Abstract
The attempt has been made to develop an “Intelligent Personal Assistant” which helps people to control device with their speech, extract information and perform tasks on their desktop. We drive our motivation from the fact that visually impaired people still have problem due to inability to read so it gets difficult for them to use laptops and computers, beside this, all other personal assistant have problems understanding Indian accent. This project has been created by keeping in mind the necessity of proper internet speed required to go online and so offline speech recognition is required. This project will listen and respond to what user will say in an effective manner via voice, just like having a conversation. This assistant has got capabilities to learn and based on experience it will perform the task in a better way, in a way assisting the user. The process includes four major steps: speech recognition, speech synthesis, content extraction and conversational agent all working in the same order in which they are mentioned. The use of neural networks has also been made to embed the learning capability helping the user to improve the task performance. Focus has also been laid on reducing the incompetence of the whole application and hence increasing the efficiency of task and time management.

I. INTRODUCTION
In this era where computers play an inevitable role in our lives, we have made an attempt to make their use much easier. A user-friendly, personal assistant has been developed to
make the use of computers more efficient and fast. An assistance is needed to the users to perform their important and tedious task. Yet somehow the existing assistants find it difficult to assist the users who have Non-British accent since they have been trained and tested on British Accent specifically. Moreover, visually disabled users have found it challenging to keep up with the upcoming improvements made in the field of Computer Science and IT resulting in a gap between general population and visually challenged people. Also, online speech recognition is sometime difficult as it is nearly impossible for every user to achieve the desired internet bandwidth and speed. So, most of the time, offline support is needed to assist the users who lack required internet support. There are two key challenges faced by personal assistant which are learning and adapting to the users preferences. In order to solve this problem in a smart and intelligent manner, the assistant agent has to thoroughly develop its behaviour based on its past experiences of the users’ actions that achieved a specific task. The approach to make IPA includes the use of Java library Sphinx-4, MaryTTS and neural networks to embed the learning capabilities.

In this paper we describe an intelligent personal assistant agent that learns from past experiences to assist users in performing specific tasks. The structure of the paper is as follows: Section II presents the Literature survey describing the related existing work. Our method to achieve this has been described in Section III. The next section IV consist of system design and architecture which is followed by Section V comprising results, providing analysis of our project and last but not least is the conclusion and future work.

II. LITERATURE SURVEY

In the year 2015, at the tenth International Conference on Industrial and Information Systems, assistance was proposed for visually disabled users. Project Nethra, was exclusively designed for visually challenged population to reach out to social media and several other Internet based services. In this paper the main focus was on the improvement of conversational agent and speech recognition module which will work offline. Earlier than this, there were many work done in this field which were dedicated to bridge this gap. Some of them are: AIML, Façade, ChatScript and RiveScript [1]. Moreover, in 2009, an intelligent personal assistant was made which had the ability to learn through its experience. The intelligent personal assistant anticipating user thus, helping them to manage their time and work efficiently [2]. This reduces the mistakes performed by the user in his daily routine tasks. The paper suggested a cognitive solution in making the life of the user easy by simplifying it [3].

III. METHODOLOGY

Our aim is to build such a personal assistant that will make easy for blind to use computer and make task easier. To implement this we are using java libraries and neural networks. Java libraries are used for speech recognition module whereas for learning ability we will use neural networks algorithms.
IV. ARCHITECTURE AND SYSTEM DESIGN

The whole process includes speech recognition, speech synthesis, content extraction and the communicating it back to the user. The speech recognition and synthesis and content extraction is done by the Sphinx library in java and text-to-speech part is done by MaryTTS. One of our main challenges is in the speech recognition part since user voice is the input and is to be detected properly using speech recognition tool. Speech is the combination of stable and dynamic states, which makes it difficult to distinguish the words. There can be two different waveform of same words. For doing this we have used Sphinx-4.

- **Sphinx-4**

  There are many speech recognition libraries in java and one of them is Sphinx-4. It has acoustic model which translates speech recordings to text making it easy and quick. Our motive behind using Sphinx-4 is that it comes with an advantage of modularity and flexibility. It is totally developed in Java programming language and therefore provide high portability. It also enables multithreading and supports flexible user interface in which newer search strategies and algorithms are used. The design of Sphinx-4 includes: **Frontend, Decoder, and Linguist.** The input is given in the form of one or more input signals into Frontend which parameterizes them into arrangement of ‘Features’. These
signals are altered, to different versions of language model, using vocabulary information form Dictionary, and other types of structural information into SearchGraph coming from one or more sets of Acoustic model, using Linguist. Actual decoding is done by the SearchManager which uses features of FrontEnd and SearchGraph to obtain results. Before going to recognition process, the Controls of the modules can be issued to any application in a way making it convenient for the recognition, having an added benefit [5]. Sphinx-4 uses different interfaces for recognition:

- LiveSpeech Recognizer
- StreamSpeech Recognizer
- SpeechAligner

For this four attributes have to be set up:

- Acoustic model.
- Dictionary.
- Grammar/Language model.
- Source of speech.

Configuration object is used to set up the first three which is then passed to recognizer. The concrete recognizer defines the way to recognize the speech source which is subsequently passed as a method parameter.

The process of speech recognition follows specific pattern and these steps are described below:

- **LiveSpeech Recognizer** – Takes input as speech from microphone.
  
  This input is stored in a format of the file and is passed as an attribute which is supplied through configuration.

- **StreamSpeechRecognizer** - takes this input in form of file.

- The next step is of alignment of text with audio-speech which is done by SpeechAligner.

- **SpeechResults** - recognizes the words from the utterances, words from time stamp and lattice.

  - **MaryTTS**

  Since communication with user is an important task so to fulfill it MaryTTS has been used. MaryTTS gives the voice to the words which comes as a results of sphinx-4 synthesis. This has been also developed in Java programming language and is integrated with Sphinx-4 [7].

### V. RESULTS

Our IPA was meant to control the applications online as well as offline and to analyze the performance of the intelligent personal assistant number of test cases were designed. These test cases include various aspects like response time and correctness of understanding speech as computed by intelligent personal assistant.
The results show that IPA is working offline to work with the desktop application. Sets of tasks were performed to analyze the working of IPA and the processing time of IPA depend on the type of task. eg. The processing time for calculation was near to 1.5s – 1.7s. Beside this we tested IPA over controlling the applications and the processing time of opening an app say notepad was 0.88s and of opening Google was 0.73s.

VI. CONCLUSION AND FUTURE WORK
We were able to describe an intelligent assistant that learns by supervision to assist users in their task performance. The main focus will be on training a more personalized and rich language model to enhance the offline speech recognition accuracy. Since most of the users have the limitation of internet speed, an attempt has been made to make an offline interaction with the IPA rather than online. The idea of an artificially intelligent virtual assistant that supports interactive conversations with user, instead of working as a data service provider is the area of prime importance for research at present. Therefore, in the future we are aiming at making this software more interactive and intelligent as the communication will be bi-directional that is from the software side as well as it will also respond back by speaking to the user. And we will try to increase the learning capacity of IPA so that it stores the user’s preferences.

VII. REFERENCES
[1] “Project Nethra - An Intelligent Assistant for the Visually Disabled to Interact with Internet Services”.
[6] The CMU Sphinx-4 speech recognition system
[9] A New language and a new voice for MaryTTS
[10] Automatic Speech Recognition – A Brief History of the Technology Development
TO CITE THIS PAPER