**VISVESVARAYA TECHNOLOGICAL UNIVERSITY**

Jnana Sangama, Belgaum- 590 018



SYNOPSIS ON

# “VOICE BASED AUTHENTICATION FOR THE BLIND”

**COMPUTER SCIENCE & ENGINEERING**

***During the Academic year 2019-20  
Submitted By***

**Name: USN:**

**Anurag G 1DB16CS023**

**C Sagar Patil 1DB16CS042**

**Harsha J K 1DB16CS056**

**Deepak S D 1DB17CS403**

Under the guidance of

**Prof. HEMALATHA M**

Designation

Dept of CSE

**

**Department of Computer Science and Engineering**

**Don Bosco Institute of Technology**

**Mysore Road, Kumbalgodu, Bengaluru-560074**

**CONTENTS**

* **Introduction**
* **Problem Statement & Motivation**
* **Aim & Objectives**
* **Relevance**
* **Modern Tools**
* **References**

**INTRODUCTION**

Automatic speech recognition, translating of spoken words into text, is still a challenging task due  
to the high viability in speech signals. For example, speakers may have different accents, dialects,  
or pronunciations, and speak in different styles, at different rates, and in different emotional states.  
The presence of environmental noise, reverberation, different microphones and recording devices  
results in additional variability.

Conventional speech recognition systems utilize Gaussian mixture model (GMM) based hidden  
Markov models (HMMs) to represent the sequential structure of speech signals. HMMs are  
used in speech recognition because a speech signal can be viewed as a piecewise stationary  
signal or a short-time stationary signal. In a short time-scale, speech can be approximated as  
a stationary process. Speech can be thought of as a Markov model for many stochastic  
purposes. Typically, each HMM state utilizes a mixture of Gaussian to model a spectral  
representation of the sound wave. HMMs-based speech recognition systems can be trained  
automatically and are simple and computationally feasible to use. However, one of the main  
drawbacks of Gaussian mixture models is that they are statistically inefficient for modelling data  
that lie on or near a non-linear manifold in the data space.

Neural networks trained by back-propagation error derivatives emerged as an attractive  
acoustic modelling approach for speech recognition in the late 1980s. In contrast to HMMs, neural  
networks make no assumptions about feature statistical properties. When used to estimate the  
probabilities of a speech feature segment, neural networks allow discriminative training in a  
natural and efficient manner. However, in spite of their effectiveness in classifying short-time units  
such as individual phones and isolated words, neural networks are rarely successful for  
continuous recognition tasks, largely because of their lack of ability to model temporal  
dependencies. Thus, one alternative approach is to use neural networks as a pre-processing e.g.  
feature transformation, dimensionality reduction for the HMM based recognition.

**PROBLEM STATEMENT & MOTIVATION**

Speech recognition is a technology that automatically finds the words and phrases that best match the input of human speech. This technology enables the creation of a wide range of applications where the output of speech recognition can be either the end result (e.g., dictation) or input for further processing (e.g., giving commands to devices based on the detected words, semantic analysis of speech in an automatic dialogue system).

The most natural application of speech recognition is dictation, where speech recognition can be used to compose letters/e-mails and other documents. For many languages, such dictation software has been quite widely used for several years already, and it works rather well. In addition to dictation the recognition of coherent natural speech can be used in many applications where speech recognition is hidden from the user, for example, in automatic transcription of audio archives, in order to enable better organization and indexing. Speech recognition is also an important part of automatic speech to speech translation, as yet a somewhat sci-fi application. Speech recognition is also used in automatic dialogue systems. Such typically automated telephone answering systems hold information about a specific domain and can be interacted with in natural speech. A user's speech is detected, analysed, appropriate queries are made or activities performed, and an answer is generated, which is then once again converted into speech with the speech synthesizer. Apple's Siri and Google Now are examples of such dialogue systems.

The development of speech recognition for a specific language is a rather resource-intensive project. Due to the low number of Estonian-speaking people, the development of Estonian speech recognition cannot be expected to be carried out for commercial purposes. For example, Google has announced that their goal is to develop speech recognition for all the languages which have more than a million speakers. Estonian language with its 900,000 speakers falls below this limit. Therefore, it is reasonable that the state supports the development of Estonian speech recognition.

To design an automatic speech recognition system that gives best recognition results for both male and female speakers. Parameters effecting speech recognition are gender, accent, emotion, age, health and noise.

**AIM & OBJECTIVES**

The aim of the project is to enhance the speech recognition technology, to make the technology available for the development of new applications, to improve the existing applications and create new ones.

When improving the speech recognition technology, the main focus is on the aspects which currently have relatively low quality. Our objectives are:

* Improved recognition quality of lower quality speech recordings (above all speech recorded through the phone)
* Better quality for spontaneous speech.
* Better handling of speakers with accent.
* Better detection of the mentions of named entities.
* Indexation of speech for the search of terms and named entities.

**RELEVANCE**

Uses of our project in real world are as follows:

Used in evolving search engines, when using search engines there can be differences between how we type our inquiries and how we verbalize the same queries. The user may have trouble expressing a phrase or their intent thus may not acquire appropriate results. With the inclusion of speech recognition in search engines, the results accuracies will be significantly increased. As speech recognition improves, there will be a significant implication on how the public views search engines generally.

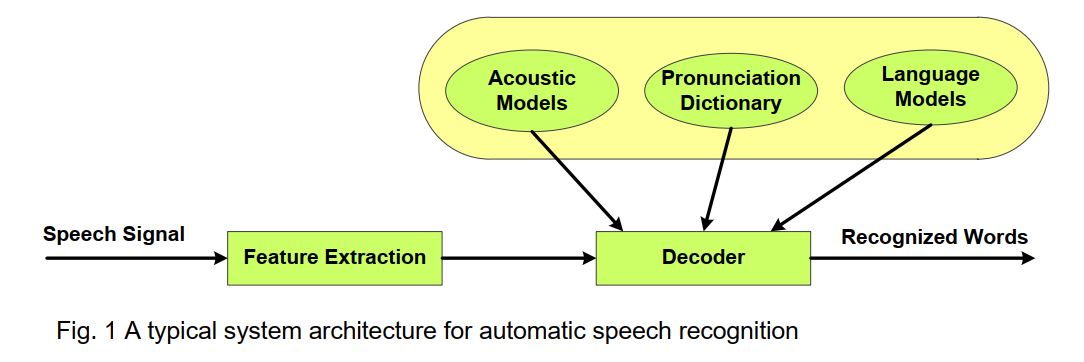
Impact in the healthcare industry***,*** the feature has its use in medical reporting by medical personnel. When it was introduced in this industry doctors had trouble using it to accomplish tasks. The system had a limited understanding of medical terminologies. Therefore, doctors had to learn on how to talk to the software. The technology was improved to be user-friendly and accurate; this was established by imperative improvements and inclusion of relevant vocabularies.

Use in service delivery***,*** customers and clients may not want to speak to a live operator. Therefore, they opt to use the speech recognition systems. This helps to make the process efficient and improves on time as it cuts on waiting time. This has its application in various airports in confirming travel schedules of the aircraft.

Automated identification, in order to avoid providing sensitive and risky personal information, institutions may opt to use speech recognition to authenticate identities of their clients. This has helped to curb fraud and phone crimes by use of voice biometrics in certain institutions like banks.

Communication in service providers, telecommunication providers use speech recognition to serve their clients who may want to receive customer care services. This consists of various questions by the software to establish the caller’s demands and then directs them to the appropriate operator for assistance.

**MODERN TOOLS**



**Y=[y0,…,yn]**

**Libraries and modules(python) used:**

* NumPy
* Pandas
* Matplotlib
* Seaborn
* Wit
* Ploty and Cufflinks
* Sklearn
* Kerass
* TensorFlow