ISSN:0975-9646

Tripti Dodiya et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (1) , 2016, 185-189

Speech Recognition System for Medical Domain

Tripti Dodiya[1], Sonal Jain[2]

[1] *GLS(I&RKD) Institute of Computer Applications, GLS University, Ahmedabad, India*

[2]*Department of Computer Science & Engineering, J K Lakshmipat University, Jaipur, India*

***Abstract*— In today's world, speech recognition has become very popular and important. Over the past 20 years, despite considerable advances in computer technology, the keyboard and video display are still the principal means of entering and retrieving data. However, as the use of computers increases, the need for interacting with the computer also grows. This demand has generated a need for an intuitive human - machine interface to accommodate the increase in the number of users. Over the recent years, advances in speech recognition technology have enabled a wide range of voice-enabled services. However, after years of research and development the accuracy of automatic speech recognition software's remains an important research challenge.**

This paper presents a brief survey on the features and applications of Automatic Speech Recognition systems and investigates the results of these systems for medical domain questions. Our results suggest that as the medical questions are quite long and complex, this domain differs from the open domain and requires additional work in automatic speech recognition to be adapted as per the domain.

***Keywords*— Automatic Speech Recognition (ASR), human-machine interface, voice enabled services.**

I. INTRODUCTION

Speech recognition (SR) is the translation of spoken words into text [1]. It can also be defined as understanding voice by computer and performing the required task. They are also known as 'automatic speech recognition' (ASR), 'computer speech recognition' (CSR) or 'speech to text' (STT). Speech recognition technology has made possible computers to follow human voice commands and understand human languages.

Question Answering (QA) systems is an extension of information retrieval that takes a question as an input from the user and gives the desired answer instead of a list of documents[15]. These systems can offer advantage to the end users when the input as well as the output is in the form of speech. Studies have shown that doctors/medical students have many questions when seeing the patients. As they are busy and less likely to use technologies that require more time, they would be willing to simply speak the questions as they arise in the clinical setting and get a fast and accurate response. Speaking a question is much faster and more natural than typing and can be used in situations like combat zones or ambulance delivery, where a standard computer interface is not available. Moreover, a speech interface can prevent errors resulting from typing the long and complex medical terminologies. In addition, smart phones are now ubiquitous and entering the questions using miniature keyboard or touchscreen can be tedious and error-prone. Under such conditions, automatic speech recognition systems can offer a more practical option.

In open domain, automatic speech recognition is already considered good enough despite the fact that word error rates (WER) for ASR systems in that domain are still quite high (around 35%). In this paper, we investigate the results of various speech recognition systems for medical domain questions.

II. BACKGROUND

Automatic speech recognition is well known task of converting acoustic signal into a string of words. In the process of speech recognition under ideal situation, the speech recognition engine recognizes all words spoken by a human, but, the performance of the speech recognition engine depends on various factors. Noisy environment, multiple users and vocabularies are the major factors considered as the depending factors for efficiency of speech recognition engine [3].

Automatic speech recognition systems involve several components from different disciplines such as statistical pattern recognition, signal processing, communication theory, combinatorial mathematics, and linguistics.

The basic model of the speech recognition system is as shown in Figure 1:

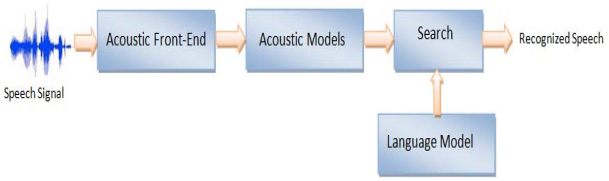

Fig 1: Basic model of speech recognition [4]

The speech recognition systems are generally used for : a) dictation wherein spoken words are converted to text and b) controlling the computer wherein software's are developed that are capable to authorize a user to operate different applications by voice [2][10].

A speech recognition system consists of: Feature extraction, Acoustic modeling, Pronunciation modeling, Decoder. The process of speech recognition starts with a speaker which creates utterances consisting of sound waves. These sound waves are further captured by microphone and converted into electrical signals. To make the signals understandable by the speech-system, the electrical signals are converted into digital form. Further, the signal is converted into discrete sequence of feature vectors, which contains only relevant information about the original words which are important for correct recognition. The feature extraction suppresses irrelevant information like information about speaker and information about transmission channel for

correct classification. The decoder helps in finding the best match from the knowledge base, for the incoming feature vectors.

There are three approaches to automatic speech recognition [2][4] :

a) Acoustic phonetic approach
b) Pattern Recognition approach
c) Artificial intelligence approach

Acoustic phonetic approach is a rule based approach[2]. It uses the knowledge of phonetics & linguistics for the search process. It has defined rules to layout the possibilities and rules to determine the sequences that are permitted. However, it has not been very successful in practical speech recognition systems.

Pattern recognition approach uses two methods: 1) training of speech patterns b) recognition of pattern by way of pattern comparison [4]. Initially, a sequence of measurements is made on the input of define 'test pattern'. The unknown pattern is then compared with each sound reference pattern. The measure of similarity between test pattern and reference pattern is computed. Finally the best matching pattern based on the similarity score is selected.

Artificial intelligence approach is a combination of acoustic phonetic and pattern recognition approach [2][4]. It uses an expert system to classify the sounds. The basic idea is to compile and incorporate knowledge from a variety of knowledge sources with the problem at hand [2].

III. TYPOLOGY OF SPEECH RECOGNITION SYSTEMS

The capability of speech recognition system is based on number of parameters like type of speech or speaking mode, dependence on speaker and the size of vocabulary [2][3]. Based on the speaking mode the ASR's are classified into:

a) Speaker dependent: they require a user to train the system according to their voice.
b) Speaker independent: they work for any speaker and are not trained for a specific user.
c) Isolated word recognizers: they accept one word or one utterance at a time. They allow the user to speak naturally.
d) Connected word systems: they require the user to speak slowly and distinctly each word with a pause.
e) Spontaneous recognition systems: they allow the user to speak spontaneously.

IV. APPLICATIONS

Despite of some limitations, speech recognition technology can be a very useful tool for a variety of applications. They can be very helpful to people with disability that are unable to write. It is also used for live subtitling on television, dictation tools in the medical and legal profession, for off-line speech-to-text conversion systems etc. They help in saving time and effort.

Today ASR are used in many applications which require human machine interface like automatic call processing in telephone networks, query based information systems providing updated travel information, weather reports, data entry, voice dictation, stock price quotations, railway reservation speech transcription, banking, automobile, avionics, differently-abled people and many more[8][10].

A. Medical

Speech recognition systems are extensively useful in medical field like maintaining electronic medical record (EMR), medical transcription and many more [9]. Medical professionals are always looking for ways to improve efficiency. Less time spent taking notes and filling out charts means more time to help patients. Using speech recognition software, doctors can dictate notes much faster than typing them. This allows more time spent diagnosing and treating patients. It also creates a digital record of a patient's history which can be later searched instead of spending time in searching countless paper documents.

B. Assistive technology for Speech/hand disability

Speech recognition is especially useful for people who have difficulty using their hands or having visual impairment wherein speech recognition programs are much beneficial for operating computers [5]. Many students, who need additional support, have difficulties in reading, writing, or spelling, due to specific learning difficulty or visual impairment. Students with hearing impairments rely on reading lips or watching the interpreter to understand what the instructor spoke. This further makes it extremely difficult to focus their attention on note-taking and the instructor (or interpreter) simultaneously. Assistive technologies like automatic speech recognition can help to enhance computer-assisted learning for students with different types of disabilities [5]. People with disabilities can benefit from speech recognition programs.

C. Military & Training

Speech recognition systems can be used in multiple areas of military like communications and control, intelligence, training etc. It is used in fighter aircraft to command the autopilot system, set radio frequencies, controlling flight display, set steer-point coordinates and weapons release parameters [10].

Speech recognition systems can also be very helpful in training for Air traffic controllers (ATC). Currently, some ATC training systems require a person to act as pseudo-pilot who interacts with the trainee and communicate with him just like the communication with pilots in real ATC situation. Speech recognition systems eliminate the need for a person as pseudo-pilot. The US Army, US Navy, and FAA as well as a number of international ATC training organizations are currently using ATC simulators with speech recognition.

V. EXISTING SOFTWARE

A. CMU Sphinx

CMU Sphinx toolkit is a leading speech recognition toolkit with various tools used to build speech applications [12]. It is jointly designed by Carnegie Mellon University, Sun Microsystems laboratories and Mitsubishi electric research laboratories. It has been built entirely on Java programming language. It is flexible and modular, supporting various types of HMM-based acoustic models, language models, and multiple search strategies [7]. CMU Sphinx toolkit has a number of packages like Pocketsphinx, Sphinxbase,

Sphinx4, Sphinxtrain for different tasks and applications. It can be downloaded from http://cmusphinx.sourceforge.net.
Sphinx4 is written in Java programming language which makes it easily portable to multiple platforms. It provides a quick and easy API to convert the speech recordings into text with the help of CMU Sphinx acoustic models. It can be used on servers and in desktop applications. Beside speech recognition, Sphinx4 helps to identify speakers, adapt models, align existing transcription to audio for time stamping and more. Sphinx4 supports US English and many other languages.

B. Dragon Naturally Speaking

Dragon Naturally Speaking is a proprietary, commercially available speech recognition software package developed by Dragon Systems and acquired by Nuance Communications [14]. The latest version, Dragon Naturally Speaking 13, supports both 32-bit and 64-bit editions of Windows. Its Mac OS version is called Dragon Dictate or Dragon for Mac. It is available in various languages like German, Italian, Spanish, Dutch, English, French, Japanese and other languages.

C. Voce

Voce is a free, open source cross platform speech synthesis and recognition library with a simple API [6]. It uses CMU sphinx for speech recognition and FreeTTS for speech handling. As sphinx 4 and FreeTTS are both written in java, voce is easily portable to various platforms. For synthesis, it takes the strings of text from applications and passes it to FreeTTS which converts them into audio output. For recognition, Sphinx continuously listens for the incoming audio, processes this data and adds recognized strings to the internal queue to be further queried by the application. Voce provides a thin layer between underlying speech interaction libraries and applications [6].

D. Julius

Julius is high-performance open source speech recognition software. It uses major speech recognition techniques, and performs a large vocabulary continuous speech recognition (LVCSR) task effectively in real-time processing [11]. Building a speech recognition system requires combining a language model (LM) and acoustic model (AM). It is written in pure C language. It runs on Linux, Windows and Mac OS. Although originally used for Japanese language it has also been applied for other languages like English, French, Korean etc.

E. Google Dictation

Dictation is free online speech recognition software by Google chrome. It uses the built-in speech recognition engine of Google chrome to translate the speech to text [13]. It can be easily accessible from www.dication.io.

VI. EVALUATION & RESULTS

We evaluated the performance of Sphinx, Voce, Dictation and Dragon Naturally Speaking systems for recognition of medical questions. The goal of this evaluation was to determine the accuracy of the spoken medical questions by the automatic speech recognition systems which can further help us in our future work towards the overall value of the question answering systems. Therefore, for the experiment, a question set consisting of 150 simple, complex, short and long medical questions was generated with the help of doctors and medical students as shown in Table 1. The domain expert was asked to categorize the questions as per their complexity and create the spoken data needed for our experiment. These categorized questions were further inputted into the speech recognition systems to evaluate the accuracy of the output.

Table 1: Sample data as input to speech recognition systems

Medical Domain Questions
What is normal blood pressure
When was penicillin invented
Give two names of steroid hormones
What is Obesity
How to calculate BMI
When was penicillin invented
Name of fluid present in joint cavity

Typically, some of the evaluated ASR systems support multiple languages. To limit the scope of this study, we selected English as the input language out of the multiple languages supported by the systems and the observations are made accordingly in this paper. These systems can also be tested for other supported languages and the observations may vary.
For Sphinx and Voce, the Grammar file had a limited word recognition, which we populated with our sample data as shown in Table 1. We found experimentally that, simple and short questions with easy terminologies were easily identified by all the systems. As per our observations, majority of the long questions were recognized incorrectly and gave a high sentence error rate (SER) by the systems. Also, as the medical domain uses complex terminologies, the speaker pronunciation could not be identified and were either substituted by some other word or a similar sounding word was inserted by the systems.
Noise is considered as a major depending factor for efficiency of speech recognition systems. Considering the fact that the doctors may use the systems in noisy situations like operating rooms, battlefields or ambulances, our experiment was conducted in a noisy environment for further evaluation. Our results show that this further reduced the efficiency of the systems.
For the evaluation, the domain expert was asked to give ratings with 5 point likert scale to the output returned by the ASR systems. The output that were rated from moderately acceptable level were considered to analyze the results. We used the sentence error rate (SER) metric to evaluate the systems. The results show that the sentence error rate (SER) for Sphinx and Voce were found to be 62% and 65.6 % respectively while for Dictation and Dragon the sentence

error rate (SER) was 46.6 % and 40 % respectively. The accuracy of the systems is shown in the Figure 2. From our observations, it was found that the tested systems performance was not remarkable. But, the proprietary systems performed slightly well as compared to the open source systems and were more adapted to the specific domain.

Table 2: Comparative study of ASR software

System observed / Features	Sphinx [7][12]	Voce [6]	Dictation [13]	Dragon Naturally Speaking [14]
Open Source / Proprietary software	Open Source	Open Source	Proprietary Software	Proprietary Software
Language used for Implementation	Java	Java	NA	NA
Sample test data	150	150	150	150
Sentences returned satisfactorily	57	52	80	90
Sentence error rate	62%	65.4%	46.6%	40.6%
Languages	English Chinese French Spanish German Russian and others	English	Arabic, Chinese, Spanish, French, German, Italian, Malay, Indonesian etc.	UK English, US English, French, German, Italian, Spanish, Dutch, Japanese

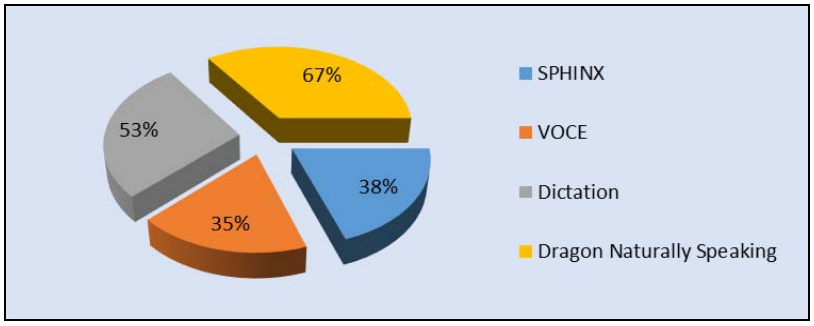

Figure 2: Evaluation results of ASR software's

VII. Conclusion

Since its inception in 1950's, there has been a considerable progress in speech recognition systems. In this paper, we have compared some of the speech recognition systems and observations listed.

We found that the existing systems did not performed well on spoken medical questions, possibly because they were tuned for other applications. This indicates the importance of speech recognition adapted for medical domain. A good example is provided by the medical term 'Paget's disease'. Even though the second term 'disease' was identified correctly, the first term 'Paget's' could not be identified and was misinterpreted by the systems.

As observed, noisy environments are a challenge for automatic speech recognition systems. Some real world situations such as operating rooms, ambulance and other noisy environments require these systems extensively but are not efficient due to the level of noise. The results of this study show that the generic ASR systems still need improvement for spoken medical questions. Domain specific adaptation is crucial if existing systems are to be applied in the medical domain.

VIII. Future Scope

Speech recognition offers real potential but also has some significant limitations. We intend to extend our investigation by testing the systems with supported multiple languages. We also need to further test them with speaker-specific voice. We also intend to build a medical domain QA system and to evaluate different ASR systems with respect to question input and the answer as the output wherein the accuracy of these systems play a vital role.

References

[1] Kuldeep Kumar, R.K.Aggarwal, "Hindi Speech recognition system using HTK", International Journal of Computing and Business Research, vol. 2, Issue 2, May 2011.

[2] M A Anusuya, S. K Katti, "Speech recognition by Machine – A review", (IJCSIS), International Journal of Computer Science and Information Security, Vol.6, No.3, 2009, ISSN 1947-5500.

[3] Samudravijaya K. Speech and Speaker recognition tutorial TIFR Mumbai 400005.

[4] Rajesh Kumar Aggarwal, Mayank Dave, "Acoustic modelling problem for automatic speech recognition system: conventional methods(0), International Journal of speech technology, Springer, Vol.14, Issue 4, pp 297-308, ISSN 1381-2416.

[5] Rustam Shadiev, Wu-Yuin Hwang, Nian-Shing Chen and Yueh-Min Huang, "Review of Speech-to-text recognition technology for

enhancing learning, Educational technology and society, 2014, 17(4), pp 65-84, ISSN 1436-4522.

[6] Tyler Streeter, "Open source speech interaction with Voce library", available as http://voce.sourceforge.net/files/VoceWhitePaper.pdf

[7] Kai-Fu Lee, Automatic speech recognition system- The development of SPHINX System, Book, ISBN 978-1-4615-3650-5.

[8] Shinya Iizuka, Kosuke Tsujino, Shin Oguri, Hirotaka Furukawa, "Speech Recognition Technology and Applications for Improving Terminal Functionality and Service Usability", NIT Docomo Technical Journal, Volume 13 No.4. 2012.

[9] Grasso, M. Automated Speech Recognition in Medical Applications. MD Computing, hi, v12, Jan. 11 1995, pp 16-23

[10] Application of Speech Recognition. Available as http://nlg.isi.edu/teaching/cs544/spring10/apps5-speech.pdf

[11] A. Lee and T. Kawahara, "Recent development of open-source speech recognition engine Julius, In Proceedings of the 1st Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC2009), pp 6, 2009.

[12] http://cmusphinx.sourceforge.net

[13] https://dictation.io

[14] http://www.nuance.com/dragon

[15] Miller T, Ravvaz K, Cimino JJ, Yu H, "An investigation into the feasibility of spoken clinical question answering", AMIA Annual Symposium Proc. 2011; page 954–959.