EXPLORING MACHINE AND DEEP LEARNING METHODS FOR ACCURATE ACCENT RECOGNITION

ABSTRACT

Accent recognition plays a critical role in enhancing the performance of Automatic Speech Recognition (ASR) systems, which often struggle with accent variations. This paper presents a comprehensive review of machine learning (ML) and deep learning (DL) techniques applied to accent recognition. It systematically examines preprocessing methods, feature extraction techniques, and classification models used in the literature. The study highlights the dominance of Mel-Frequency Cepstral Coefficients (MFCC) as a feature extraction method and discusses the effectiveness of models such as Gaussian Mixture Models (GMMs), Support Vector Machines (SVMs), and deep architectures like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. Furthermore, the paper identifies research gaps, including the lack of standardized datasets for low-resource languages and the need for robust models that generalize across diverse accents. Future directions such as cross-lingual accent classification, generative models, and explainable AI are also discussed.

EXISTING SYSTEM

The existing accent recognition system typically follows a three-stage pipeline: preprocessing, feature extraction, and classification. Preprocessing involves steps like silence removal and normalization to clean the audio signal. Feature extraction primarily relies on MFCCs, often augmented with delta and double-delta coefficients. Classification is performed using a range of models, from classical methods like GMM and SVM to deep learning architectures such as CNNs and RNNs. Hybrid models like GMM-UBM and CNN-SVM have also been employed to improve accuracy.

Disadvantages of the Existing System:

1. Heavy Dependence on MFCCs: While effective, MFCCs are sensitive to noise and may not capture all accent-specific nuances, especially in low-frequency ranges.

2. Limited Generalization Across Languages: Most systems are trained on high-resource languages like English, leading to poor performance on low-resource or cross-lingual accents.
3. Lack of Real-World Robustness: Models are often evaluated in controlled settings and struggle with background noise, speaker variability, and channel distortions in real-world scenarios.

PROPOSED SYSTEM

The proposed system aims to address the limitations of existing approaches by integrating multi-modal data, self-supervised learning, and explainable AI techniques. It leverages advanced DL architectures such as transformers and attention-based models, trained on diverse and augmented datasets. The system also incorporates cross-lingual and few-shot learning strategies to improve generalization across accents and languages.

Advantages of the Proposed System:

1. Enhanced Robustness: Use of multi-modal features (e.g., visual, textual) and data augmentation improves performance in noisy and real-world environments.
2. Cross-Lingual Capability: Incorporation of meta-learning and self-supervised models like Wav2Vec 2.0 enables recognition of accents across multiple languages with limited data.
3. Explainability and Adaptability: Integration of Explainable AI (XAI) tools provides insights into model decisions, aiding in the development of more transparent and adaptable ASR systems.

SYSTEM REQUIREMENTS

- **H/W System Configuration:-**
- Processor - Pentium –IV
- RAM - 4 GB (min)
- Hard Disk - 20 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA

SOFTWARE REQUIREMENTS:

- **Operating system** : Windows 7 Ultimate.
- **Coding Language** : Python.
- **Front-End** : Python.
- **Back-End** : Django-ORM
- **Designing** : Html, css, javascript.
- **Data Base** : MySQL (WAMP Server).