waheed Ali Abro[5], Shah Pisand Chandio[6], Muhammad Zahid Tunio[7], Dr. Hyder Abbas Musavi[8], Asadullah Shah[9]

[*,1]Department of Computer Science, Dawood University of Engineering and Technology Karachi 74800, Pakistan.
[2]Department of Information and Computing, University of Sufism and Modern Sciences, Bhitshah Pakistan.
[3]Kulliyyah of Information and Communication Technology, Islamic International University Malaysia.
[4]Department of Computer Science and IT Benazir Bhutto Shaheed University Lyari Karachi.
[5]Department of Petroleum and Gas Engineering, DUET, Karachi Pakistan.
[6]Government Degree College K.N.Shah Dadu Sindh Pakistan.
[7]Deptt of Computer Systems Engineering Dawood University of Engineering and Technology Karach Pakistan
[8]Department of Electronic Engineering Mehran University of Engineering & Technology, SZAB Campus Khairpur Mir's Sindh, Pakistan.
[9]Kulliyyah of Information and Communication Technology, Islamic International University Malaysia.

Abstract

Interacting with machines by human voice is called speech recognition systems. These speech recognition systems are built separately for each of the human language. The challenge of the implementing these systems are the changing rules and regulations of the human language. So, it is very much necessary to build a separate system for specific language. Many of the Latin languages, Indian languages and others have their own speech recognition system. Sindhi an official language of Sindh province is still lacking this system as no speech recognition system have been found (to the best of our knowledge). For the development of any speech recognition system a basic element of database is necessary where the rules and regulations have been defined so that the speech recognition system can work upon. A comprehensive database has been created for the purpose of Sindhi speech recognition for the Sindhi language and for the purpose various parts of Sindh have been visited to collect speech samples spoken by various subjects around Sindh province so that the versatile database for the Sindhi SRS can be built. The subjects were requested to speak selected single word, double words, sentences and the continuous speech. An analysis of the comprehensive speech database is presented in this paper.

Key words: *Speech Synthesis, recording environment, Acoustic Model, Datasets.*

1. INTRODUCTION

Human speech is considered the way of communication among various generations and group of people especially the area driven tribes. Many of the groups use spoken language to communicate with in the group of human being and this is reason many of the spoken languages are available in the world. For the communication with computers a human spoken language is preferred as it is one of the fastest methods [1-4].

Many of the research areas are engaged to communicate with computers and speech recognition is one of the areas where the human being is performing interacting to computers by human spoken languages. In order to achieve the goal many of the language rules and regulations are required to be understood and taught to the computers [1]. The main challenge of the speech recognition is considered the variance of the language and the computer must understand the various rules and regulations of specific language which means a speech recognition system for one language cannot be used with the other and each of the language need a separate algorithms and methodology [2].

The speech recognition can save a lot of time and play a vital role in the human society especially when many of the languages are applied and multilingual systems are required [3]. The country like Pakistan has many of the languages spoken and there is a need of multilingual system which may support many of the languages can be incorporated in a single system. As a matter of resources there is a need of an expert of computer science and a language expert [5].

Many of the languages are equipped with their speech recognition system and their speech recognition systems are near to the perfection level along with a good accuracy. Latin languages such as English, German and French possess a good level

of accuracy in the SR systems whereas the much of the work has been done on Indian languages such as Hindi [11], Bengali [2], Punjabi [8], Gujarati [34], Marathi [7], Malayalam [6], Khannada [5], Tamil [35], Urdu [9][13], Telugu [36], Garhwali[4] and others [3]. Many other Indian languages [3] are in process of building speech recognition system [4]. Much of the work on Indian languages have been done by HP labs India and IBM research lab [1][15].

For the sake of recognition system of any language a database is the prerequisite so that an efficient recognition is based on the available samples collected of that particular language [4]. There is very small work done on Pakistani languages reasoning many aspects including financial support and others [30]. To the best of authors there is no work has been done on Sindhi language which is rich in history and culture. Much of the Sindhi computing research such as corpus construction [10], Sindhi text segmentation [23], Sindhi Entity Recognition System [24][25], Sindhi OCR [26][28][29][30][31][32], Sindhi Information retrieval from images [27] is availablebut Sindhi speech recognition needs a lot of attention. This paper presents the speech database creation along with steps of sample collection from various resources. The paper also presents the recognition of selected words.

2. RELATED WORK

A speech database is required for the process of speech recognition an element that provides the platform for testing and training of speech recognition systems. Many of the databases are available for various languages. The larger size of the database indicates more samples for testing and training. The sample collection varies in size and available in various forms such as single word, double word, individual sentences and continuous sentences. Various speech databases have been created in different languages by various institutions. Different environments have been used for voice recording purpose such as noise free environment and noisy environment inside or outside for the particular purpose. Much of the combined speech databases have been developed in different languages in Indian institutes such as Marathi, Telugu, Hindi, Garhwali, Punjabi, Kannada, Marathi, Tamil & Telugu and so on.

In Algerian Arabic Speech Database (ALGASD) through Modern Standard Arabic (MSA) speech corpus is presented. A total of 300 native speakers haven been requested for the collection of samples. Speakers were chosen from 11 regions of Algeria. Three age groups were selected including adult speakers between 18 & 30 years of age, speakers of middle age from 30 to 45 years and speakers more than 45 years. In more than twenty-two countries the Arabic Language is used as an official language with nearly three hundred millions of Arabic speakers. The Algerian official language is MSA. The Algerian speech database is speech collection which focuses and brings lights on important features of pronunciation which are related to some regional and social variations of Algerian speakers [1].

Another speech database of Bengali Speech Database is Bengali Continuous Speech recognition is presented with the continuous samples for independent speakers. Assurance is provided by another research that speech corpora are the core of Automatic Speech Recognition (ASR). For the development phonetically balanced text corpora. The selection techniques and optimal process were used in this study. The unique sentences around 7500 and 19640 unique words spoken by 70 male and 40 female speakers have been used for the purpose of text corpus development. The database has been divided into two parts by age groups. Two age groups have been described comprising of 20 to 40 years and then 60 to 80 years for older age groups. Three types of Phonemes have been used namely phone, mono phone, tri phone and labelled the speech corpora have been developed. Hidden Markov Model Toolkit is used as a classifier in this paper [2].

After the review of some databases another database Indian speech database have been discussed in this paper. In this paper the total no of speech languages and various databases have been created by the various institute. The recording environment was noise free, the number of speakers were 1500 and the database created by TIFR Mumbai and IIT Bombay. The recording devices for the voice sample collection were used as cell phone and voice recorder. The application of database was speech recognition for Agriculture purpose and the Marathi language was used to create database. Another database is also created by C-DAC Noida which was also noise free environment and depends upon 30 speakers. The recording devices were used standrad Mics for the creation of Speech Recognition System for travel domain database which was created in Hindhi language. Another database was created by the Government of P.G college Rishikesh, which was based on noisy environment, no of speakers were 100, recording deviceswere used Standrad Microphones. The database application was created for speech recognition system and language was used as Garhwali. In the 13th differen languages the databases have been created including Marathi, Telugu, Hindhi, Punjabi and Garhwali etc. The different databases have been created in the various environments, different recording devices were used. The database were created for different fields just as agriculture , speech recognition and travel agent [3]. Another speech database have been presented which is Garhwali Speech Database in this paper the author highlited Automatic Speech Recognition System and on the basis of enhancement of ASR that will be used for Garhwali language. Garhwal region is based on local dialect of Garhwali Hindhi which is spoken in Uttarkhand state of India. For the enhancement of ASR the primary goal of Garhwali speech database is to provide that facility. The definition or classification of independent of the speaker is that through the ASR system any person who speaks

Garhwali words and sentences fluently that identifies quickly and easily by the system. In Garhwali language Devanagari script is also used. For character represntation there are additional constants, zero and one vowel. A bunch of consoents is represetnted by 100 of characters and some ligatures. In the Garhwali speech database the procedure for recording were utilized on the different microphones with the help of PRAAT speech software in the front of computer and all scattred words and sentences were used. The directional microphone was kept at the distance of 1 meter from the mouth therefore information was influenced through room acoustic the computer noise fan near to the computer, that was the recording procedure which made in such a way the microphone by cardioids was placed at the fifteenth cm from the mouth. The noise speech for rational at the same time recording and for the development of vigorous speech recognition system which were consumed for high quality speech [4]. Other than speech database have also been discussed which is Kannada speech recognition system that solving the problems of said database. The design algorithm were identified and work can be done through the continuous state in the form of speaker dependent and Hidden Markov Model are used in this database as classifier.

An acoustic speech signal to text is prepared and made by computer by the process of Automatic Speech Recognition (ASR). To enable computers to translate spoken words into text and commands and accepts speech input, and this is aim of to develop techniques for speech recognition. Since 1950 the speech recognition problem was being actively discussed and studied and it was thought that it is natural why one should be processed to study speech recognition [5]. In the Emotion Speech Recognition System, the most important part considered to be is Emotional speech database. For specific emotion and identification from speech samples, quality is an essential factor and it is the basic requirement. There is no built-in standard database in Malayalam language. For emotional speech database this task has been fabricated. Four basic sentimental categories such as natural anger, happy and sad are included in developed emotional dataset which is acted with the help of no-professional actors. Speakers of the ages between 18 to 50 years of male and female genders have been chosen for samples collection. In this work of database creation 08 female and 08 male have participated. Each speaker was given four emotions in 20 sentences in this task. They were trained well before recording. The recording work was done by microphone in a very high quality studio of author and after that the samples were stored in the databases' [6].

The Automatic Speech Recognition System have been developed and elaborated in this paper and the first step for the development is speech database. The computer speech has been used as potential as a mode of interaction. The computer has been mobilized as to talk and understand just like human being. The research workers have struggled to develop the system for the analysis and categorization of speech signal. The computer scientists since 1960 have been making research to understand, interpret and record human voice. It is commonly understood that computer system is very useful in various sectors of social life such as health care, education and government services to understand the spoken language. The existing data in digital world if one wants to study and digest a specific language it is easily accessible. The Marathi people of central and western India speak Marathi language which is an Indo-Aryan language. The fluent speakers of Marathi language are about 68 millions of people in all over the world. It is similar to the National language of India whereas Devanagari script is used in Marathi language. Utilizing the website related to agriculture the data for text corpus was collected and generated. For the websites the blogs articles which were published were chosen. From the websites 100 of the words were selected, then classified according to their groups and categories such as cash crops, fruits, vegetables, grains, fertilizers, esticides, diseases and equipment's [7].

3. SINDHI SPEECH DATABASE

Various speech databases have been checked and analyzed for the suitability but none of the speech database is suitable and usable for the development of Sindhi speech recognition system. Reviewing of the literature suggested a separate speech database is needed to train and test the speech recognition system for Sindhi. As there have been no speech recognition systems available for Sindhi Language (to the best of our knowledge), in this regard there is a need to create a speech database from the scratch.

3.1 Sample Collection

Speech samples were collected from various sources for the aim of collecting data along with different properties in various formats. The continuous speech, individual sentences, single and double words are considered in database. With the collection of large corpus of Sindhi words which are spoken by different subjects all around the cities of Sindh has been used for the speech database creation**.** In the next section the various properties and comprehensive analysis of the data collection are given. For this purpose, the speakers were assigned different words and sentences to speak. We carefully selected the sentences and words for example single words, double words, sentences and the continuous text extracted from various sources including news and lessons. These words are selected from the names of cities in the province of Sindh, names of countries, male and female names common in Sindh province and other sources. In Table.1 illustrates the single word names.

Table 1: Names as Single word

هڪ لفظ وارا ماڻهن جا نالا		
علي	ماجد	حمل
منير	عمر	ميداد
چڪور	حميد	پناهو
هوشو	اقبال	ملوڪ
آصف	ستار	سارن
سومر	اصغر	سدورو
ڪرشنا	ڪمال	سونهن
آمر	سڪندر	پورڻ
اسد	رمضان	چيو
هڪاڻ	نواز	موڻي
سلمان	نور	پنهور
سادق	ظفر	ڪرڙ
سرمد	آچر	جمع
احسان	جابر	سسي
عثمان	شاهو	ڪاڻيرو
سمير	نبين	خميسو

Table 2: Names as Double word

بن لفظن وارا ماڻهن جا نالا		
اختيارعلي	ذاهد علي	الهه ڈنو
امجد علي	جميل احمد	الله رکهيو
اجمل حسين	عبدالرشيد	الهه ورايو
واحد بخش	غلام قاسم	اماالدين
فدا حسين	آفتاب حسين	محمدحنيف
غلام حسين	علي بخش	امامبخش
عبدالغفار	ادريس خان	عليابکير
غلام محمد	حسين بخش	عليااصغر
غلام علي	گل محمد	مارقعلي
خان محمد	قيصر خان	صداقتعلي
محمد خان	غلام ياسين	محمدعلي
غلام عباس	حماد حسن	عليرضا
غلام مصطفي	الهه داڈا	ايازعلي
علي خان	غلام سرور	ذاڪرعلي
غلام حيدر	پوتاخان	جاويدعلي
عبدالستار	قاد داد	صادقعلي
سومرخان	رميز عباس	اسدحسين
سوداخان	خداداد	ڪاشفعلي

From various subjects the double word names are chosen for the collections of samples of the speech. In Table 2 the selected double word names are given. The speech samples which were spoken by different subjects from the cities of Sindh are contained in the database are shown in Table 3. Table 4 shows the countries names were spoken by subjects.

Table 3:Names of Cities in Sindh

سنڌجيشهرنجانالا		
ڪراچي	خيرپورميرس	ڪنڊيارو
حيدرآباد	پيرجوگوٺ	سيتا
جامشورو	سڪرنڊ	ڈوڪري
سيهوڻ	گمبٽ	روهڙي
دادو	ٽنڈو محمد خان	ڪوٽڙي
خانپور	اسلامڪوٽ	پياروگوٺ
ٽل	سمارو	خيرپورناٿشاه
ميهڙ	ميرپرماڻيلو	نوشهروفيروز
نصيرآباد	مٺي	سانگهڙ
وگڻ	ٽنڈوالهيار	رتوديرو
لاڑڪاڻو	شهداد پر	ميهراب پور
قمبر	سجاول جوڻيجو	خداباد
شهداد ڪوٽ	ٺٽا	آمري
واره	ميرپرخاص	پيٽارو
مورو	ڈهرڪي	راڻيپور
نوابشاه	عمرڪوٽ	ٽنڈوجام
سن	شڪارپور	ٽريميرواه
ناروشاه	سکر	ڪنڊو سعد خان

Table4:Names of Countries in Sindhi

ملڪن جا نالا		
پاڪستان	سويزيلينڈ	بيلجم
هندستان	جرمني	ايشيا
ايران	سومالييا	روانڈا
افغانستان	بيلارس	رشيا
آسٽريليا	بنگلاديش	ترڪي
آمريڪا	فرانس	تاجڪستان
جاپان	سنگاپر	اندورا
چين	ڪينيڈا	تاجڪستان
شام	اسپين	البانيا
سريلنڪا	موروڪو	بيلجم
رومينيا	ڪينيا	ٽونگا
اومان	انڈونيشيا	توگو
پولانڈ	اٽلي	تنزانيا
يمن	بهرين	تيونسا
ناجيريا	ميڪسيڪو	لنڊن
نيوزلينڈ	مالي	سوڊان
نيپال	موناڪو	امريڪا
برما	سعودي	ملائيشيا

In Table 5 the speech data samples the selected sentences are presented. The sentences were selected carefully so that the various sounds and characters for Sindhi language can be represented. The data collection was performed in following districts shown in Table 6.

Table5: Some Sentences in Sindhi

1.	تنهنجو نالو ڇا آهي؟
2.	اڻون حيدرآباد ۾ رهندي آهيان
3.	اڻون ڪراچيءَ ۾ پڙهندو آهيان
4.	توهان سان ملي ڏاڍي خوشي ٿي
5.	هو ڪهڙي اسڪول ۾ ويندو آهي؟
6.	سڀ جو مالڪ هڪ آهي
7.	پلي ڪري آيا
8.	علي اسڪول جو هوم ورڪ پيو ڪري
9.	اهو سنڌي هوليءَ جو مضمون آهي
10.	ڇا هو منهنجي مدد ڪندو

Table 6: List of the Selected Districts

S.No	The list of ten districts of Sindh		سنڌ جي 10 ضلعن جا نالا	
1	Jacobabad	(Northern Sindh)	(اترسنڌ)	جيڪب آباد
2	Shikarpur	(Northern Sindh)	(اترسنڌ)	شڪارپور
3	Larkana	(Northern Sindh)	(اترسنڌ)	لاڑڪاڻو
4	Dadu	(Lower North)	(هيٺيون اتر)	دادو
5	Naushraro Firoze	(Central Sindh)	(وچولو سنڌ)	نوشهروفيروز
6	Umarkot	(East Sindh)	(اوپر سنڌ)	عمر ڪوٽ
7	Tharparkar	(East Sindh)	(اوپر سنڌ)	ٿرپارڪر
8	Badin	(Southern Sindh)	(ڏکڻ سنڌ)	بدين
9	Thatta	(Southern Sindh)	(ڏکڻ سنڌ)	ٺٽا
10	Karachi	(west Karachi)	(اولهه)	ڪراچي

3.2 Procedure for Creation Database

For the purpose we have selected 10 districts of Sindh province covering upper, middle and south regions of the province. The different sources of recording were used such as laptop, online recorder, smartphones, voice recorder and other resources have been used. For the data collection we have selected ten districts from province of Sindh and the detail has been given in Table 6. The recorded material has been stored in the directory form along with information file so that the researchers performing the recognition process can easily identify the accuracy of the system. The recorded files were processed and extracted information with the help of MATLAB Software so that the coding information or feature extraction information can be extracted automatically and stored in a file. The original audio signals along with information file is shown in Figure 1.

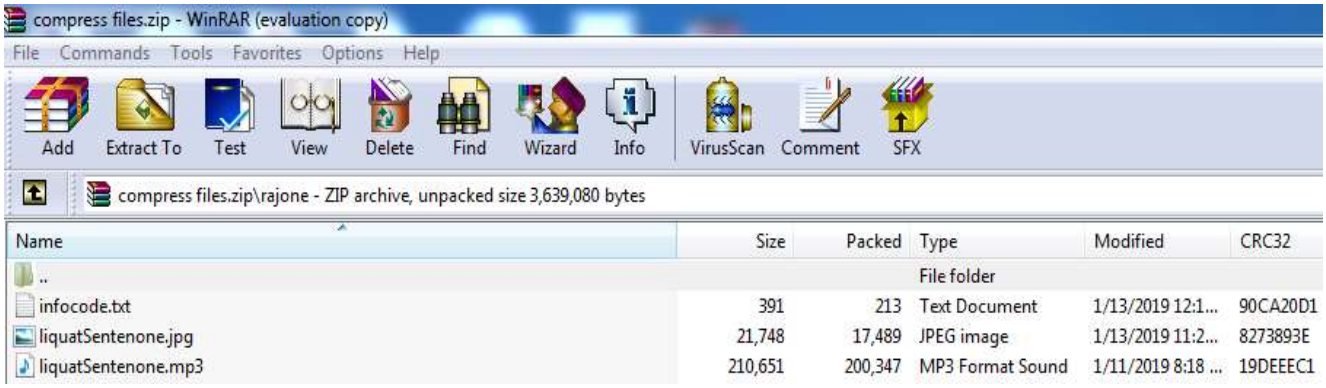

Fig 1: Generated files types

One out of hundreds of the samples collected for the purpose of database is given in Figure 2 where Mr liaquat Ali Rajper has spoken 17 individual sentences which were arranged into three files type as analogue signals, MP3 and info or in wave form showing the properties in Figure 2. The text file contains the information about the spoken speech in form of technical and non-technical information. The technical information contains the properties such as number of channels, sample rate, total number of samples, duration bit rate and other information which may be helpful in the process of speech recognition. The non-technical information may help in comparing accuracy and other parameters such as the name of subject, duration, title, environment, recording device and others.

Fig 2: List of combined 3 types of files

The wave forms have been extracted in each file so that the process of speech recognition can be helped out. The wave form generated from the speech of Mr Liaquat Ali Rajper spoken is shown in Figure 3.

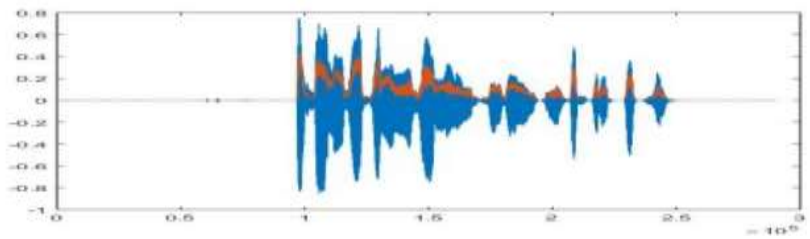

Fig 3 : Wave form of Liaqat Rajper Speech sample

3. Travelling

For the collection of data samples of Sindhi speech database, we had chosen 10 districts of Sindh. The criterion of the choice is that these 10 districts were taken from the central, South, East and North of the province. In the process of voice recording we did face many problems for example non accessibility of proper persons and non-availability of environment which was not noise free sometimes. For the collection of data there was a big challenge of travelling to specific places such as convenience, environment and the time of the respected subjects of various walk of life. While going to collect voice data we have taken one or two villages or towns from each district. The speakers were divided age wise, native and non-native, male and female and then the data has been collected. We started the data collection from author's district west Karachi and in this district, we selected Orangi town taluka which is the biggest one and its population is about 30 lacs.

We started from Karachi and continued to visit various parts of Sindh province, requested subjects and recorded speech samples. From taluka of Karachi west, we have chosen two villages Mangho pir of Orangi Town and Baloch Goth. The reason of selecting Baloch Goth was that the different persons of Sindh having different accents from each other are living in this Goth. In the starting phase we selected both male and female subjects, native and non-native subjects and requested to read single word, double words, sentences and a continuous speech in Sindhi text. The detail of the total samples recorded is given in Table 7.

Table 7: Details of Database samples

District Name	No of Speaker	Single Words	Double Words	Sentences	Continuous speech	Total Recording time
Larkana	10	200	230	50	10	06:48:33
Dadu	20	250	270	44	10	19:20:00
Karachi	25	275	300	65	10	27:36:06
Jacababad	15	210	220	55	10	14:52:00
Shikarpur	10	150	200	53	10	09:00:00
N-Feroze	15	150	200	60	10	13:41:05
Umarkot	10	200	230	30	10	09:25:34
Tharparker	12	170	190	52	10	10:44:23
Thatta	20	250	265	58	10	20:54:39
Badin	15	150	250	45	10	14:23:26
Total Time of collected samples						146:45:46

An overall one hundred and forty-six hours and forty-five minutes' audio samples have been collected for the purpose from ten districts of the Sindh province. The single word and double word combination have been given to the subjects and their speech or spoken samples were collected. The maximum number of samples were collected from Karachi district as it is the native city of the authors, so the more efforts have been exerted in this district Karachi.

4. Proposed Speech Recognition System

This thesis, presents a novel speech to text recognition system for the Sindhi language. The system performs all actions via different steps. Furthermore, we explain all steps into the following.

Random Voice: The users generate the random input to the system. For instance, an input is a voice, however, the voice accent could be changed by the users (e.g., young people, children, male and female).

Speech Analysis: It is filtration in the system, where noise, interruption, and related variance of input data purified. For example, a user give an input to the system in the noisy and roaming environment. Definitely, it will gain an extra disturbance variance in the original data. Therefore, filtration in the system is an effective way to pass original data to the remaining components in the system.

Match Attributes: There are many attributes in the considered system related to the user's data. We define the attributes as a tuple (e.g., input type, user accent, speech device, given accuracy range, and error threshold value). All the given attributes are strict constraints in the process, therefore it is necessary to meet the requirements of all attributes during processing phase.

Robust Processing: The system process the data based on the interactive data without any delay. However, existing studies have exploited queuing system for the data, it takes an extra time to process the data in the system.

Get Attributes: In the system, the attributes are defined as a tupe as mentioned above. Nonetheless, the proposed algorithm sophistically to be met the requirements of users data based on the given attributes.

Speech Recognition: is the aptitude of a engine or plan

to recognize sentences and in verbal speech and change them to a machine-readable format.

Proposed Model Improved Hidden Markov Model: The existing HMM does not covers the random inputs in the model, therefore, it is required to modify the input according to the proposed system. The system exploits many fields as a hidden filed such as random input. The decision process follows the discrete and continuous input too, but it is not concern to the system. The final component is the output phase, where internal system to be generated output based on the given proposed algorithm. However, it is not necessary users get the same output on different devices, it is uncontrollable factors due to the heterogeneity of the devices.

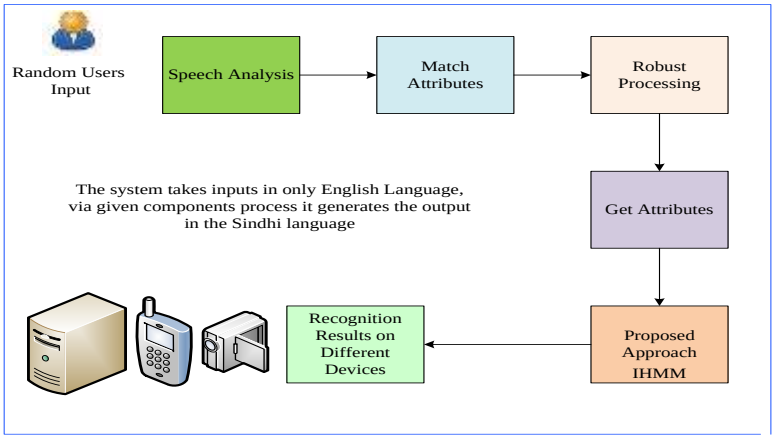

Fig 4: Proposed Improved Hidden Markov Model Based Speech Recognition System

IHMM Algorithm

The Improved Hybrid Markov Model is a discrete model which accepts input and match to the given Sindhi database of words. There are many parameters to be matched for the voice recognition system during process. The parameters are defined as tuple (DB, a, b, c, d, e, in, s, ac).

Algorithm 1: IHMM Scheme

Input: DB, a, b, c, d, in, s, ac,

Z=null, initial solution of users

Step-1: for ($in \in DB$) then

Step-2: If ($a \le b \,\&\, c \le d$)

Step-3: All given constraints must be equal to 1

Step-4 $x_{a,b,c,d} = 1$

Step-5 call the original solution Z

Step-6: if ($a > b \,\&\, c > d$)

Step-7 still the solution to be improved

Step-8 Call local search to improve the original solution Z

Step-9 call Algorithm 2

Step-10 $Z \to Z^*$ optimize initial by the local search

Step-11 if ($Z \le Z^*$) remain initial solution stable

Step-12: Calculate the error rate e by Boltzmann constant

Step-13: End main loop

Where DB is a database of Sindhi words, and a shows an accuracy level of voice, b denotes the voice level of accent, c illustrates that voice ratio of input, and d demonstrates the originality of voice. We assume that, all inputs and constraints

are given in the advance to the system. Therefore, each state has proper action, i.e., the original text accuracy to be calculated based on the given constraints. The proposed algorithm is defined as below.

1. Step-1 takes input of different voices, database and constraints with set of states and actions. Each action shows the solution Z with different states.
2. Step-2 verifies that, all constraints to be satisfied when input voice match to the database.
3. When all constraints must be equal to 1, when they are satisfied by the system as defined in step-3
4. X is a binary, which shows the status of all constraints during initial solution as defined in step-4.
5. Z is an original solution in step-5, if all constraints are not satisfied, then initial solution still to be improved by the local search algorithm as defined in step-6, 7, 8 and 9.
6. Z^* is the improved solution of original search via proposed local search method as defined in step-10 and 11.

System still calculate the iteratively error rates of inputs until original to be optimally Improved for all users explained in step=10 to 12. Step-13 will terminate the all condition when original solution is improved. Boltzmann constant is a method which calculates the error rates of the system.

7. Here e ensures that, all users could get satisfied results during voice to text recognition process in IHMM.

Local Search Scheme

The proposed local search scheme will improve existing solution with new improved from search space. Therefore, this thesis exploited exhaustive search to avoid from local optima (e.g., initial worst) in the thesis. The initial could be improved by the proposed local search scheme because of less of accuracy, thesis show the strength of local search via experimental results. The global improvement in the system is done via variable neighbour search, which search the many solution and choose the better one random than existing from the entire space. There are few techniques for searching mechanism, for instance, random searching, sequencing searching, and mid-searching to find the best solution. However, we are using random searching which choose the optimal solution which is optimal than existing one. It is less time taking and the complexity of the algorithm is improved than existing local search methods. The proposed local search algorithm for the optimal solution as explained as follows.

1. The algorithm takes initial solution as an input, and some related parameters such as iterations, and cooling parameters like simulated annealing method.

2.In step-1, the iteration starts with zero, and system will search the optimal from search space N as defined in step-2.

3.The rate of change of between solutions is calculating in the step-3. It means, system match initial solution with new solution, may be some solution return as a worst solution. However, system only chooses optimal solution among new solution as defined in step-3 and 4.

4.From step-5 to 8, system ensures that the new solution will be optimal solution and not stuck with the local optima. The obtained solution is a global optimal solution in the proposed local search method.

5.If there is not optimal solution in the search space, the initial solution will remain in the system.

6.Whereas, step-10-13 encounter the given iteration with given values, and return the optimal solution to the system.

The complexity proposed local search scheme is calculated via time and space. It is equal to. Here log represents the exhaustive search, and n represents the number of matching initial solution, and m represents the new obtained solutions via search space.

Algorithm 2: Local search scheme

Input: Z

Output: Z*

tmp=1000 ;

α =Cooling-factor

tr =100 number of iteration

Begin

Step-1: While (tmp>0)

Step-2: choose best Z* solution from search space N

Step-3: $\Delta = \frac{f(Z^*) - f(Z)}{tmp}$

Step-4: If ($\Delta < 0$)

Step-5: $Z \rightarrow Z^*$

Step-6: ($f(Z) < f(Z^*)$)

Step-7: replace original solution with new one

Step-8: $Z^* \leftrightarrow Z$

Step-9: Else

Stay with current solution

Step-10 $tr = tr - 1$

Step-11: $1 - \alpha$

Step-12 Return Z*

Step-13: End Main

5. CONCLUSION

After reviewing and comparing we have concluded that the requirement cannot be accomplished by available databases, in this regard a comprehensive Sindhi speech database must be created to achieve the goal of Sindhi speech recognition. We have created Sindhi speech database in accordance with the state-of-the-art approach followed by various researchers of the speech recognition. Analysis of the various available databases has been given and compared so that a comprehensive and large speech database can be built. For the collection of different dialects and accents we have gone to various areas of Sindh to collect voice samples for providing database a versatility. It was time taking and required patience to request various subjects as the time and appointment was not so easy and faced many problems. The remarkable and intense efforts were made with multiple voices to build this large database for Sindhi speech. To facilitate other research scholars a very new window in Sindhi language computing will be opened for speech to text recognition system.
The speech database contains sentences, double and single words and continuous speech having different accents of various subjects selected from parts of Sindh province and the samples in audio are thousands in numbers. With the combination of samples of sound further, the recording database can be increased. The database is adaptive and can be enhanced by adding more samples whereas the database can be extended to multilingual speech recognition systems by adding more samples from other languages and collecting various samples all around Pakistan.

6. REFERENCES

[1].Selouani, Sid Ahmed, and Malika Boudraa (2010)."Algerian Arabic speech database (ALGASD): corpus design and automatic speech recognition application." *Arabian Journal for Science and Engineering* 35, no. 2C: 158.

[2]. Das, Biswajit, Sandipan Mandal, and Pabitra Mitra (2011). "Bengali speech corpus for Continuous Automatic SpeechRecognition System." In *Speech Database and Assessments (Oriental COCOSDA), International Conference on*, pp. 51-55. IEEE.

[3]. Shrishrimal, Pukhraj P, Ratnadeep R. Deshmukh, and Vishal B. aghmare (2012)."Indian language speech database: A review." *International journal of Computer applications* 47.5: 17-21.

[4]. Upadhyay, R. K., and M. K. Riyal (2010). "Garhwali speech database." *Proceedings of O-COCOSDA.*

[5]. Punitha, P., and G. Hemakumar (2014). "Speaker dependent continuous Kannada speech recognition using HMM." In *Intelligent Computing Applications (ICICA), 2014 International Conference on*, pp. 402-405. IEEE.

[6]. Rajisha, T. M., A. P. Sunija, and K. S. Riyas (2016). "Performance analysis of Malayalam language speech emotion recognition system using ann/svm." *Procedia Technology* 24: 1097-1104.

[7].Shrishrimal, P. P., R. R. Deshmukh, and Vishal B. Waghmare (2013). "Development of isolated words speech database of Marathi words for agriculture purpose." *Asian Journal of Computer Science & Information Technology* 2, no. 7.

[8]. Singh, Parminder, and Gurpreet Singh Lehal (2010). "Corpus based statistical analysis of punjabi syllables for preparation of punjabi speech database." *International Journal of Intelligent Computing Research (IJICR)* 1, no. 3.

[9]. Becker, Dara, and Kashif Riaz (2002). "A study in Urdu corpus construction." In Proceedings of the 3rd workshop on Asian language resources and international standardization-Volume 12, pp. 1-5. Association for Computational Linguistics.

[10]. Rahman, Mutee U (2010). "Towards Sindhi corpus construction." In *Conference on Language and Technology, Lahore, Pakistan*.

[11]. Samudravijaya, K., P. V. S. Rao, and S. S. Agrawal (2000). "Hindi speech database." In *Sixth International Conference on Spoken Language Processing*.

[12]. Campbell, Nick ((2004)). "Databases of expressive speech." *Journal of Chinese Language and Computing* 14, no. 4: 295-304.

[13]. Raza, Agha Ali, Sarmad Hussain, Huda Sarfraz, Inam Ullah, and Zahid Sarfraz (2010). "An ASR system for spontaneous Urdu speech." *the Proc. of Oriental COCOSDA* : 24-25.

[14]. Anumanchipalli, Gopalakrishna, Rahul Chitturi, Sachin Joshi, Rohit Kumar, Satinder Pal Singh, R. N. V. Sitaram, and S. P. Kishore (2005). "Development of Indian language speech databases for large vocabulary speech recognition systems." In *Proc. SPECOM*.

[15]. Campbell, Nick (2000). "Databases of emotional speech." In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*.

[16]. Minematsu, Nobuaki, Yoshihiro Tomiyama, Kei Yoshimoto, Katsumasa Shimizu, Seiichi Nakagawa, Masatake Dantsuji, and Shozo Makino (2004. "Development of English speech database read by Japanese to support CALL research." In Proc. ICA, vol. 1, pp. 557-560.

[17]. Campbell, Nick (2002). "The recording of emotional speech: JST/CREST database research." In Proc LREC.

[18]. Chourasia, Vishal, K. Samudravijaya, and Manohar Chandwani (2005). "Phonetically rich Hindi sentence corpus for creation of speech database." Proc. O-Cocosda : 132-137.

[19]. Champion, Colin, and S. M. Houghton." Application of Continuous State Hidden Markov Models to a classical problem in speech recognition." Computer Speech & Language 36 (2016): 347-364.

[20]. Chadha, Neha, R. C. Gangwar, and Rajeev Bedi. "Current Challenges and Application of Speech Recognition Process using Natural Language Processing: A Survey." International Journal Journal of Computer Applications 975 (2015): [21].Gopalakrishna Anumanchipalli, Rahul Chitturi, Sachin Joshi, Rohit Kumar, Satinder Pal Singh, R. N. V. Sitaram, S. P. Kishore. 2005. Development of Indian Language Speech Databases for Large Vocabulary Speech Recognition Systems. In Proceedings of International Conference on Speech and Computer (SPECOM), Patras, Greece.

[22].Chalapathy Neti, Nitendra Rajput, Ashish Verma. 2002 A Large Vocabulary Continuous Speech Recognition system for Hindi. In Proceedings of the National conference on Communications, Mumbai, pp. 366-370.

[23].Shaikh, N. G. Mallah, and Z.Shaikh, (2009), Character Segmentation of Sindhi, an Arabic Style Scripting Lan-guage, using Height Profile Vector, Australian Journal of Basic & Applied Sciences 3(4),4160-4169.

[24].D. N. Hakro, S. A. Awan Z.A. Bhutto, M. Memon, M. Hameed (2017), 'Handling ambiguities in Sindhi Entity Recognition (SNER), Sindh University Research Journal (Science Series) 49 (3), 513-516.

[25]. D. N. Hakro, Hakro, M. A., & Lashari, I. A. (2017). Sindhi Named Entity Recognition (Sner). The Government-Annual Research Journal of Political Science., 5(5).

[26].D. N. Hakro, M. Memon, S. A. Awan Z.A. Bhutto, M. Hameed (2016), 'Isolated Optical Character Recognition', Sindh University Research Journal (Science Series) 48 (4), 839-844.

[27], Urooba Zaki, M. Memon, D. N. Hakro, K.U.R. Khoumbati, M. Hameed, M. Ahmed Zaki, G.Nabi (2019), 'Implementation Challenges in Information Retrieval System', Sindh University Research Journal (Science Series) 51 (2), 339-344. http://doi.org/10.26692/sujo/2019.6.55.

[28]. Awan, S. A., Abro, Z. H., Jalbani, A. H., & Hameed, M. (2018). Handwritten Sindhi Character Recognition Using Neural Networks. Mehran University Research Journal of Engineering and Technology, 37(1), 6.

[29]. Solangi, Y. A., Solangi, Z. A., Raza, A., Shaikh, N. A., Mallah, G. A., & Shah, A. (2018, November). Offline-printed Sindhi Optical Text Recognition: Survey. In 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS) (pp. 1-5). IEEE.

[30].Hakro, D. N. (2015). Enhanced Segmentation and Feature extraction approaches for Sindhi Optical Character Recognition (Doctoral dissertation, Ph. D. Thesis, University Science Malaysia, Malaysia).

[31].A.A. Chandio, M. Leghari, D. N. Hakro, S. Awan, A.H. Jalbani. (2016), 'A Novel Approach for online Sindhi Handwritten Word Recognition Using Neural Network', Sindh University Research Journal (Science Series) 48(1), 213-216.

[32]. D. N. Hakro, I. A. Ismaili, A. Z. Talib. Z. Bhatti. G. N. Mojai. (2014), 'Issues and Challenges in Sindhi OCR', Sindh University Research Journal (Science Series) 46(2), 143-152.

[33].Shaikh, H., Mahar, J. A., & Malah, G. A. (2013). Digital Investigation of Accent Variation in Sindhi Dialects. Indian Journal of Science and Technology, 6(10), 5429-5433.

[34].Patel, H. N., & Virparia, P. V. (2011). A Small Vocabulary Speech Recognition for Gujarati. International Journal of Advanced Research in Computer Science, 2(1).

[35].Chandrasekar, M., & Ponnavaikko, M. (2008). Tamil speech recognition: a complete model. Electronic Journal «Technical Acoustics, 20.

[36].Mannepalli, K., Sastry, P. N., & Suman, M. (2016). MFCC-GMM based accent recognition system for Telugu speech signals. International