SPEECH RECOGNITION TECHNIQUES AND METHODS: A SURVEY

[1] Madhuri Gupta, [2]Dr. Megha Mishra,

[1]Research Scholar, [2]Assistant Professor
[1]Department of Computer Science,SSGI,Bhilai,INDIA,
[2]Department of Computer Science,SSGI,Bhilai,INDIA

Abstract: This paper presents a deep view of what is Speech, types of speech, Speech recognition, techniques, algorithm, pros and cons of different methods and techniques. This paper emphasis on HMM method, ASR methods, and different tools available to recognize speech signal. This report shows the effectiveness ratio of different methods. Importance of neural network and different techniques useful for speech recognition and why artificial intelligence needed for accomplish speech recognition techniques. At the end this paper will show the comparison of each technique with different parameters.

***Index Terms*: Speech Recognition, Neural Network, Artificial Intelligence, and ASR.**

I. INTRODUCTION

In the present world the science and technology are working on making secure and intelligent systems. As you can see the modern developments from the last 2 decades are mostly in the fields of mobile technology, automation, automobile, and infrastructure. In these entire categories one thing is common that is making everything more smart, secure and intelligent. For that several technologies are working together to achieve this goal. Now you can find Google with more intelligent features like automatic spelling check, Google map, Google speaker, Google translator, and many more. Facebook is also using artificial intelligence features like they recognize faces of common friends. Modern websites are more users friendly, easy to use and more interactive. All this only possible due to advance research in the field of Artificial Intelligence and supporting areas like neural networks, and speech recognition etc. In this paper we broadly focus on the artificial intelligence, different techniques available for speech recognition, comparison of different techniques and the applications of speech recognition in modern era.

II. WHAT IS ARTIFICIAL INTELLIGENCE

According to the father of Artificial Intelligence, John McCarthy, it is *"The science and engineering of making intelligent machines, especially intelligent computer programs".*

Artificial Intelligence (AI) is a way of making a computer, a computer-controlled robot, or a software think intelligently, in the similar manner the intelligent humans think.

AI is accomplished by studying how human brain thinks and how humans learn, decide, and work while trying to solve a problem, and then using the outcomes of this study as a basis of developing intelligent software and systems. [1][8].

Different Components of AI: AI is not the single domain it contributes to different area of science, psychology, biology, social science, and computer science etc. As shown below on Figure 1.

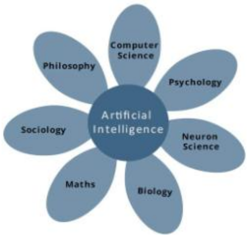


Figure 1: Components of AI

This paper mainly focuses on Speech Recognition which is the application part of AI. In the next sections this paper presents the deep view of it. As Shown in Figure 2.

III. SPEECH RECOGNITION

Definition: The fundamental aspect of speech recognition is the translation of sound into text and commands. Speech recognition is the process by which computer maps an acoustic speech signal to some form of abstract meaning of the speech [3].

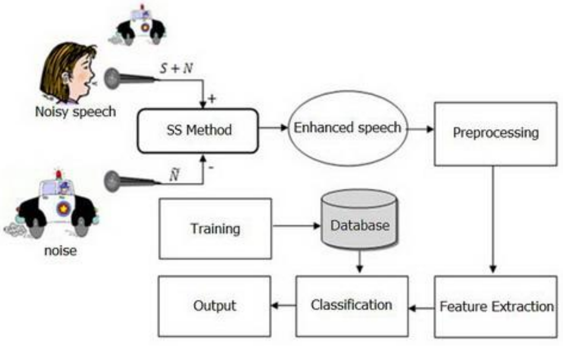


Figure 3: Basic Structure of Speech Recognition

3.1 Structure of Speech Recognition:

The structure of Speech recognition as shown in the figure 3.1 describes the steps involved for the recognition process

- Step 1: Collection of speech signal, it may contain noise.
- Step 2: Using Signal Synthesizer de noise the signal.
- Step 3: Then Preprocessing has been done.
- Step 4: Feature extraction
- Step 5: Training of Database
- Step 6: Classification
- Step 7. Output

3.2History of Speech Recognition

Table1: History of Speech Recognition

S.No	Researcher	Discovery	Techniques	Year
1.	Harvey Fletcher	First Speech Recognition	Speech Perception	1932
2.	three Bell Labs researchers	single-speaker digit recognition	Power spectrum of each utterance	1952
3.	Gunnar Fant	source-filter model of speech production	Useful model of speech production.	1960
4.	Raj Reddy	continuous speech recognition	DTW algorithm, continuous Speech	Late 1960's
5.	Leonard Baum	the mathematics of Markovchains	Markov chains	Late 1960's
6.	James Baker and Janet Baker	speech recognition	HMM	Late 1960's
7.	DARPA	beam search	Linear predictive coding and cepstral analysis.	1971
8.	Fred Jelinek	activated typewriter called "Tangora"	Statistical modeling techniques like HMMs	Mid 1980's
9.	Gunnar Fant	source-filter model of speechproduction	Useful model of speech production	1960
10.	Katz	back-off model	multiple length n-grams	1987
11.	Xuedong Huang	Sphinx-II	speaker-independent, large vocabulary, continuous speech recognition	1992

12.	Lernout & Hauspie	Dragon Systems	the typical commercial speech recognition system was larger than the average human vocabulary	2001
13.	IBM, a team led by BBN with LIMSI and Univ.Pittsburgh, CambridgeUniversity	Speech-to-Text (EARS)	Global Autonomous Language Exploitation (GALE)	2002
14	National Security Agency	speech recognition for keywordspotting	database to find conversations of interest	2006
15	Google	GOOG-411, a telephone based directory service	produced valuable data that helped Google improve their recognition systems	2007
16	Geoffrey Hinton	acoustic modeling	use of deep feedforward	2009
17	Hinton et al. and Deng et al	Gaussian mixture model/Hidden Markov model	neural nets	2009-2010

IV. TECHNIQES AND METHODS

Speech recognition techniques Approaches can be classified on the basis of:

- Acoustic Phonetic
- Pattern Recognition
- Artificial Intelligence

4.1 Acoustic phonetic

The basic step in this approach is a spectral analysis of the speech signal with feature detection method which converts the spectral measurements to a set of features that explain the major acoustic properties of different phonetic units. The next steps are segmentation and validation process. In segmentation the speech signal is segmented into stable acoustic regions, followed by attaching one or more phonetic labels to each segmented region, resulting in a phoneme lattice characterization of speech. In the validation process determines a valid word from the phonetic label sequences produced by the segmentation to labeling. The following table 2 broadly gives the major speech recognition techniques & methods:

Table 2: Major Speech Recognition Techniques & Methods

Techniques	**Method**	**Recognition Function**	**Typical parameters**
Acoustic Phonetic Approach	Phonemes/ Segmentation and Labeling	Probabilistic Lexical Access Procedure	Log Likelihood ratio
Pattern Recognition			
• Template	Speech Samples, Pixels and curves	Coordination distance measure	Classification Error
• DTW	Set of sequence of spectral vector	Dynamic wrapping	Dissimilarity Measure
• VQ	Set of	Optimal	Euclidian

	Spectral Vector	Algorithm	Distance
• Statistical	Features	Clustering Function	Classification Error
Neural Network	Speech Perception/ rules/Features/Units/procedure	Network function	Mean Error Function
Support Vectors Machine	Kernel Based machine	Maximal Margin hyper plane	Minimizing a bound on Generalization error
Artificial Approach	Knowledge Based		Word error Probability

4.2 Pattern Recognition

This technique involves two important steps which are namely, pattern training and pattern comparison. The basic feature of this technique is that it uses a well formulated mathematical algorithm and consistent speech pattern representation.

The whole idea is shown in the figure 4 with block diagram which also shows the methods involved namely template approach and stochastic approach.

Figure 4: Block Diagram of Pattern Recognizer

4.2.1. Template Based Approach

It is the simplest method of pattern recognition technique in which a collection of prototypical speech patterns are stored as reference patterns representing the dictionary of speaker’s words. And then recognition process carried out by matching an unknown spoken utterance with each of these reference templates and selecting the category of the best matching pattern. Most commonly templates for entire words are constructed.

4.2.2. Stochastic Approach

This method uses probabilistic models to deal with uncertain or incomplete information. This uncertainty can be arises from many sources for example, confusable sounds, speaker variability, contextual effects and homophones words. So this method is more suitable for speech recognition.

4.3 Dynamic Time Warping (DTW)

It is an algorithm for measuring similarity between two sequences which may vary in time or speed. This method is more suitable for video, audio, and graphics indeed, any data which can be turned into linear representation.

4.4. Vector Quantization (VQ)

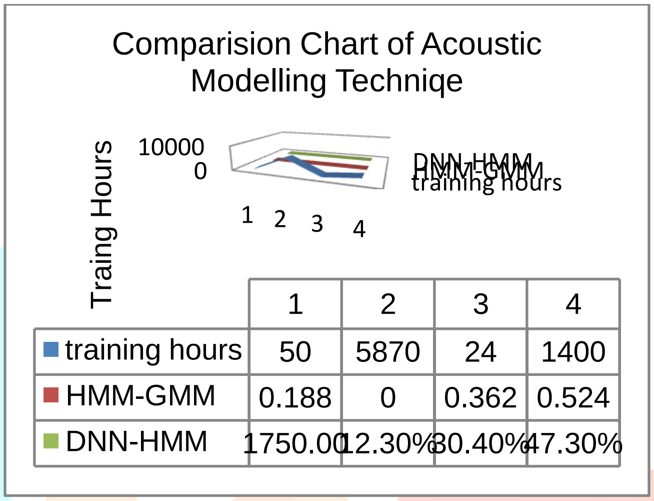
This method is commonly applied to ASR. Majorly it is useful for speech codes i.e. efficient data reduction. The codebook is prepared for the collected data which further used as reference for evaluation.

4.5. Artificial Intelligence approach (Knowledge Based approach)

In this approach a hybrid of the acoustic and pattern recognition approach has been followed. For information regarding linguistic, phonetic and spectrogram knowledge based approach is useful. Researchers are more focused on human speech processing which require knowledge representation of the artificial intelligence approach [4].

IV. COMPARISION OF VARIOUS ACOUSTIC MODELLING TECHNIQUES

In the below figure 5.1 shows the 3d comparison chart of various acoustic modeling techniques. In this figure comparison has been done on different categories such as HMM-GMM, DNN-HMM and the training hours and the data set of different broad companies such as English Broadcast news, Google voice inputs, Bing voice search and You Tube[5][6].



	1	2	3	4
🟦 training hours	50	5870	24	1400
🟥 HMM-GMM	0.188	0	0.362	0.524
🟩 DNN-HMM	1750.00	12.30%	30.40%	47.30%

Figure 5: 3D Chart Comparison of percentage word error of various acoustic modeling Techniques

Where data sets

1 = English Broadcast news

2 = Google voice inputs

3 = Bing Voice Search

4 = You Tube

VII. CONCLUSION

This paper has presented various uses of ASR in modern world. After discussing various Speech recognition techniques, history, and comparisons this paper conclude that in today's world there is lot more to be discover in the field of speech recognition. Lots of discovery already has been done with using different techniques. Yet there is hope to create a new algorithm, new language recognition, different parameters discovery, and new technique to recognize speech signal.

VIII. REFERENCES

[1]M.A.Anusuya, S.K.Katti "Speech Recognition by Machine: A Review" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009.

[2]Suman K. Saksamudre, P.P. Shrishrimal, R.R. Deshmukh " A Review on Different Approaches for Speech Recognition System" International Journal of Computer Applications (0975 – 8887) Volume 115 – No. 22, April 2015.

[3] KIRAN.R, NIVEDHA.K, PAVITHRA DEVI "VOICE AND SPEECH RECOGNITION IN TAMIL LANGUAGE" 978-1-5090-6221-8/17/$31.00_c 2017 IEEE".

[4]. Anjali Garg,Poonam Sharma "Survey on Acousting modeling and feature extraction for Speech Recognition"@IEEE 2016.

[5] JON BARKER & RICARD MARXER, EMMANUEL VINCENT, SHINJI WATANABE"THE THIRD 'CHIME' SPEECH SEPARATION AND RECOGNITIONCHALLENGE: DATASET, TASK AND BASELINES" ©2015 IEEE.

[6]D. Mostefa, N. Moreau, K. Choukri, G. Potamianos, S. M.Chu, A. Tyagi, J. R. Casas, J. Turmo, L. Cristoforetti, F. Tobia,et al., "The CHIL audiovisual corpus for lecture and meeting analysis inside smart rooms," Language Resources and Evaluation, vol. 41, no. 3-4, pp. 389–407, 2007.

[7] S. Renals, T. Hain, and H. Bourlard, "Interpretation of multiparty meetings: The AMI and AMIDA projects," in Proceedings of the 2nd Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA), 2008, pp. 115– 118.

[8] M. Lincoln, I. McCowan, J. Vepa, and H. Maganti, "The multichannel Wall Street Journal audio visual corpus (MC-WSJAV): specification and initial experiments," in Proceedings of the 2005 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2005, pp. 357–362.