Impact Factor Value 3.441 e-ISSN: 2456-3463
International Journal of Innovations in Engineering and Science, Vol. 2, No.7, 2017
www.ijies.net

Voice based interaction System with artificial character and AIML

Meghavi M Prajapati[1], Sumit Chaudhary[2]

[1] *Student, M.Tech (Computer Engineering),Indrashil Institute of Science & Technology, Kadi.*
[2] *Head of Department (Computer Engineering),Indrashil Institute of Science & Technology,Kadi.*

Abstract – *We humans have five sense and the sixth is the one we have internally. Being a human we can talk and put over opinions to others. You can read this paper and understand the information as I have written it over the time. Talking and chatting is the best communication tool found since discovery of the universe. We have ears to listen, tongue to speak and eyes to see. All this things together make you a perfect human. We grow by learning and nobody teach us how to talk, how to walk, run and do natural things [1]. It is true that talking and recognizing speech is very easy for us. But when it comes to the computers they are not comparative with the human, Computers are evolved so much since last few decades, now a day's robotics and computers can do such things which are near to impossible for human. This paper talks about the computers having the artificial intelligence, computers that can understand human, computers that can listen to the person and give voice based feedback. Same as talking to the real person [3]. I will be discovering the basics of the Artificial intelligence and the need of that along with the chat but integration in the computers.*

Keywords *–* ***Chat Bot, Artificial Intelligence, AIML, Real-time Chat, Human Behavior, facial expression, ALICE***

1. INTRODUCTION

In many science fiction movies, we see people talking with computers and robotic machines which we think that may all these fake and unbelievable in the current time. Speech recognition was one of the very difficult task when it comes to the computers and robots, voice recognition can be useful in many ways in computer science [4]. The speech recognition systems can be used in computers for dictation and in controlling the other computer programs, which may be very helpful for handicapped people. The auto responsible voice interaction can be used in the communication, which we can see in current mobile systems, normally abbreviated as IVR (Interactive Voice Response).

This kind of technologies can be implemented in the computers to talk in a way that people can feel that they are talking with human not the computer. This can be achieved with the help and power of Artificial Intelligence. We can make computers learn itself. Computer can keep understanding the user's feedback and can build their own brain (memory). At some point the computers will be that much powerful to understand everything which may be human cannot even think of.

In this paper proposed system has the conversational agent in the computer program is simulated. It has intelligent conversation skills. This bot is based on the AIML. The AIML is Artificial intelligence Mark-up Language which provides a XML based brain mechanism to the Chabot.

Speech recognition applications that have emerged over the last few years and after the evolution of the Google things the voice dialling ("Call home"), Voice search ("What is taj mahal ?") search for Pizza ("Pizza shops near me") simple data input (e.g., entering a text typing), Preparation of structured documents (e.g.,an architectural report) and spoken audio search (e.g. find a meaning of emotions in Spanish language) [2].

In this paper a connected word speech recognized as a computer command which will be implemented and performed as per the defined operations, this part of artificial intelligence and some others which basically depend on human voice are come under "speech based pattern recognition" applications. Some others of this type of applications are:

Speech Recognition: used to know the contents of the speech.

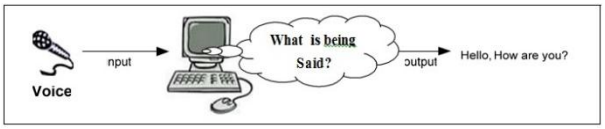

FIGURE 1: SPEECH RECOGNITION

Speaker Recognition: used to know the person who is talking? This can be used for the user authentication system for unlocking various computer programs.

FIGURE 2: SPEAKER RECOGNITION

LanguageIdentification:usedtoknowthespokenlanguage.

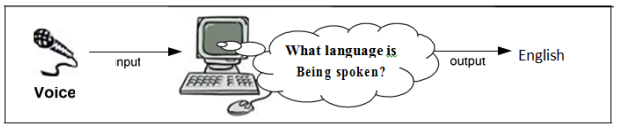

FIGURE 3: LANGUAGE IDENTIFICATION

2 LITERATURE SURVEY

Artificial intelligence (AI) is intelligence of the machines. It is human like behavior proposed by the machines and computers. In computer science, an ideal "intelligent" machine is a flexible software agent that controls its environment and performs operations that can maximize the chance of success. Colloquially, the term "Artificial Intelligence" is applied when a machine mimics "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving".

AIML: Artificial Intelligence Mark-up Language AIML (Artificial Intelligence Mark-up Language) is an XML-compliant language that's easy to learn, and makes it possible for you to begin customizing an Alicebot or creating one from scratch within minutes. It is XML based language, if you can understand the XML then you can easily build the AIML. The AIML can perform as a brain of the agent which can supply necessary outputs to the given input.

The most important tags of AIML are:

<aiml>: the tag that begins and ends an AIML document

<category>: the tag that marks a "unit of knowledge" in an Alicebot's knowledge base

<pattern>: used to contain a simple pattern that matches what a user may say or type to an Alicebot <template>: contains the response to a user input

There are also 20 or so additional more tags often found in AIML files, and it's possible to create your own so-called "custom predicates". Right now, a beginner's guide to AIML can be found in the AIML Primer.

The free A.L.I.C.E. AIML includes a knowledge base of approximately 41,000 categories. Here's an example of one of them:

```
<category>
  <pattern>WHAT ARE YOU</pattern>
   <template>
     <think><set name="topic">Me</set></think>
     I am the latest result in artificial intelligence,
              with greater speed and accuracy.
   </template>
</category>
```

(The opening and closing <aiml> tags are not shown here, because this is an excerpt from the middle of a document.)

Everything between <category> and </category> is -- you guessed it -- a category. A category can have one pattern and one template. (It can also contain a <that> tag, but we won't get into that here.)

The pattern shown will match only the exact phrase "what are you" (capitalization is ignored).

But it's possible that this category may be invoked by another category, using the <srai> tag (not shown) and the principle of reductionism.

In any case, if this category is called, it will produce the response "I am the latest result in artificial intelligence..." shown above. In addition, it will do something else interesting. Using the <think> tag, which causes Alicebot to perform whatever it contains but hide the result from the user, the Alicebot engine will set the "topic" in its memory to "Me". This allows any categories elsewhere with an explicit "topic" value of "ME" to match better than categories with the same patterns that are not given an explicit topic. This illustrates one mechanism whereby a botmaster can exercise precise control over a conversational flow.

Voice recognition (in many contexts also known as automatic speech recognition, computer speech recognition or erroneously as Speech recognition) is the process of converting a speech signal to a sequence of words in the form of digital data, by means of an algorithm implemented as a computer program. The recognized words can be the final results, as for applications such as commands & control, data entry, and document preparation. They can also serve as the

input to further linguistic processing in order to achieve speech understanding.

3-Sound Recording And Word Detection

Microphone Features

The microphones take input from the user and save it or forwards this is the analogue signal. Default sampling rate of the microphone is 44100 samples per second, at a size of 16 bits per sample and dual channel.

The Concept

It is very important to detect when a word is spoken or when a command is spoken. The system detects the silence continuously, it checks for the silence coming from the microphone. Anything other than silence is considered as a spoken words. The system uses the pattern present in the sound signal to detect the voice input form the user.

The Method

For word detection a method or a function is called from the software program having switch statement compared by the commands in XML files and executes its body. This XML files can have our custom commands. Every time when the word is detected it is compared against the list of the commands in the XML if the match is there the respective command is executed.

Training Models

For speech recognition, the system needs to know how the words are pronounced by particular person. Each individual has different rate of speech and different speaking styles. So for the better understanding computer needs to know the user's voice and pronunciation method. For this we need to train the system. During the training, using the data given by the user, the system generates acoustic model and language model. These models are later used by the system to map a sound to a word or a phrase. We can use pre built training wizards such as Microsoft training wizard for the windows systems to make this training very simple and user friendly.

Process Flow:

In this paper proposed system has the conversational agent in the computer program is simulated. It has intelligent conversation skills. This bot is based on the AIML. The AIML is Artificial intelligence Mark-up Language which provides a XML based brain mechanism to the Chabot.

This paper does not justify the use of the connected devices such as computer peripheral, machinery and computer software control. This paper propose a design of a Chabot with avatar and voice interaction to make a conversation more alive.

This system does not analyse the user's emotions and reaction. This system's Avatar cannot be directly modified or created without proper change in the API. This application is based on AIML Chabot with custom commands.

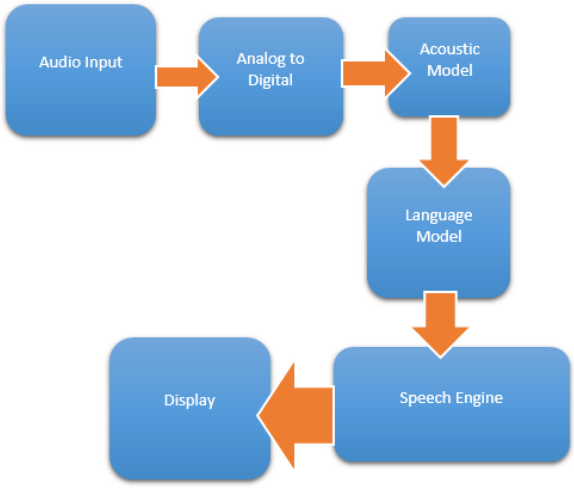

Figure 4: Process Diagram

We can create custom native language commands also, the languages can be Hindi, Gujarati and other native languages. This paper also demonstrate the use of a computer programs using voice command, you can fully operate the computer using the native voice commands. The assistant used in this system is the MS-Agent character which can be changed whenever needed.

Audio Input

The audio input devices such as microphone can be used to get the input from the users, we human speak in analogue signal, and the microphones takes the analogue input and pass it to analogue to digital converter.

Figure below shows typical microphone which can record voice and convert the voice signals to the digital form using Analogue to Digital converter.

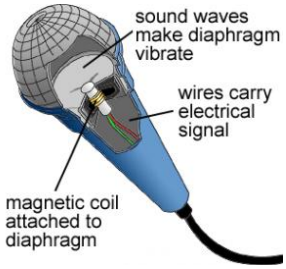

Figure 5: Typical Microphone

ANALOGUE TO DIGITAL CONVERTER

We human speak in analogue signal while computers can only understand digital input, hence the conversion of the voice signal from analogue to digital is done for further processing.

FIGURE 6: ANALOGUE TO DIGITAL CONVERTER

ACOUSTIC MODEL

Whatever we speak is kept as acoustic model which means that the input will not understand the actual language here. For ex. If we say "how are you" the acoustic synonym for this can be "haaou aare yoouw", or something else which will not be any language word.

FIGURE 7: ACOUSTIC MODEL

The input from the microphone is directly parsed as the acoustic words and series of redundant characters.
But later it can be parsed to desired language using the language model.

LANGUAGE MODEL

Language model takes the input as the acoustic words and then extracts the words in the language to process them. It actually identifies the words of real language from the given input. Here the language can be English or other supported languages.

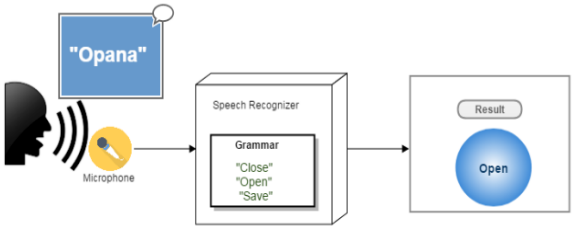

FIGURE 8: LANGUAGE MODEL

SPEECH ENGINE

After processing the language model there is need for extraction of the commands, the commands are the keywords which are associated with some actions, such as "close that" can be used to close the current program, This can be achieved with the help of the Speech engine which maps the given input with the commands and can further display the results.

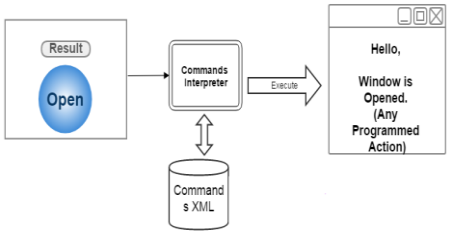

FIGURE 9: SPEECH ENGINE

DISPLAY

The display can be anything where user can see the feedbacks and output for the given input. The user can see the agent responding with the voice output or they can simply see the console screen written with the output. Here in this paper we have the Microsoft Agent characters to provide the feedbacks to the users with gestures and voice.
Microsoft Agent is a technology developed by Microsoft which employs animated characters, avatars and cartoons. In the artificial intelligence system this kind of characters can be used for the interaction with human. People can talk to them and can interact with them, it feels like they are actually talking to the real person.
In background the software program is the responsible for handling the MS-Agent characters and it will provide necessary inputs to the character.
There are several different characters available from Microsoft and we can also design our own character for the system to interact.
Following are the some of the characters that can be used with this system.

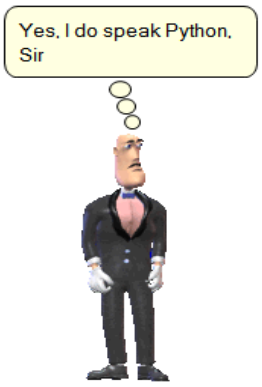

FIGURE 10. MICROSOFT AGENT CHARACTER

4 IMPLEMENTATION

User speaks the word in his/her accent. There may different accents for different people. Speech Recognizer has the Grammar which can define the list of possible synonym that can be extracted from particular spoken stream.

For additional speech-related capabilities, both native and managed interfaces are provided: a COM-based Microsoft Speech API (SAPI), and the Microsoft .NET Framework 3.0 System. Speech. Recognition and System Speech. Synthesis namespaces. Windows Vista provides SAPI version 5.3, although applications that were developed with the use of the SAPI 5.1 SDK should be forward-compatible with SAPI 5.3 on Windows Vista.

Using SAPI the grammar is loaded from the files stored as XML. Normal XML file contains the list of possible grammar commands and associate ID with each which is later use to execute that particular command.

There may different XML files with different commands and categories. To provide more accurate response there are more than one grammar files are available.

After extraction of particular command the system must execute the appropriate task associated with the command. Providing the capability of hands-free computing the system must perform certain task automatically after receiving the commands form the users. This is achieved using simple command executor module. Command executor/interpreter is responsible only for matching the NAME of the RULE in the XML grammar file for which the grammar recognition is already done.

```
    <P>up</P>
  </RULE>
  <RULE NAME="Down" ID="RID_Down" TOPLEVEL="ACTIVE">
    <P>down</P>
  </RULE>
  <RULE NAME="Right" ID="RID_Right" TOPLEVEL="ACTIVE">
    <P>right</P>
  </RULE>
  <RULE NAME="Left" ID="RID_Left" TOPLEVEL="ACTIVE">
    <P>left</P>
  </RULE>
  <RULE NAME="Enter" ID="RID_Enter" TOPLEVEL="ACTIVE">
    <L>
      <P>enter</P>
      <P>run</P>
      <P>O K</P>
    </L>
  </RULE>
  <RULE NAME="Escape" ID="RID_Escape" TOPLEVEL="ACTIVE">
    <L>
      <P>escape</P>
      <P>cancel</P>
    </L>
  </RULE>
```

FIGURE 5. GRAMMAR XML FILE

INITIALIZING THE SPEECH RECOGNITION CONTEXT

By creating object of RecoContext we can handle the Recognition events. Here in below given code we have created handler for the recognition to get the phrases in the running context.

```
objRecoContext =
new SpeechLib.SpSharedRecoContext();
objRecoContext.Recognition+=
new_ISpeechRecoContextEvents_RecognitionEventHa
ndler (RecoContext_Recognition);
```

LOADING GRAMMAR FROM FILE

Loading grammar from specific file limits the unnecessary and wrong recognition. The grammar file will include the rule and names associated with that. All the grammar files have their commands associated with them, these commands are only the phrase that can be recognized when the SAPI is loaded from that particular grammar file. Grammar files are typically XML files.

Following code snippet shows the loading of XMLCommands.xml file in the SAPI with load Option dynamic.

```
grammar.CmdLoadFromFile(appPath+FileName,Speech
Lib.SpeechLoadOption.SLODynamic);

grammar.CmdSetRuleIdState(0,SpeechRuleState.SGDS
Active);

//save as previous grammar for commands list
if (FileName != "XMLCommands.xml")
        previousGrammar=FileName;
```

Following are the content of the grammar xml file:

```
<RULE NAME="Next" ID="RID_Next" TOPLEVEL="ACTIVE">
  <P>next </P>
</RULE>
<RULE NAME="MyComputer" ID="RID_MyComputer" TOPLEVEL="ACTIVE">
  <P>My computer</P>
</RULE>
<RULE NAME="RightClick" ID="RID_RightClick" TOPLEVEL="ACTIVE">
  <P>Right click </P>
</RULE>
<RULE NAME="Back" ID="RID_Back" TOPLEVEL="ACTIVE">
  <P>go back </P>
</RULE>
```

RULE defines the name and id of the Phrase.

P tag defines the actual phrase which is recognized by users.

5 ACKNOWLEDGMENTS

I would like to take this opportunity to bestow my acknowledgement to all my friends and people who have directly or indirectly availed me in making my project feasible and to turn it up into successful piece of work.

I am very thankful to Mr.Sumit Chaudhary for creating the light environment which well suited for the outgrowth of the student personality. His insinuations will help a lot in sailing my ship in the professional world.
This research was really very fruitful for me, it has helped me to gather information about various aspects of working of trade and has broadened my vision on the applicability and the implementation of this system.
Lastly, I am very grateful to my family members and friends for moral support given throughout the study.

CONCLUSION

Voice recognition (in many contexts also known as automatic speech recognition, computer speech recognition or erroneously as Speech recognition) is the process of converting a speech signal to a sequence of words in the form of digital data, by means of an algorithm implemented as a computer program. The recognized words can be the final results, as for applications such as commands & control, data entry, and document preparation. They can also serve as the input to further linguistic processing in order to achieve speech understanding. This project will cover the "connected word, speaker independent and small vocabulary" Voice recognition. The concept will compose of three phases:

1) Training phase: In this phase, a number of words will be trained to extract model for each word.
2) Recognition phase: In this phase, a sequence of connected word is entered by microphone and the system will try to recognize these words.
3) Operation phase: The system performs the necessary action on behalf of user by extracting the commands from the input given by the users.

REFERENCES

[1] A. Augello, A. Santangelo, S. Sorce, G. Pilato, A. Gentile, A. Genco, and S. Gaglio, "A multimodal interaction guide for pervasive services access," in Pervasive Services, IEEE International Conference on. IEEE, 2007, pp. 250-256.

[2] A. Augello, G. Pilato, A. Machi, and S. Gaglio, "An approach to en hance chatbot semantic power and maintainability: Experiences within the frasi project," in Semantic Computing (ICSC), 2012 IEEE Sixth International Conference on. IEEE, 2012, pp. 186-193

[3] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/ contrib. /supported/ IEEEtran/

[4] F. Mikic, J. C. Burguillo, M. Llamas, D. Rodrfguez, E. Rodrfguez et aI., "Charlie: An aiml-based chatterbot which works as an interface among ines and humans," in EAEEIE Annual Conference, 2009. IEEE, 2009, pp. 1-6.

[5] A. Augello, A. Santangelo, S. Sorce, G. Pilato, A. Gentile, A. Genco, and S. Gaglio, "A multimodal interaction guide for pervasive services access," in Pervasive Services, IEEE International Conference on. IEEE, 2007, pp. 250-256.

[6] D. De Orlando and F. Giovanni, "An integrated system, with natural language management, for the monitorinn activities in e-Iearning en vironments," in Complex, Intelligent and Software Intensive Systems, 2008. CISIS 2008. International Conference on. IEEE, 2008, pp. 965-970.

[7] G. Pilato, A. Augello, G. Vassallo, and S. Gaglio, "Sub-symbolic semantic layer in eyc for intuitive chat-bots," in Semantic Computing, 2007. ICSC 2007. International Conference on. IEEE, 2007, pp. 121- 128.

[8] F. Mikic, J. C. Burguillo, M. Llamas, D. Rodrfguez, E. Rodrfguez et aI., "Charlie: An aiml-based chatterbot which works as an interface among ines and humans," in EAEEIE Annual Conference, 2009. IEEE, 2009, pp. 1-6.

[9] S. du Preez, M. Lall, and S. Sinha, "An intelligent web-based voice chat bot," in EUROCON 2009, EUROCON'09. IEEE. IEEE, 2009, pp. 386-391.

Biography

Meghavi M Prajapati is studying Masters of Engineering in Indrashil Institute of Science & Technology (Saraswati College of engineering & technology). She has completed her bachelor degree in Computer science from the same college. She has started her career in computers from the year 2009. During her diploma she has learnt a lot about the computers and then after she has started graduation in computer engineering. She is very quick learner and self-motivated person. Meghavi had participate in many events and presentations during her curriculum one of them is the GSM and CMS which tells about the Mobile computing with GSM technologies and CMS for the content management systems. She is very quick learner and confident personality. Meghavi has developed LocateThem which is a mobile application used for tracking the users through their mobile phones. This systems was developed during her education period and it was a big hit in the college.

Sumit Chaudhary is working as Head of Department in CSE department at Indrasheel Institute of Technology, Kadi, Rajpur, Gujarat. He obtained his M-Tech (Computer Engineering) with Hons. from Shobhit University and B-

Tech (Computer Science) from SCRIET, Meerut (U.P.). He has been in teaching from more than four years.

During this short period of time, he has been supervised several dissertation of M.Tech. Students. He has been member of several academic and administrative bodies. During his teaching he has been coordinated many Technical fests and National Conferences at Institute and University Level. He has attended several seminars, workshops and conferences at various levels. His many papers are published in various national, international journals and conferences. His area of research includes Wireless Sensor Network (WSN), Network Security, Neural Network, Artificial Intelligence and MANET (Mobile Ad-Hoc network).