iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

AUTOMATIC PRONUNCIATION MISTAKE DETECTOR USING MACHINE LEARNING

Mr. M. Rajkumar [1]
Assistant Professor, Department of Computer Science and Engineering,
J.B. Institute of Engineering and Technology, Hyderabad.
MD.SANA [2]
B. LAVANYA [3]
K. KINNERA [4]
N. SUMA [5]
UG Students, Department of Computer Science and Engineering,
J.B. Institute of Engineering and Technology, Hyderabad.

ABSTRACT
Machine learning is an artificial intelligence technology that educates systems to identify patterns from data and generate inferences or decisions without explicit programming. This approach is similar to teaching machines how to learn and think autonomously. Traditional English education emphasizes on writing and grammar instead of vocal language learning. As a result, while many Chinese students perform well on written exams, few can speak English successfully in everyday situations. With the increase in around the world relations, the importance placed on language as an instrument of communication has underlined the value of spoken language training. To help students practice their pronunciation outside of school, an English pronunciation detection tool is presented. Only proper pronunciation can help users convey their ideas clearly. As a result, the primary outcome of oral language learning is the development of pronunciation.

Keywords:
Automatic pronunciation, mistake detection, machine learning, speech recognition, error detection, natural language processing, acoustic features, language models, deep learning, neural networks, feature extraction, speech processing, phoneme recognition, AI-based detection, pronunciation evaluation, mispronunciation, speech-to-text, voice recognition, learning algorithms, automatic correction, speech analysis

INTRODUCTION

Machine learning is an artificial intelligence technology that educates systems to identify patterns from data and generate inferences or decisions without explicit programming. This approach is similar to teaching machines how to learn and think autonomously. Traditional English education emphasizes on writing and grammar instead of vocal language learning. As a result, while many Chinese students perform well on written exams, few can speak English successfully in everyday situations. With the increase in around the world relations, the importance placed on language as an instrument of communication has underlined the value of spoken language training. To help students practice their pronunciation outside of school, an English pronunciation detection tool is presented. Only proper pronunciation can help users convey their ideas clearly. As a result, the primary outcome of oral language learning is the development of pronunciation. Machine learning is akin to educating machines to act intelligently. By analysing a vast array of examples, these machines can discern patterns and subsequently make informed decisions about new data based on their acquired knowledge. The development of machine learning algorithms often utilizes frameworks like TensorFlow and Py Torch, which streamline the process of building solutions. As globalization progresses, English has gained a lot of attention as the most widely spoken language. The development of oral ability necessitates extensive oral practice, which is expected of all students studying English. Furthermore, throughout the practice phase, wrong feedback must be delivered in a timely and appropriate manner. Chinese students generally practice their pronunciation by listening to a tape, reading it out, and acting it out, similar to how they utilize a common language repeater. It was difficult to determine the relationship between the machine's speech and the children's reading because they received no feedback from the system during practice. With the growing trend of globalization, English, as the most widely used language, has gotten a lot of attention. Every student in English learning must do a lot of oral practice, which is an essential part of developing oral ability. Simultaneously, during the practice process, there must be timely and appropriate corrective feedback. Chinese students usually practice their

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

pronunciation by listening to are cording, reading it, and imitating it, such as by using a widely used language repeater. Because students received no feedback during the practice period, it was difficult to determine the relationship between the machine's speech and the students' reading.

OBJECTIVES

The objectives of an automatic pronunciation mistake detector using machine learning are:
1. Identify Pronunciation Errors: The system should be able to recognize when a word or sound is pronounced incorrectly by comparing it to correct pronunciation examples.
2. Provide Feedback: It should give users immediate feedback, indicating where their pronunciation mistakes are and how to improve.
3. Learn from Data: The system should improve over time by learning from a large dataset of correct and incorrect pronunciations, allowing it to detect a wider range of errors.
4. Support Language Learning: Help language learners improve their pronunciation by identifying mistakes and offering suggestions for improvement.

METHODOLOGY

The methodology for developing an automatic pronunciation mistake detector using machine learning can broken down into several key steps:
1. Data Collection: Collect a large dataset of spoken words or sentences, including both correct and incorrect pronunciations. This dataset could include various accents, speaking speeds, and contexts. The data should be labeled to indicate which words or sounds are pronounced correctly and which ones are incorrect.
2. Preprocessing the Audio: Convert the recorded speech into a format suitable for machine learning, typically by extracting features like Mel-frequency cepstral coefficients (MFCC), spectrograms, or pitch Normalize the audio to ensure consistency in terms of volume and background noise.
3. Model Training: Use the preprocessed audio data to train a machine learning model, such as a neural network or support vector machine (SVM). The model learns to classify pronunciation as correct or incorrect based on features extracted from the audio. Common approaches involve supervised learning, where the model is trained with labeled data (correct or incorrect pronunciation).
4. Error Detection: Once trained, the model is used to identify pronunciation errors by comparing the spoken input to the expected pronunciation in the dataset. It can detect errors like mispronunciations of vowels, consonants, intonation, stress patterns, etc.
5. Feedback Generation: When the model detects a mistake, it generates feedback for the user. This could be in the form of visual (e.g., a color-coded error), auditory (e.g., a message or sound), or text-based feedback (e.g., spelling or phonetic corrections).
6. Evaluation and Improvement: Continuously evaluate the model's accuracy using a test dataset that it has not seen before. This helps to measure how well the model generalizes to new inputs. Fine-tune the model by adding more data, improving feature extraction methods, or adjusting the learning algorithm to increase accuracy.
7. Deployment: Once the model achieves satisfactory performance, it can be deployed in language learning applications, mobile apps, or platforms where users can practice and improve their pronunciation.
8. Continuous Learning: Implement feedback loops where the system can learn from new data, improving over time as it encounters more variations of speech.

RESULTS AND DISCUSSION

1. Improved Accuracy: With the use of deep learning algorithms, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and more recently Transformer models, pronunciation mistake detection has become significantly more accurate. These models can analyse speech and compare it to a native speaker's baseline.
2. Phoneme-Level Detection: Machine learning models can detect mispronunciations at the phoneme level (the smallest unit of sound), which is crucial for more granular detection of pronunciation mistakes. For example, mispronunciations like swapping "r" with "w" can be caught and flagged.
3. Real-Time Feedback: Systems are increasingly capable of providing real-time feedback. This allows users, especially language learners, to receive immediate suggestions for correcting their pronunciation mistakes.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

4. Robustness to Accents: With enough diverse training data, modern models can accommodate a wide range of accents and dialects, improving the generalizability of pronunciation mistake detection. This is especially helpful for global applications.
5. Error Classification: Some models are designed to not just detect errors but also classify them into categories, such as substitutions, omissions, additions, or distortions. This allows for more precise feedback, which can aid in focused learning.
6. Contextual Awareness: Advanced systems now integrate contextual awareness, where the pronunciation mistake is judged not just based on individual sounds but also on how they fit within the context of surrounding words. This improves the detection of both phonetic errors and natural flow issues in speech.
7. Multilingual Capabilities: There is increasing support for multilingual systems, where pronunciation mistakes can be detected across different languages, supporting users in learning multiple languages simultaneously.

ACKNOWLEDGEMENT

At outset we express our gratitude to the almighty lord for showering his grace and blessings upon us to complete this Major Project. Although our name appears on the cover of this book, many people have contributed in some form or the other to this project Development. I could not have done this Project without the assistance or support of each of the following. First of all I am highly indebted to Dr . P. C. KRISHNAMACHARY, Principal for giving us the permission to carry out this Major Project. I would like to thank Dr. G. SREENIVASULU , Associate Professor & Head of the Department of COMPUTER SCIENCE AND ENGINEERING, for being moral support throughout the period of the study in the Department. I am grateful to Mr. M .RAJKUMAR, Assistant Professor COMPUTER SCIENCE ENGINEERING, for his valuable suggestions and guidance given by him during the execution of this Project work. I would like to thank the Teaching and Non-Teaching Staff of the Department of Computer Science & Engineering for sharing their knowledge with me.

CONCLUSION

In conclusion, this speech recognition and analysis system offers a comprehensive solution for evaluating and improving pronunciation accuracy. By combining advanced techniques in speech recognition, phoneme comparison, and pitch analysis, the system provides users with detailed feedback on their spoken language performance. Through meticulous text preprocessing and textual similarity calculation, it accurately assesses the alignment between spoken words and reference text. Additionally, the incorporation of pitch analysis ensures that not only the words but also the intonation matches the reference, enhancing the overall assessment of pronunciation fidelity. The dynamic adjustment of similarity scores based on pitch disparities further refines the evaluation process, accounting for variations in tonal accuracy. The system's feedback generation mechanism categorizes pronunciation accuracy into distinct levels, empowering users with actionable insights for targeted improvement efforts. Moreover, efficient audio processing and temporary storage mechanisms contribute to the system's seamless operation, ensuring a smooth user experience. Overall, this system serves as a valuable tool for language learners, speech therapists, and individuals seeking to enhance their pronunciation skills, facilitating effective communication and language proficiency development. Pronunciation Error Detection and Correction demonstrates a practical approach to real-time pronunciation error detection and correction. Combining speech recognition, the CMU Pronouncing Dictionary, and text-to speech capabilities, the script offers a valuable tool for language learners and individuals aiming to enhance their spoken communication skills. While the focus of this implementation centers on individual word pronunciation, it sets the foundation for more advanced and comprehensive error detection and correction systems

REFERENCES

[1]Bandar Ali Al-Rami and Yousef Houssni Zrekat, “A framework for pronunciation error detection and correction for non-native Arab speakers of English language”, ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print) © 2023 by the authors; licensee Growing Science, Canada.doi: 10.5267/j.ijdns.2023.5.004
[2]Yuhua Dai, “An Automatic Pronunciation Error Detection and Correction Mechanism in English Teaching Based on an Improved Random Forest Model”, Journal of Electrical and Computer Engineering Volume 2022,Article ID 6011993, 2022.
[3]Lei Chen , Chenglin Jiang, Yiwei Gu, Yang Liu ,and Jiahong Yuan, “Automatically Detecting Reduced formed English Pronunciations by Using Deep Learning”,Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022), pages 22 – 26 July 15, 2022 c 2022 Association for Computational Linguistics.
[4]Liang Meng, Jinal Upadhyay, Sumit Kumar et al., “Nonlinear Network Speech Recognition Structure in a Deep Learning Algorithm”, Computational Intelligence and Neuroscience Volume 2022, Article ID 6785642.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

[5]Zhan Zhang, Yuehai Wang, Jianyi Yang, "Text-conditioned Transformer for automatic pronunciation error detection", https://doi.org/10.1016/j.specom.2021.04.004 Received 26 August 2020; Received in revised form 4 11 February 2021; Accepted 18 April 2021 Available online 24 April 2021 0167-6393/© 2021 Elsevier B.V. All rights reserved.
[6]Fengming Jiao, Xin Zhao, Jiao Song, "A Spoken English Teaching System Based on Speech Recognition and Machine Learning", iJET – Vol. 16, No. 14, 2021 Jitang College of North China University of Science and Technology, Tangshan, China, 2021
[7]Zhang Gang, "Quality evaluation of English pronunciation based on artificial emotion recognition and gaussian mixture model", Journal of Intelligent & Fuzzy Systems 40 (2021) 7085– 7095 DOI: 10.3233/JIFS 189538 IOS Press, Henan Medical College, Henan Zhengzhou, China 2021.
[8]Zhang Shufang, "Design of an Automatic English Pronunciation Error Correction System Based on Radio Magnetic Pronunciation Recording Devices", Journal of Sensors Volume 2021, Article ID 5946228, 12 pages https://doi.org/10.1155/2021/5946228
[9] Cai Xiao-ming, "Automatic Detection of English Pronunciation Errors based on Statistical Pattern Recognition", December 25-27, 2020 Harbin, China IEEE International Conference on Industrial Application of Artificial Intelligence (IAAI), 2020.
[10] Rahib Abiyev, John Bush Idoko, Murat Arslan, "Reconstruction of Convolutional Neural Network for Sign Language Recognition", Proc. of the 2nd International Conference on Electrical, Communication and Computer Engineering (ICECCE) 12-13 June 2020, Istanbul, Turkey, 2020