iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40989

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

A Literature Review on Speech Recognition and Cyber Security Systems

Anish K[1], Arvind Rao[2], Sagar S Bhat[3], Vinod B Durdi[4]
[1, 2, 3, 4]*Department of TCE, Dayananda Sagar College of Engineering, Bangalore, India*

Abstract: ***People use communication to convey their thoughts, ideas, and feelings, as well as to understand the feelings of others. It is a dynamic, methodical, adaptable, commercial and continuous process.***
Visual, text, spoken, written, gestural, and other modalities of communication are among them. These types can frequently be converted into other formats. When a person cannot understand or recognize the type of communication being used, this step becomes extremely crucial. The communication gap between the specially abled i.e. blind, deaf, mute or differently abled i.e. people with disorders like dyslexia, arthritis etc. and the rest of the world can be closed by converting to a form that the rest of the world understands. The paper that follows focuses mostly on Speech to Text conversion using Deep Learning Algorithms like LPC, MFCC, PLP which provides an innovative, real-time, natural, and user-friendly means of interacting with a computer that is more familiar to humans. The converted speech data can be made full duplex by transmitting the data to the cloud. Hence the converted speech data has to be sent to the cloud over a wireless channel. Thus, all converted data can be posted to the cloud over a wireless channel and retrieved by network clients and users in the network can retrieve them using a web application designed specifically for this purpose. Client in the network communicate utilizing the internet and website applications. Hence, they need to be secure and protected. This has necessitated the development of a robust data concealing method. To secure the messages generated by this system, it employs a dynamic approach to cryptography. The following paper provides an overview of the various methodologies mentioned.
Keywords: Speech Recognition Algorithms (LPC, MFCC, PLP), Wireless Communication, Encryption, Decryption, ESP8266, Text To Speech, Cloud Computing.

I. INTRODUCTION

Speech is a very natural way of communicating ideas and information between people. It helps to convey our ideas and interact with people [1]. The Speech to Text (STT) system detects the words and voice waveforms and phrases in audio input by a person and converts it into readable text format. This system is really helpful for human to human, machine to human, human to machine communication. This modern computerized technology has become more prevalent with many using STT tools to overcome different challenges. Most recently, the field has benefited from advances in deep learning and big data [2]. The STT system consists of various operations to filter the noise, analyze and structure the data. Various models like Hidden Markov Model, Maximum Entropy Model etc.[3], algorithms like LPC, PLP, MFCC [4][5][6]. are used to improve the performance of the system.
For models, both acoustic modeling and language modeling are important modern components based on speech recognition algorithms. Markov (HMM) hidden models are most widely used in many applications[3]. Language modeling is used in many other natural language processing applications such as document fragmentation or mathematical translation. The language model originates in the field of natural language processing. It predicts the next word in the sequence using the given set of words with the help of various tools and neural network algorithms.

A. Linear Predictive Coding (LPC)

It is the most widely used method of audio signal processing and speech processing to represent the spectral digital signal envelope in a compressed manner. It makes use of precise prediction model information to speech synthesis. It consists various of steps like Preemphasis, Frame Blocking, Windowing, Auto Correlation Analysis and LPC Coefficient Parameter[4]. LPC algorithm is also applied to the input signal where in a preemphasis filter is applied after the sampling process. A smooth spectral shape of the speech signal is processed after the filtering.
The Text-to-Speech (TTS) system on the other hand converts text into voice using a speech synthesizer [1]. It usually converts language text into US & UK English accents and these synthetic speeches can be understood by a person with average communication skill. It main steps involved are text analysis, phonetic analysis and prosodic analysis.

The text analysis is mainly concerned with text normalization and linguistic analysis. The phonetic analysis deals with Grapheme-to-phoneme Conversion. Grapheme can be defined as the smallest unit in writing system. The TTS system helps in converting letters (grapheme sequence) to their pronunciations (phoneme sequence). Prosodic Analysis method deals with the pitch of signal and duration attachment. In this paper the second section analyzes the existing speech recognition algorithms, such as LPC and MFCC, as well as cryptography algorithms, such as DNA cryptography and bit rotation, are researched and analyzed. In the third section different algorithms are compared in terms of various aspects such as power, memory, efficiency, and so on. Finally, in last the section paper's conclusion is offered.

II. REVIEW ON SPEECH PROCESSING, DEEP LEARNING, CRYPTOGRAPHY AND CLOUD COMPUTING

The following literature survey is carried out with reference to systematic analysis of Speech processing techniques.

A. Itunuoluwa Isewon, et al. [2].

In this paper, a Text-to-speech synthesizer is used to convert text into spoken word, by analyzing and processing the text using Natural Language Processing (NLP). It then makes use of use of Digital Signal Processing (DSP) technology to convert this processed text into synthesized speech representation of the text. They have developed text-to-speech synthesizer in the form of a simple application that converts inputted text into synthesized speech and reads out to the user which can then be saved as an mp3. file. The Text-to-speech system follows a series of steps from the NLP module to DSP module. In the NLP module, it produces a phonetic transcription of the text read, together with prosody. In the DSP module, it transforms the symbolic information it receives from NLP into audible and intelligible speech. It gets the text as the input and then a computer algorithm which is called TTS engine analyses the text, pre-processes the text and synthesizes the speech with some mathematical models. The TTS synthesis consists of two main phases. The first is text analysis, where the input text is transcribed into a phonetic representation, and the second one is the generation of speech waveforms, where the output is produced from this phonetic and prosodic information.

The next two papers carried out evaluation of various cyptographic techniques that paves way for secure communication.

B. Deepraj Pradhan, et al. [7]

This paper uses a "Bit Rotation" technique to encrypt the given data. The propped technique follows three steps they are Generating Key, Encryption Technique and Decryption Technique. Generating Key is done by using random number generators of the size no less than 512 bits i.e., at least 154 digits or 77 pairs of unsigned integers.The key is divided into array of pairs and each pair is generated separately. The Key is given as key= [56, 12,..... , n] where the first digit of each pair is the location of the character and second digit is the number of bits to shift. Encryption technique is carried out in three steps, first by dividing the plaintext to blocks. The plain text is divided into blocks of 10 bytes. The formula to calculate the number of blocks required for text is 'ceiling (length (TEXT)/10)'. If the bits are not 10 bits, then they are padded with "_". Then a circular right shift rotation is performed on the byte where bits were shifted. The position of bytes ranges from 0 to 9 and bits to shift ranges from 1 to 8. Since a byte consists of 8 bits only, 0 and 9 will invert the bits instead of shifting. So, a random key generated will have its bits rotated according to the method mentioned above and then all the blocks are concatenated. Third one is Decryption Technique which is a reverse process of Encryption Technique, hence exact opposite of encryption is done. Here the bits are shifted to the left and the from the current position the bits are carried to the end position. And like encryption technique the position of a byte ranges from 0 to 9 and bits to shift ranges from 1 to 8. Since a byte consists of 8 bits only, 0 and 9 will invert the bits instead of shifting. Hence here the output obtained will be blocks of 10 bits where padding is removed if present and then the words are concatenated. After testing it with RSA and AES technology the proposed system was considered to be slow but has got high chi-square value compared to that of RSA and AES algorithm. Hence according to the paper this proposed system can be considered as slow but more secured communication can be carried out using this algorithm

C. Bahubali Akiwate, et al. [8].

This paper presents a new data security technique called DNA cryptography which is more secure and reliable data security approach than the current existing one which is based on ASCII character set. Here text, messages, audio and video all are encrypted using the Unicode technique to reach more users worldwide. DNA cryptography uses various technology like PCR (Polymerize Chain Reaction), DNA synthesis, DNA Digital Coding. This paper uses DNA digital coding technique can be applicable in various areas like card/debit card payments, email, SMS (Short Message Service) encryption where users want to have more secure communication.

Here the DNA algorithm has been implemented using NetBeans IDE environment, which encrypts and decrypts the characters, text file, image file and audio file by using Java language which selects text or file that contains data, convert data into ASCII equivalent and then into Unicode characters, then this Unicode is converted into hexadecimal, then to binary and then to DNA digital code. DNA digital coding is a technique in which the two state binary levels such as combinations of 0 and 1 can be initialized with the DNA digital codes that uses four nucleotides (A, T, G, and C).

The 4 nucleotides can be used as a key combination and can be assigned for letters as long as 64 bits and each time it generates new combination of letters.

The letters are decrypted by converting their key combination into DNA digital coding and then the words are split into 4 parts and then the original message is obtained in binary which is converted into hexadecimal and then to Unicode. From Unicode the ASCII characters of the letter are taken and then the original text message is obtained. According to this paper the time taken for encryption and decryption for Plain Text of size 5 letter is 8325.816ms and 5.346728ms, Text file of size 10KB is 5529.8784ms and 5.346728ms, Image of size 90KB is 7397.661ms and 5223.019ms, Audio of size 490KB is 14580.631ms and 2243.8176ms respectively.

The characteristics and properties of wireless communication systems have been studied and described in the next paper.

D. Sasikumar Gurumurthy, et al. [9].

This paper gives an introduction to wireless communication and applications in fast growing part of the dynamic filed of electronic communication. It gives a brief glimpse of the past history of wireless technology. Different elements of a wireless communication system, setting up of radio frequency channel link for mobile internet access, different types of mesh network topologies are discussed.

The emergence of distributed applications and middleware architectures have seen new business opportunities in terms of third-party service provisioning (3PSP) of user demands and offered services. Middleware technology enables to manage heterogeneity by abstracting details of lower-layer communication protocols for application programmers.

The next wireless protocol is of the Bluetooth standard that is designed to be an open standard for short-range systems. The features of long-range wireless Ethernet bridge are specifications, roaming between multiple hubs, roaming operation, roaming license and wireless modems.

In this paper, several recent business and technological trends in the ICT industry and their consequences for performance analysis are addressed. The paper concludes by highlighting that the use of wireless communication technologies delivery of television and internet service is exploding rapidly.

Building and creating an online Web Application is crucial process, this paper gives the description of the same.

E. Maha A. Sayal, et al. [10].

Modern web applications can retrieve data from the cloud or any external APIs. This paper demonstrates about building web applications in the cloud, providing multiple interfaces for it and how to choose appropriate service from Amazon Web Services for an application.

Cloud computing is a technology in which a remote server and internet have ability to maintain various data and applications. A typical web application has three layers: presentation, application and database. The proposed system of a n-tier architecture is built on of four –tier successive layers.

Here, three different client layers are provided: one for basic HTML and JavaScript, another for iPhone and iPad clients and a third for standard desktop clients.

These three layers each can present the information to the user and set directly on the client's machine. Under these client layers is the filter layer that enables a developer to abstract the authentication and authorization from the representation and application layers.

The developer has to sign up for using different services provided by Amazon Web Services. Later, choose the right environment for boto (python interface for AWS) compatible with downloaded Amazon tools and connected with amazon s3 cloud storage service.

Finally, the application is processed with amazon EC2. This application’ s efficiency can also be increased significantly by integrating with deep learning algorithms like MFCC, PLP, LPC etc. [4][5][6]

III. COMPARISION OF DEEP LEARNING TECHNIQUES FOR SPEECH PROCESSING IN CYBER SECURITY

The analysis of the different algorithms in relation to various system parameters is carried out in Table 3.1.

Table 3.1

Algorithm	Performance Parameters	Remark
LPC – Linear Prediction Coding [4]	Computational Power	The LPC algorithm utilizes approx. 24.3% of total power, which is significantly less than those of other algorithms.
	Memory Consumption	It occupies 2.2 GB of RAM, which is quite a lot.
	Efficiency	The efficiency of this algorithm is 70%.
PLP – Perceptual Linear Prediction[5]	Computational Power	This algorithm uses 72 % of total resources, which is much higher than other algorithms.
	Memory Consumption	The total RAM consumed by this algorithm is 1.6 GB
	Efficiency	The proposed algorithm is accurate up to the extent of 55%.
MFCC – Mel Frequency Cepstrum Coefficient [6]	Computational Power	This algorithm uses 44% of the CPU, which is ideal.
	Memory Consumption	This algorithm uses up 3 GB of RAM
	Efficiency	This approach can improve the system's efficiency by 90%.
DNA Cryptography [7]	Computational Power	This algorithm requires 30% of the resources, which is a substantial number.
	Memory Consumption	This approach consumes only 490 KB of RAM
	Time Period	Average duration to run this algorithm is 3.832 seconds
	Efficiency	The efficiency of this method is 85.2 percent.
Circular Bit Rotation [8]	Computational Power	This approach requires 15% of the CPU, which is considerably less than other methods.
	Memory Consumption	It takes up 13 MB of RAM, which is a significant amount of memory
	Time Period	The proposed approach takes a maximum of 3.832 seconds to perform.
	Efficiency	This method has an accuracy of 73.6%

III. CONCLUSION

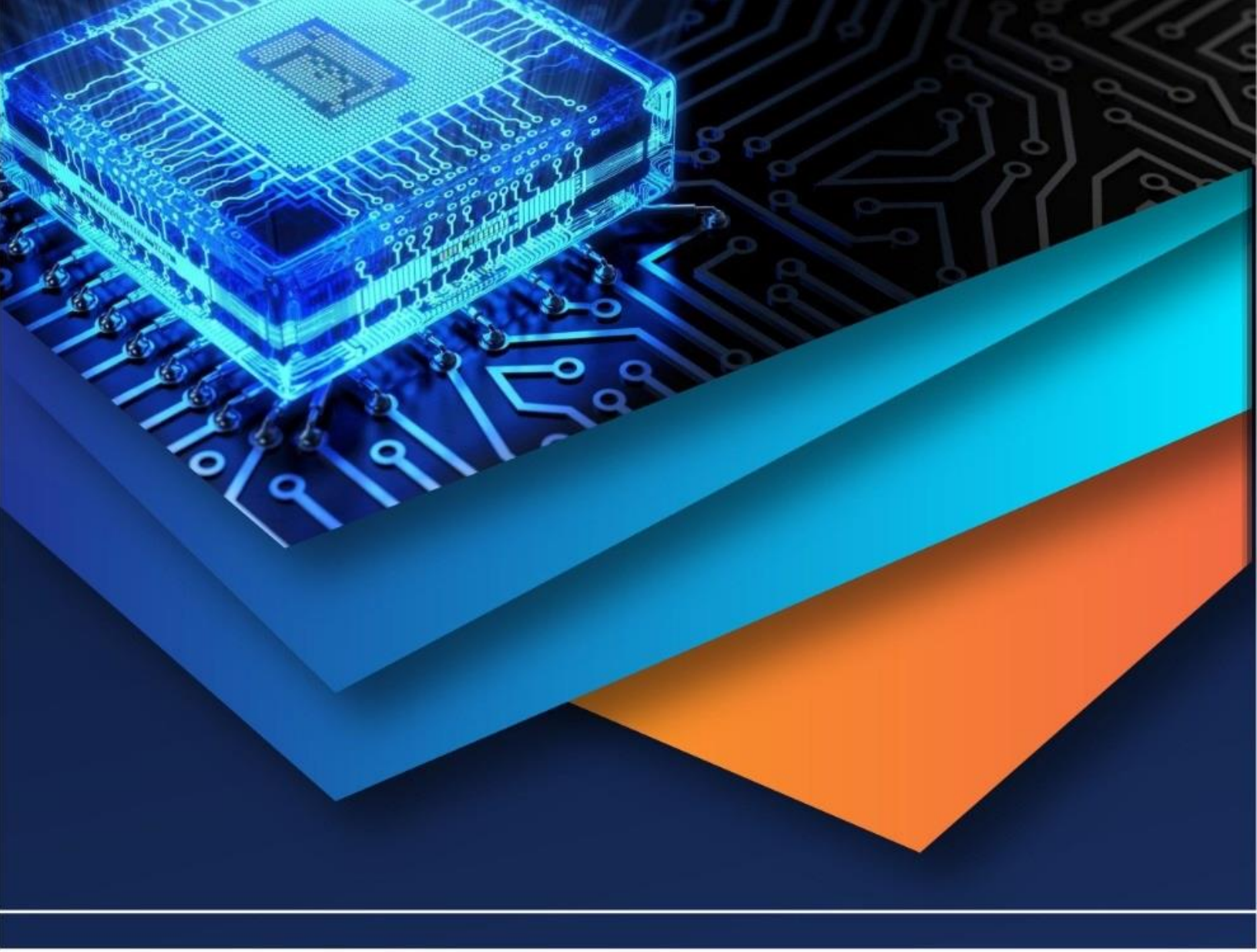
Based on review of existing procedures and strategies for converting speech to text and vice versa, it is obvious that this is accomplished through the use of mathematical modelling and inherent functions. The techniques like LPC, MFCC etc. have been shown to be accurate, rapid, and easy to use. The process of converting audio to text has been simplified. The most efficient algorithm among those mentioned above for STT process is MFCC algorithm.Using MFCC algorithm a system can be designed having capacity to reach out to people with a variety of disabilities [6]. This has prompted more research into the deep learning paradigm. Along with reaching out to many users the privacy of the data also has to be maintained Thus providing data security is just as critical as maintaining communication. It is the service provider's responsibility to provide a strong encryption method to protect the data. This type of security is also used while transferring STT converted data for military and defense-related applications where confidentiality becomes a top issue. The survey also revealed some interesting facts about cryptographic techniques like encryption and decryption, can also make use of cloud computing for memory storage and access. This process of storing the encrypted data on the cloud reduces the load on local devices. These data that are stored in cloud can be retrieved using web application which also has a system for audio conversion of local, regional, and native languages endemic to a given geological site, this project can also benefit people from various geological places.

IV. ACKNOWLEDGEMENT

The researchers would like to acknowledge their colleague for the excellent contributions and discussions on the topics mentioned above.

REFERENCES

[1] Simon DobriSek, et al., (2019), “HOMER: a Voice-Driven Text-to-Speech System for the Blind IEEE International Symposium on Industrial Electronics (ISIE), July 2019.
[2] Itunuoluwa Isewon, et al., (2018), Design and implementation of text to speech Conversion for Visually Impaired People " International Journal of Applied Information Systems (IJAIS)
[3] Rev,” All You Need to Know About Automatic Speech Recognition Transcription Models” from https://www.rev.com/blog/guide-to-speechrecognitiontranscriptionmodels#:~:text=They%20provide%20a%20way%20of,Random%20Fields%2C%20and%20Neural%20Networks
[4] Hyung-Suk Kim ,”Linear Predictive Coding is All-Pole Resonance Modeling”, Center for Computer Research in Music and Acoustics, Stanford University
[5] H Hermansky , “,” Perceptual linear predictive (PLP) analysis of speech”, from https://pubmed.ncbi.nlm.nih.gov/2341679/
[6] uday200, “MFCC Technique for Speech Recognition”, from https://www.analyticsvidhya.com/blog/2021/06/mfcc-technique-for-speech-recognition/
[7] Deepraj Pradhan, (2020), et al., " Cryptography Encryption Technique Using Circular Bit Rotation in Binary Field " 8th International Conference on Reliability, Infocom Technologies and Optimization, 2020
[8] Bahubali Akiwate (2018), et al., “A Dynamic DNA for Key-based Cryptography”, 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems, 2018
[9] Sasikumar Gurumurthy, (2019), et al., " Recent Trends In Wireless Technology " National conference on "Network Technologies" [NCNT-2019 January]
[10] Maha A. Sayal ,(2016), et al., ., " Building web applications using Cloud Computing” International Journal of Software and Web Sciences-(IJSWS)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY