

# Speech Recognition System

Pragati Gupta<sup>1</sup>, Shubham Kumar<sup>1</sup>, Ravi Prakash<sup>1</sup>

1. B.tech students, Department of Computer Science, IMS Engineering College, Ghaziabad-201009, Uttar Pradesh

\*\*\*

**Abstract** - This paper is demonstrating to convert the audio signals to perform the task. Speech recognition is one of the fastest growing technology nowadays. In this paper, we aimed at developing the speech recognition system as a helping tool for the differently able people. This paper demonstrates to convert the speech into English text. The conversion of speech into text is made by the speech recognizer. It can be used at various places with many possible solutions. There are around 20% people who are suffering from many disabilities. There are people who are blind, some cannot use their hands effectively and for illiterates, for them this system could be very helpful. This system will also be helpful for the enterprises where most of the work is to type. This system can recognize the audio signals and convert into text it can perform some operations, such as open calculator, open Google chrome etc. ; it also enables a user to perform operations such as "save, open, exit" a file by providing voice input. Likewise this system can perform some operations. At the initial level effort is made to provide help for basic operations as discussed above, to perform more operation this software can be updated and enhanced further.

This paper presents a method to design a speech to text then performs a task accordingly using .net framework using Visual Studio.

**Key Words:** Speech, text, desktop application, recognition, using .net framework

## 1. INTRODUCTION

Speech is the primary means of communication between people. Speech is the most common method of exchanging thoughts among humans. The clearness of speech and accent are the important part to convey the message correctly for the real conversation. Speech can be processed in two forms Speech Synthesis and Speech recognition. The speech can also be artificially produced. The artificial production of speech is called as Speech Synthesis. The word 'Synthesis' is defined by the vocabulary as the combining of the constituent elements of separate material or abstract entities into a single or unified entity the separating of any material or abstract

entity into its constituent elements. The recognition of speech signals by the system is called as Speech Recognition. The process of mapping an acoustic speech signal to text is called as Speech Recognition. In this the computer receives the user's speech and interprets what is said. This allows the user to control the computer (or some aspects of it) by voice, rather than having to use the mouse and keyboard, or alternatively just dictating the contents of a document. There are possibilities of one of the two things that could be possible in Speech Recognition. First approach is *Command and Control* (can also be abbreviated as CnC, or simply SR) in this the application can understand running text should match with the list of references at the end of the paper.

and follow simple commands that it has been educated about in advance.

Second approach is *Dictation* (can also be abbreviated to DSR). In this the engine has to identify arbitrary spoken words, therefore it is more complex, and also need to decide which spelling of similarly sounding words is required. The context information is developed based on the preceding and following words to try and help decide. CnC is sometimes referred to as context-free recognition, because this context analysis is not required with Command and Control recognition.

Dictation speech recognition is speaker-dependant. It means different people's enunciation, accent, pitch and many such factors, varies from person to person. For the decent results, recognizer requires a speaker profile to be set up.

On the other hand, command and control speech recognition is usually not speaker-independent.

This paper describes about the Speech Recognition System which is in particular CnC application i.e. Command and Control. This system is developed on .net framework. It runs on Microsoft Visual Studio 2015. API is provided by Microsoft that allows developers to use speech recognition and speech synthesis engines in windows applications. Speech-to-text conversion is done with the help of Speech Recognition engine, while speech synthesis provides access to text-to-speech conversion engine. The SAPI (Speech API) can be seen as an interface between the Application and SR/TTS engines.

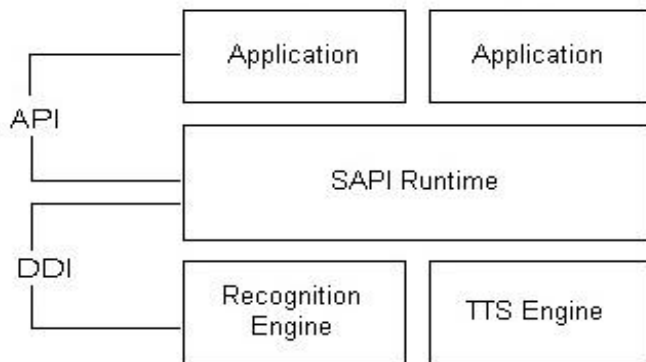


Fig 1: SAPI Overview [4]

Microsoft developed an API called as SAPI in order to implement speech applications. SAPI can be implemented either via .net framework namespaces that wrap the functionalities of SAPI, or via SAPI COM component. COM, abbreviated for Component Object Model. In order to use the API's functionalities number of interfaces are represented by COM.

SAPI acts like a connecting bridge between engines and application, while .net framework handles the interaction between the application and SAPI. The .net framework namespace (System.Speech) handles the interaction between the application and SAPI. This namespace comes along with .net framework version 3.0 and above and provides certain number of classes to interact with speech recognition engine.

## 2. LITERATURE SURVEY

Literature review on speech recognition systems genuinely demands the very first attention towards the process of converting sound waves into electrical impulses which was discovered by Alexander Graham Bell. The first speech recognition system developed by Davis et al for recognizing telephone quality digits spoken at normal speech rate. This effort for automatic recognition of speech was basically centered on the building up of an electronic circuit for recognizing ten digits of telephone quality [1]. To get a 2-dimensional plot of formant 1 versus formant 2 spoken utterances were analyzed. A circuit was designed for determining the highest relative correlation coefficient between a set of new incoming data and each of the reference digit patterns for the pattern matching. To perform well for the speech of different speakers it was observed that circuit adjustment helps the recognition system. To display the recognized spoken digit an indication circuit was built. There are many approaches to speech recognition that have evolved afterwards which had a major significance on finding speech sounds and providing appropriate labels to these sounds. In last five decades various approaches and types of speech recognition systems came into existence. This development has led to an extraordinary impact on the

development of speech recognition systems for various languages worldwide. Successive transformations of acoustic micro structure of speech signal into its implicit phonetic macro-structure can be viewed as automatic speech recognition. In other words, a speech recognition system is a speech-to-text conversion wherein the output of the system displays text corresponding to the recognized speech[1].

## 3. PROPOSED WORK

There are many applications for speech recognition system. Different applications are made from using different framework such as java, mat lab, sphinx, .net and many more. There are some approaches to ASR :

### 3.1 Acoustic-Phonetic approach

Hemdal & Hughes [2] took the basis of finding speech sounds and providing labels to them and proposed that there exist a fixed number of distinctive phonetic units in spoken language which are broadly characterized by a set of acoustics properties varying with respect to time in a speech signal[1].

According to this approach, the message bearing elements of speech are to be extracted explicitly with the determination of relevant binary acoustic properties such as nasality, frication, voiced-unvoiced classification and continuous features such as formant locations, ratio of high and low frequencies. This approach hasn't provided a reasonable platform for commercial applications. This approach is implemented in sequence: Spectral analysis, Features detection, Segmentation & Labelling, Recognizing valid word. Linguistic constraints are applied to access the lexicon for word [1].

### 3.2 Pattern recognition approach

The first person to propose this approach was Itakura (1975) and also got appreciable support from Rabiner & Juang (1989,1993) for its further acceptance among the researchers. For the last six decades this approach has become the ruling method for speech recognition. The two essential steps in this approach are pattern training and pattern comparison.

## 4. METHODOLOGY

In this we have made this using .net framework

### 4.1 Technology used

1. ASP.NET: ASP.NET is a web development platform, which provides a programming model, a comprehensive software infrastructure and various services required to

build up robust web applications for PC, as well as mobile devices.

2. C#: C# is an elegant and type-safe object-oriented language that enables developers to build a variety of secure and robust applications that run on the .NET Framework. You can use C# to create Windows client applications, XML Web services, distributed components, client-server applications, database applications, and much, much more

#### 4.2 Software Used

**Microsoft Visual Studio:** Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs for Microsoft Windows, as well as web sites, web applications and web services. Visual Studio uses Microsoft software development platforms such as Windows API, Windows Forms, Windows Presentation Foundation, Windows Store and Microsoft Silverlight. It can produce both native code and managed code.

#### 4.3 Working

A speech recognition application will typically perform the following basic operations:

1. Initialize the speech recognizer.
2. Create a speech recognition grammar.
3. Load the grammar into the speech recognizer.
4. Register for speech recognition event notification.
5. Create a handler for the speech recognition event [2].

#### 4.4 Advantages

- It is user efficient.
- Voice recognition of different notepad commands such as open save and clear.
- Open different windows softwares, based on voice input.
- Requires less consumption of time in writing text.
- Provide significant help for the people with disabilities
- Lower operational costs.

### 5. CONCLUSION

This Project work of speech recognition started with a brief introduction of the technology and its applications in different sectors. The project part of the Report was based

on software development for speech recognition. After the testing work, advantages of the software were described and suggestions for further enhancement and improvement were discussed.

#### REFERENCES

- [1]. International Journal of Computer Applications (0975 – 8887) Volume 41– No.8, March 2012
- [2].[https://msdn.microsoft.com/en-us/library/office/hh361683\(v=office.14\).aspx](https://msdn.microsoft.com/en-us/library/office/hh361683(v=office.14).aspx)
- [3].<http://www.chiark.greenend.org.uk/~owend/speech/proposal/>