SPEECH RECOGNITION IN MOBILE ENVIRONMENT FOR MALAYALAM LANGUAGE

Dr. Cini kurian, Department of Computer Applications

As many of the modern devices are designed and produced user-friendly for the convenience of general public. , speech recognition applications are becoming common and useful in this day and age Speaking/communicating directly with the machine to achieve desired objectives, make usage of modern devices easier and convenient. Although may interactive software applications are available, the use of these applications are limited due to language barriers. Hence development of speech recognition systems in local languages will help anyone to make use of this technological advancement. In India, speech recognition systems have been developed for many indigenous languages, however very less work has been done in Malayalam Language. Towards this goal, this work makes an attempt to develop Speech Recognition System for Malayalam language and has successfully developed a system with many features that are vital and unique for Malayalam Language. This would pave way for advanced studies for future researchers in this area.

Speech is the vocalized form of human communication. Speech is natural, easy, fast, hands-free and do not require technical knowledge. Human beings are comfortable with speaking directly with computers rather than depending on primitive interfaces such as keyboards and pointing devices. The primitive interfaces like keyboard and pointing devices require certain amount of skill for effective usage. Use of mouse requires a good hand-eye coordination. Physically challenged people find it difficult to use computer. It is difficult for blind people to read from monitor. Moreover current computer interface assumes a certain level of literacy from the user. It expects the user to have certain level of proficiency in English apart from typing skill. Speech interface helps to resolve these issues.

Interaction with computer through a convenient and user-friendly interface has always been an important technological issue. Machine-oriented interfaces restrict the computer usage to a minuscule fraction of the population, who are both computer literate and conversant with

written English. Computers which can recognize speech in native languages enable common man to make use of the benefits of information technology.

Speech recognition system keeps elderly, physically challenged especially blind people closer to the Information technology revolution. Speech recognition benefits a lot in manufacturing and control applications where hands or eyes are otherwise occupied. It has large application for use over telephone, including automated dialing, telephone directory assistance, spoken database querying for novice users, voice dictation systems like medical transcription applications, automatic voice translation into foreign languages etc. Speech enabled applications in public areas such as; railways, airport and tourist information centers might serve customers with answers to their spoken query. Such tantalizing applications have motivated research in automatic speech recognition(ASR) since 1950's. Great progress has been made so far, especially since 1970's, using a series of engineered approaches that include template matching, knowledge engineering, and statistical modeling . Yet computers are still nowhere near the level of human performance at speech recognition, and it appears that further significant innovation requires serious research/studies.

Automatic speech recognition has tremendous potential in Indian scenario. Although literacy rate of India is above 65%, less than 6% of India's total population uses English for communication. Since the internet has become universal, common man now mainly depend the same for any sort of information and communication. Therefore it is imperative that the about 95% of our population cannot enjoy the benefits of this internet revolution. If these information is available in local languages, India could also be benefited by this technology revolution and could stand along with developed countries.

It would be a vital step in bridging the digital divide between non English speaking people and others. Since there is no standard input for Indian languages, it eliminates the key board mapping of different fonts. In Indian scenario, where there are about 1670 dialects of spoken form, speech recognition technology has wider scope and application .

Malayalam is one among the 22 official languages spoken in India with about 38 million speakers. It belongs to the Dravidian family of languages and is one of the four major languages of this family with a rich literary tradition . The majority of Malayalam speakers live in Kerala, one of the southern states of India and in the union territory of Lakshadweep. The language has 37

consonants and 16 vowels. There are different spoken forms in Malayalam although the literary dialect throughout Kerala is almost uniform .

Various dimensions of speech recognition

A speech recognition system's accuracy depends on the condition under which it is evaluated. Under narrow conditions almost any system can attain human-like accuracy, but it is much harder to attain good accuracy under normal environment. Hence the accuracy of any system can vary along with the following dimensions.

- **Vocabulary size and confusability**

Generally, it is easy to classify a small set of words but as the vocabulary size increases, classification becomes a complex issue. For example, the 10 digits "zero" to "nine" can be recognized perfectly , but vocabulary sizes of 200, 5000 and 100000 may have error rates of 3%, 7% and 45% . It is hard to classify the words of a vocabulary which contains confusing words although it is small in size.

- **Speaker dependence vs. Speaker independence**

Speaker dependent speech recognition system is dependent on knowledge of the speaker's particular voice characteristics. This system learns the characteristics of the speaker's voice through voice training (or enrollment). This type of system must be trained on a specific user before being able to recognize what has been said. This type of system works well if there is only one user speaking to the system and it cannot be used for general purpose. Speaker independent systems are generally able to recognize speech from a variety of contexts, speakers etc. This type of system is used in all general purpose recognizers. A complete speaker independent system, is hard to achieve since it needs rigorous training from all type of speakers (age wise and gender wise), all dialects etc. Error rates are typically 3 to 5 times higher for speaker independent system than for speaker dependent ones .

- **Isolated, Connected and Continuous speech recognition**

Isolated and Connected speech recognition is relatively easy because word boundaries are detectable and the words tend to be clearly pronounced. Continuous speech is more difficult because word boundaries are unclear and their pronunciations are more corrupted by co-

articulation. In a typical evaluation, the word error rate for isolated and continuous speech were 3% and 9% respectively .

- **Read vs. spontaneous speech**

Speech can be either read from prepared scripts, or that is uttered spontaneously. Spontaneous speech is difficult to recognize, because it tends to be peppered with disinfluencies like "uh" and "um", false starts, incomplete sentences, stuttering, coughing, and laugher; and moreover, the vocabulary is essentially unlimited. On the other hand, read speech can be handled more easily since it does not contain unexpected words.

- **Adverse conditions**

A system's performance can also be degraded by a range of adverse conditions . These include environmental noise (eg. noise in a car or a factory); acoustic distortions (e.g. echoes, room acoustic); different microphone (e.g. close-speaking, unidirectional or telephone); limited frequency bandwidth (in telephone transmission); and altered speaking manner(shouting, whining, speaking quickly etc.).

Challenges in building a mobile based ASR:

ASR is a branch of Artificial Intelligence (AI) and is related with number of fields of knowledge such as acoustics, linguistics, pattern recognition etc . Speech is the most complex signal to deal with since several transformations are occurring to it at semantic, linguistic, acoustic and articulatory levels. In general, the challenges of a mobile based ASR task includes, i) the acoustic variability that results from changes in training and testing environment, ii) intra-speaker variability due to changes in the speaker's physical and emotional state, iii) inter-speaker variability that results from differences in accent, dialect, vocal tract size and shape, iv) channel variability due to different kinds of telephone apparatus with varying microphones and transmission quality.. Moreover as the speech may come from an uncontrolled environment, the background noise could degrade the input speech further. The constraints that the network imposes on the bit rate of the transmitted signal, the limitations imposed by the computing capabilities of the device on the complexity of the signal processing front-end and the decoder, compounded with the potential exposure of the user to more intense and challenging acoustic

environments, make the problem of ASR in mobile environments more susceptible to performance degradation than fixed network speech recognition applications.

The best recognition performance in a telephone based ASR can, in principle, be achieved by using speaker-dependent models specifically tailored to the vocal characteristics of each user of the system using a large amount of training data from each user. This is, however, practically infeasible. On the other hand, a reasonable performance can be obtained by using a speaker-independent model trained on a huge corpus that captures a wide spectrum of speakers, environments, channels and application domains. However, such a huge training data is often unavailable.

Speech is the most complex signal to deal with since several transformations occurring at semantic, linguistic, acoustic and articulator levels. In addition to this, the following factors make this area the most challenging.

- Context variability -: Some words having different meaning and usage may have same phonetic realizations. Few examples are: write vs. right, four vs. for. Since phonetic realization is one of the key factors for speech recognizers, these types of words are of enormous challenge for speech recognizers.
- Co-articulation Effect -: This is caused by different acoustic realization for the same phoneme. e.g. Jeep vs. peak. Here same phonemes /ee/ has different acoustic realization. This type of variability is difficult to model.
- Speaker Variability -: This variability is due to variations in vocal tract size, vocal cord vibration, physical characteristic like age, sex etc.
- Speaking rate variation (words/minute) and emotional rate changes are other two factors which affect the speech quality/modeling.
- Environmental variability is one of the most rigorous challenges since performance of speech recognition degrades due to mismatched conditions.
- Speech has undergone to many transformations, from the time it is delivered from mouth till it is converted to digital form. This factor affect the speech quality in a prominent way. Adverse conditions; as specified in section 1.2 and additive noise like fan, AC,

another speakers voice and channel distortions are other factors which have to be taken into account while building ASR system.

- Automatic speech recognition technology needs knowledge from multidisciplinary areas like Acoustics, Linguistics, Biology, Physiology, Cognitive Science, Intelligence, Artificial Intelligence, Electrical Engineering, Computer science, Digital signal processing, Mathematics and Statistics.

Objectives of the project

The primary objectives of this research is to initiate the process of developing Malayalam Speech Recognition in mobile environment to explore its wide opportunities for practical applications. To achieve this goal, the current research work concentrated on the following goals.

i) *To create speech database, pronunciation dictionary and transcription file*

As already referred . the main constraints in the speech recognition research is the lack of the speech corpus, pronunciation dictionary, and its transcription file. For filling this gap, we propose to collect/create the above mentioned files /database for all the recognition tasks as well as for analysis of unique phoneme features.

Contributions

The following are the different contributions of this work

- This work would be a great contribution to the future speech recognition research of Malayalam language.
- This work will be a vital contribution from Malayalam language in the process of developing a multilingual speech recognition system for Indian languages in mobile environment.
- Creation of pronunciation dictionary, transcription and speech corpus, make the speech recognition area comparatively tough and boredom. This problem persists in Malayalam language speech recognition research. Hence this work will be of great encouragement for speech recognition research.