IJNRD.ORG ISSN : 2456-4184

INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG

An International Open Access, Peer-reviewed, Refereed Journal

AUTOMATIC PRONUNCIATION ANALYSIS USING ARTIFICIAL INTELLIGENCE TECHNOLOGY

[1] Dr.P.Thiyagarajan , S Kavinkumar[2], S Kavinraja[3], P Manoj Prabhakar[4]

[1]*Associate professor,* [2] *student,* [3] *student,* [4] *student*

Department of Computer Science and Engineering,Paavai Engineering College (Autonomous)
Pachal, Namakkal, Tamil Nadu., India.

Abstract: The Automatic Pronunciation analysis (APA) is an innovative language learning tool designed to revolutionize the way individuals acquire and perfect their pronunciation skills. Built upon advanced speech recognition and machine learning technologies, offers a real-time, personalized, and effective solution for learners of all levels. This groundbreaking system operates by capturing the user's spoken language input and meticulously analysing it in comparison to the desired target pronunciation. Leveraging deep neural networks, acoustic modelling, and phonetic analysis, then can swiftly identify and assess pronunciation errors, enabling users to pinpoint areas of improvement. The APA provides immediate, constructive feedback, highlighting specific mispronunciations and offering visual representations for enhanced learning. Key features of the APA include its adaptability to a wide range of languages and dialects, making it a versatile tool for learners from diverse linguistic backgrounds. It can be integrated seamlessly into language learning applications, e-learning platforms, or used as a standalone tool for self-improvement or by educators and speech therapists. In this project, explore the APA's architecture, capabilities, and its transformative potential in the realm of language learning and education

IndexTerms-CNN,NLP,UML,DNN,RNN,SSD Algorithm

INTRODUCTION

In an increasingly interconnected world, effective communication in a global context has become a critical skill. Language proficiency and accurate pronunciation are integral components of this skill, enabling individuals to engage with diverse cultures, forge international relationships, and succeed in today's multicultural environments. For language learners and educators, the quest for impeccable pronunciation has long been a fundamental challenge. Pronunciation analysis represents a groundbreaking solution to this challenge. Language learning has evolved with advancements in technology, and stands at the forefront of this evolution. This innovative tool harnesses the power of artificial intelligence, speech recognition, and machine learning to provide real-time, personalized feedback on pronunciation accuracy. This introduction sets the stage for a comprehensive exploration of the APMD, its underlying technologies, and its potential to transform the way we approach language learning. By seamlessly integrating cutting-edge speech analysis techniques, offers a new dimension in pronunciation improvement, empowering individuals to enhance their linguistic proficiency with unprecedented precision and efficiency. As we delve into the intricacies of the speech processing, it becomes evident that this tool has the potential to redefine language learning paradigms and facilitate effective cross-cultural communication on a global scale. The pursuit of linguistic proficiency is a journey fraught with challenges, chief among them being the acquisition and refinement of accurate pronunciation. Pronunciation, the art of articulating words and sounds correctly, is not only pivotal for clear and effective communication but also for building cultural empathy and understanding. Incorrect pronunciation can lead to misunderstandings and hinder one's ability to connect with speakers of a different language or dialect. The traditional approach to pronunciation improvement often involves human teachers and extensive practice, which, while effective, can be time-consuming and costly. Furthermore, access to experienced language instructors is not always readily available, limiting the opportunity for focused, personalized learning. This is where the APMD steps in as a transformative solution. By harnessing the power of state-of-the-art speech recognition technology, the APA bridges the gap between traditional language learning and cutting-edge artificial intelligence. It provides users with an opportunity to receive real-time feedback on their pronunciation, allowing them to identify and rectify mistakes as they happen. This immediate feedback loop not only accelerates the learning process but also boosts learners' confidence by enabling them to track their progress in a tangible and visible manner. The APA 2 is versatile, accommodating a wide array of languages and dialects, thereby catering to the needs of learners from various linguistic backgrounds.

RESEARCH METHODOLOGY

[1] EXISTING SYSTEM Automatic pronunciation in the context of TTS involves converting written text into spoken language with correct pronunciation, intonation, and natural-sounding speech. A rulebased system for pronunciation mistake detection is designed to identify and correct mispronunciations in speech or text-to-speech (TTS) applications by applying predefined linguistic rules. These rules encompass various aspects of phonology and phonetics, including phoneme pronunciation, stress patterns, syllable divisions, and contextual variations. The system maintains comprehensive dictionaries mapping words to their correct pronunciation and employs phonetic transcription, such as the International Phonetic Alphabet (IPA), to represent words accurately. When a user input or TTS output deviates from these rules, the system flags it as a pronunciation mistake and offers feedback or correction suggestions based on the established rules. It's a valuable tool for improving the correctness and naturalness of speech synthesis, but may have limitations when dealing with complex or rare pronunciation variations. Combining rule-based systems with data-driven techniques can enhance overall performance. Machine learning can be defined as artificial intelligence algorithms that can infer and predict from data to mimic the way humans learn. There are various machine learning algorithms that are capable of solving classification, regression and clustering tasks. In our work, popular methods suitable for classification problem are emphasized which are support vector machine (SVM), k-Nearest neigbour (k-NN), decision tree (DT) and naïve Bayes SVM is a supervised learning approach. Kernel functions can also be used depending on the type of data during the operation of the algorithm. In this way, both linear and nonlinear classification operations can be performed. It is aimed to separate all data with a hyperplane. However, if the data cannot be fully separated, they cannot be classified with a single plane. Therefore, different kernel functions are used. A margin is determined around the hyperplane. Whether this margin is large or small directly affects the classification performance. Margin can be controlled with the “C” hyperparameter. The larger the C, the narrower the margin. Also, if the model is overfit, C needs to be reduced. In this work linear kernel and 0.02 used as C parameters. 9 k-NN is basically based on the determination of the class of the data whose class is unknown, according to the nearest “k” neighbor from the data in the training set. As a result of performing a distance measurement between the test data and the training data, the nearest “k” nearest neighbors are determined. Then, the class value of the tested data is determined according to these labels. Basic purpose of decision trees is to divide the data set into smaller subgroups that are more visually understandable within the framework of certain rules (decision rules). Since the output of the algorithm is a flowchart that looks like a tree visually, it is called a decision tree. There are 4 basic structures on a decision tree: root node, nodes, branches and leaves (terminal node). The root node is where classification process starts from this point. If the observations are in a homogeneous structure, they will naturally be in the same class and the classification process will end without branching the root node. In heterogeneous observations, the root node divides into two or more branches according to the best quality that divides the observations into classes and creates new nodes. The last non-branching node of the tree is the terminal node and represents the classes to which the observations are assigned. Naïve Bayes classification is based on Bayes theorem. It is used to estimate the probability that a particular set of features belongs to a particular class. It aims to select the decision with the highest probability using probability calculations. Each attribute is considered independent from other attributes in the class. different class based on various attributes. Naïve Bayes classifiers are extremely fast compared to more complex methods.

FEASIBLITY STUDY

The purpose of this chapter is to introduce the reader to feasibility studies, project appraisal, and investment analysis. Feasibility studies are an example of systems analysis. A system is a description of the relationships between the inputs of labour, machinery, materials and management procedures, both within an organisation and between an organisation and the outside world. During the planning and execution stages of an audit, it’s important to have a clear understanding of what the objectives of the audit include. Companies should strive to align their business objectives with the objectives of the audit. This will ensure that time and resources spent will help achieve a strong internal control environment and lower the risk of a qualified opinion. Objectives of Feasibility Study • To explain present situation of the automation. • To find out if a system development project can be done is possible. • To find out whether the final product will benefit end user. • To suggest the possible alternative solutions.

TECHNICAL FEASIBILITY

Technical Feasibility assessment focuses on the technical resources available to the organization. It helps organizations determine whether the technical resources meet capacity and whether the technical team is capable of converting the ideas into working systems. In technical feasibility the following issues are taken into consideration.

• Whether the required technology is available or not
.• Whether the required resources are available - Manpower- programmers, testers & debuggers, Software and hardware

Once the technical feasibility is established, it is important to consider the monetary factors also. Since it might happen that developing a particular system may be technically possible but it may require huge investments and benefits may be less. For evaluating this, economic feasibility of the proposed system is carried out.

FEASIBLITY STUDY

The purpose of this chapter is to introduce the reader to feasibility studies, project appraisal, and investment analysis. Feasibility studies are an example of systems analysis. A system is a description of the relationships between the inputs of labour, machinery, materials and management procedures, both within an organisation and between an organisation and the outside world. During the planning and execution stages of an audit, it's important to have a clear understanding of what the objectives of the audit include. Companies should strive to align their business objectives with the objectives of the audit. This will ensure that time and resources spent will help achieve a strong internal control environment and lower the risk of a qualified opinion. Objectives of Feasibility Study • To explain present situation of the automation. • To find out if a system development project can be done is possible. • To find out whether the final product will benefit end user. • To suggest the possible alternative solutions.

TECHNICAL FEASIBILITY

Technical Feasibility assessment focuses on the technical resources available to the organization. It helps organizations determine whether the technical resources meet capacity and whether the technical team is capable of converting the ideas into working systems. In technical feasibility the following issues are taken into consideration. • Whether the required technology is available or not • Whether the required resources are available - Manpower- programmers, testers & debuggers, Software and hardware Once the technical feasibility is established, it is important to consider the monetary factors also. Since it might happen that developing a particular system may be technically possible but it may require huge investments and benefits may be less. For evaluating this, economic feasibility of the proposed system is carried out.

PROBLEM DEFINITION

Data Collection: Gather a diverse dataset of spoken language samples covering various accents, dialects, and linguistic contexts. This dataset serves as the foundation for training the AI models. Feature Extraction: Extract relevant features from the collected speech data. This may include acoustic features such as pitch, duration, intensity, and spectral features, as well as linguistic features like stress patterns and phonetic context. Model Selection: Choose appropriate machine learning models for the task. Common choices include cosine similarity. Consider the complexity of the pronunciation analysis task and the available computational resources when selecting models. Training: Train the selected models using the extracted features and labelled data. The training process involves optimizing the model parameters to minimize a chosen objective function, such as mean squared error or cross-entropy loss. Evaluation: Evaluate the trained models using a separate validation dataset to assess their performance. Integration: Integrate the trained models into an application or platform where users can input spoken language samples for analysis. This may involve developing a user interface for data input and visualization of analysis results. Feedback Mechanism: Implement a feedback mechanism to provide users with insights into their pronunciation accuracy and areas for improvement. This could include visualizations of pronunciation errors, suggested corrections, and personalized recommendations for practice. Scalability and Maintenance: Ensure that the system is scalable to handle large volumes of data and users. Regular maintenance and updates may be required to keep the system performing optimally and to incorporate improvements based on user feedback and advancements in AI technology.

FRAMEWORK CONSTRUCTION

Automatic Pronunciation Analysis involves the use of technological tools to assess and analyze spoken language, particularly focusing on pronunciation accuracy. Automatic Pronunciation Analysis (APA) refers to the process of evaluating and providing feedback on the pronunciation of speech utterances using computational techniques. This approach leverages technologies such as speech recognition, signal processing, and machine learning to assess the accuracy and fluency of spoken language. This process often employs Speech Recognition Technology, which transcribes spoken words into text and can assess pronunciation by considering acoustic features like pitch and intonation. In this module, we can create the framework for admin and user. Admin can view the user details and also provide score details about user speech

READ THE SPEECH DATA

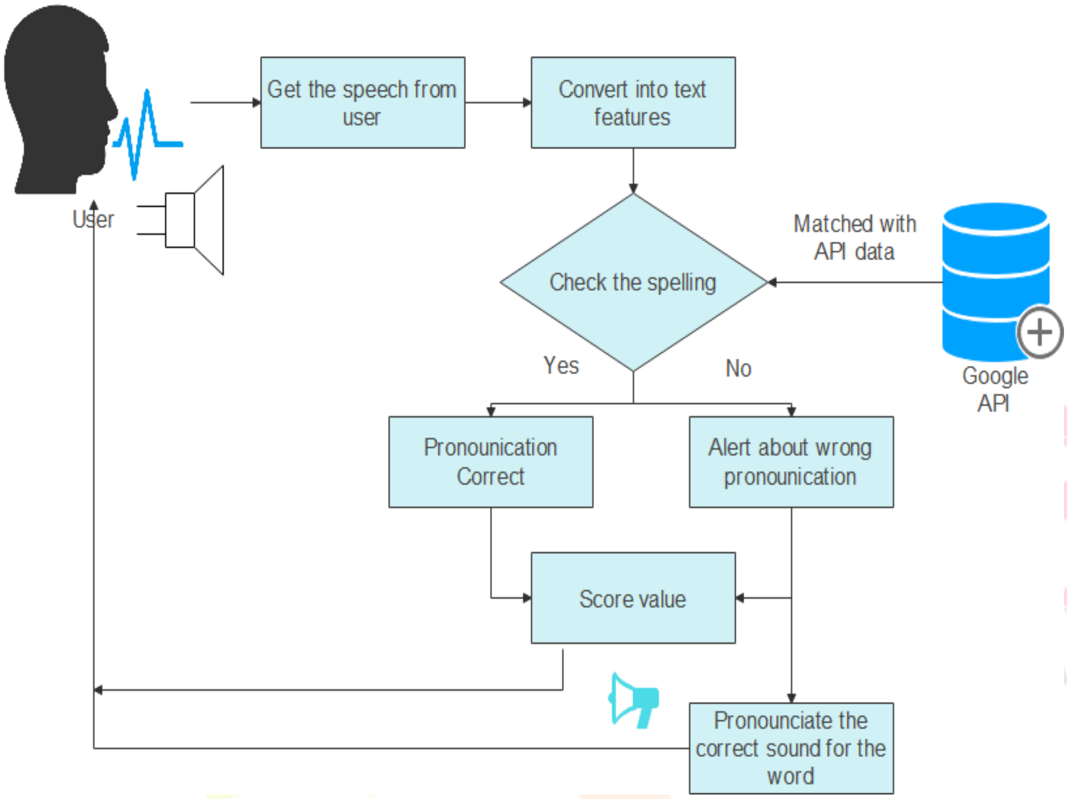
Spell checkers and correctors are either stand-alone applications capable of processing a string of words or a text, or as an embedded tool which is part of a larger application such as a word processor. Various search and replace algorithms are adopted to fit in the domain of spell checking and correcting. Spelling error detection and correction are closely related to exact and approximate pattern matching respectively. In this module, user speech about word can be read from voice device. Speech can be any persons and any pitch. Speech Recognition is a popular Python library that provides simple and easy-to-use functions to work with speech recognition. It supports multiple speech engines, including Google Web Speech API. Spell checking identifies words that are valid in some language, as well as the misspelled words in the language. Spell correcting suggests one or more alternative words as the correct spelling when a misspelled word is identified. Spell checking involves non-word error detection and spelling correction involves isolated-word error correction. Isolated word error correction refers to spell correcting without taking into account any textual or linguistic information in which the misspelling occurs whereas context-dependent word correction would correct errors involving textual or linguistic context.

FEATURES EXTRACTION

Extracting speech features using Mel-Frequency Cepstral Coefficients (MFCC) involves several steps. Initially, the speech signal undergoes preprocessing, including noise reduction and framing to segment it into short, overlapping frames. Each frame is then multiplied by a window function to mitigate spectral leakage. Following this, a Fast Fourier Transform (FFT) is applied to convert the framed signal into the frequency domain. Subsequently, the power spectrum of each frame is computed and passed through a bank of Mel filters, which are designed to mimic the human auditory system's frequency response. The energies obtained from the Mel filters are then logarithmically compressed to account for the non-linear perception of sound intensity by humans. Finally, Discrete Cosine Transform (DCT) is employed to decorrelate the log filterbank energies and extract the MFCCs, which serve as compact representations of the spectral characteristics of the speech signal. These MFCCs can then be utilized for various applications such as speech recognition, speaker identification, and emotion detection. Feature extraction in the context of speech processing involves converting raw speech signals into a set of relevant features that can be used for various speech-related tasks such as speech recognition. MFCCs are widely used for speech and audio signal processing. They represent the short-term power spectrum of a sound and are particularly effective for speech recognition.

SYSTEM ARCHITECTURE

A system architecture or systems architecture is the conceptual model that defines the structure, behaviour, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviours of the system. System architecture can comprise system components, the externally visible properties of those components, the relationships (e.g. the behaviour) between them. It can provide a plan from which products can be procured, and systems developed, that will work together to implement the overall system. There have been efforts to formalize languages to describe system architecture; collectively these are called architecture description languages (ADLs)



CONCLUSION

In conclusion, cosine similarity-based spell checking offers a robust and contextually informed approach to identifying and correcting misspelled words in text documents. By utilizing word embeddings and cosine similarity metrics, this method leverages the semantic and syntactic relationships between words to provide more accurate and relevant suggestions for corrections. Through the comparison of word vectors and the ranking of potential candidates based on their similarity scores, cosine similarity-based spell checking enhances the accuracy and effectiveness of spell-checking algorithms. This approach not only improves the quality of spell correction but also enhances the overall user experience by reducing the need for manual intervention and offering more contextually appropriate suggestions. As such, cosine similarity-based spell checking represents a valuable tool for improving the accuracy and readability of text documents across a wide range of applications and domains.

FUTURE

Integration of multimodal data sources, such as images or speech inputs, alongside text data could enhance the spell-checking capabilities. For example, incorporating visual context from images or audio context from speech inputs could provide additional clues for identifying and correcting misspellings. And implementing real-time spell-checking capabilities in text editors, web browsers, or communication platforms would provide immediate feedback to users as they type, helping to prevent spelling errors before they occur.

REFERENCES

[1] Acquah, Emmanuel O., and Heidi T. Katz. "Digital game-based L2 learning outcomes for primary through high-school students: A systematic literature review." Computers & Education 143 (2020): 103667.

[2] Ballard, Kirrie J., Constantina Markoulli, and Penelope Monroe. "A Phonological Longitudinal Evaluation of Tablet-Based Child Speech Therapy with Apraxia World." (2021).

[3] Bashori, Muzakki, et al. "'Look, I can speak correctly': learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology." Computer Assisted Language Learning (2022): 1-29.

[4] Baevski, Alexei, et al. "wav2vec 2.0: A framework for self-supervised learning of speech representations." Advances in neural information processing systems 33 (2020): 12449-12460

[5] Cummings, Alycia, Kristen Giesbrecht, and Janet Hallgrimson. "Intervention dose frequency: Phonological generalization is similar regardless of schedule." Child Language Teaching and Therapy 37.1 (2021): 99-115.

[6] Chen, Li-Wei, and Alexander Rudnicky. "Exploring Wav2vec 2.0 Fine Tuning for Improved Speech Emotion Recognition." ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023.

[7] Li, Ping, and Yu-Ju Lan. "Digital language learning (DLL): Insights from behavior, cognition, and the brain." Bilingualism: Language and Cognition 25.3 (2022): 361-378.

[8] Shi, Jiatong, Nan Huo, and Qin Jin. "Context-aware goodness of pronunciation for computer-assisted pronunciation training." arXiv preprint arXiv:2008.08647 (2020).

[9] Tejedor-Garcia, Cristian, et al. "Using challenges to enhance a learning game for pronunciation training of English as a second language." IEEE Access 8 (2020): 74250-74266.

[10] Zou, Di, Yan Huang, and Haoran Xie. "Digital game-based vocabulary learning: where are we and where are we going?." Computer Assisted Language Learning 34.5-6 (2021): 751- 777

[1]