International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Keyword Based Speech Synthesis to Control Activities of Mobile Robot

Anil Kumar[1], K.R. Shylaja[2], Ravinandan M E[3]

[1]M.Tech IV semester, Dept. of CSE, Dr. Ambedkar Institute of Technology, Bangalore, Karnataka, India

[2]Associate Professor, Dept. of CSE, Dr. Ambedkar Institute of Technology, Bangalore, Karnataka, India

[3]Teminnova Technologies Private Limited, Bangalore, Karnataka, India

Abstract: ***Human-Machine Interface (HMI) is a part of machine which handles the Human-Machine Interaction; machines can be interacted using either Speech Recognition, Gestures or Brain-Machine Interface etc. In this paper a novel idea of controlling the activities of Mobile robot using Speech Recognition has been proposed. This approach is named as Keyword based Speech Recognition because certain set of keywords are used to control the mobile robot.***

Keywords: Human-Machine Interface (HMI), Speech Recognition, Mobile Robot, Sphinx, Bluetooth Communication

1. Introduction

HMI is a medium for information exchange and mutual communication between electromechanical system's and the user. It allows the user to complete settings through touchable images, Speech, Gestures, keys on the user-friendly window. This not only offer's fast and convenient control of manufacturing automation, but also has replaced traditional controlling panel's which need extensive wiring.

In the **Existing System** physical devices like mouse keyboard are required to communicate with computers or any other electronic devices. These devices are not portable because a physical connection with the system is all time required, this will hinder the usage of systems at all places and at all times.

In the **Proposed System** an alternative way to communicate with computer and robots has been proposed, this work mainly focuses on creating a different environment where we can communicate with computers in a human friendly way for example controlling system through speech. This doesn't require any specific hardware and is operable anytime anywhere.

1.1 Speech Recognition

Speech recognition, or more commonly known as automatic speech recognition (ASR), is the process of interpreting human speech in a computer [2].Where it can also be defined as ASR as *the* building of system for mapping acoustic signals to a string of words [3].

The main goal of speech recognition is to get efficient ways for humans to communicate with computers. This can be an important application for physically disabled, lawyer etc. Another application can be environmental control, such as turning on the light, controlling the TV etc.

We feel that speech recognition is important, not because it is 'natural' for us to communicate via speech, but because in some cases, it is the most efficient way to interface to a computer [2].

2. Procedure

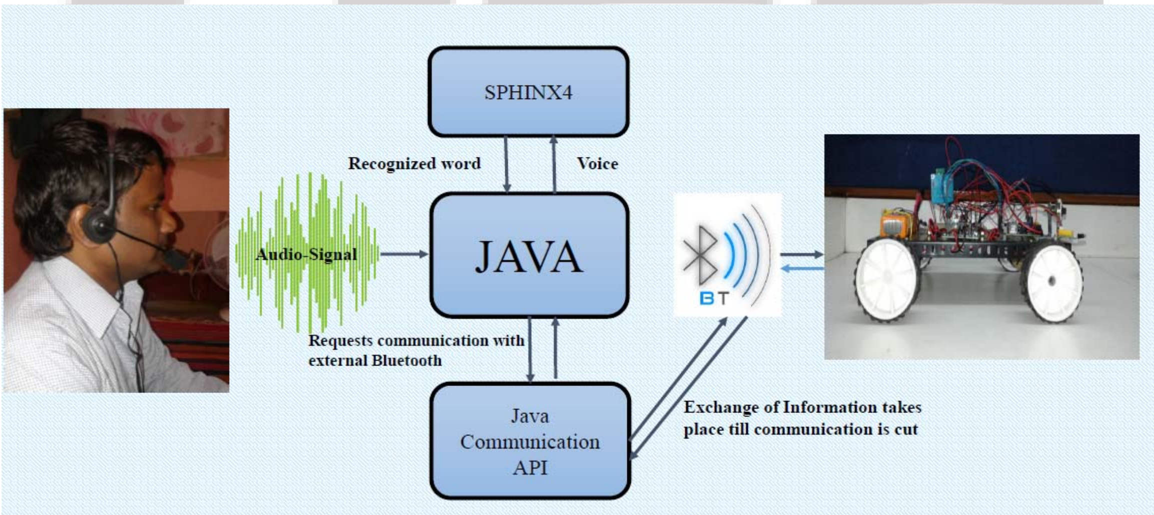

Figure 1: Functional flow diagram

Licensed Under Creative Commons Attribution CC BY

Figure 1 shows a simplified architecture of speech recognition. Here we give speech input in form of certain keywords. We have a grammar file written which will have all these keywords stored in it. The input keyword is matched against the grammar file. If a match is found it will perform the desired action which it is meant for. For example if a keyword called "front" is used then robot moves forward till the next instruction is specified. This architecture mainly consists of three major parts i.e.

1) Sphinx-4 library which helps in identifying the spelled word.
2) Mobile Robot which performs the locomotion based on user input.
3) Bluetooth primarily used to communicate instructions to the robot.

2.1 Sphinx-4

Sphinx-4 is a pure Java speech recognition library. It is very flexible in its configuration, and helps in speech recognition tasks by detecting the words based on its phonics. It recognizes the words by comparing with the stored dictionary of words.

The Sphinx framework has been formed with a high degree of flexibility and modularity. Figure 2 shows the overall architecture of the system. There are three main modules in the Sphinx framework: the **FrontEnd,** the **Decoder**, and the **Linguist**. The FrontEnd takes one or more input signals and arranges them into a sequence of Features.

The Linguist translates any type of standard language model, along with pronunciation information from the Dictionary and structural information from one or more sets of AcousticModels, into a SearchGraph. The SearchManager in the Decoder uses the Features from the FrontEnd and the SearchGraph from the Linguist to perform the actual decoding, generating Results. At any time prior to or during the recognition process, the application can issue Controls to each of the modules, effectively becoming a partner in the recognition process.

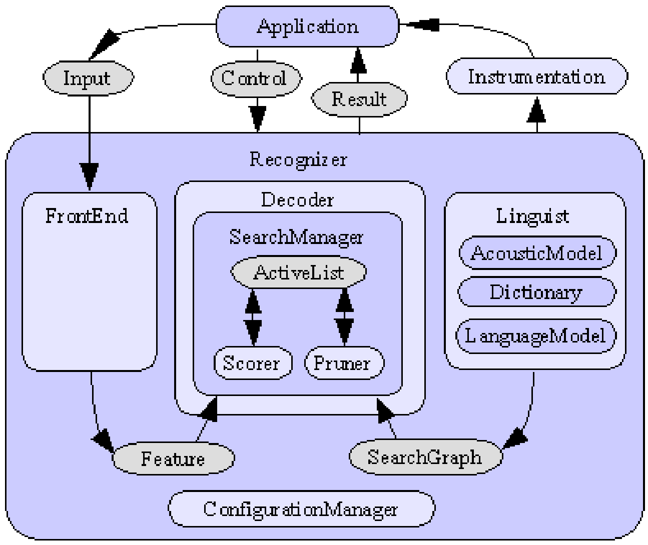

Figure 2: Sphinx-4 Decoder Framework

2.1.1 Frontend

The purpose of the FrontEnd is to arrange an Input signal (e.g., audio) into a sequence of output Features. As shown in the Figure-3, it performs Digital Signal Processing and comprises one or more parallel chains of replaceable communicating signal processing modules called DataProcessors.

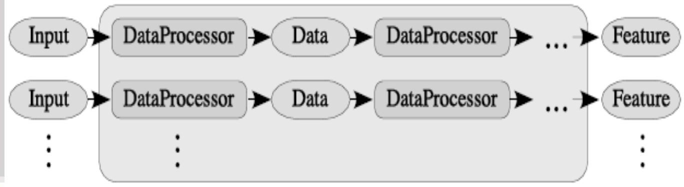

Figure 3: FrontEnd

2.1.2 Linguist

- **Acoustic model**: Acoustic model contains a representation of the sound, often created by training using lots of acoustic data.
- **Dictionary:** Dictionary is responsible for determining how a word is pronounced.
- **Language Model:** Language Model contains a representation of the probability of occurrence of words.
- **Search Graph:** This graph structure produced by the linguist according to certain criteria (e.g. Grammar) using knowledge from the dictionary, the acoustic model and the language model. SearchGraph is the primary data structure used during the decoding process which is illustrated in Figure-4.

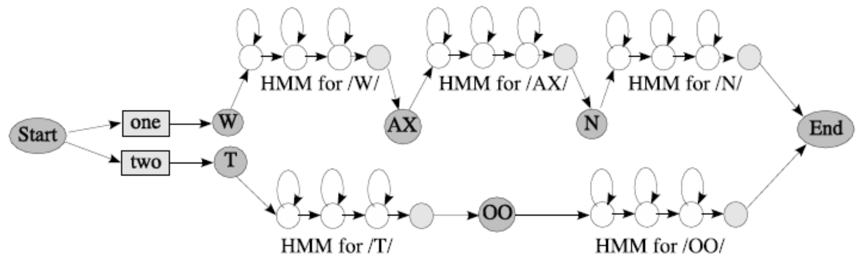

Figure 4: The SearchGraph is a directed graph composed of optionally emitting Search States and Search StateArcs with transition probabilities. Each state in the graph can represent components from the LanguageModel (words in rectangles), Dictionary (sub-word units in dark circles) or AcousticModel (HMMs).

Licensed Under Creative Commons Attribution CC BY

2.1.3Decoder

Decoder contains the search manager.

- **Search Manager:** search manager performs search using certain algorithm used e.g. SimpleBreadth First Search Manager, Word Pruning Breadth Search Manager, Bushder by Search Manager, Parallel Search Manager etc. also contains the feature score and the pruner.
- **Active List:** A list of tokens representing all the states in the search graph that are active in the current feature frame.
- **Scorer:** scores the current feature frame against all the active states in the Active List.
- **Pruner:** pruner the activelist according to certain strategies.

2.2 Mobile robot

It is two wheeled robot driven by the DC motors at two corners at the back. Three obstacle sensor placed at the front, one being the distance sensor is placed at the centre and two IR proximity sensor in right and left side of the bot. Figure 5 shows the top view of the Arduino robot and Figure 6 shows the Block diagram of the robot, and Table 1 specifies the complete specifications of the robot.

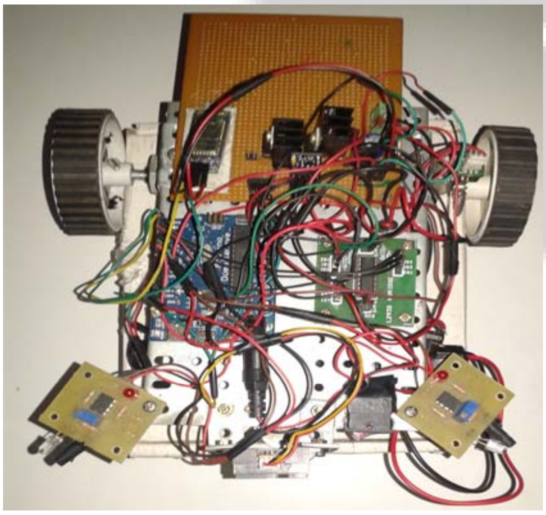

Figure 5: Mobile bot top view

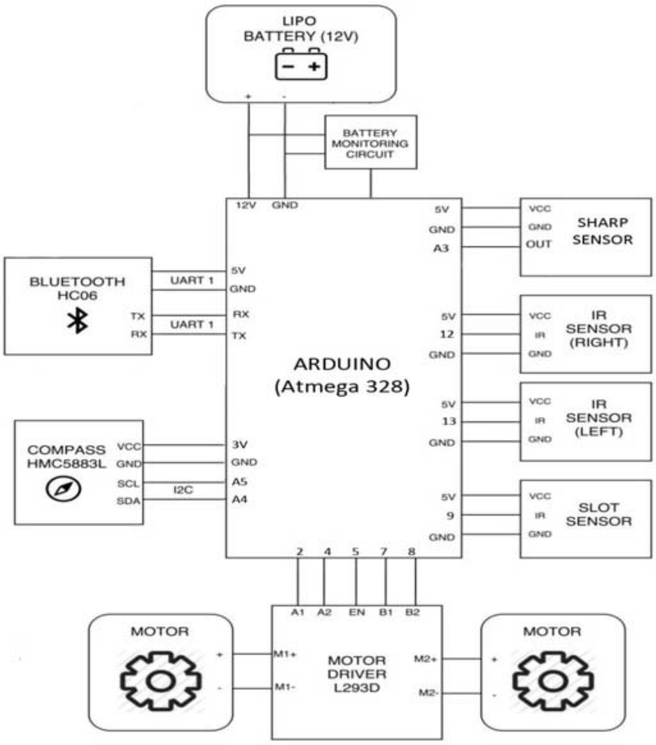

Figure 6: Block diagram of Mobile bot

Table 1: Hardware specifications of bot

Specification	*Mobile bot*	
	Type	*Part name*
Primary Controller	Arduino	Atmega 328
Secondary Controller	-	-
Communication	Bluetooth	Linvor
Obstacle detector	Infrared	SHARP
Motor driver	Dual H-bridge	L293D
Distance sensor	Infrared	slot sensor
Proximity sensor	Infrared	-
Battery	LiPo	11.1V,2000mAH
Voltage regulator	-	-

3. Experiments and Results

Many test trials were run with different kind of test environments. The details of the tests are as given below.

3.1 Test in normal office environment

In the normal office environment almost all the words were recognized correctly. Out of 10 words 7-8 words were recognized. So accuracy was 70-80%.

3.2 Test on traffic road

In this environment due to the external noise the accuracy was less compared to the first scenario. Out of 10 words 5-6 words were recognized and hence the accuracy reduced to 60%.

3.3 Test in exhibition environment

In this case we were surrounded by some 400-500 people, this that mass environment the accuracy was about 60% same as the results of traffic road.

Licensed Under Creative Commons Attribution CC BY

4. Conclusion

In this paper we could successfully recognize the keywords spoken by the user using spinix-4 library. Through these keywords we were able to control the mobile robot. We had a challenge while making the robot work in noisy environment as the accuracy was less. The proposed idea is very easy to implement and special training is required for the user.

Speech recognition is been done in English language by specifying keywords from English.

5. Future work

Increasing accuracy is the major concern and extending recognition of words from other languages can be done in future. This extension will make the idea even more user friendly.

References

[1] Human-Machine Interaction (HMI): A Survey Technical Report: CES-508 James Cannan and Huosheng Hu School of Computer Science & Electronic Engineering University of Essex.

[2] Why is Speech Recognition Difficult? Markus Forsberg Department of Computing Science Chalmers University of Technology markus@cs.chalmers.se February 24, 2003

[3] J. H. M. Daniel Jurafsky. *Speech and Language Processing, An introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, Upper Saddle River, New Jersey 07458, 2000.

[4] Sphinx-4: A Flexible Open Source Framework for Speech Recognition Willie Walker, Paul Lamere, Philip Kwok, Bhiksha Raj, Rita Singh, Evandro Gouvea, Peter Wolf, Joe Woelfel.

Author Profile

Anil Kumar received the B.E. degree in Information Science Engineer from Banglore Institute of Technology, Banglore in 2012 and pursuing MTech degree in Computer Science and Engineering from Dr. Ambedkar Institute of Technology, Bangalore in 2012-2014. Having high interest in robotics.

Mrs. K.R. Shylaja received MTech degree in Computer Science from Visvesvaraya Technological University. She is currently perusing her Ph.D from JNTU, Kakinada, Andhra Pradesh, India. She has 14 years of experience as lecturer and Asst. Professor, Currently she is an Associate Professor, Department of Computer Science & Engineering, Dr. Ambedkar Institute of Technology, Bangalore.

Ravinandan M E received B.E degree in Computer Science Engineer from Bangalore University, MTech degree in Computer Science & Engineering from VTU. He had served as a scientist in DRDO and GE Healthcare. Currently he is MD & CEO of Teminnova Technologies Private Limited and also a Founder Member of Organization for Rare Diseases India.

Licensed Under Creative Commons Attribution CC BY