Journal Of Harmonized Research (JOHR)

Journal Of Harmonized Research in Engineering
2(1), 2014, 246-250

ISSN 2347 – 7393

Original Research Article

A STUDY ON ISSUES IN SPEECH RECOGNITION SYSTEMS

Mausmita Devi, Laba Kr. Thakuria, Purnendu Bikash Acharjee, Pranhari Talukdar

Department of Instrumentation and USIC
Gauhati University

Abstract

Speech recognition or which is more commonly known as automatic speech recognition is the process of converting an acoustic waveform into the text form which is known to the user. So, ASR also sometimes can be defined as the speech to text conversion process [2]. Application of speech recognition system more commonly defined in terms of three tasks: signal modeling, network searching, and language understanding [8]. But this paper only concern with the ASR to understand the Assamese language only. Any how the main aim of ASR is to build an interface for computers which is truly useful for users. But, it still has several problems that prevent it from being used exclusively as a method of transcription. Some of problems of ASR are the accent problem of the user or it may be for the wrong pronunciation of the word. Some problems may be caused because of the homonyms. Homophones are the words that sound similar but have different orthography. In Assamese language there are number of words are acts like homophones. One of the basic speech recognition problems may also be caused due to quality of hardware being used. This work represents an attempt to discuss the probable issues related to build ASR of Assamese language.

Key words: phonemes, acoustic, homophones, ASR, issues.

Introduction

Speech is the most natural way of communication [4]. Speech recognition software has greatly improved since it was first invented, but still it has some problems that prevent it from being used exclusively as a method of speech to text conversion. Several issues may incur for an effective ASR. These issues are very difficult problems for automatically recognizing speech with computer as an interface, and the reason for this is the complexity of the human language. In this paper we will try to figure out some of the issues that make ASR very difficult. It is very important to quantify the issues to incorporate

For Correspondence:
thakurialaba@gmail.com
Received on: March 2014
Accepted after revision: March 2014
Downloaded from: www.johronline.com

them in an Automated Speech Recognition (ASR) system.

Assamese Language

Assamese is an Indo-Aryan language spoken by the Assamese people in general. The mixed Aryan culture and the mongoloid culture gave birth to a new culture. So, every community from this region always exhibits their indigenous culture with diversity. It is the link language for the people living in Assam and adjoining states of Arunachal Pradesh, Meghalaya, and Nagaland etc. This language has come from Sanskrit as its offshoot, through different stages of development.

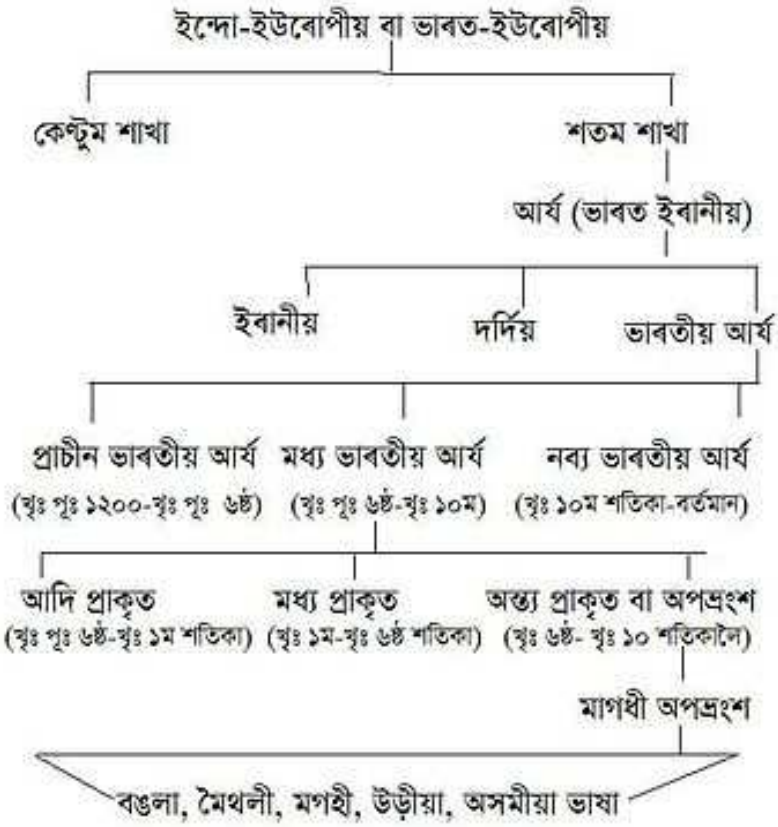

Fig-1: Proto Indo-European Family Tree

Phonological Structure of Assamese Language

The ASSAMESE phonemic structure consists of eight vowels, ten diphthongs, twenty-one consonants and two semi vowels [10]. The ASSAMESE vowels and consonants are shown in the tables below.

Table 1: Vowels in ASSAMESE.

	Front	**Central**	**Back**
Close	i		u û
Half-close	e ê		o ô
Half-open	ɛ		ɔ
Open		A	

Table 2: Consonants in ASSAMESE

Nature of Articulation	Bi-labial	Alveolar	Palatal	Velar	Glottal
Plosive	p b	T d		K g	
	ph bʱ	th dʱ		Kh gʱ	
Fricative		s z		X	ɦ
Nasal	M	n		Ŋ	
Lateral		l			
Rolled		r l			
Semi Vowel	W		J		

Why We Need Speech Recognition

The main goal of speech recognition is to get efficient ways for humans to communicate with computers [7]. Personal computers that can be made voice-controlled and used for dictation. This can be an important application for physically disabled persons. Speech recognition is important, not because it is 'natural' for us to communicate via speech, but because in some cases, it is the most efficient way to interface to a computer. Speech is a useful and effective communication medium with machines, especially where keyboard input is awkward or impossible [6]. Sometimes key board acts as a barrier between computer and the user. This is especially true for rural areas. ASR is an attempt towards reducing the gap between the computer and the people of rural places from Assam, by allowing them to use Assamese language, the most common language being used by the people in rural areas. Speech recognition will, indeed, play a very significant role in promoting the technology in the rural areas of Assam. Also commercially speech recognition products are increasingly in demands as alternate input devices for computers, particularly by persons with physical disabilities

Different issues related to ASR

The various key issues or probable issues are analysed in speech to text convertor for Assamese language. While speaking a particular language, the vocal tract of human may vary widely in terms of their accent, pronunciation, articulation, roughness, nasality, pitch, volume, and speed [3]. All these sources of variability make speech recognition, a very complex problem. So, these issues or the problem may be avoided in our STT for Assamese language. For improving our STT for Assamese language we can take consideration of the following points:

In terms of Speaker's variability:

All humans speak differently [6]. They exhibit different way of delivering their speech. Humans also communicate their emotions via speech. Regional dialects also introduce some features of pronunciation and vocabulary which is different according to their geographical area. Sometimes speaker does not know the context of the words being spoken, because of inaccurate spelling which can lead to text that has no punctuation.

In terms of Software and Hardware variability:

Speech recognition problems also incur due to involvement of low quality of hardware used to actually input the sound, because its results have a large impact on how the software will interpret the speech. For example, if a microphone is not sensitive enough or is over sensitive, then the audio information that it results is difficult for the software to decipher. In other words when a microphone is very sensitive that the speech is distorted, resulting the recognition nearly useless.

Noise:

A similar problem may come from background noise that can be very problematic to separate out from the main speech and can results inaccurate translations when included in the speech processing [5]. In other words noise can be defined as unwanted information in the

speech signal. In ASR we have to identify and filter out these noises from the speech signal.

Body Language:

A human speaker does communicate with speech, along with their body signals like hand waving, eye movements, postures etc. This information is completely missed in case of ASR. But recent research area multimodality studies the body language along with vocal speech to improve the human-computer communication.

Spoken and Written language are different:

Assamese is the easternmost member of this New Indo-Aryan (NIA) subfamily spoken mainly in the Brahmaputra Valley of Assam. Assamese is, therefore, a composite language into which words of both Indo-Aryan and Indo-Chinese origins have made their way. Besides, other Pre-Aryan and non-Aryan influences are discernible not only in loan-words but also in print of grammar, syntax, and pronunciation. Spoken language has for many years been viewed just as a less complicated version of written language, with the main difference that spoken language is grammatically less complex. In ASR, we have to identify and address these differences.

Amount of data and search space

Communication with a computer introduces a large amount of speech data every second [4]. This has to be matched to group of phones which results words or sentences. Groups of phones that build up words and words build up sentences. The quality of the input, and thereby the amount of input data, can be regulated by the number of samples of the input signal, but the quality of the speech signal will, of course, decrease with a lower sampling rate, resulting in incorrect analysis. We can also minimize our lexicon, i.e. set of words [7]. This introduces another problem, which is called out-of-vocabulary, which means that the intended word is not in the lexicon. An ASR system has to handle out-of vocabulary in a robust way.

Homophones ambiguity

Homonyms are two words that are spelled differently and have different meanings but sound the same. Assamese is a very strong vocabulary language. It has numbers of homophones which exactly sounds alike but possesses different meanings. For example. ""চিন"" (/chin/) means "sign" and "চীন" (/cheen/) has meaning "country name". Likewise কলা (/kalaa/) has meaning "black" and on the other hand ক'লা (/kalaa/) has meaning "said". However, extensive training of systems and statistical models that take into account word context can greatly improved their performance.

Word boundary ambiguity

When a sequence of groups of phones is put into a sequence of words, we sometimes encounters word boundary ambiguity. Word boundary ambiguity occurs when there are multiple ways of grouping phones into words.

Overlapping of speech

Many times system has difficulty in separating simultaneous speech. If we try to employ recognition technology in conversations or meetings where people frequently interrupt each other or talk over one another the ASR will get extremely poor results.

Conclusion

The purpose of this paper is to analyse and provide information about the issues or problems related to Automatic Speech Recognition system. But one thing is certain, ASR is a challenging task. ASR will continue to improve if we handle the issues. It seems quite sure that to do a perfect ASR, the above mentioned issues or problems should handle carefully.

Acknowlegment

Firstly, we would like to express my sincere gratitude and heartfelt thanks to all the members of Department of Instrumentation and USIC, Gauhati University. We would not have been able to complete the research work and shape it in the form of the research paper without their consistent advice, and never ending enthusiasm,

positivity, encouragement, support and understanding. We are very fortunate for having an opportunity to work with them from which we benefited enormously.

References

- N. M. Ben Gold. Speech and Audio Signal Processing, processing and perception of speech and music. John Wiley & Sons, Inc., 2000.
- J. H. M. Daniel Jurafsky. Speech and Language Processing, An introduction to Natural Language Processing, Computational Linguistics, and Speech recognition. Prentice Hall, Upper Saddle River, New Jersey 07458, 2000.
- Douglas O'Shaughnessy, Speech Communication Human and Machine. IEEE Press, 1987.
- Sami Lemmetty (`1999), Review of Speech Synthesis Technology, MPhil Thesis, submitted to Department of Electrical and Communications Engineering, Helsinki University of Technology.
- Lawrence Rabiner,Biing-Hwang Juang and B.Yegnanarayana, Fundamentals of Speech recognition. Pearson Education, 2009.
- Pran Hari Talukdar. Speech Production, Analysis and coding Pno-18
- Rabiner, L. R. (1995) Proc. Natl. Acad. Sci. USA 92, 9911-9913
- O. Fujimura, "Syllable as a unit of speech recognition," in IEEE Trans. Acoust., Speech, Signal Processing, vol. 23, February 1975, pp. 82–87
- Elenius and G. Takács "Phoneme Recognition with an Artificial Neural Network"
- Banikanta Kakati. Assamese,its formation and development, 3rd Ed 1972