

# A New Speech Recognition Model in a Human-Robot Interaction Scenario Using NAO Robot: Proposal and Preliminary Model

Hussain A. Younis

*School of Computer Sciences, Universiti Sains  
Malaysia, 11800 USM, Penang, Malaysia  
& College of Education for Women, University of  
Basrah, Basrah, Iraq.*  
hussain.younis@uobasrah.edu.iq

A.S. A. Mohamed

*School of Computer Sciences,  
Universiti Sains Malaysia, Minden,  
11800, Penang, Malaysia.*  
sufri@usm.my

M. N. Ab Wahab

*School of Computer Sciences,  
Universiti Sains Malaysia, 11800  
USM, Penang, Malaysia.*  
mohdnadhir@usm.my

R. Jamaludin

*Centre For Instructional Technology & Multimedia,  
Universiti Sains Malaysia, Minden, 11800, Penang,  
Malaysia*  
rozinah@usm.my

Sani Salisu

*School of Computer Sciences, Universiti Sains  
Malaysia, 11800 USM, Penang, Malaysia &  
Information Technology Federal University Dutse,  
Jigawa, Kano, Nigeria.*  
sani.salisu@fud.edu.ng

**Abstract**— There are several terms for speech recognition. Auto speech recognition (ASR), speech-to-text, and computer speech recognition are all terms used to describe speech recognition. A single user's voice it is necessary to distinguish between speech recognition and voice recognition. The first is to translate speech into text, such as, the audible voice and concept (human speech), and the second is to define only sound, such as, animal sound, car, etc. There is no algorithm that is specifically designed for this field; instead, techniques such as N-grams and neural networks are used to explain and treat this type. Natural Language Processing (NLP), Hidden Markov Model (HMM), and Speaker Diarization (SD). The last type would be addressed in my work. Natural language processing is a computational technique that can be used and applied to various levels of linguistic analysis (dare, deep analysis) to represent natural language in a useful or more representation. It is still possible to improve current recognition and identification systems in order to achieve greater accuracy. A new approach has been proposed that distinguishes speech in four stages: speech recognition, tokenization, extracting features of speech from texts, and part speech: The three patterns of Name Entity Recognition (NER), followed by the possibility of implementing the proposed model It achieved more accurate and applied results in an educational environment by using a NAO-robot.

**Keywords**— *Speech Recognition, NAO-robot, Natural Language Processing (NLP), Model.*

## I. INTRODUCTION

Since the 19th century, due to technical advances, robots have been performing more precisely than humans at lower cost. Creative solutions such as 3D printers and the ability of these manufacturing robots to self-learn can replace human workers [1]. Artificial Intelligence (AI) is a branch of computer science that makes an intelligent machine, or what we can say to make a computer system behave like a human being. In other words, artificial intelligence makes a computer or robot perform at above the level of human computing capability in terms of accuracy, capability, and speed. Artificial intelligence programmers a machine to simulate human activity. It is necessary to have access to the Oxford dictionary definitions of learning, teaching, translation, and

before starting the subject. Learning (education) is defined as “a process of teaching, training and learning, especially in schools, colleges or universities, to improve knowledge and develop skills”, while teaching is defined as “the work of a teacher”. The translation is defined as “the process of changing something that is written or spoken into another language”. Dictionary is defined as “a book or electronic resource that lists the words of a language in an alphabetical order and explains what they mean or provides a word for them in a foreign language”.

The robot connects learning with practical life because most of the projects presented in the competitions are real examples that the learners live in their daily lives, such as, automatic teller machines and smart doors, which contribute to students' learning through understanding, application, and solutions to problems experienced by society, through the use of scientific research strategies [2]. Natural linguistic programming is the art of changing technology, the art of investigating positive and cooperative communication between others and bringing others closer together, as well as the art of communicating and talking with others.

This paper is organized as following. Section: II presents about the Speech Recognition and the related works about. Section III Overview of Automatic Speech Recognition (ASR) approaches. Section IV the method used respectively result and discussion in Section V. Finally, conclusions and Future works with Section VI.

## II. LITERATURE REVIEW

The subject of speech recognition was launched at the Bell laboratory in the year 1950. The main objective was to ensure that the device (speech reception) was effective and capable of hearing, analyzing, understanding, recognizing and distinguishing speech and of knowing the identity and characteristics of the speaker. Speech recognition systems have gone through four main stages: 1) Analysis of the data 2) Extraction of the feature 3) The modelling (4) Testing [3].

The speech recognition mechanism is such that the speech signal of a client is a device or that the consumer or human is received by the system via a microphone, that the system evaluates that signal and that useful information is derived from that signal. Speech recognition has proven to be one of the fields of text processing and is also referred to as Automatic Speech Recognition. i.e., converting speech signals into a series of words and processing them (speech to text) using a computer algorithm [4]. A study of the design and implementation of the voice recognition system, which included 150 samples, used the Google Application Programming Interface (API) with a recommendation of up to 80% accuracy. Used integrated speech with the Internet of Things (IOT) control electrical appliances and doors in homes, departments and government institutions using Raspberry. The study focused on simulating the three methods of Digital Signal Processing (DSP), Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM) [1].As shown in Fig.1.

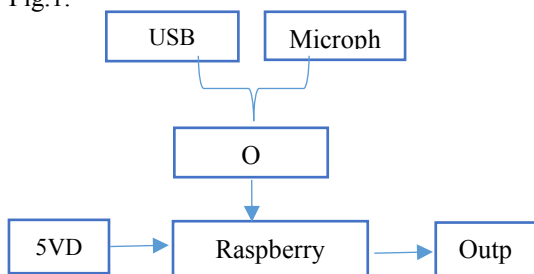


Fig. 1. Block diagram for the speech recognition system [1].

The results showed accuracy in the following Table I

TABLE I. ACCURACY OF THE SAMPLE TAKEN[1]

No	Accuracy of the sample taken		
	Test Condition	Number of Accuracy	Percentage (%)
1	A condition of crowded place with background noise.	97	64.67
2	A condition of silent place with little or no background noise.	123	82.00
3	A condition , such that, the speaker’s voice was low.	137	91.33
4	A condition, such that, the speaker’s voice was loud..	110	73.33

The Natural language processing is widely used in the field of speech and text processing, through which systems can manipulate speech and texts, the study of text-to-speech conversion that is read to the user and has a portfolio in the form of MP3, means converting the natural language processing and then using digital signals [1].

Speech is the basic mode of communication for the intelligent human system. NLP has a role to play in dealing with many of its aspects of natural language programming. Natural language programming, therefore, is the use of a computer that is essential to the understanding of the student language, a game to be a concept of speech recognition and processing[5].

A study presents the design of a mobile application that tracks the navigation path using natural software languages and Automatic Speech Recognition intends to create a user's preferred path by converting voice streams to text through the speech recognition process with the help and use of the Sensex Pocket Librarian followed by natural programming language processing with Framework Stanford Core (NPL) The results of the study reached the production of an effective approach by translating the directions of natural language into an automated conceptual coordination that follows the waves of the human deity that depend on the navigation of voice.

Speech recognition is a system that is useful for deaf and stupid people who also have visually impaired people. There are a number of useful applications from speech to text recognition.[3]. The following points can be summarized:

1. Vocally Handicapped,
2. Games and Education,
3. Learning for Visually Impaired,
4. Voice Enabled E-mail,
5. Man-Machine Communication,
6. Telecommunication and Multimedia,
7. Automobile control systems,
8. Medical and industrial systems.

Some people spoke a special language that was understood in Southeast Asia. It is written in Jawi and is a writing language used in the countries of Southeast Asia in Western Australia, Indonesia and Penang Island in Malaysia. It is similar to the Arabic text, and six letters were developed and added to make it a new writing language called Jawi. The study link to the generation of a student assistant to read Jawi writing. microphone, linear speech recognition used by linear [6]. Other studies, as shown in the Table II , used a variety of algorithms.

TABLE II. ALGORITHM USED IN STUDY

No.	Details			
	Ref.	Main purpose	Algorithm used	FPGA
1	[6]	Jawi character Speech-to-text engine	LPC Neural Network	Xilinx FPGA
2	[7]	Online Speech To text engine	LPD HMM	Altera development board
3	[8]	Helping blind people to access computer	14 to 8 Bit c conversion state machine	Spartran FPGA

A study demonstrated the development of a speech recognition model in a noisy educational environment that used noise cancellation technology with voice over speech data using certain machine learning-based neural networks. The discovery of the main words and continuous speech, and the creation of lectures with and without noise, and the possibility of recording and returning to them at a different time. In addition, he used a number of techniques to get rid of the noise: Prominent Feature Extraction Techniques used by ASR Systems and Noise Cancellation Techniques for Speech Signals [9].

The general architectural concept is the process of implementing and designing a specific work, or the mechanisms by which the functional units are organized within a limited computer system. The architecture is evaluated using three main criteria: performance, costs and the maximum size of the data programmer. The representative

would then work on each part or specific architectural subject to be proposed, inserted and manipulated.

The speech recognition architecture consists of the following parts acoustic front-end, acoustic model, lexicon (dictionary), language model and decoder [4],[9]. As for previous studies in the field of applying NAO robots in the educational environment, they were reviewed in our previous work in detail [11],[13]. And studies in the same field [14],[16], [22],[24].

### III. OVERVIEW OF AUTOMATIC SPEECH RECOGNITION (ASR) APPROACHES

Speech recognition has played an important role in the last 25 years, in which most areas of life have been covered, including education, industry, commercial and medicine. There are three types of Automatic Speech Recognition (ASR) Approaches are the artificial intelligence approach, the pattern-recognition approach and the acoustic-phone approach of each type will be briefly addressed.

#### A. Pattern-Recognition Approach

This pattern is primarily based on two important parts: model training and pattern comparison. The main aspect of the training is the identification of the unknown based on the conformity of the patterns, but the stage in which the patterns are compared is the process of making a direct comparison between the unknown speech and each pattern that can be used during the training phase. The main advantage of this type of crash to develop a good sports framework test and to create a consistent speech pattern for the purpose of comparison is to train in one form a statistical model or speech template that can be applied to a voice. A word for it. A phrase. Character and so on.

#### B. Acoustic-phonetic approach

This pattern is a model for understanding and preparing rules by understanding speech recognition systems, and its voice function is the process of transferring sounds from a speaker (microphone, phone, and computer).

#### C. Artificial intelligence approach

This pattern is known as the combination of two patterns: the pattern-recognition approach and the acoustic-phonetic approach. The acoustic and pattern recognition approach is the restoration of pattern recognition techniques and the identification of sound-based approaches that have been studied [4],[17].

It represents each class of it. They are identified by one or more templates in order to improve the way in which the speakers/contrast text display of the model squares contrast and calculate the distances between the sequence of the speech manuscript and its patterns, and to get rid of the mismatch between the patterns tested by other study patterns [17]. The achieved a wide pattern match. Uses hidden Markov (HMM) models with one or more templates. In which word sequences and category patterns are calculated on the basis of the mismatch between patterns, tests and inscriptions. Writing using certified and spurious (and unsupported) versions of patterns to calculate distances.

#### D. Principles of Automatic Speech Recognition

The basics of speech recognition are started from several phases, starting with the generation of text through the middle stages of speech production and acoustic processing, and the

stage of ink is concluded. Meet the first two phases of text generation and speech production corresponds to the source of information, the third stage of acoustic processing corresponds to the channel, while you meet the fourth stage of linguistic decoding corresponds to decoder. In the state-of-the-art approach, both the production of human speech and the recognition process are modelled through four stages: text generation, speech production, acoustic processing and linguistic decoding [17].

The speaker is a transducer that works on converting energy into textual speech. Represent the source of the information. Depending on the theory of information transfer, it begins by converting the word cues (A) to the sound control sequence (Z) with the probability  $P(A, Z)$ , in a noise environment and then decoding into a sequence of magnitude (A) the goal of decoding in word strings The decoded audio surveillance sequence represents an extreme later possibility:

$$\check{A} = \arg \max A P(A|Z) \quad (1)$$

Using the Bayes law

$$\check{A} = \frac{\arg \max A W P(Z|A)P(A)}{P(Z)} \quad (2)$$

Since  $P(Z)$  is A independent

$$\check{A} = \arg \max A P(Z|A)P(A) \quad (3)$$

$P(Z|A)$  is commonly referred to as an acoustic model as it calculates the likelihood of a series of acoustic findings conditioned by the word string. The second term,  $P(A)$ , is commonly referred to as the language model, because it represents the probability associated with a postulated sequence of terms. Such, language models can introduce both syntactic and semantic language constraints, as well as, re-language limitations. Sometimes, when only syntactic restrictions are used, the language model is referred to as grammar. Hidden Markov Models (HMMs) and statistical language models are commonly used as acoustic and language models.

Natural language processing is a sub-field of linguistics, computer science, and artificial intelligence (AI). Natural Language Processing is normally used for interactions between computers and human languages involving the processing and analysis of large amounts of natural language data. Like machine learning or deep learning, it is the subset of AI that enables computers to understand, interpret and manipulate human language by matching the processed spoken words with the modelling provided, which consists of the labelled dictionary.

The Natural language processing is a computational technique that can be used and applied to different levels of linguistic analysis (dare, deep analysis) to represent the natural language in a useful representation or more[19]. Artificial intelligence and language are closely related that they cannot be separated. The language is an introduction and code for every work that includes artificial intelligence. Artificial intelligence has been used in a number of applications, such as, speech excellence, image distinction, language recognition, image recognition, machine translation, general education and machine education robots, analysis, extraction and conformance, while in the fields of education, medicine, health, business, commercial, telecommunications and sports.

Speech recognition is modern technology, and one of the most important themes of our time is gradually growing. Its applications became real and applied to the real world, including the use of speech in computers and devices. Its uses have entered many areas, including computer and smart phone applications. For e.g., (iPhone, Bing voice search on winPhone, and Google Now on Android), this mechanism has also been used to help people with special needs and people with visual impairments, and systems have been designed to be useful for people who are deaf and dumb and others [20],[21]. Controls doors and household and electrical appliances [1]. It is used in the field of automatic response by companies, car control systems and others. As a result, our study will introduce the use of a new speech recognition model using this SAT data. This study may also address the ambiguity of speech recognition and NLP.

#### IV. METHODS

This part included the methodology adopted to meet the objectives of this research study. The methodology is categorised into three (3) phases, including preliminary study and analysis, framework design and implementation and evaluation. Each of these phases has an important procedure to be followed in this work. The first phase of this research describes the related literature, based on visualization approaches, methods, techniques and tools used to visualize traceability requirements, in order to help design the proposed new approach for this study. The second phase of this study describes all the design and implementation techniques that will be used to develop the proposed tool. The third phase gives an explanation of the evaluation in which experimental information would be used to present results, conclusions and future works. All procedures that is used to complete this research are shown in Fig. 2. which presents the research methodology.

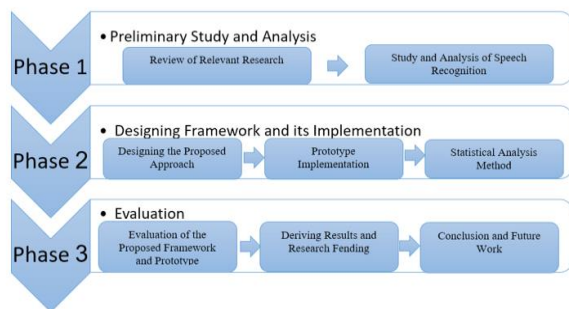


Fig. 2. Research Methodology

#### Proposed Framework

Speech recognition, also known as *Automatic Speech Recognition*, computer speech recognition or speech-to-text, is a capability that enables a programmer to process human speech in a written format. While, it is commonly confused with voice recognition, speech recognition focuses on the translation of speech from a verbal format to a text one, while voice recognition only seeks to identify the voice of the individual user.

Our study addressed the construction of a proposed system that could introduce speech and convert it into a text that includes tokenization, speech parts and lematization in the Second phase. The Third phase includes the introduction of

sound through input devices and the process of converting it to text. Extract the features of speech to text stopwords, duplicate word, question marks & punctuation. Finally, the fourth phase include Name Entity Recognition. It would be discussed in more detail in this study. As shown in this Fig.3.

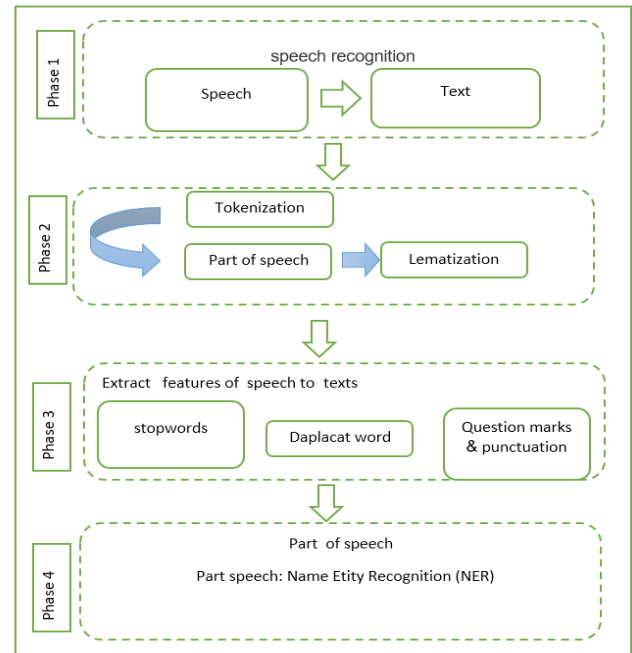


Fig. 3 the Proposed Framework Design.

#### A. Phase 1: Speech Recognition

It is the first and the main stage of the telemetry. It is the stage at which a voice is introduced through a microphone, a computer, or any device that inputs a voice into a telemetry system (a voice is in the form of a concept, such as, a word, a sentences a phrase, etc.) and then a voice is converted into a text for that purpose. The treatment and the process of segmentation of sentences. So, is meant to break the text into sentences that have a clear meaning. Speech recognition is the process of converting speech into text.

Each sentence is a specific idea for any text in natural programming languages that makes it easier for the computer to understand and manipulate it. So, the computer understands the text as a whole from the interconnectedness of these sentences.

#### B. Phase 2: Tokenization

The second stage is the tokenization phase. It is the stage of cutting sentences entered in the system (model) to the computer into words that can be processed in a way that is understandable to the computer to perform many operations, including flattening, processing and cutting.

#### C. Phase3: Extract Features of Speech to Texts.

This stage is found to be essential, fundamental and very important. We represent the extraction of important and distinctive characteristics and can be represented by the following three points:

1. Stopwords according to the input language used. Example (English, Arabic, Malaysian, Hindi, Chines. etc.) For example, stopwords in English language {‘ourselves’, ‘hers’, ‘between’, ‘yourself’, ‘but’, ‘again’, ‘there’, ‘about’, ‘once’, ‘during’, ‘out’, ‘very’, ‘having’,



'with', 'they', 'own', 'an', 'be', 'some', 'for', 'do', 'its', 'yours', 'such', 'into', 'of', 'most', 'itself', 'other', 'off', 'is', 's', 'am', 'or', 'who', 'as', 'from', 'him', 'each', 'the', 'themselves', 'until', 'below', 'are', 'we', 'these', 'your', 'his', 'through', 'don', 'nor', 'me', 'were', 'her', 'more', 'himself', 'this', 'down', 'should', 'our', 'their', 'while', 'above', 'both', 'up', 'to', 'ours', 'had', 'she', 'all', 'no', 'when', 'at', 'any', 'before', 'them', 'same', 'and', 'been', 'have', 'in', 'will', 'on', 'does', 'yourselves', 'then', 'that', 'because', 'what', 'over', 'why', 'so', 'can', 'did', 'not', 'now', 'under', 'he', 'you', 'herself', 'has', 'just', 'where', 'too', 'only', 'myself', 'which', 'those', 'i', 'after', 'few', 'whom', 't', 'being', 'if', 'theirs', 'my', 'against', 'a', 'by', 'doing', 'it', 'how', 'further', 'was', 'here', 'than' }

For example, stopwords in Arabic language.

[ 'إذ', 'إذما', 'إذني', 'أف', 'أقل', 'أكثر', 'ألا', 'إلا', 'التي', 'الذي', 'الذين', 'اللاتي', 'اللاتي', 'اللاتي', 'اللذان', 'اللذين', 'اللواتي', 'إلى', 'إليك', 'إليكم', 'إليكن', 'أم', 'أما', 'أما', 'إما', 'أن', 'إن', 'إنا', 'أنا', 'أنت', 'أنتم', 'أنتما', 'أنتن', 'إنما', 'إنه', 'أني', 'أني', 'أه', 'أها', 'أو', 'أولاء', 'أولئك', 'أوه', 'أي', 'أيها', 'إي', 'أين', 'أينما', 'إيه', 'بخ', 'بس', 'بعد', 'بعض', 'بك', 'بكم', 'بكن', 'بلى', 'بما', 'بماذا', 'بمن', 'بنا', 'به', 'بها', 'بهم', 'بهما', 'بين', 'بي', 'بين', 'بيد', 'تلك', 'تلكم', 'تلكما', 'ته', 'تي', 'تين', 'تينك', 'ثم', 'ثمة', 'حاشا', 'حاشا', 'حتى', 'حيث', 'حيثما', 'حين', 'اخلا', 'دون', 'ذا', 'ذات', 'ذاك', 'ذان', 'ذاتك', 'ذلك', 'ذلكم', 'ذلكما', 'ذلكن', 'ذه', 'ذو', 'ذوا', 'ذواتا', 'ذواتي', 'ذي', 'ذين', 'ذينك', 'رئت', 'سوف', 'سوى', 'شتان', 'عدا' ]

2. Frequent words: refers to repeated words in the sound after conversion to text. It has appeared more than once.
3. Question marks, periods, signs and symbols: they are not important.

*D. Phase 4: Part of Speech: Name Entity Recognition (NER)*

In the third phase, an evaluation is conducted to validate the effectiveness of the proposed visualization framework.

Here, the basic part of building and creating a system (model) begins with the parts of speech by category of speech, whether it is a noun, adjectives and verbs. etc. As shown in Fig. 4. It is carried out by performing the following operations:

1. Training section models of speech in a large group of sentences.
2. Predict the use of speech on the basis of sentences and words that you have previously passed on.
3. After processing sentences, the prediction of speech is correctly reached.

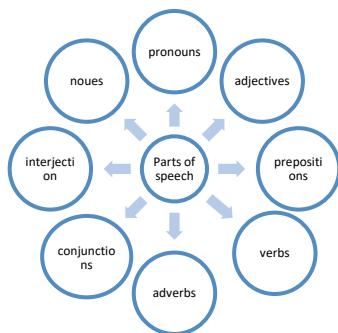


Fig. 4. Parts of speech

V. RESULTS AND DISCUSSION

The study sets out the research methodology for a Model for speech recognition. It fully discussed all methods and procedures for carrying out work, including the proposed approach of the research study, the method of data collection and the conduct of the work, together with its method of analysis, implementation techniques and evaluation method to be used in this work, where a number of questions will be defined and used to communicate with users view-point in order to obtain their understandability and interaction feedback. We started designing the models to depend on the four stages and introduced many lectures on a sample of students (27 students) from Stars of Science National Elementary Mixed School in Iraq and will include (25 students) from Universiti Sains Malaysia, as shown in Fig. 5. Our study concludes with a model capable of analyzing the system and achieving accuracy in the identification of speech recognition as opposed to previous studies.



Fig.5. Training in a class at Stars of Science National Elementary Mixed School

VI. CONCLUSIONS AND FUTURE WORKS

There are several terms for speech recognition. Auto speech recognition, speech-to-text, and computer speech recognition are all terms used to describe speech recognition. The Educational Robotics approach is multidisciplinary. ER includes architecture, assembly, and the utilization of robots based on engineering principles, computing, mathematics, physics, and others. Hence, ER has been used to teach Science, Technology, Engineering, and Mathematics (STEM), as well as, various levels of computer science pedagogy. The use of robots in teaching deals with many problems and challenges facing the educational process today. However, we think robots can play a role in this, as robots can provide a personalized instruction directed at the individual, and we can imagine that in the future the classroom has a number of robots that are helping students who need more support, or who want to learn more. This shows that the robot is actually a part of most aspects of life. Within the context of this realm, this paper presents a research review literature on robotics in education in view of the possibility of using robots in the classrooms to improve the human condition.

The study's goal is to propose and design an alternative that recognises speech by utilising language coordinates and capabilities. After completing the four phases, natural programming is used to perform speech recognition, tokenization, extracting features of speech to texts, and part speech: Name Entity Recognition proposed a mechanism to apply the method to a NAO-robot with high accuracy, and the

correct results were obtained. Furthermore, many challenges arose during the work stages that were resolved, avoided, or addressed.

The future working, the creation of the system varies from one designer to the next, as well as, from one programmer to another. To solve a specific problem, the strategy is to think, innovate, analyse, and summarise. The analysis is meant to be the process of analysing programme requirements in order to select and build a programme structure. The proposed model was designed in the Python programming language and printed on two types of quantitative and qualitative data. It can be applied to another sample and a different environment, depending on the type of sound, the input devices, the mechanism, the method of processing, the amount of data flow, the type of rollers, nuts, mechanisms, and the used techniques, and the system can be used in another environment the educational environment in emergency and changing circumstances.

#### ACKNOWLEDGEMENT

This project is supported by the RUI grant (1001 / PMEDIA / 8016096), Universiti Sains Malaysia. Author acknowledges the contributions from Centre for Instructional Technology & Multimedia, Universiti Sains Malaysia and the School of Computer Sciences at Universiti Sains Malaysia.

#### References

- [1] A. Abdulkareem, T. E. Somefun, Oji K. Chinedum, F. Agbetuyi , "Design and implementation of Speech Recognition system integrated with internet of things", *International Journal of Electrical and Computer Engineering (IJECE)* Vol. 11, No. 2, April 2021, pp. 1796–1803, DOI: 10.11591/ijece.v11i2.pp1796-1803.
- [2] C. Bartneck, "The end of the beginning: A reflection on the first five years of the HRI conference", *Scientometrics*, vol. 86, no. 2, pp. 487–504, 2011, doi: 10.1007/s11192-010-0281.
- [3] P. Khilari and B. V. P., "A Review on Speech To Text Conversion Methods," *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 4, no. 7, pp. 3067–3072, 2015, [Online]. Available: <http://ijaracet.org/wp-content/uploads/IJAR CET-VOL-4-ISSUE-7-3067-3072.pdf>
- [4] K. S and C. E., "A Review on Automatic Speech Recognition Architecture and Approaches," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 9, no. 4, pp. 393–404, 2016, doi: 10.14257/ijcip.2016.9.4.34.
- [5] P. Mukherjee, S. Santra, S. Bhowmick, A. Paul, P. Chatterjee, and A. Deyasi, "Development of GUI for text-to-speech recognition using natural language processing," 2018 2nd Int. Conf. Electron. Mater. Eng. Nano-Technology, IEMENTech 2018, pp. 1–4, 2018, doi: 10.1109/IEMENTECH.2018.8465238.
- [6] Z. Razak, S. R. Sumali, M. Yamani, I. Idris, and I. Ahmedy, "To-Text Engine for Jawi Character," no. Csr, pp. 565–568, 2010.
- [7] Samad, S.A., A. Hussain, and L.K. Fah. "Pitch Detection of Speech Signals using the Cross-Correlation Technique." *TENCON 2000.Proceedings.* 2000 p. 283 - 286.
- [8] M.M. Syiam, H.M. Klash, I. Mahmoud, and S.S. Haggag. "Hardware implementation of neural network on FPGA for accidents diagnosis of the multi-purpose research reactor of Egypt." *Proceedings of the 15th International Conference on Microelectronics*, p. 326 – 329, 2003.
- [9] P. Blunt and B. Haskins, "A Model for Incorporating an Automatic Speech Recognition System in a Noisy Educational Environment," *Proc. - 2019 Int. Multidiscip. Inf. Technol. Eng. Conf. IMITEC 2019*, pp. 1–7, 2019, doi: 10.1109/IMITEC45504.2019.9015907.
- J. Padmanabhan and M. J. J. Premkumar, "Machine learning in automatic speech recognition: A survey," *IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India)*, vol. 32, no. 4, pp. 240–251, 2015, doi: 10.1080/02564602.2015.1010611.
- [10] H.A. Younis, A.S. A. Mohamed, R. Jamaludin, & Ab Wahab, M. N. Survey of robotics in education, taxonomy, applications, and platforms during COVID-19. *Computers, Materials and Continua* vol. 67 687–707, 2021. doi:10.32604/cmc.2021.013746.
- [11] H. A.Younis, R. Jamaludin, M.N.A. Wahab and A.S.A Mohamed " The review of NAO robotics in Educational 2014-2020 inmCOVID-19 Virus (Pandemic Era): technologies, type of application, advantage, disadvantage and motivation" , *Journal of Physics: Conference Series.IOP journal*,2020, doi: 10.1088/1757-899X/928/3/032014.
- [12] Y. Diyas, D. Brakk, Y. Aimambetov, and A. Sandygulova, "Evaluating peer versus teacher robot within educational scenario of programming learning," *ACM/IEEE Int. Conf. Human-Robot Interact.*, vol. 2016-April, no. Fig. 1, pp. 425–426, 2016, doi: 10.1109/HRI.2016.7451788.
- [13] O. Mubin, M. Alhashmi, R. Baroud, and F. S. Alnajjar, "Humanoid robots as teaching assistants in an arab school," *ACM Int. Conf. Proceeding Ser.*, pp. 462–466, 2019, doi: 10.1145/3369457.3369517.
- [14] K. Yoshino and S. Zhang, "Construction of Teaching Assistant Robot in Programming Class," *Proc. - 2018 7th Int. Congr. Adv. Appl. Informatics, IIAI-AAI 2018*, pp. 215–220, 2018, doi: 10.1109/IIAI-AAI.2018.00049.
- [15] E. Vrochidou, A. Najjoua, C. Lytridis, M. Salonidis, V. Ferelis, and G. A. Papakostas, "Social Robot NAO as a Self-Regulating Didactic Mediator: A Case Study of Teaching/Learning Numeracy," in 2018 26th International Conference on Software, Telecommunications and Computer Networks, *SoftCOM 2018*, pp. 93–98, doi: 10.23919/SOFTCOM.2018.8555764.
- [16] Sadaoki Furui, "speech and speaker recognition evaluation ", *Sadaoki Furui Department of Computer Science, Tokyo Institute of Technology ,Study 1.Tokyo, Japan,2007.*
- [17] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26, no. 1, pp.43–49. 1978.
- [18] T. V. Ngo, T. Le Ha, P. T. Nguyen, and L. M. Nguyen, "Combining Advanced Methods in Japanese-Vietnamese Neural Machine Translation," *Proc. 2018 10th Int. Conf. Knowl. Syst. Eng. KSE 2018*, pp. 318–322, 2018, doi: 10.1109/KSE.2018.8573329.
- [19] J. Li, L. Deng, R. H.-Umbach and Yifan Gong, —*Robust Automatic Speech Recognition: A Bridge to Practical Applications*, Academic Press, 2015.
- [20] B. Sanjana, J.RejinaParvin. "Voice Assisted Text Reading System for the Visually Impaired Using Tts Method", *IEEE. 10 Humanit. Technol. Conf. R10-HTC 9*, 1–6 , 2016.
- [21] Z. Sun, Z. Li, and T. Nishimori, "Development and assessment of robot teaching assistant in facilitating learning," *Proc. - 6th Int. Conf. Educ. Innov. Through Technol. EITT 2017*, vol. 2018-March, pp. 165–169, 2018, doi: 10.1109/EITT.2017.47.
- [22] H. .A. Younis a, b, A.S.A Mohamed, M.N.A. Wahab and R. Jamaludin, "A review of the applicability of robots in education with Neuro-linguistic programming (NLP), 2021.
- [23] F. Tanaka, T. Takahashi, S. Matsuzoe, N. Tazawa, and M. Morita, "Telepresence robot helps children in communicating with teachers who speak a different language," *ACM/IEEE Int. Conf. Human-Robot Interact.*, pp. 399–406, 2014, doi: 10.1145/2559636.2559654.