

Article

Not peer-reviewed version

A method for automatically detecting errors in an embedded English speech teaching system

[Soumya Majumdar](#)*, Karan Das, Omar Abdullah Chowdhury, Ashutosh Prasad Yadav, Bapun Sahoo

Posted Date: 9 April 2024

doi: 10.20944/preprints202404.0611.v1

Keywords: Teaching System, Error Detection, Automated detectors, Online Learning, Spoken English, Speech Teaching System, Artificial Intelligence, Speech Recognition



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

A Method for Automatically Detecting Errors in an Embedded English Speech Teaching System

Soumya Majumdar ^{1,*}, Karan Das ², Omar Abdullah Chowdhury ², Ashutosh Prasad Yadav ² and Bapun Sahoo ²

¹ IIT Kharagpur, Kharagpur, India

² JIS University, Agarpara, Kolkata, India; karandas.cse.jisu@gmail.com (Karan Das), rohan.sk.509@gmail.com (Omar Abdullah Chowdhury), ashutoshprasadyadav007@gmail.com (Ashutosh Prasad Yadav), bapun8351@gmail.com (Bapun Sahoo)

* Correspondence: soumya.majumdar@jisuniversity.ac.in

Abstract: This enables the chip to regulate the voice recognition function of the gadget, leading to a more natural and human-like process. Integrating this technique into the English language voice recognition system can greatly improve its precision in identifying the speech of a particular person. This research seeks to investigate and create an automatic error detection technique for an embedded speech instruction recognition system for English using artificial intelligence. The study commences with a conventional introduction to artificial intelligence, thereafter delving into a comprehensive investigation of the speech recognition algorithm. The MATLAB programme is used to determine the accurate number of words recognised by the system and its accuracy rate. The embedded teaching recognition system for English is then tested in different contexts to assess its performance. Several tests are performed to compare and analyse the results, which indicate that the integrated English voice recognition system achieves an accuracy rate of over 90% and has a low error rate in a calm environment. In a bustling environment with numerous auditory stimuli, the accurate recognition rate typically exceeds 60%.

Keywords: teaching system; error detection; Automated detectors; Online Learning; Spoken English; Speech Teaching System; artificial intelligence; speech recognitio

1. Introduction

Due to rapid advancements in computer and artificial intelligence technologies, the use of intelligent mechanical equipment is increasing. Voice recognition control is a crucial intelligent control approach as it directly impacts the intelligence of mechanical equipment. The advancement of artificial intelligence has contributed to the progress of English phonetic education in China. Consequently, the English education sector is actively adopting the integrated English pronunciation teaching system to enhance the quality and standard of teaching. This system is well-suited for application in the field of intelligent control due to its exceptional skills in English pronunciation recognition and control.

Although there is a growing interest in artificial intelligence and its application in English teaching, there has been minimal research on the automatic error detection method for speech recognition systems used in teaching embedded English. This work seeks to address this deficiency by investigating and developing an automated detection approach for identifying faults in the recognition of this system. This research holds importance in both theoretical and practical contexts.

1.1. Previous Works

Prior investigations on the subject have produced intriguing discoveries. Misirov stated that artificial intelligence has influenced English pronunciation teaching, whereas Vong and Kaewurari recommended a cognitive approach to teaching that improves critical thinking skills. Yin focused on

studying the embedded grammar-assisted teaching strategy, which was found to be effective in teaching English to senior high school students. Gashaw performed acoustic analysis on speech recordings and discovered that native Amharic samples necessitated extended pronunciation durations. Chika's research emphasised the significance of employing various indicators for systematic phonetic assessment, whereas Thompson discovered that engaging in English conversations can enhance pronunciation.

1.2. New Method

Nevertheless, this work presents a novel approach by implementing the integrated English pronunciation teaching system into English education research within the domain of artificial intelligence. The primary focus is on the research and development of an embedded English speech teaching recognition system's automatic mistake detection approach. Multiple sets of comparative tests are conducted using MATLAB software to assess the voice recognition and processing capabilities of intelligent mechanical devices. This work has the capacity to offer useful insights into technology related to artificial intelligence.

2. Related Work

One of the primary techniques employed for identifying faults in English language was rule-based. This strategy involves using a pre-established set of well designed criteria to identify errors in written text. An example of this is the method developed by Kim and Attardi (2007), which utilised rules to detect errors in subject-verb agreement, word morphology, and verb tense. Although rule-based techniques can effectively identify defects in specific areas, they need significant manual effort for their development and maintenance.

Modern techniques for detecting errors in English utilise statistical models to automatically identify problems in text. Usually, these approaches rely on large volumes of labelled data to train machine learning models for identifying errors. The CoNLL-2014 shared task on grammatical error correction, as outlined by Ng et al. (2014), employed a dataset of English essays to train machine learning models with the aim of identifying errors in spelling, grammar, and punctuation. Further studies by Chollampatt and Ng (2018) and Junczys-Dowmunt et al. (2018) have examined the use of neural network models in detecting errors in English.

Multiple research have investigated hybrid approaches to mistake detection in English, which entail the combination of rule-based and statistical methodologies. An example of this is the method developed by Rozovskaya and Roth (2010), which utilised rules to detect errors in punctuation and capitalization, along with a statistical model to find flaws in word choice and verb tense.

Traditionally, error detection methods in English have evolved from rule-based approaches to including statistical and hybrid methodologies. Statistical approaches have shown promise in automatically identifying errors in text, but they require large annotated datasets for training the models. Integrating rule-based and statistical methods in hybrid approaches may provide a more efficient option for identifying English errors in natural language processing (NLP). However, additional research is required to examine the benefits and drawbacks of these approaches and to create error detection systems that are more precise and effective.

3. Theory

AI, or Artificial Intelligence, is the application of computer systems to replicate human brain functions, including reasoning, identifying, planning, learning, understanding, thinking, and problem-solving. Previously limited to humans, these capabilities are now attainable by AI, which can simulate human brain activity using acquired knowledge. Indeed, artificial intelligence has the potential to exceed human intelligence in certain activities. The advent of AI has disrupted conventional notions and expedited the advancement of human understanding and learning.

Emerging technologies have consistently played a significant role in driving educational reform within the field of education. Artificial intelligence is not an exception. It has enhanced the

convenience and efficiency of teaching, while also fostering equity, accessibility, and democratisation in education. Teaching software that utilises artificial intelligence has the ability to perceive, listen, communicate, learn, understand, and even replicate human emotions. This enables users to connect with computers in a natural and seamless manner using different communication modalities. Therefore, it is expected that AI will have a substantial influence on administration, teaching methods, and educational philosophy. Artificial intelligence is considered one of the three major scientific and technological advancements of the 20th century, alongside space and atomic energy technology.

4. Automated Error Detection Method for English Embedded Teaching recognition System

By referring to relevant literature [9], one can acquire the mathematical model of the method and the corresponding mathematical model of the speech recognition algorithm. Figure 1 depicts the structure of the system, where the sound frequency is sent via the impulse mode together with noise. The model must address the output problem in the communication channel to ensure the successful conversion of the speech signal.

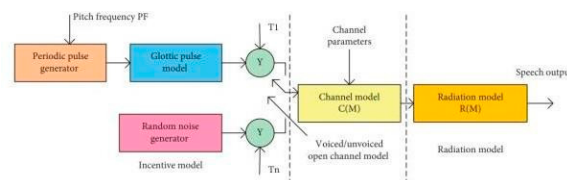


Figure 1. The construction of the voice recognition algorithm is based on a mathematical model.

$$A_Y = R[Y^3(E)] + \sqrt{ABU}[A(E - e_0)],$$

$$L_Y = R[Y^4(E)] - 3R^2[Y^2(E)]N.$$

4.1. Two-Stage Recognition Algorithm

Using a single-word template is unsuitable for an embedded English speech recognition system since it is sensitive, causing longer detection time and increased space usage as the number of candidate words grows, resulting in longer waiting times. Consequently, it is hardly utilised. In contrast, the Monophony model offers greater convenience, consumes less physical space, and exhibits superior reading efficiency. However, its descriptions of pronunciation lack precision, resulting in a lower rate of identification. Regarding pronunciation performance, the Triphone and Syllable models exhibit higher accuracy. However, because to their several models and states, they are slower and require more storage space.

This study utilises a two-stage search technique to specifically tackle the problems related to space occupancy and reading speed. In the initial stage, a model and static recognition network are employed to generate several potential word entries. In the subsequent stage, another recognition network is constructed using the multiple candidate entries obtained from the previous stage. In order to acquire accurate reading and obtain the final recognition result, an additional model is employed. The quantity of entries processed during the second stage is comparatively limited, resulting in an accelerated reading velocity. Furthermore, the memory usage of the identification system is reduced as the second-stage reading can utilise the same space resources as the first stage. The fundamental procedure of this programme is illustrated in Figure 2.

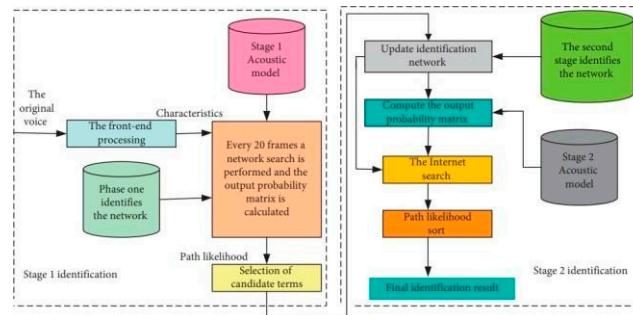


Figure 2. Basic framework of a two-level search algorithm.

4.2. System Hardware Part Design

The hardware components of the embedded English speech recognition control system can be identified by studying the relevant literature provided in Figure 3. The literature also indicates that the central processing unit (CPU) of the system employs a highly efficient microchip for data processing. The chip's architecture is specifically built with an instruction set that improves the execution of dispatched instructions and enhances the overall performance of the chip. The enhanced efficiency and performance of the chip in executing instructions and conducting operations make it highly versatile in multiple sectors.

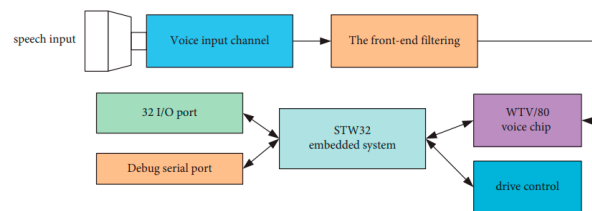


Figure 3. System hardware part structure.

4.3. Principle of Embedded Speech Recognition Technology

Previous research has identified four ways for embedded voice recognition: vocal tract modelling, speech knowledge-based approach, pattern matching, and statistical model and artificial neural network-based approach [14]. Pattern matching has four primary stages: feature extraction, template training, template classification, and decision-making. Figure 4 depicts a schematic representation of the pattern matching methodology.

Figure 4 illustrates the conversion of the speech signal into a signal flow that passes via a radio facility and is then combined with the input of the recognition system. Once the preprocessing stage is over, the speech signal's features are retrieved, and a necessary model is built on top of it, referred to as the instruction procedure [15]. The most suitable embedded speech recognition model for read speech is determined by employing a search and matching technique that compares the attributes of the prepared speech template with those of the read speech signal. The identification process encompasses the complete procedure of aligning the recently obtained characteristics with the established template. Ultimately, the computer acknowledges and understands the spoken words by means of this specific format.

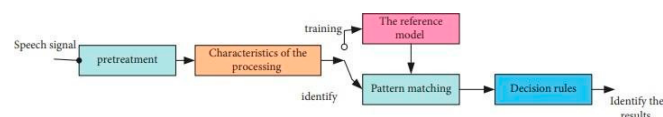


Figure 4. Principle of pattern matching in speech recognition system.

5. Results and Discussion

We performed a simulation test using MATLAB software to assess the sensitivity and error automatic checking function of the embedded English recognition system, according to the specific requirements. The objective was to assess the system's efficacy in precisely identifying English speech for educational purposes. This study comprises two studies, with the initial experiment assessing the precision of speech recognition errors. The results of our thirty tests on six speeches are shown in Table 1. The initial experiment's findings demonstrate that the embedded speech recognition system consistently obtained a recognition rate of 80%, thereby satisfying the design specifications of the intelligent recognition system. In addition, each word was accurately identified more than 25 times during the thirty tests, showcasing the system's outstanding performance with a minimal rate of errors. This makes it well-suited for doing research on English phonetic instruction.

Table 1. First experiment accuracy test results.

Identification of statement	Correct recognition times	Accuracy(%)
Happy	26	84
Cry	25	87
Yes	26	84
Fly	27	94
Dog	30	95
Like	24	82

The experiment may lack sufficient rigor, as the second simulation experiment was conducted using MATLAB software, and the results are presented in Table 2.

Table 2. Second experiment accuracy test results.

Identification of statement	Correct recognition times	Accuracy(%)
Happy	27	87
Cry	26	84
Yes	25	87
Fly	27	94
Dog	26	84
Like	28	91

It is evident that the difference between the results of the first and second experiments is minor, with the overall word accuracy being greater than 80%. This verifies the scientific rigor of both trials and amply proves the viability of the embedded English voice recognition system.

5.1. Test of Embedded English Speech Recognition System in Different Environments

Two experiments were conducted, one in a calm environment and the other in a noisy one, to assess the impact of varied settings on the deviation value of the integrated English recognition system. A total of thirty tests were conducted on the identical set of six words in order to guarantee the scientific precision and genuineness of the experiment. Both tests utilised the MatLab software. Figure 5 displays the findings from the initial experiment.

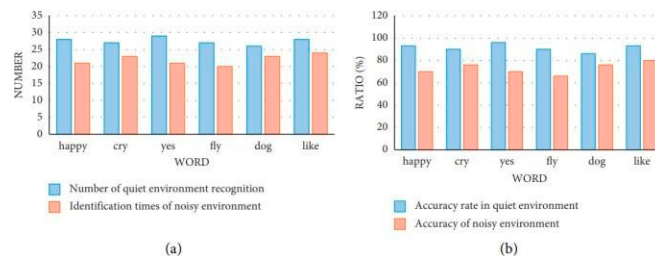


Figure 5. The first accuracy test results in different environments. (a) Number of correct word recognition, (b) word recognition accuracy.

Figure 5 shows that the English speech instruction recognition system, when integrated, achieves a word identification accuracy rate of over 80% in a quiet environment. Nevertheless, in a cacophonous setting, the system's ability to accurately recognise words diminishes, yet it still maintains a rate of accuracy that exceeds 66%. These findings suggest that the error rate of the embedded English teaching recognition system might be influenced by the surrounding environmental conditions. However, the precision of the recognition system remains astonishingly high, and additional refinement and investigation could potentially decrease its mistake rate even lower.

In order to enhance the intricacy and variety, the previous experiment conducted in a noisy environment was replicated by incorporating a distinct noise element. The experiment employed both the two-stage detection strategy and the fixed energy threshold method, and the results are presented in Table 3.

Table 3. Test accuracy rate of fixed threshold method and two-end detection method.

Test voice	Correct recognition times	Accuracy (%)
Original voice	92.76	93.25
Add car noise(15 dB)	88.74	89.77
Add car noise(10 dB)	67.98	74.79
Add white noise(15 dB)	86.68	88.68
Add white noise(10 dB)	63.55	75.35
Add sound of people talking (15 dB)	87.66	89.55
Add sound of people talking (10 dB)	78.35	79.24

The findings presented in Table 3 demonstrate that the two-stage detection method outperforms the endpoint detection method using a fixed energy threshold, regardless of whether the environment is silent or contains several noise components. The implementation of the two-stage detection technique results in a significantly higher rate of identification [19]. The two-stage detection technology is highly accurate and scientific, making it well-suited for similar studies.

6. Conclusion and Future Scope

Varying environmental conditions affect the accuracy of the embedded English speech instruction recognition system. The system demonstrates optimal performance in both word recognition tests and sentence recognition tests when operating in a noise-free environment. Nevertheless, when there is noise, the accuracy of recognition fluctuates based on the specific type and intensity of the noise. In general, the system's accuracy in identifying objects is significantly

reduced in loud environments compared to calm ones, however there are exceptions to this pattern. As artificial intelligence technology continues to evolve, future improvements to the embedded English spoken instruction recognition system may lead to enhanced performance in detecting errors. The English voice recognition system has exceptional accuracy. However, the presence of noise-related elements can result in a reduced identification rate compared to a calm context. In order to evaluate the impact on the system's performance, it is essential to consider both the specific type of noise and the corresponding decibel level.

In summary, the study emphasises the capacity of integrated English speech teaching recognition systems to enhance language education and offer immediate feedback to learners. With continued progress in artificial intelligence and voice recognition technology, it is feasible to enhance the performance of these systems in various settings.

References

1. S. Misirov, "The peculiarities of teaching English pronunciation in elementary classes (GRADES)[J]," *Scientific Bulletin of Namangan State University*, vol. 1, no. 2, p. 63, 2019.
2. S. A. Vong and W. Kaewurai, "Instructional model development to enhance critical thinking and critical thinking teaching ability of trainee students at regional teaching training center in Takeo province, Cambodia," *Kasetsart Journal of Social Sciences*, vol. 38, no. 1, pp. 88–95, 2017.
3. Y. Yin, "Microclassroom design based on English embedded grammar compensation teaching," *Mathematical Problems in Engineering*, vol. 2021, no. 1, pp. 1–9, 2021.
4. Gashaw, "Rhythm in Ethiopian English: implications for the teaching of English prosody[J]," *International Journal of Education and Literacy Studies*, vol. 5, no. 1, pp. 2202–9478, 2017.
5. F. Chika, E. Sayoko, and A.-Y. Reiko, "An evaluation of English pronunciation of Japanese EFL learners using multiple metrics[J]," *Journal of the Phonetic Society of Japan*, vol. 22, no. 2, pp. 39–43, 2018.
6. T. Isaacs and L. Harding, "Pronunciation assessment," *Language Teaching*, vol. 50, no. 3, pp. 347–366, 2017.
7. E. M. Weisberg, L. C. Chu, and E. K. Fishman, "The first use of artificial intelligence (AI) in the ER: triage not diagnosis," *Emergency Radiology*, vol. 27, no. 4, pp. 361–366, 2020.
8. D. Bedzow Ira Leeds, "Artificial intelligence (AI) and halakhic responsibility for physicians to consult data[J]," *Studies in Judaism, Humanities, and the Social Sciences*, vol. 3, no. 1, pp. 90–103, 2020.
9. S. Bi, R. Chen, K. Zhang, Y. Xiang, and H. Wang Lin Yang, "Differentiate cavernous hemangioma from schwannoma with artificial intelligence (AI)," *Annals of Translational Medicine*, vol. 8, no. 11, p. 710, 2020.
10. D. Calaprice-Whitty, K. Galil, W. Salloum, and B. Jimenez, "Improving clinical trial participant prescreening with artificial intelligence (AI): a comparison of the results of AI-assisted vs. standard methods in 3 oncology trials[J]," *Therapeutic Innovation and Regulatory Science*, vol. 54, no. 1, pp. 1–6, 2020.
11. J. Gregory, S. Welliver, and J. Chong, "Top 10 reviewer critiques of radiology artificial intelligence (AI) articles: qualitative thematic analysis of reviewer critiques of machine learning/deep learning manuscripts submitted to JMIR," *Journal of Magnetic Resonance Imaging*, vol. 52, no. 1, pp. 248–254, 2020.
12. F. Schuur, M. Rezazade, M. H. Rezazade Mehrizi, and E. Ranschaert, "Training opportunities of artificial intelligence (AI) in radiology: a systematic review," *European Radiology*, vol. 31, no. 8, pp. 6021–6029, 2021.
13. Koo, Z. Xiang, U. Gretzel, and M. Sigala, "Artificial intelligence (AI) and robotics in travel, hospitality and leisure," *Electronic Markets*, vol. 31, no. 3, pp. 473–476, 2021.
14. E. Kotter and E. Ranschaert, "Challenges and solutions for introducing artificial intelligence (AI) in daily clinical workflow," *European Radiology*, vol. 31, no. 1, pp. 5–7, 2021.
15. K. H. Ng and J. H. D. Wong, "A clarion call to introduce artificial intelligence (AI) in postgraduate medical physics curriculum," *Physical and Engineering Sciences in Medicine*, vol. 45, no. 1, pp. 1–2, 2022.
16. M. K. Tripathi, A. Nath, T. P. Singh, A. S. Ethayathulla, and P. Kaur, "Evolving scenario of big data and Artificial Intelligence (AI) in drug discovery," *Molecular Diversity*, vol. 25, no. 3, pp. 1439–1460, 2021.
17. V. Arvind, D. Kaji, J. Kim, J. M. Caridi, S. K. Cho, and K. Samuel, "Artificial intelligence (AI) can predict postoperative complications better than traditional statistical testing following anterior cervical discectomy and fusion (ACDF)," *The Spine Journal*, vol. 17, no. 10, pp. S145–S146, 2017.
18. M. F. Byrne, N. Chapados, F. Soudan et al., "Su1614 artificial intelligence (AI) in endoscopy-deep learning for optical biopsy of colorectal polyps in real-time on unaltered endoscopic videos," *Gastrointestinal Endoscopy*, vol. 85, no. 5, pp. AB364–AB365, 2017.
19. M. KauKau, "Der Aufstieg von Artificial Intelligence (AI) und seine Auswirkungen auf die Rechtsanwaltschaft - existenzbedrohung oder willkommene Entlastung?" *Computer und Recht*, vol. 37, no. 7, pp. 498–504, 2021.
20. S. Makridakis, "The forthcoming Artificial Intelligence (AI) revolution: its impact on society and firms," *Futures*, vol. 90, no. jun, pp. 46–60, 2017.

21. J. A. Belfor, I. S. Sena, D. K. B. d Silva, and B.E. Fd LopesKoga JúniorSantos, "Competências pedagógicas docentes sob a percepção de alunos de medicina de universidade da Amazônia brasileira," *Ciência & Saúde Coletiva*, vol. 23, no. 1, pp. 73–82, 2018
22. M. C. Pennington and P. Rogerson- Revell, "Using technology for pronunciationteaching, learning, and assessment," *English Pronunciation Teaching and Research*, vol. 2019, pp. 235–286, 2019, Chapter 5.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.