Available online at http://journalijces.com

INTERNATIONAL JOURNAL OF CURRENT ENGINEERING SCIENCES

REVIEW ARTICLE

ISSN: 2320- 9046

International Journal of Current Engineering Sciences- Vol. 5, Issue, 7, pp. 52-55 July, 2016

AUTOMATIC SPEECH RECOGNITION BY MOBILE ROBOT – REVIEW

Suraj Kumar and Pankaj Mohindru

Department of Electronic Technology, Punjabi University, Patiala

ARTICLE INFO

Article History:

Received 21st April, 2016
Received in revised form 8h May, 2016
Accepted 5th June, 2016
Published online 28th July, 2016

Key words:

Modeling techniques, robotics, and Speech recognition.

ABSTRACT

The main objective of speech recognition technology is to achieve better accuracy of communication between human and machine. This paper discussed all aspects, principles and methods of classification of speech recognition system, the use of speech recognition in robotics, definition of various types of speech classes, Speech representation and various feature extraction techniques, speech classifiers, Database and their performance evaluation.

INTRODUCTION

Speech is a common way of communication for human. Research in speech recognition was motivated by people desire to interact with machine through voice. After research from decades and also advancement in the computer technology it became possible for human to interact with machine through voice. ASR makes possible for machine to understand and identify human voice commands and react accordingly. The main objective of the speech recognition technology is to make an efficient and high accuracy machine that identify and understand human language and react according to human voice accurately. Speech recognition is a technology that converts human voice to text [1]. With the rapid advancement in computer technology ASR technology has been also advancing day by day and found in many application like in automation, banking, secure voice access system, in robotics etc. speech recognition is application of pattern recognition [2]. There is number of issue that is needed to be tackle. Speech recognition system classified as discrete and continuous.

Types of speech

A. ***Isolated word speech***- isolated word means single word at a time. Isolated word is form of 'command and control' application, in which machine is capable of understand and identifying single word and respond according to command [2]. Isolated word contains silence and background noise. For isolated word system there is need to be accurately detection of the end points of a spoken word to achieve high accuracy of system [4].

B. ***Connected word speech***- it is similar to isolated word that are spoken together with minimum pauses between them. This technology is very useful in recognizing digit strings and alphanumeric strings and catalog ordering [3].

C. ***Continuous speech***- continuous speech means naturally spoken sentences. Continuous speech is very difficult to recognize, because word boundaries are unclear.

D. ***Spontaneous speech***- it is very difficult, because it tends to be interspersed with dis-fluencies like "um" and "uh", false starts, incomplete sentences, stuttering, mumbling. The complex computation is required for this task and also there is no clear method for efficiently and accurately determine spontaneous boundaries [5].

Approach to Speech Recognition-Acoustic variability is more difficult to model, partly because it is so heterogeneous in nature. Consequently, research in speech recognition has largely focused on efforts to model acoustic variability. Past approaches to speech recognition have fallen into three main categories:

I. ***Template-based approaches***-Where it is compared to the unknown speech against a group Words in advance, to find similarity between two. This has the advantage of using the word

Corresponding author:* **Suraj Kumar
Department of Electronic Technology, Punjabi University, Patiala

accurate models altogether. But it also has the drawback the patterns are fixed in advance, so the differences in expression can only be using many forms of the word, which eventually becomes impractical style.

II. ***Knowledge-based approaches***-Which "experts" knowledge to speak directly to the question - is encoded in this system. It is clearly important speech difference modeling. But such expert knowledge unfortunately necessitated the access to be impractical, and gets automatically instead sought ways to learn and difficult to use with success [8].

III. ***Statistical-based approaches***-Which are modeled statistically significant differences in speech (for example, through the hidden Markov models or HMMs), using the automatic learning procedures [7]. The main drawback of statistical models is that they must be submitted prior model assumption, which is likely to be inaccurate, which cripples system performance. We will see that the neural networks help to avoid this problem [5]

Methodology of Speech Recognition

The fig 3.1 shows the block diagram of speech recognition system. The basic principle of speech recognition includes training and testing of speech database to recognizer output. The both training and testing include various processes that are discussed as following;

Input speech signal- input speech signal is a collection prerecorded audio signal. That is further used for speech recognition process. Input speech signals typically a highly sampled speech data [5].

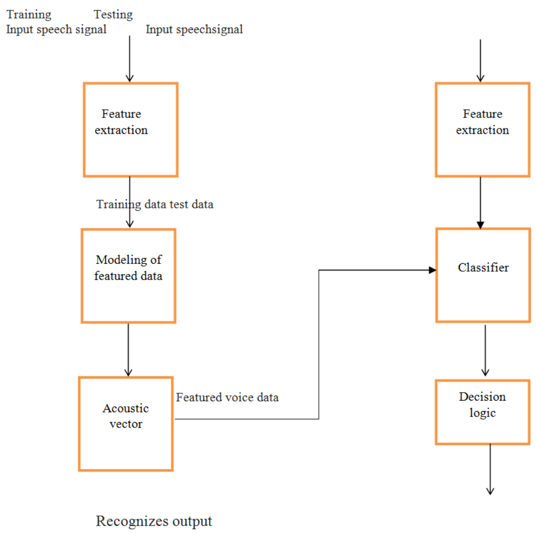

Fig 1. Block diagram of speech recognition system [1]

Feature extraction- it is a very important step in speech recognition process. Input speech signal is a time varying signal so some features extraction is performed to reduce the variability, and also eliminating noise, silence etc. there are various feature extraction techniques used in speech recognition technology such as LPC, MFCC, VQ are the name of few. LPC (linear predictive coding) - it is algorithm which used to approximate the recent history of raw speech input [7].MFCC (Mel frequency cepstral coefficient) – cepstral analysis compute the logarithmic values of input speech signal [9].

***Feature classification* –** feature classification is process of computing the similarity between the test data and training data to give final recognizer output. There are number of feature classification technique like ANN (artificial neural network), HMM (Hidden Markov model)**,** GMM (Gaussian mixer model) etc.

Hidden Markov Model- The hidden Markov model is a set of states linked by transitions; it begins in the initial state designated. Each discrete time step, and the transition is Taken into a new state, after which it is to create a single code that came out the state [6]. Selection Transitional phase and output are both random code, the prospect judged by distributions. The HMM it can be thought of as a black box, where the sequence of output symbols generated during the observed time, the objective of the states visited the sequence over time is hidden from sight. This is why it's called Hidden Markov Model [8]. HMMs have a variety of applications. When applied to HMM speech recognition, the interpretation states acoustic models, referring to what is likely to hear the sounds during the corresponding segments of speech; while transitions provide time constraints, indicating how states may follow each other in sequence. Because speech always goes forward in time and the changes in the application of speech always go forward (or make and self- ring, and let's have the state's two arbitrary).

Artificial Neural Network - Neural network made up for a possible large number of simple processing elements (calledUnits, contract, or nerve cells) that affect the behavior of some people through a network of excitatory Or inhibitors of weight. Each unit is simply calculates the likely amount of the non-linear input, Itdistributes outgoing connections to other units results [6]. A set of training consist typical values that are assigned to units of the input and or output designate. As patterns the view of the training set, learning rule amends the strengths of the weights so the network gradually learns the training set. This basic paradigm can be fleshed out in many different ways, so that different types of networks can learn to calculate the implicit function of input and output vectors, or automatically block data entry, or the generation of compact .the representation of the data, or the availability of content addressable memory and performance of the finished pattern. There are number of different algorithm of artificial neural network such as single layer Perceptron model, multilayer Perceptron model, Kohenen self-organizing method, deep neural network, radial basic function etc.

***Vector Quantization* -** the system identify the language, and a unique representation of each speaker is an effective way through a process known As vector quantization which is based on the principle of mass Coding. Vector quantization is the process for mapping the carriers bound for a large space for a limited number of areas in this space and it knows all of the region as a block can be the middle of a term known as code word [9]. All code words are known as the codebook. Thus each group or represent cluster in a different category of words that signal is Data compression dramatically, however, still represents accurately. The system would be a very large and complex mathematically if feature vector does not quantized. The course starts test or determine by finding Euclidean distance between the speaker feature vectors for unknown model ofknown speakers in the database. To identify unknown speaker minimum distance between the unknown speaker and codebook of feature vector is calculated [11].

$$\begin{pmatrix} =0, \text{ if } v_i = v_j \\ \\ >0, \text{ otherwise} \end{pmatrix}$$

The equation shows that the similarity is measured in the form ofv_i, v_j. There is no. of mathematical process to calculate similarity distance.

In the social context, people are used to speak in their natural language so they expect that robot understand and identify constrained spoken language. To do so the most important part is to design a robot according to the people instructions that they want to do from robot. For example, if we give ‘run on green light’ it will need to provide color vision. Continuous testing with user instruction is most important thing in design process of robot. The instruction design should not limit to individual user it should be like that other member also able to use it. Speech recognition system is taking interest in robotics for the same reason. After decades of research, speech recognition system has been advancing and getting mature in the voice user interface area. Researchers are still working to overcome issues of speech recognition system and number of research work is going in the field of voice user interface.

Authors	Parameter of speech	Methods of classification	Performance	Apllication
1.E.Chandra,C. Sunitha [7]	Continuous Speech	RBF NEURAL NETWORK	94.82%	identify speaker identity
2.Steve,Herve,Nelson[12]	Connected speech	CI-HMM	92%	Context independent MLP-HMM hybrid system
3.Yifan Gong [10]	speech in noisy environment	HMM Vector Equalization	98.3%	multi-speaker system

Application

There is some practical application of speech recognition; AT & Bell developed a VRCP in 1992. The system is small word vocabulary speech recognition system [1].

Charles Schweb developed a first large scale speech recognition system ‘the stock quotation system’.

The US telecommunications network developed a voice recognition system in 2000 that provides customer service, voice dialing, change address and check number and other services. At the same time china also developed a speech recognition system called Voice value added services (Cell-VVAS) [10].

Speech recognition in robotics

Over the past few years, great advancement has been made in both speech recognition and synthesis. it makes possible to make voice only controlled robot. The concept of robotics is to make an automatic machine which can reduce the human labor work. To provide control commands to robot and guide line for work, human should capable to communicate with robot, this concept accomplish the robotics to introduce voice user interface with robot. In the past eras, keyboard, joysticks, keypad are the guiding tools for interface to machine. Now there are number of techniques are introduce in human machine interface field; speech recognition techniques is an interesting tools for the researchers to human machine interaction field. Because of speech recognition technology it made possible to transferred voice data to machine, for example, if the user give the order “go to office ”the ASR system will convert that order into corresponding event. The challenge of human robot interface is that to make an efficient and autonomous system which can understand instruction of the no-voice user and also understand natural language accurately. The voice interface is newly added in human robot interaction.

Comparative Analysis on The basis of past researches

A speech recognition analysis is given in table below, on the basis of past researches. This comparisons results they are reported, which speech system they used and performance. This table contains different five researches in the field of speech recognition.

CONCLUSION

Interaction between human and mobile robot is an important, interesting and challenging field in human robot interaction. Robot controlled by voice getting more interest to make it more users friendly in the social content. Speech recognition makes possible for researcher to add natural language and multi- language communication with mobile robot. The main objective of human robot interface to makes an efficient and accurate service robot that will help people in everyday life in an efficient manner. This review present the use of speech recognition in voice over user interface for interacting with mobile robot and investigate the use of natural language as a user interface for interaction with robot. In this review we have described different techniques of speech recognition and the basic principle and classification methods of speech recognition.

References

1. JianliangMeng, Junwei Zhang and Haoquan Zhao, “Overview of the Speech Recognition Technology”, 2012 Fourth International Conference on Computational and Information Sciences, 978-0-7695-4789-3/12$26.00©2012 IEEE.
2. Giuseppe Riccardi, “active learning: Theory and applications to automatic Speech Recognition”, IEEE transactions on speech and audio processing, Vol. 13, No. 4, July 2005.

3. ZahinKaram, William Campbell, “A new Kernel for SVM MIIR based Speaker recognition”, “MIT Lincoln Laboratory, Lexington, MA, USA.
4. SantoshK. Gaikwad, BhartiW. Gwali and PravinYannawar, “A Review on Speech Recognition Technique”, *International Journal of Computer Applications* (0975 – 8887) Volume 10– No.3, November 2010.
5. Joe Tebelskis,” Speech Recognition using Neural Networks”, a thesis submitted for the degree of doctor of philosophy School of Computer Science Carnegie Mellon University Pittsburgh, Pennsylvania.
6. WouterGevaert, GeorgiTsenov, ValeriMladenov*,* Senior Member, IEEE*,”* Neural Networks used for Speech Recognition*”,* journal of automatic control, university of belgrade, vol. 20:1-7, 2010.
7. Chandra and C. “A review on Speech and Speaker Authentication System using Voice Signal feature selection and Extraction”, 2009 IEEE *International Advance Computing Conference* (IACC 2009) Patiala, India, 6-7 March 2009.
8. Joseph Picone, “continuous speech recognition using hidden Markov models”, IEEE Assp magazine, July 1990.
9. Anjali Jain, 2O.P. Sharma,”A Vector Quantization Approach for Voice Recognition Using Mel Frequency Cepstral Coefficient (MFCC): A Review**”,** IJECT Vol. 4, Issue Spl- 4, April - June 2013.
10. Yifan Gong, “Speech recognition in noisy environments: A survey”, Speech Communication 16(199.5) 261-291,0167-6393/95/$09.50 0 1995 Elsevier Science B.V.
11. José Leonardo Plaza-Aguilar, David Báez-López, Luis Guerrero-Ojeda and Jorge Rodríguez Asomoza, “A Voice Recognition System for Speech Impaired People”, Proceedings of the 14th International Conference on Electronics, Communications and Computers (CONIELECOMP’04) 0-7695-2074-X/04 $ 20.00 © 2004 IEEE.
12. Todd A. Stephenson*,* Mathew Magimai Doss and HervéBourlard*,* “Speech Recognition with Auxiliary Information”, IEEE transactions on speech and audio processing, vol. 12, no. 3, May 2004.
13. Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and DongYu, “Convolutional Neural Networks for Speech Recognition”, IEEE/ACM Transactions on Audio, Speech, and Language Processing, VOL. 22, NO. 10, OCTOBER 2014.
14. Steve Renals, Nelson Morgan, Herve Bourlard and Michael Cohen, “Connectionist Probability Estimators in HMM Speech Recognition”, IEEE Transactions on Speech and Audio Processing, VOL. 2, NO. 1, PART 11, JANUARY 1994.

****