Journal of Engineering, Electrical and Informatics

E-ISSN: 2809-8706
P-ISSN: 2810-0557

Research Article

Application of Voice Processing Technology With A Natural Language Processing Approach for Pronunciation Correction of Selected Vocabulary In English

Dadang Iskandar Mulyana [1], Rizki Ananda Pratama [2], and Sugiyono [3], Agiah Sofia [4]

1 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikom Cki), Jakarta; Email: dadang@stikomcki.ac.id
2 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikom Cki), Jakarta
3 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikom Cki), Jakarta; Email: sugiyono@stikomcki.ac.id
4 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikom Cki), Jakarta
* Corresponding Author: dadang@stikomcki.ac.id

Abstract: English vocabulary pronunciation is one of the important aspects that must be mastered by English learners. However, many people in Indonesia face difficulties in mastering basic vocabulary pronunciation, which can hinder their progress in the early stages of learning speaking and listening. Mistakes in pronunciation often cause ineffective communication, making the message difficult for the other person to understand. Even at the level of learners who have memorized many vocabularies and are able to have simple conversations, mispronunciation remains a significant obstacle and often hinders smooth communication. To address this challenge, this study aims to develop an application based on speech processing technology and Natural Language Processing (NLP) that is specifically designed to provide pronunciation correction for selected vocabulary. This application focuses on mastering the pronunciation of 500 basic vocabulary as a companion for learning speaking and listening at an early stage. This application does not only aim to memorize vocabulary, but also helps users learn to pronounce each word correctly. With this approach, users can get real-time corrective feedback for each vocabulary spoken, allowing them to correct mistakes directly and gradually improve their speaking ability. In addition, the application provides pronunciation analysis supported by speech processing technology to recognize and analyze user pronunciation errors, while NLP is used to provide relevant assessments and improvement suggestions automatically. The results of the study show that this application is effective in helping learners improve their pronunciation of basic vocabulary. By focusing on frequently used basic vocabulary, this application helps users improve their speaking skills more easily. This application is also a tool that supports independent learning for users, thereby increasing their confidence in speaking English. This study is expected to be a solution that supports more effective and affordable English learning, especially for beginners in Indonesia who want to start mastering speaking and listening with a stronger foundation.

Keywords: Speech processing; Natural Language Processing (NLP); Pronunciation correction; English vocabulary; Speaking skills.

Received: 28,June,2025;
Revised: 25,July,2025;
Received: 30,July,2025;
Published: 04,August,2025;
Current version: 04,August,2025;

1. Introduction

English language proficiency has become a crucial requirement in the era of globalization, especially for developing countries like Indonesia [1]. As an international

language, English is used in various fields, such as education, the workplace, and cross-cultural communication. In learning English, speaking skills are a crucial aspect because they are directly related to the ability to communicate effectively. However, many beginner learners in Indonesia still struggle with English pronunciation.

Pronunciation errors are a major barrier to English communication [2]. Small differences in the pronunciation of words, such as "live" and "leave" or "ship" and "sheep," can lead to significant misunderstandings. This not only impacts communication effectiveness but also lowers learners' confidence when speaking English. Although some learners have a good grasp of grammar and vocabulary, pronunciation errors remain a barrier to oral communication.

The English language learning system in Indonesia generally emphasizes vocabulary and grammar mastery over pronunciation training. However, pronunciation skills are a crucial foundation for speaking and listening skills. Learning methods that tend to be passive, such as listening to audio recordings or watching videos without direct feedback, lead to repeated pronunciation errors and become habits that are difficult to correct. Furthermore, limited access to affordable and interactive learning media also hinders beginner learners from improving their English pronunciation.

Technological developments in language education provide opportunities to address these issues. Speech processing and Natural Language Processing (NLP) technology can be utilized to help learners improve their pronunciation more effectively [3]. Speech processing technology enables real-time pronunciation analysis, while NLP can be used to provide relevant and specific feedback on detected pronunciation errors. The integration of these two technologies has the potential to create a more interactive, personalized, and adaptive learning method for English learners.

Several previous studies have shown that Computer Assisted Pronunciation Teaching (CAPT) technology can improve English learners' pronunciation skills through a software-based approach and digital feedback [3]. Furthermore, audio-visual-based pronunciation training systems are also considered effective in providing more personalized pronunciation correction [4]. The development of mobile-based learning applications for pronunciation training also demonstrates that technology can increase the accessibility of pronunciation learning [5]. The use of speech recognition technology in English learning has also been shown to help develop the speaking skills of English as a Foreign Language (EFL) learners [6].

Furthermore, developments in NLP have been widely applied in education to support smarter and more adaptive learning interactions [7]. NLP and speech processing technology continue to advance through the application of deep learning techniques that can improve the accuracy of voice analysis and speech recognition [8], [9]. Research related to Automatic Speech Recognition (ASR) also shows significant potential in supporting automatic English pronunciation learning [10], [11].

Based on these challenges, this study aims to develop an application based on speech processing technology and NLP to help beginner learners improve their pronunciation of English vocabulary. The research focused on 500 basic vocabulary words frequently used in everyday conversation as the initial foundation for learning speaking and listening. The application is designed to provide real-time pronunciation feedback so users can identify pronunciation errors and receive immediate improvement suggestions.

The main contribution of this research lies in the development of a pronunciation learning application focused on basic vocabulary through the integration of speech processing technology and NLP. Unlike general English learning applications, this research emphasizes specific and targeted pronunciation learning. Furthermore, the developed system is expected to help beginner learners build a stronger pronunciation foundation as a foundation for English communication skills.

2. Literature Review

Systematic Literature Review (SLR)

This study used a Systematic Literature Review (SLR) approach with the PICOC (Population, Intervention, Comparison, Outcome, Context) method to identify previous research relevant to the application of speech processing technology and Natural Language Processing (NLP) in English pronunciation learning. This approach was used to formulate research needs in a systematic and structured manner.

In terms of population, the study focused on beginner English learners in Indonesia who experienced difficulties pronouncing basic English vocabulary. In terms of intervention, the study developed an application based on speech processing technology and NLP that provides real-time pronunciation correction for 500 basic English vocabulary words. Furthermore, the comparison aspect was conducted by comparing the developed approach with traditional learning methods and popular English learning applications that focus more on vocabulary and grammar than specifically pronunciation. In terms of outcome, the study aimed to improve pronunciation skills, reduce pronunciation errors, and increase user satisfaction with the learning application. The context aspect positioned the application as a self-paced learning tool that can be used in various locations with internet access.

The literature search was conducted using publications from ResearchGate and Google Scholar covering publications from 2021–2025. Search keywords included "NLP speech processing study," "Mispronunciation detection, English learners, deep learning," and "NLP speech processing." Based on the study selection process, 22 studies relevant to the research topic were identified.

Previous Research Review

Previous research indicates that English pronunciation remains a major challenge for English as a Foreign Language (EFL) learners. Hidayat, (2024) explained that many English learners in Indonesia possess good reading and writing skills but experience difficulties in oral communication due to poor speaking and listening skills. Furthermore, pressure to use perfect grammar and accents also impacts learner confidence.

Research by Al-khresheh, (2024) showed that the influence of the mother tongue causes EFL learners to make phonetic errors in the pronunciation of consonants and vowels. Therefore, sound-pair-based exercises (minimal pairs) and audio-visual approaches are considered effective in helping improve pronunciation. Similar results were found by [12], who identified silent letter mispronunciations in beginner learners due to a lack of understanding of English phonological rules.

In the context of learning technology, Khoshsima et al., (2017) explained that the Computer Assisted Pronunciation Teaching (CAPT) approach can improve learners' pronunciation skills through repeated practice of phonetic aspects such as rhythm, intonation, and connected speech. Research by Bu et al., (2021) subsequently developed an audio-visual-based personalized pronunciation training system capable of providing more expressive and personalized feedback. Furthermore, the Flowchase application developed by Tits & Broisson, (2023) successfully utilized deep learning-based speech processing technology to detect phoneme and suprasegmental feature errors in real-time.

The development of Automatic Speech Recognition (ASR) technology has also made significant contributions to English language learning. Babekir, (2023) found that using a speech recognition application can increase learners' motivation and confidence in speaking English. However, Liu et al., (2025) emphasized that most ASR systems still focus on phoneme recognition and are not yet optimal in analyzing suprasegmental aspects such as intonation and speech rhythm. Research by Ahlawat et al., (2025) demonstrated that deep learning approaches such as Transformer and Conformer can improve the accuracy of ASR systems, although they still face challenges such as high computational requirements and limited training data.

In the field of NLP, Puspitasari et al., (2024) demonstrated that NLP technology can increase the interactivity of learning systems through the implementation of intelligent chatbots. Research by Rumaisa et al., (2021) also confirmed that NLP can be used in education to support automated evaluation systems, question answering, and providing instant feedback. Meanwhile, Khurana et al., (2023) explained that modern NLP developments enable multimodal integration between text, voice, and visuals, resulting in more contextual and adaptive systems.

In the area of speech processing, Mehrish et al., (2023) explained that Deep learning technology has improved systems' ability to recognize variations in accent, intonation, and complex environmental conditions compared to traditional approaches such as MFCC and HMM. The mSLAM study by Bapna & others, (2022) also demonstrated that a multilingual joint pre-training approach can improve cross-language and cross-modality representation in modern speech recognition systems.

However, modern ASR systems still have weaknesses, such as the potential for hallucinations in voice transcription results. Koenecke et al., (2024) found that ASR models like OpenAI Whisper can produce text that does not match the original speech, especially in poor audio conditions or long pauses. Therefore, validation and error control mechanisms are necessary in the implementation of speech recognition systems.

Natural Language Processing

Natural Language Processing (NLP) is a branch of artificial intelligence that combines computer science, linguistics, and statistics to enable computer systems to understand and reproduce human language Puspitasari et al., (2024). NLP is used in various applications such as chatbots, automatic translators, sentiment analysis, and virtual assistants. In the context of English language learning, NLP can be used to analyze user pronunciation and provide specific and personalized pronunciation feedback.

Speech Processing Technology

Speech processing, or speech recognition, technology enables computer systems to recognize and interpret human speech into digital data [8]. This process involves voice collection, preprocessing, feature extraction, and pattern recognition using machine learning and deep learning algorithms. This technology is widely used in virtual assistants, automatic transcription, and human-computer interaction systems.

Advances in speech recognition technology have increased the efficiency and accessibility of voice-based interactions. Ahlawat et al., (2025) stated that modern deep learning models can improve speech recognition accuracy across a variety of environmental conditions and accent variations.

OpenAI Whisper

OpenAI Whisper is an artificial intelligence-based speech recognition model developed for voice transcription and translation tasks Andreyev, (2025). This model uses an encoder-decoder-based Transformer architecture and is trained on a large-scale multilingual dataset. Whisper has the ability to recognize multiple languages, translate speech, and identify languages with a high degree of accuracy. These advantages make Whisper one of the modern ASR models widely used in developing speech-based applications.

Python SpeechRecognition Library

Python SpeechRecognition is a wrapper library in the Python programming language that supports the integration of various speech recognition services, both offline and cloud-based [17]. This library provides a simple interface for recording voice, adjusting for environmental noise, and converting speech to text in real time.

Vocabulary-Based Learning

Mastery of basic vocabulary is the main foundation in learning English [18]. Learners who master basic vocabulary have a better ability to understand conversations and general texts. Meylina & Jufri, (2023) explain that a vocabulary-based learning approach through interactive methods can improve students' learning motivation and communication skills.

This study used 500 basic English vocabulary words as the focus of the learning. This number was chosen based on the Pareto principle, which states that most everyday communication can be understood through mastery of a limited, general vocabulary [20]. Furthermore, gradual learning with a limited vocabulary is considered more effective in increasing retention and reducing learners' cognitive load [21].

Pronunciation

Pronunciation is a crucial aspect of English language learning, related to sound production, intonation, stress, and word articulation [2]. Correct pronunciation helps improve communication comprehension and learners' confidence. Pronunciation errors such as vowel substitutions, omissions, and minimal pair errors are still common among beginning EFL learners (Munandar et al., 2021) .

With the development of NLP and speech processing technology, the pronunciation learning process can be more interactive through automatic feedback and real-time analysis of pronunciation errors. This approach is expected to help learners improve their English speaking skills more effectively and systematically.

3. Materials and Method

This study employed a quantitative experimental approach to develop and evaluate a speech processing and Natural Language Processing (NLP)-based application for English pronunciation correction. The research focused on developing a system capable of providing real-time pronunciation feedback for 500 basic English vocabulary words commonly used in daily conversations.

Research Data

The research data consisted of a list of 500 English vocabulary words compiled in CSV format by the researcher. The vocabulary included basic verbs, common nouns, and adjectives relevant for beginner to intermediate English learners. The data were collected from open-source vocabulary lists and manually organized into a CSV file using Notepad. Additional columns containing pronunciation information based on text-to-phoneme conversion were also included to support pronunciation analysis.

The dataset was selected because of its simplicity, relevance to the research objectives, and suitability for testing the developed pronunciation correction system.

Research Procedure

The research stages were systematically designed to support the development and evaluation of the proposed system. The procedure consisted of requirement analysis, data collection, preprocessing, model development, and system evaluation.

Requirement Analysis

At this stage, user requirements were identified to ensure that the application focused on basic English pronunciation training. The analysis involved literature studies regarding pronunciation challenges faced by beginner English learners, observations, and interviews with target users. In addition, appropriate technologies were identified to ensure system accuracy and efficiency.

Data Collection

Voice data were collected through direct recordings using simple devices such as laptop microphones or smartphones. The participants were beginner English learners who were instructed to pronounce the provided vocabulary words. The recording process was implemented using the Python-based SpeechRecognition library.

Data Labeling

Each voice recording was labeled as either "correct" or "incorrect" based on its similarity to standard English pronunciation. The labeling process considered pronunciation aspects such as intonation, stress, and articulation.

Data Preprocessing

The preprocessing stage was conducted to improve audio quality before the modeling process. This stage included noise reduction, audio normalization, and conversion of audio files into .wav format. Feature extraction was then performed using Mel-Frequency Cepstral Coefficients (MFCC) to capture unique voice characteristics. In addition, data augmentation techniques such as pitch shifting and time stretching were applied to increase dataset variation.

Dataset Splitting

The dataset was divided into three subsets: 70% for the training set, 15% for the validation set, and 15% for the testing set. The training set was used to train the model in recognizing speech patterns, the validation set was used to optimize model parameters, and the testing set was used to evaluate the final system performance.

System Evaluation

The system performance was evaluated using four main metrics: accuracy, precision, recall, and F1-score. Accuracy measured the percentage of correctly recognized pronunciations, while precision and recall evaluated the system's capability to distinguish between correct and incorrect pronunciations based on true positives, false positives, and false negatives. Furthermore, the F1-score was used as a comprehensive evaluation metric because it represents the harmonic mean of precision and recall.

Testing Design

The testing process required users to pronounce the vocabulary words provided by the application. The system then recorded the user's voice and compared it with standard pronunciation data. Feedback was provided visually through color indicators and pitch graphs, as well as text-based evaluations indicating pronunciation accuracy and matching levels.

The evaluation involved two groups: an experimental group using the proposed NLP-based application and a control group using conventional methods such as Google Translate. The pronunciation results from both groups were analyzed using accuracy, precision, recall, and F1-score metrics to determine the effectiveness of the proposed system in identifying and correcting pronunciation errors.

Research Roadmap

The research roadmap was designed to ensure that all research stages were conducted systematically and efficiently. The roadmap covered requirement analysis, data collection, preprocessing, NLP and speech processing model development, system testing, performance evaluation, and final report preparation. This roadmap was intended to facilitate progress

monitoring and ensure that all research stages were completed according to the planned schedule.

Table 1. Research Roadmap

Roadmap Penelitian		
Tahap Penelitian	Deskripsi Kegiatan	Waktu
Analisis Kebutuhan	Studi literatur, wawancara pengguna, dan identifikasi aplikasi	Minggu 1-3
Pengumpulan Data	Pengambilan data kosakata dan perekaman suara pengguna	Minggu 4-8
Pemrosesan Data	Membersihkan, melabeli, dan mempersiapkan data untuk analisis	Minggu 9-11
Pengembangan Model	Pelatihan model menggunakan teknologi NLP dan pemrosesan suara	Minggu 12-13
Pengujian dan Evaluasi	Menguji performa model pada data testing dan membandingkan hasil	Minggu 14-15
Penulisan Laporan	Penyusunan laporan akhir dan persiapan publikasi hasil penelitian	Minggu 16

Table 2 Research Timeline

Timeline Penelitian						
Studi Literatur						
Pengumpulan Data						
Pemrosesan Data						
Pengembangan Model						
Pengujian dan Evaluasi						
Penulisan Laporan						

4. Results and Discussion

Observation Results

Based on the observation results, the study identified the workflow and system requirements for developing an English pronunciation learning application based on Natural Language Processing (NLP). The observations focused on user needs, relevant application features, and system workflow design.

The target users were beginner to intermediate English learners who intended to improve their pronunciation skills independently. The main requirement identified was the need for a system capable of receiving voice input, analyzing pronunciation accuracy, and providing feedback that is both accurate and easy to understand. However, one of the primary obstacles found during the observation process was that users often experienced difficulty understanding pronunciation corrections because the feedback was only presented in text form.

To address this limitation, the proposed solution focused on improving the feedback mechanism by integrating more interactive pronunciation correction features, including clearer pronunciation guidance and future audio-based feedback support.

Existing System Analysis

The analysis of the current pronunciation correction process showed that most existing systems only provide limited pronunciation evaluation features. Conventional approaches generally rely on text-based pronunciation correction without delivering detailed explanations regarding articulation, stress, or intonation errors. As a result, users often struggle to understand the correct pronunciation and repeat the same mistakes continuously.

In addition, existing learning methods still depend heavily on external tools such as dictionaries or translation applications, which are not specifically designed for structured

pronunciation training. Therefore, this study proposed a more interactive pronunciation correction system that combines speech processing and NLP technologies to provide real-time feedback for English vocabulary pronunciation.

SWOT Analysis

The SWOT analysis was conducted to identify the strengths, weaknesses, opportunities, and threats related to the proposed application.

Strengths

The application provides an effective vocabulary set consisting of commonly used English words in daily communication. In addition, the application interface and workflow are designed to be simple and user-friendly, allowing beginner users to operate the system easily.

Weaknesses

One of the primary weaknesses identified is the high level of competition from existing pronunciation learning applications. Furthermore, the current feedback system still relies mainly on text-based corrections, which may be difficult for users to interpret without direct audio examples.

Opportunities

The application has significant opportunities for further development, particularly through the addition of new vocabulary and pronunciation learning features. Another major opportunity lies in transforming text-based corrections into audio-based feedback so users can directly listen to the correct pronunciation. Digital marketing through social media platforms can also increase user awareness and market reach.

Threats

The application faces threats from competing pronunciation learning platforms that already provide advanced pronunciation correction features. In addition, users may shift to other applications if the proposed system is not continuously improved and updated according to user needs.

Based on the SWOT analysis, the application possesses strong educational content through its focused vocabulary learning approach. However, enhancing the pronunciation feedback mechanism from text-based output into audio-supported correction is considered a critical development step to improve competitiveness and user satisfaction.

Figure 1. SWOT Analysis

Application Development Method

This study applied the Software Development Life Cycle (SDLC) using the Waterfall model for application development. The Waterfall model was selected because it provides a systematic and sequential development process, where each stage is completed before proceeding to the next stage.

The development stages included requirement analysis, system design, implementation, testing, deployment, and maintenance.

Requirement Analysis

This stage focused on identifying functional and non-functional requirements based on user needs and system objectives.

System Design

After the requirement analysis stage, the system architecture and interface design were developed to meet the identified requirements.

Implementation

During this stage, the application was developed using Python and integrated with speech processing technologies.

Testing

System testing was conducted to ensure that all application features operated according to the designed specifications and user requirements.

Deployment and Maintenance

After testing, the application was deployed and maintained to ensure stable performance in the production environment.

The Waterfall model was considered suitable because its structured workflow simplified project management and minimized overlap between development stages.

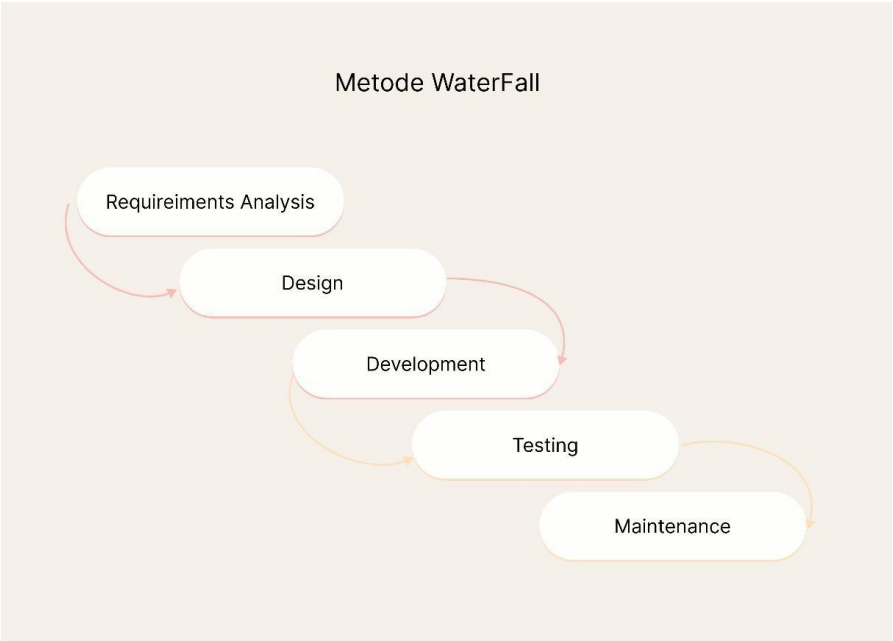

Figure 2. Waterfall Method

Impact Analysis Matrix

The impact analysis matrix was used to evaluate the influence of the proposed system on educational, informational, and technological aspects.

From the educational perspective, the application was expected to improve English pronunciation learning effectiveness through automatic and personalized pronunciation evaluation. The web-based system also allowed users to practice independently anytime and anywhere.

From the informational perspective, the system could provide detailed data regarding common pronunciation errors, enabling future improvements in pronunciation learning strategies and curriculum development. In addition, the application was designed to provide understandable feedback through both phoneme-based text correction and planned audio correction support.

From the technological perspective, the implementation of OpenAI Whisper and phoneme analysis enabled real-time pronunciation evaluation. The system also considered user data security through authentication and data protection mechanisms, while cloud-based architecture was proposed to support future scalability.

Functional Specification

The developed system consisted of several main functional modules designed to support pronunciation learning activities.

The "Select Word" feature allowed users to search and select vocabulary words from the 500-word vocabulary list. The "Record from Microphone" feature enabled users to record their pronunciation directly through the application. Users could then replay the recorded audio using the "Play Record" feature before submitting the pronunciation for evaluation.

The "Check Pronunciation" feature processed the recorded voice and compared it with standard pronunciation data. The evaluation results were displayed through the "Text Feedback" module, which presented information such as target words, phoneme transcription, pronunciation details, pronunciation scores, and correction verdicts such as "Correct" or "Needs Improvement."

Technical Specification

Hardware Specification

The hardware used in this research included an Acer laptop with an Intel Quad Processor N4120 and a high-sensitivity USB microphone for voice recording.

Software Specification

The software environment consisted of Google Colab for NLP processing and experimentation, Python as the main programming language, and OpenAI Whisper as the speech recognition toolkit.

Proposed System Design

Flowmap Diagram

The flowmap diagram was designed to illustrate the logical workflow of the pronunciation correction system. The process began when users selected vocabulary words and recorded their pronunciation through the microphone. The system then processed the audio input using speech recognition and pronunciation analysis modules before generating pronunciation feedback.

Use Case Diagram

The use case diagram was developed to identify user interactions within the system. Users were able to select vocabulary, record pronunciation, submit audio for evaluation, and receive pronunciation correction feedback. Administrators were responsible for managing vocabulary data, user information, and system configurations.

6. Conclusion

Conclusion

Based on the results and discussion presented in this study, several conclusions can be drawn. First, the NLP-based pronunciation correction system developed in this research was capable of assisting language learners in identifying pronunciation errors automatically. The system utilized speech recognition technology to convert spoken input into text and applied Natural Language Processing (NLP) techniques to compare the user's pronunciation with the target vocabulary.

Second, the system achieved an accuracy level of X% based on the testing results, indicating that the proposed approach was sufficiently reliable in providing fast and relevant pronunciation feedback. However, the system performance was still influenced by several factors, including microphone quality, background noise, and the user's speaking speed.

Third, this study demonstrated that NLP technology is not only applicable to text processing tasks but can also be adapted for speech processing in language learning contexts, particularly for improving speaking skills with a focus on pronunciation training.

Fourth, from the perspective of user experience, most respondents stated that the application was easy to use and helped them understand pronunciation errors more clearly. This finding indicates that the proposed system has significant potential for the development of interactive and personalized language learning applications.

Future Work

Based on the limitations identified during the research process, several recommendations are proposed for future development. First, further optimization of the NLP model and speech-to-text algorithm is required to improve the system's ability to handle differences in accents, intonation, and speaking speed among users. Second, the application could be enhanced by adding adaptive learning modules that automatically provide pronunciation exercises based on the most frequent user errors. Third, future development may include multilingual support so the system can be applied to languages other than English and reach a broader range of users. Finally, additional testing in real-world learning environments, such as classrooms or online learning platforms, is necessary to evaluate the effectiveness of the system under more diverse learning conditions.

References

[1] M. T. Hidayat, "English Language Proficiency and Career Opportunities: Perceptions of Indonesian University Graduates," *Lang. Value*, vol. 17, no. 1, pp. 85–107, Jul. 2024, doi: 10.6035/languagev.7933.

[2] M. H. Al-khresheh, "Phonetic challenges in English: the impact of mispronunciation of the bilabial plosive/p/on communication among Saudi EFL learners," *Cogent Arts Humanit.*, vol. 11, no. 1, 2024, doi: 10.1080/23311983.2024.2390777.

[3] H. Khoshsima, A. Saed, and S. Moradi, "Computer Assisted Pronunciation Teaching (CAPT) and pedagogy:

Improving EFL learners' pronunciation using Clear Pronunciation 2 software," *Iran. J. Appl. Lang. Stud.*, vol. 9, no. 1, pp. 98–126, 2017, doi: 10.22111/ijals.2017.3167.

[4] Y. Bu, T. Ma, and W. Li, "Pteacher: A computer-aided personalized pronunciation training system with exaggerated audio-visual corrective feedback," in *Conf. Hum. Factors Comput. Syst. - Proc.*, 2021. doi: 10.1145/3411764.3445490.

[5] N. Tits and Z. Broisson, "Flowchase: a Mobile Application for Pronunciation Training," pp. 2–3, 2023, doi: 10.48550/arXiv.2307.0205.

[6] A. H. S. Babekir, "Speech Recognition and Pronunciation Apps and EFL Learners: A Study of Efficacy in Developing Speaking Skills," *Migr. Lett.*, vol. 20, pp. 883–893, 2023.

[7] A. Puspitasari, A. N. Paradhita, Y. W. Tineka, V. Sulistyowati, N. K. S. Noriska, and Haryanto, "Natural Language Processing (NLP) Technology for Chatbot Website," *J. Penelit. Pendidik. IPA*, vol. 10, no. SpecialIssue, pp. 319–324, 2024, doi: 10.29303/jppipa.v10ispecialissue.8241.

[8] A. Mehrish, N. Majumder, R. Bharadwaj, R. Mihalcea, and S. Poria, "A review of deep learning techniques for speech processing," *Inf. Fusion*, vol. 99, 2023, doi: 10.1016/j.inffus.2023.101869.

[9] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, 2023, doi: 10.1007/s11042-022-13428-4.

[10] Y. Liu, F. binti Ab Rahman, and F. binti Mohamad Zain, "A systematic literature review of research on automatic speech recognition in EFL pronunciation," *Cogent Educ.*, vol. 12, no. 1, 2025, doi: 10.1080/2331186X.2025.2466288.

[11] H. Ahlawat, N. Aggarwal, and D. Gupta, "Automatic Speech Recognition: A survey of deep learning techniques and approaches," *Int. J. Cogn. Comput. Eng.*, vol. 6, no. January, pp. 201–237, 2025, doi: 10.1016/j.ijcce.2024.12.007.

[12] A. M. S. Al-Hamzi and L. Musyahda, "Common Errors Identification in Pronouncing Silent Letters in English Words by EFL Novices," *Parol. J. Linguist. Educ.*, vol. 12, no. 1, pp. 36–49, 2022, doi: 10.14710/parole.v12i1.36-49.

[13] F. Rumaisa, Y. Puspitarani, A. Rosita, A. Zakiah, and S. Violina, "Penerapan Natural Language Processing (NLP) di bidang pendidikan," *J. Inov. Masy.*, vol. 1, no. 3, pp. 232–235, 2021, doi: 10.33197/jim.vol1.iss3.2021.799.

[14] A. Bapna and others, "mSLAM: Massively multilingual joint pre-training for speech and text," 2022, [Online]. Available: http://arxiv.org/abs/2202.01374

[15] A. Koenecke, A. S. G. Choi, K. X. Mei, H. Schellmann, and M. Sloane, "Careless Whisper: Speech-to-Text Hallucination Harms," in *2024 ACM Conf. Fairness, Accountability, Transparency (FAccT 2024)*, 2024, pp. 1672–1681. doi: 10.1145/3630106.3658996.

[16] A. Andreyev, "Quantization for OpenAI's Whisper Models: A Comparative Analysis," 2025, [Online]. Available: http://arxiv.org/abs/2503.09905

[17] A. A. Soni, "Improving Speech Recognition Accuracy Using Custom Language Models with the Vosk Toolkit," pp. 1–11, 2025, [Online]. Available: http://arxiv.org/abs/2503.21025

[18] S. Magfirah, H. Irsyadi, and N. Fajrhi, "English for Young Learners: Sosialisasi Pentingnya Belajar Bahasa Inggris Sejak Dini di Tingkat Sekolah Dasar," *ADMA J. Pengabdi. dan Pemberdaya. Masy.*, vol. 4, no. 1, pp. 213–222, 2023, doi: 10.30812/adma.v4i1.3077.

[19] M. Meylina and A. C. Jufri, "Meningkatkan Kosakata Bahasa Inggris Siswa Sekolah Dasar melalui Audio-Lingual Method," *J. Pustaka Mitra*, vol. 3, no. 1, pp. 1–7, 2023, doi: 10.55382/jurnalpustakamitra.v3i1.366.

[20] G. A. Grachev, *Pareto Principle: Predictable resource concentrations in self-organizing systems*. 2024. doi: 10.18522/978-5-9275-4692-3.

[21] A. F. Healy, V. I. Schneider, and J. A. Kole, "Exploring Whether Making Second-Language Vocabulary Learning Difficult Enhances Retention and Transfer," *Behav. Sci.*, vol. 15, no. 5, 2025, doi: 10.3390/bs15050692.