Industrial Engineering Journal
ISSN: 0970-2555
Volume : 55, Issue 01 : 2026

VOCALEASE: AI FOR VOICE TRAINING AND ACCENT SMOOTHING

Prof. Sanjeevani R. Kale, Neha S. Balsaraf, Aditi D. Nikalje, Rasika R. Santape, Saloni D. Patil, Shruti G. Gangane Computer Science and Engineering Department, PRPCEM

Abstract
In today's globally connected world, effective communication plays a vital role in personal, academic, and professional success. However, many individuals face challenges related to unclear pronunciation, lack of vocal control, and non-native accents that hinder fluent speech delivery. VocalEase AI is an intelligent voice training and accent improvement system designed to address these issues through advanced artificial intelligence and speech processing techniques. The system utilizes machine learning algorithms to analyze a user's speech patterns, detect pronunciation errors, and provide personalized feedback for improvement. By incorporating real-time speech recognition, acoustic analysis, and natural language processing, VocalEase AI guides users through interactive voice exercises tailored to their linguistic background and learning pace. The platform also integrates deep learning models for accent modification, helping users achieve clearer and more confident communication. With a user-friendly interface and adaptive learning modules, VocalEase AI offers a comprehensive solution for enhancing vocal clarity, articulation, and accent accuracy. This project aims to make professional-level voice and accent training accessible to everyone, promoting confidence and global communication competence through technology-driven learning.

Keywords: *Accent Smoothing, ASR, Neural Networks, VAEs, Speech Processing, Deep Learning, Accent Recognition, Machine Learning, Inclusive Voice Technology, Multilingual Speech.*

1. Introduction

Voice plays a central role in how people communicate, shaping the way ideas, emotions, and information are shared. In many professions—such as teaching, public speaking, customer support, broadcasting, and online content creation— the clarity and quality of speech can greatly influence confidence, performance, and the listener's engagement. Yet, a large number of individuals struggle with issues like unclear pronunciation, lack of fluency, strong regional accents, or low vocal confidence. Although traditional voice training and speech therapy can help, they often require significant time, money, and access to trained experts. This creates a strong need for a more accessible and personalized solution that can help users improve their vocal skills whenever and wherever they want.

Advancements in Artificial Intelligence (AI), Machine Learning (ML), and speech processing technologies have opened the door for automated voice training systems. Today, speech recognition models, deep learning algorithms, and audio signal processing tools can analyze human voice with impressive accuracy. These technologies can measure elements such as pitch, tone, rhythm, stress, phoneme clarity, and overall fluency—allowing machines to provide feedback that once only a trained voice coach could offer.

VocalEase is developed with this goal in mind. It is an AI- driven voice training application that provides an intelligent, interactive, and easy-to-use platform for improving vocal clarity and pronunciation. The system records the user's speech, processes the audio, extracts key acoustic and linguistic features, and evaluates them using machine learning models trained on high-quality speech datasets. From this analysis, VocalEase gives personalized feedback, identifies incorrect pronunciations, suggests corrections, and provides tailored practice exercises. Over time, the system tracks the user's progress, helping them improve through consistent and adaptive learning.

The idea for VocalEase is inspired by the increasing need for affordable and effective voice improvement tools—especially now, when digital communication, virtual meetings, and online learning have become everyday norms. Many existing apps focus only on basic language learning or simple speech repetition, offering limited analysis or feedback. VocalEase aims to fill this gap by combining advanced speech processing with a user-friendly experience, making professional-level voice training available to a much wider audience.

This research paper explains the design, architecture, implementation, and evaluation of the VocalEase system. It discusses the algorithms used for speech processing, the structure of the AI models responsible for pronunciation assessment, and how the system performs in terms of accuracy and user experience. The results highlight the potential of AI as a practical tool for voice training and show how VocalEase can function as a helpful virtual voice coach for students, professionals, and anyone wishing to enhance their speech quality.

2. Objectives

1. To help people speak more clearly by analyzing their voice and showing where they can improve.
2. To identify common pronunciation mistakes and guide users on how to correct them.
3. To support users in developing a more neutral or consistent accent.
4. To improve overall confidence while speaking in daily conversations or professional settings.
5. To give personalized feedback based on each person's unique speaking style.
6. To provide a comfortable space where users can practice speaking without fear or judgment.
7. To help non-native speakers reduce mother-tongue influence and sound more fluent.
8. To use AI and machine learning so the system becomes smarter and more accurate over time.
9. To track progress regularly and show users how their speech is impressive.

3. Literature Review

Accent recognition is a critical area in speech processing and speaker identification, where machine learning (ML) techniques play a central role in identifying speaker specific traits from audio data. ML models learn patterns from labeled speech features to classify accents accurately. Early studies predominantly used Support Vector Machines (SVMs) due to their robustness in high-dimensional spaces. Hanani et al. (2013) demonstrated that SVMs combined with Mel- frequency cepstral coefficients (MFCCs) could effectively differentiate between native and non-native English speakers, achieving reasonable classification accuracy.

Accent variability remains a major challenge for ASR, motivating research in accent smoothing to convert accented speech into a neutral form. Traditional GMM/HMM models were limited, while deep learning architectures (CNNs, LSTMs, BLSTMs, and CNN-LSTM hybrids) significantly improved accent classification. Generative models like CycleGANVC, StarGAN-VC, and VAEs enable accent conversion and disentanglement of accent from speaker/linguistic features, producing more natural, accent- neutral speech. Recent advances include transformer-based multimodal models (e.g., AVHuBERT) that fuse audio–visual cues, and Explainable AI (XAI) for interpreting model decisions. Novel frameworks such as the intra-native accent shared feature (NAI) model leverage CNN-LSTM networks to capture shared accent traits, while data augmentation with voice conversion further boosts recognition under limited resources.

Machine Learning and Deep Learning Approaches for Accent Recognition: A Review [1] The study reveals that financial literacy significantly influences investment awareness and saving behavior, with individuals having higher financial literacy levels showing better awareness.

M. A. Dar and P. Jagalingam(2025) ML frameworks improved accent recognition [2]. Berrak Sisman And Junichi Yamagishi, "An Overview of Voice Conversion and Its Challenges: From Statistical Modeling to Deep Learning" VOL. 29, 2021[3] Findings reveal that environmental concern, health consciousness, and product quality significantly influence consumer purchase decisions, with eco-friendliness emerging as the strongest factor.

Yeshanew Ale Wubet And Deepak Balram, Intra-Native Accent Shared Features For Improving Neural Network-Based Accent Classification And Accent Similarity Evaluation" IEEE ACCESS.2023.3259901[4] finds that digital payments are widely adopted due to convenience, speed, and security, with user satisfaction largely dependent on ease of use and trust in the system. uses a descriptive research approach, gathering data via structured questionnaires to examine consumer perceptions and satisfaction regarding digital payment system.

Chenfeng Miao And Qingying Zhu, "EfficientTTS2: Variational End-To-End Text- To-Speech Synthesis And Voice Conversion, VOL. 32, 2024 [5] results indicate that convenience, price, product variety, and customer service strongly impact consumer preferences, with convenience and competitive pricing being the most influential factors. The study adopts a descriptive research design and collects primary data through a structured questionnaire from respondents to analyze factors influencing online shopping behavior.

Q. Shao, P. Guo, J. Yan, P. Hu, and L. Xie, "Decoupling and Interacting Multi-Task Learning Network for Joint Speech and Accent Recognition," arXiv preprint, arXiv:2311.07062, Nov. 2023[6]. Improving the speech enhancement model with discrete wavelet transform sub-band features in adaptive FullSubNet," IEEE Signal Processing Letters, Mar. 28, 2025[7] Decision Trees and Random Forests have also been explored for accent recognition. Decision Trees are interpretable and efficient, building rules based on acoustic features such as pitch, formants, and intensity. Random Forests, as ensemble methods, reduce overfitting and improve robustness, making them suitable for real-time applications and datasets of moderate size.

4. Methodology

The proposed Accent Detection and Pronunciation Training System follows a modular, multi-stage methodology designed to guide users from registration to speech analysis and learning improvement. The system architecture includes User- Side Modules and Admin-Side Modules, each optimized for specific operations.

1. System Initialization and User Management

The methodology begins with a structured onboarding sequence comprising a Landing Page, User Registration, and Login Authentication. New users create an account, after which the system verifies credentials and redirects them to a personalized dashboard. All user data—including profile information, speech tests, translation logs, and quiz history— is securely stored in the system database.

2. User Dashboard Module

Once authenticated, users access the Dashboard, which serves as the central hub for monitoring progress. The system retrieves and visualizes key metrics such as:

- Speech test history
- Translation activity
- Quiz performance
- Tutorial completion rate

Based on these metrics, the platform automatically generates recommended learning steps to enhance

the user's accent and pronunciation skills.

3. **Translation Module**

The Translator component enables users to convert text from English to other languages such as French. The module integrates text-to-speech (TTS) technology to generate accurate pronunciation audio for the translated output. Users can input text, listen to the correct accent, and practice through repeated playback, supporting accent acquisition and comparison.

4. **Speech Accent Detection Module**

The core of the system is the Speech Analysis Module, where users speak into the microphone, and the system processes the audio using accent classification algorithms. The module evaluates:

- Accent type
- Pronunciation clarity
- Fluency and speaking pace

The system generates instant feedback to help users improve articulation. An optional "Speak & Translate" feature detects accent characteristics while simultaneously providing translated output.

5. **Quiz Module**

To reinforce learning, the platform offers short quizzes on pronunciation, accent recognition, and language comprehension. User responses are evaluated, scored, and stored. The system maintains historical quiz performance to track progress and adapt recommendations.

6. **Profile Management Module**

The Profile module displays personal details such as name, email, and profile photo. It also includes a comprehensive activity history covering:

- Accent detection attempts
- Translation sessions
- Quiz scores

Users can update their information or reset their password through secure profile management.

7. **Tutorials Module**

A series of beginner-friendly tutorials provides structured training on accent basics, pronunciation improvement, and vocal exercises. Users can access instructional videos at any time, helping them practice autonomously and enhance learning outcomes.

8. **Session Management**

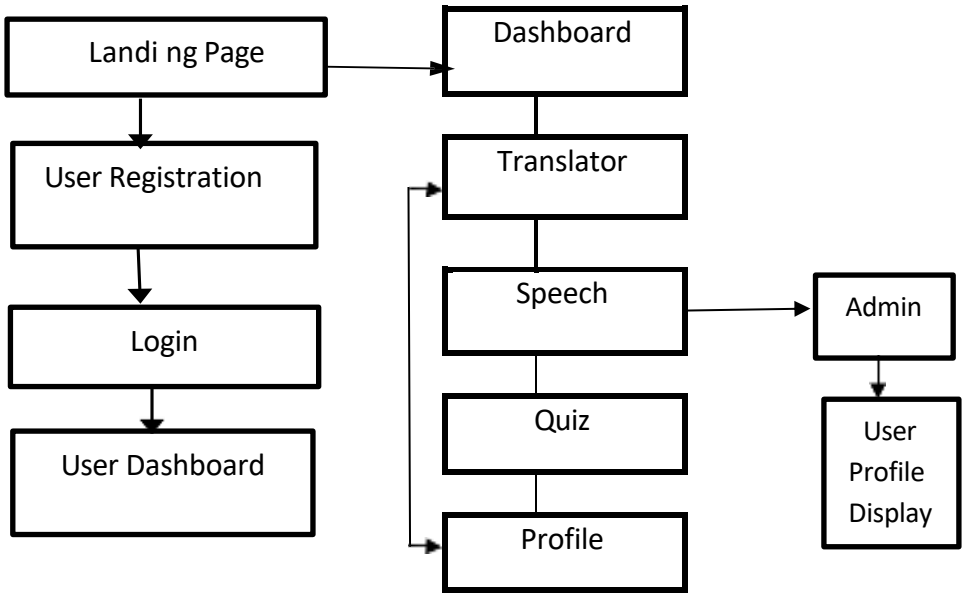
The Logout function terminates the user session securely and redirects them to the login page, ensuring proper authentication flow and data privacy.

9. **Admin Module**

- The administrative interface enables system administrators to monitor and manage registered users.
- Administrators can view user details such as name, email, and registration date, and track platform usage statistics, including the number of speech tests, quizzes attempted, and tutorials accessed.
- This module supports system maintenance, usage auditing, and backend management.
- This module supports backend analytics and maintains system integrity.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 55, Issue 01 : 2026

Fig 4.1 : Accent Smoothing Methodology



5. Conclusion

The VocalEase project shows how modern technology can genuinely help people improve their speaking and communication skills. Many learners struggle with pronunciation, accent influence, and confidence, but VocalEase makes this process easier by offering real-time guidance in a simple and supportive way. By listening to each user's voice and analyzing their speech patterns, the system provides personalized feedback that meets their specific needs.

This approach not only makes learning less intimidating but also helps users practice comfortably at their own pace. Through AI and machine learning, the tool becomes smarter over time, offering more accurate suggestions and serving as an affordable, flexible alternative to traditional voice coaching.

Beyond correcting pronunciation, VocalEase focuses on building confidence, which is equally important for effective communication. The platform creates a safe space where users can practice without fear of judgment, allowing them to gradually become clearer and more fluent speakers. Overall, the project demonstrates how technology can empower human expression and make communication training accessible to learners, professionals, and anyone who wants to improve their speech. With future possibilities such as emotional tone detection and broader language support, VocalEase has the potential to grow into an even more powerful and inclusive tool. Ultimately, it highlights that when technology and human needs come together, learning becomes more meaningful, engaging, and transformative.

References

[1] P. S. Yawale, "Applications of AI," Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET), vol. 12, no. IV, Apr. 2024, Art. no. IJRASET60822, doi: 10.22214/ijraset.2024.60822.

[2] M. A. Dar and P. Jagalingam, “Machine Learning and Deep Learning Approaches for Accent Recognition: A Review,” IEEE Access, vol. 11, pp. 1-1, Jan. 2025.
[3] Yi Zhou, Student Member, IEEE, Xiaohai Tian, Member, IEEE, And Haizhou Li, Fellow, IEEE, “Language Agnostic Speaker Embedding For Cross-Lingual Personalized Speech Generation”, November 8, 2021.
[4] Muzaffar Ahmad Dar And Jagalingam Pushparaj, “Machine Learning And Deep Learning Approaches For Accent Recognition: A Review” IEEE ACCESS.2025.3552935.
[5] Yeshanew Ale Wubet And Deepak Balram, “Intra-Native Accent Shared Features For Improving Neural Network-Based Accent Classification And Accent Similarity Evaluation” IEEE ACCESS.2023.3259901.
[6] Y. A. Wubet, D. Balram, and K.-Y. Lian, “Intra-native accent shared features for improving neural-network-based accent classification and accent similarity evaluation,” IEEE Access, vol. 11, pp. 32176–32186, 2023.
[7] Y. Iqbal, T. Zhang, T. S. Gunawan, A. Pratondo, X. Zhao, Y. Geng, M. Kartiwi, N. Saleem, and S. Bourouis, “A hybrid speech enhancement technique based on discrete wavelet transform and spectral subtraction,” IEEE Transactions on Audio, Speech, and Language Processing, Feb. 27, 2025.
[8] Z.-T. Wu and J.-W. Hung, “Improving the speech enhancement model with discrete wavelet transform sub-band features in adaptive FullSubNet,” IEEE Signal Processing Letters, Mar. 28, 2025.
[9] Y. Iqbal, T. Zhang, Y. Geng, M. Fahad, X. Zhao, S. U. Rahman, and A. Iqbal, “Discrete wavelet transform and spectral subtraction based speech enhancement algorithm for hearing aid application,” IEEE Access, Apr. 4, 2024.
[10] J. Ball, “Voice activity detection (VAD) in noisy environments,” Proc. IEEE Conf. Electrical and Computer Engineering, Johns Hopkins University, Baltimore, USA, Dec. 10, 2023.
[11] Q. Shao, P. Guo, J. Yan, P. Hu, and L. Xie, “Decoupling and Interacting Multi-Task Learning Network for Joint Speech and Accent Recognition,” arXiv preprint, arXiv:2311.07062, Nov. 2023.
[12] L. Beringer, S. Jassim, and H. A. Abdulmohsin, “Accent Classification Using Machine Learning Techniques: A Review,” International Journal of Computer Information Systems and Industrial Management Applications, vol. 17, pp. 421–451, May 2025.
[13] W. Wang and H. Liu, “Study on Recognition and Classification of English Accents Using Deep Learning,” De Gruyter Brill, 2023.
[14] Y. Iqbal, T. Zhang, T. S. Gunawan, A. Pratondo, X. Zhao, Y. Geng, M. Kartiwi, N. Saleem, and S. Bourouis, “A hybrid speech enhancement technique based on discrete wavelet transform and spectral subtraction,” IEEE Transactions on Audio, Speech, and Language Processing, Feb. 27, 2025.
[15] Z.-T. Wu and J.-W. Hung, “Improving the speech enhancement model with discrete wavelet transform sub-band features in adaptive FullSubNet,” IEEE Signal Processing Letters, Mar. 28, 2025.
[16] Y. Iqbal, T. Zhang, Y. Geng, M. Fahad, X. Zhao, S. U. Rahman, and A. Iqbal, “Discrete wavelet transform and spectral subtraction based speech enhancement algorithm for hearing aid application,” IEEE Access, Apr. 4, 2024.
[17] J. Ball, “Voice activity detection (VAD) in noisy environments,” Proc. IEEE Conf. Electrical and Computer Engineering, Johns Hopkins University, Baltimore, USA, Dec. 10, 2023.
[18]. Muzaffar Ahmad Dar And Jagalingam Pushparaj, ‘‘Machine Learning And Deep Learning Approaches For Accent Recognition:A Review’’IEEE ACCESS.2025.3552935. [2]. Chenfeng Miao And Qingying Zhu, “EfficientTTS2: Variational End-To-End Text- To-Speech Synthesis And Voice Conversion,VOL. 32, 2024.
[19].Yeshanew Ale Wubet And Deepak Balram, Intra- Native Accent Shared Features For Improving Neural Network- Based Accent Classification And Accent Similarity Evaluation” IEEE ACCESS.2023.3259901.
[20].Berrak Sisman And Junichi Yamagishi, "An Overview of Voice Conversion and Its Challenges: From Statistical Modeling to Deep Learning" VOL. 29, 2021.
[21]. G. Droua Hamdani,‘‘Design of accent classifier based on spee ch rhythm features,’’Multimedia

Tools Appl., vol. 82, no.14, pp. 21715–21728, Jun 2023.
[22]Yaroslav Getman, Nhan Phan, Ragheb Al-Ghezi1, (Graduate Student Member, IEEE), Ekaterina Voskoboinik1, Mittul Singh1,2, Tamás Grósz 1, Mikko Kurimo1, Giampiero Salvi 3,4, (Member, IEEE), Torbjørn Svendsen 3, (Life Senior Member, IEEE), Sofia Strömbergsson 5, Anna Smolander6, And Sari Ylinen 6, "Developing An AI-Assisted Low- Resource Spoken Language Learning App For Children", 10 August 2023.
[23]. Y. A. Wubet, D. Balram, and K.-Y. Lian, ''Intra-native accent shared features for improving neural network-based accent classification and accent similarity evaluation,'' IEEE Access, vol. 11, pp. 32176–32186, 2023.