Evaluation of AI System's Voice Recognition Performance in Social Conversation

Sweta Kumari Barnwal
Assistant Professor
Department of Computer Science
ARKA JAIN University
Jamshedpur, Jharkhand, India.
sweta.b@arkajainuniversity.ac.in

Dr. Pooja Gupta
Associate Professor
Department of IT
PIET, Parul University
Vadodara, Gujarat, India.
pooja.gupta17524@paruluniversity.ac.in

2022 5th International Conference on Contemporary Computing and Informatics (IC3I) | 979-8-3503-9826-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/IC3I56241.2022.10073242

***Abstract*— Artificial intelligence (AI) refers to a machine's capacity for thought and learning (AI). It's also a field of study that tries to "intelligential" computers. The speech of a certain speaker can be recognized, distinguished from other speech, and authenticated using specialist software and systems that use voice recognition. The world has advanced significantly thanks to the combination of voice recognition and AI. Conversational AI is expected to transform the sector. Another advantage of technology is that it may be used to assist law enforcement in identifying criminals based on their speech. Large-scale social events may in the future be used to aid researchers in a range of subjects. Additionally, it might have an impact on the fields of healthcare and education. Int his paper the investigation of scope of the AI systems in the field of voice recognition system in various services and its challenges are discussed in this paper. The brief evaluation provides the conclusions are made with the crisp findings in the conclusion section.**

Keywords*— *AI, Voice Recognition, Speech Recognition, Conversational AI

I. INTRODUCTION

Artificial intelligence (AI) is the ability of a machine to reason and acquire knowledge on its own (AI). This topic of study includes research into the "intelligent" programming of computers. Voice recognition software and systems that reliably recognize, distinguish, and authenticate the speech of a particular speaker. Speech recognition technology and artificial intelligence (AI) have altered society worldwide.

A. AI

Machines may simulate human intellect. Machine vision, expert systems, natural language processing, and voice recognition are a few examples of uses for AI. Many people consider John McCarthy to be the inventor of artificial intelligence. Artificial intelligence is the ability of a machine to learn, think, and solve problems similarly to a human. Artificial intelligence is the capacity for thought and learning in a computer programme or other system (AI). It's also a field of study that tries to "intelligentize" computers. They are independent and don't need any programming to function.

1) AI Subtypes: AI can be categorized as one of the numerous subtypes of AI. An artificial intelligence (AI) that is more capable of doing tasks that are more similar to those performed by a human being would be considered more sophisticated.

2) Related Devices: These are the earliest AI systems, and they have few capabilities. These gadgets can imitate a variety of stimuli. Memory-based functionality cannot be used on these devices. These robots are unable to use their prior knowledge to direct their present behaviors because they lack the ability to "learn." These machines could only be controlled using a limited number of inputs or input combinations. They can't employ memory to better their processes.

3) Minimal Memory: In addition to their normal reactive capabilities, limited memory machines can also make decisions based on past data. This category includes the vast majority of current AI applications. For instance, in a deep learning system, training data is kept in memory as a reference model for future problems. An image recognition AI can be trained, for instance, using a huge number of images and the labels that go with them. Every time they scan an image, AIs use their "learning experience," which helps them identify new pictures more accurately. Chatbots and self-driving cars are both powered by AI that has limited memory and virtual assistants.

4) Theories and Mind: The next two types of AI, like the first two, are still at the concept stage or under development. The notion of mind AI, which is now being tested by scientists, is the next stage of artificial intelligence. Theory of mind-level AI will be able to identify the desires and emotions of the individuals it is interacting with. However, the development of more AI domains, in particular artificial emotional intelligence, is necessary to reach Theory of Mind-level AI. In order to truly grasp human wants, AI machines must first recognize humans as unique individuals whose thinking are shaped by a multitude of factors.

5) Self-Aware: Artificial Intelligence (AI) that has gained self-awareness is known as self-aware AI, or simply AI that is self-aware. AI researchers' ultimate goal is to develop this form of AI, which will take decades or centuries to develop. It is also capable of recognizing and evoking emotions in others.

Interacts with, this sort of AI will have its own emotions, needs, beliefs as well as goals. Those who are concerned about the dangers of AI are concerned about this form of AI as well. Although the growth of self-awareness has the potential to propel humanity forward, it also has the capacity to bring about disaster. Self-preservation may be a concept in an AI's mind when it becomes aware of itself, and this could either directly or indirectly lead to humanity's demise. It would be easy for a creature like this to outmaneuver any human being's intellect and create complex schemes to conquer the world.

B. Voice Recognition

The speech of a certain speaker can be recognized, distinguished from other speech, and authenticated using specialist software and systems that use voice recognition shown in Figure 1.

Figure 1. Speech Recognition Process

Voice recognition analyses a person's speech biometrics, including their natural accent and voice's frequency and flow. Additionally, speech recognition-enabled computers are made to recognize the voice of the speaker. Before being able to recognize the speaker, speech-recognition systems need to be trained to recognize the speaker's voice, accent, and tone. Most often, this is accomplished by a series of printed phrases and sentences that the user must read aloud into an internal or external microphone. Voice and speech recognition software are intimately related to one another.

C. Voice Recognition and AI

The use of virtual assistants like Alexa, Cortana, Google Assistant, and Siri has given us a wealth of knowledge regarding voice recognition and conversational AI. With the use of new technologies, users may speak to computers and other gadgets. The applications for voice recognition are listed below.

1) Typical Applications: Voice Search: In 2019, there will be a 10% increase in the number of Americans who regularly use voice assistants. Another study found that 71% of consumers would rather use voice searches than typing to conduct a search. Since the introduction of Siri and Google voice search, voice-activated software has grown in popularity.

Voice to Text: Speech recognition makes hands-free computing possible. Users are not required to type any of the text in emails, reports, or other documents. For instance, if you're using Google Chrome, you can utilize the speech typing and voice command features in Google Docs to write an email. YouTube can automatically subtitle films by using automatic translation and speech recognition (YouTube).

Applications for Smart Homes That Respond to Voice Commands: The majority of smart home applications carry out some action in response to voice commands. A popular function of smart home appliances is speech recognition, especially when taking into account the following: The number of smart devices that can be operated by voice assistants tripled between 2018 and 2019. Smart home gadgets are the main reason that 30% of Amazon Echo and Google Home owners said they bought the devices.

2) Applications for Business Functions: Customer service: Using AI to improve customer service is essential. Speech recognition can be implemented in a contact centre at a fraction of the expense of hiring a full staff of customer service agents. IVR (Interactive Voice Response): One of the earliest speech recognition programmes, IVR enables users to voice-command their way to the right agents or the solution to their issue. Analysis of tens of thousands of customer-agent phone calls can reveal call patterns and issues. Pre-sales: If you've ever had a sales call with an SDR, you'll be familiar with the questions they ask to ascertain whether or not you're a suitable fit for their offering. This procedure can be automated using voice bots.

Vocal biometrics for security: Vocal biometrics uses a person's distinctive biological traits, such as their voice, to verify them. Speech recognition technology can be used to replace processes that require users to provide their personal information in order to be authenticated. Due to tedious login procedures, forgotten, and stolen passwords, voice biometrics improves overall customer experience by reducing consumer annoyance.

3) Business Applications: Automotive: In-car voice recognition technologies are now standard equipment in the majority of new cars. These gadgets are made to stop drivers from looking down at their phones while operating a vehicle. Thanks to this technology, drivers may now easily use voice commands to make calls, change radio stations, and play music.

Academically, vision is the primary method of learning for children with farsightedness, accounting for 80% of the learning process. Speech recognition technology might help

students with blindness or vision problems. A user's pronunciation of a language can also be assessed using speech recognition software like Duo lingo. Pronunciation evaluation, a tool for computer-assisted language learning, aids students in getting their pronunciation right.

Media / Marketing: If the user is familiar with the subject, dictation software can let them produce 3000–4000 words of content in 30 minutes, including articles, speeches, books, notes, and emails. Despite their shortcomings, these tools can be useful for first draughts.

4) Health Services: MD Note-taking: Doctors don't need to worry about recording their patients' symptoms when they undergo examinations. Medical transcription software uses speech recognition to record patient diagnosis notes. The reduced average appointment times made possible by this technology allow doctors to see more patients during office hours.

Diagnosis: A person's mental state can be ascertained using voice analysis. These models allow for the prediction of depressive or suicidal thoughts in a patient.

II. Review

Researchers from the fields of AI, voice recognition [13-15], and automatic speech recognition contributed their knowledge and ideas to the assessment of AI System performance in social discourse. The basic books that will guide our investigation are listed below first.

According to Andrzej Cichocki et al. [1], future Artificial General Intelligence (AGI) systems will need to be able to comprehend multiple facets of human intelligence, such as ethical, social, emotional, and attention intelligence. They touched on a variety of AI problem categories in their exploration of different human intelligences and learning preferences. They categorised various AGIs according to their cognitive abilities or capacities and gave working definitions for each.

Mai Ngoc Anh et al. [2] developed Vietnamese speech recognition modules and managed a redundant manipulator using artificial intelligence (AI) techniques. The first deep learning model was developed to recognize and translate speech information into input signals for an inverse kinematics task involving a 6-degrees-of-freedom robotic manipulator. Building and practice were used to tackle the inverse kinematics puzzle. The second deep learning model was constructed using the workspace, the joint variables' constraints, and the system's geometric structure. Using Python, deep learning models were created. The Vietnamese speech recognition module's usability and the accuracy of artificial intelligence methods were demonstrated using deep learning networks.

Hongli Zhang [3] has created a speech keyword retrieval system by fusing multimodal information and combining an attention mechanism. They employed a bag of words technique to extract text from MFCCs. Context features for the attention mechanism were created by combining audio and text information. Timestamps that might include keywords were preferred. Context attributes of an attention mechanism were employed to set a discriminator confidence level. Cross-modal speech-text retrieval was carried out using the modal classifier, which was utilized to recognize modal information. Later tests revealed that their technology performed better than alternative approaches.

Rashid Jahangir and his associates [4] examined data matrices to ensure their generalizability in voice databases. Using these metrics on the training data prevented model over fitting. Furthermore, DL methods for SI have been put into practice using software frameworks and hardware elements like GPUs. Many of these frameworks have been made available as open-source initiatives. These frameworks were selected by developers as the best choice for their projects due to their qualities. They also explored unsolved problems in the realm of speaker recognition systems.

According to Anjali I.P et al. [5], significant progress has been made in machine learning, statistical data-mining, and pattern recognition technologies, which has helped to make speech interfaces more adaptive and common. Concerns concerning the barriers that can limit the successful implementation of acoustically robust natural interfaces have been raised in response to the growing demand for voice interfaces. Finally, they highlighted the scientific research and technological developments required for high-performance real-time voice recognition, which transformed human-computer interaction.

There are many obstacles to overcome before speech synthesizing technology may be used to diagnose neurological health, including the need for extensive trans-diagnostics and longitudinal studies. [6] A study team under the direction of Daniel Low et al.

The significance of adopting voice recognition systems was investigated, per Habib Ibrahim et al. [7]. Their research concentrated on automatic voice and speech recognition, which they compared to prior studies. A literature analysis indicates that the most widely used software tool is the HTK system, which was created by a team at Cambridge University under the direction of Steve Young for automatic voice recognition.

Nishtha H. Tandel et al. [8] did a thorough analysis of the literature on speaker detection and voice comparison techniques based on deep learning and more conventional techniques. The topic of publicly accessible datasets used by researchers for speaker identification and voice comparison came up throughout the conversation. Both beginners and specialists in the field of voice identification and comparison might benefit from their succinct study. Finally, the employment of CNN and Siamese NN, which was well-liked for classification issues, was considered for voice comparison.

Research is being done on the application of AI voice assistants in education, according to George Terzopoulos and colleagues [9]. To use them effectively in the learning process, there are a number of obstacles to get over. This was a challenge since there was no voice assistants that could converse in every language used today. On the other hand, voice assistants don't have many of the necessary security measures and protection filters that students might use in class. Teachers must be informed about and motivated to use these technologies for them to be successful in the classroom.

Ashok Kumar et al. [10] covered voice recognition in a study that summarized a typical speech system and described a number of procedures. Many of the consistent qualities or attributes of the spoken stream were what made the speech system's robustness possible. Local language voice recognizers must be developed when speech recognition technology improves. Multilingual voice recognition was mentioned as a recent development in the field. Although there has been much research and development into foreign languages, it is crucial to use this technology in local languages in order to increase its effectiveness and utility for native speakers.

In the past, voice assistants were thought to be restricted to cloud services by Polyakov E.V. et al. [11], but current study disproves this. This makes it possible for more people to work in fields like security, IIoT and IoT systems, smart home systems, healthcare, and others where using cloud technology might be challenging.

To highlight the state-of-the-art in voice recognition, researchers Lucas Debatin et al. [12] did a detailed literature review. They were guided in their work by a series of search questions, search phrases, and inclusion and exclusion criteria.

III. Problem Statement

The prior studies lacked precision and effectiveness. They fall short of expectations. They are not applicable in a real-life situation. Conversational AI needs to receive more attention, and it needs to be understood how important it is in many areas of life. To put it another way, previous studies have lacked accuracy and effectiveness. They don't perform well enough. They are useless in the actual world. Conversational AI needs to be given more attention so that we can appreciate its value in a number of contexts. The variety of applications for AI and speech recognition must be emphasized.

IV. Proposed Model

The proposed work takes voice pitch into account when detecting and categorizing voices shown in Figure 2. In order to forecast the voice category, the voices of men, women, and children have been collected, and a training model has been created. Research has examined voice recognition technology combined computer learning. The improvement of AI system performance during voice recognition in social conversation has been the focus of research.

Figure 2. Proposed Model

V. Results And Discussion

Considering filter mechanism and advanced learning mechanism proposed work has provided more accuracy along with better performance.

A. Simulation of Performance

Proposed model is consuming less time as compared to previous models that is simulated in table I and Figure 3.

TABLE I. Simulation Of Performance During Voice Recognition

Number of Voice Samples	Normal Voice Recognition	Hidden Markov Model	Proposed Model
1	58.43	38.9533	23.372
2	176.8808	117.92	70.7523
3	269.603	179.735	107.8412
4	71.6463	47.7642	28.6585
5	50.4433	33.6288	20.1773
6	100.1911	66.7941	40.0764
7	326.4872	217.658	130.5948
8	664.3264	442.884	265.7305
9	778.4097	518.939	311.3639
10	592.5507	395.033	237.0203

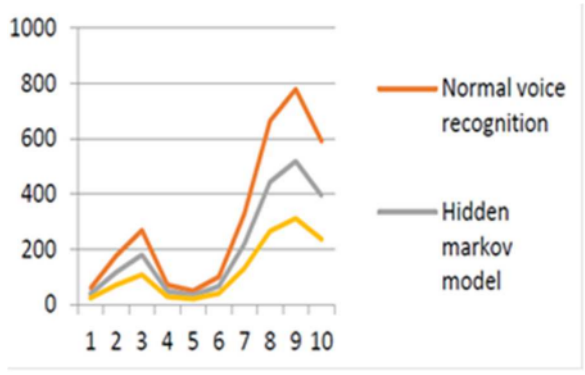

Figure 3. Comparison of Performance of Proposed Work with Previous

B. *Simulation of Accuracy*

Proposed model is providing more accuracy as compared to previous models that is simulated in Table II and Figure 4.

TABLE II SIMULATION OF ACCURACY DURING VOICE RECOGNITION

Number of Voice Samples	Normal Voice Recognition	Hidden Markov Model	Proposed Model
1	80.512296	82.03123	85.837422
2	80.13751	82.329633	86.71526
3	81.788054	84.419895	83.379448
4	81.000127	81.492722	90.564326
5	81.021764	81.632902	83.359405
6	81.316425	83.887297	90.125804
7	81.417991	82.98415	86.0187
8	80.098988	83.278039	82.5626
9	80.608185	81.954043	80.854822
10	80.880226	82.275212	86.681716

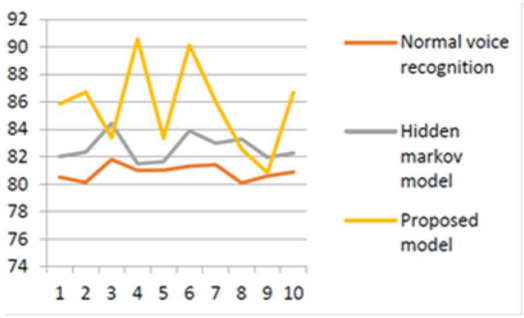

Figure 4. Comparison of Accuracy of Proposed Work with Previous

VI. FUTURE SCOPE

The scope of artificial intelligence is very high. When AI is integrated with voice recognition, applicability increases by multiplier effect. Conversational AI is upcoming revolution. It will aid in medical field too. Also, it has the potential to contribute towards maintenance of law and order, by aiding in tracing criminals by their voice. Social conversations in huge gatherings maybe used in future to aid in researches in various areas. In other words, Artificial intelligence has a wide range of applications. Integrating of AI with voice recognition has its multiplier impact increases, so does its usefulness. Conversational AI is set to revolutionize the industry. Moreover, it will be useful in the medical field as well. As a further benefit, technology might be used to help police track down criminals based on their speech. In the future, large-scale social gatherings may be used to help researchers in a variety of fields.

REFERENCES

[1] Cichocki and A. P. Kuleshov, "Future Trends for Human-AI Collaboration: A Comprehensive Taxonomy of AI/AGI Using Multiple Intelligences and Learning Styles," Comput. Intell. Neurosci., vol. 2021, 2021, doi: 10.1155/2021/8893795.

[2] M. N. Anh and D. X. Bien, "Voice Recognition and Inverse Kinematics Control for a Redundant Manipulator Based on a Multilayer Artificial Intelligence Network," J. Robot., vol. 2021, no. Dl, 2021, doi: 10.1155/2021/5805232.

[3] H. Zhang, "Voice Keyword Retrieval Method Using Attention Mechanism and Multimodal Information Fusion," Sci. Program., vol. 2021, no. 2, 2021, doi: 10.1155/2021/6662841.

[4] R. Jahangir, Y. W. Teh, H. F. Nweke, G. Mujtaba, M. A. Al-Garadi, and I. Ali, "Speaker identification through artificial intelligence techniques: A comprehensive review and research challenges," Expert Syst. Appl., vol. 171, no. January, p. 114591, 2021, doi: 10.1016/j.eswa.2021.114591.

[5] M. H. Farouk, "Speech Recognition," SpringerBriefs Speech Technol., pp. 27–29, 2014, doi: 10.1007/978-3-319-02732-6_6.

[6] D. M. Low, K. H. Bentley, and S. S. Ghosh, "Automated assessment of psychiatric disorders using speech: A systematic review," Laryngoscope Investig. Otolaryngol., vol. 5, no. 1, pp. 96–116, 2020, doi: 10.1002/lio2.354.

[7] H. Ibrahim and A. Varol, "A Study on Automatic Speech Recognition Systems," 8th Int. Symp. Digit. Forensics Secur. ISDFS 2020, 2020, doi: 10.1109/ISDFS49300.2020.9116286.

[8] N. H. Tandel, H. B. Prajapati, and V. K. Dabhi, "Voice Recognition and Voice Comparison using Machine Learning Techniques: A Survey," 2020 6th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2020, pp. 459–465, 2020, doi: 10.1109/ICACCS48705.2020.9074184.

[9] G. Terzopoulos and M. Satratzemi, "Voice assistants and smart speakers in everyday life and in education," Informatics Educ., vol. 19, no. 3, pp. 473–490, 2020, doi: 10.15388/infedu.2020.21.

[10] Kumar and V. Mittal, "Speech recognition: A complete perspective," Int. J. Recent Technol. Eng., vol. 7, no. 6, pp. 78–83, 2019.

[11] E. V. Polyakov, M. S. Mazhanov, A. Y. Rolich, L. S. Voskov, M. V. Kachalova, and S. V. Polyakov, "Investigation and development of the intelligent voice assistant for the Internet of Things using machine learning," Moscow Work. Electron. Netw. Technol. MWENT 2018 - Proc., vol. 2018-March, pp. 1–5, 2018, doi: 10.1109/MWENT.2018.8337236.

[12] L. Debatin, A. H. Filho, and R. L. S. Dazzi, "Offline speech recognition development: A systematic review of the literature," ICEIS 2018 - Proc. 20th Int. Conf. Enterp. Inf. Syst., vol. 2, no. Iceis 2018, pp. 551–558, 2018, doi: 10.5220/0006788005510558.

[13] Kumar K, Anand S, Yadava RL. Advanced DSP Technique to Remove Baseline Noise from ECG Signal. Int J Electron Comput Sci Eng. , , vol. 1, no. 3, pp. 1013-1019, 2012.

[14] Kumar K, Tanya Aggrawal, Vishal Verma, Suraj Singh, Shivendra Singh, Dr. Lokesh Varshney, "Modeling and Simulation of Hybrid System", IJAST, vol. 29, no. 4s, pp. 2857 -2867, Jun. 2020.

[15] Kumar K, Varshney L, Ambikapathy, Vrinda, Sachin, Prashant , Namya. Soft Computing and IoT based Solar Tracker. International Journal of Power Electronics and Drive System (IJPEDS). Vol 12, No 3: September 2021. doi.org/10.11591/ijpeds.v12.i3.pp1880-1889.

[16] Venkatesan, C., Karthigaikumar, P., & Varatharajan, R. J. M. T. (2018). A novel LMS algorithm for ECG signal preprocessing and KNN classifier based abnormality detection. Multimedia Tools and Applications, 77(8), 10365-10374.