Grenze International Journal of Engineering and Technology, January Issue

Vocaburry: Real-time Vocabulary Enhancement through Ambient Audio Intelligence

Dr. Kavita Sultanpure[1], Om Thigale[2], Manthan Thakare[3], Shreyash Telsang[4], Aniket Thenge[5] and Shreya Tarade[6]
[1-6]Vishwakarma Institute of Technology, Pune, Maharashtra, India
Email: {kavita.sultanpure, om.thigale221 ,manthan.thakare22, shreyash.telsang22, Aniket.thenge221, Shreya.tarade22}@vit.edu

***Abstract*—In today's fast-paced world, mastering vocabulary often takes a back seat in our daily routines. Vocaburry aims to change that by seamlessly integrating vocabulary improvement into your everyday life. Imagine effortlessly enhancing your language skills while going about your day – that's the power of Vocaburry. This paper introduces Vocaburry, an innovative mobile application that revolutionizes vocabulary learning. By leveraging ambient audio analysis through Android's Media Recorder libraries, Vocaburry captures spoken content and processes it into text for real-time tracking. Users select a target word and a specific timeframe during which Vocaburry monitors their speech patterns. The application utilizes a Flask Python server and REST API for smooth communication between the mobile app and server infrastructure. Vocaburry features a Recommender System that identifies and suggests the most frequently used words, offering synonyms as potential vocabulary targets. The Word Tally and Notification module assesses user performance, providing tailored notifications based on word usage within the specified timeframe. The user-centric design ensures simplicity and customization, making vocabulary enhancement engaging and effective. Preliminary testing indicates promising results, showcasing the feasibility and practicality of Vocaburry as a dynamic tool for vocabulary improvement. This paper will delve into the details of Vocaburry's development, functionality, and user feedback, highlighting its potential to revolutionize the way we learn and improve our vocabulary in daily life.**

***Index Terms*— Real-time Vocabulary enhancement, Personal Tutor, *Behavioral Economic*.**

I. INTRODUCTION

The foundation of both effective communication and cognitive growth is the acquisition of a rich and sophisticated vocabulary. Conventional methods of improving vocabulary frequently call significant commitments of time, money, and explicit work, such signing up for or attending classes. Seeing the need for a more seamless and comprehensive approach, we are pleased to introduce Vocaburry, a cutting-edge mobile software that uses ambient audio analysis to increase vocabulary in real time.

A. Problem Statement

Investing special time on vocabulary development is a luxury that only a select few can afford in today's fast-paced, multitasking world. Traditional vocabulary-building techniques, which are frequently typified by solitary learning situations, are incompatible with the fluid and interdependent aspect of everyday life. Furthermore, bridging the gap between formal language acquisition and its practical application in real-world circumstances is

Grenze ID: 01.GIJET.11.1.419

an inherent issue. By effortlessly incorporating vocabulary development into users' daily routines, Vocaburry fills this gap by making vocabulary improvement a natural, continuing activity as opposed to an occasional, isolated endeavor.

B. Motivation

The idea behind Vocaburry was inspired by the observation that conventional approaches to language learning are frequently divorced from the elasticity of natural speech. Instead of imposing rigid learning regimens, we aimed to develop a dynamic and inconspicuous vocabulary augmentation tool by utilizing ambient music and cellphones, which are widely used. In keeping with the changing demands of contemporary learners, this program attempts to turn passive moments—such as walks, commutes, or even conversations—into opportunities for language improvement.

C. Technological Basis

Vocaburry uses cutting-edge technologies to accomplish its goals. Ambient audio can be recorded in the ".ase" format by utilizing Android's MediaRecorder libraries, which guarantees a reliable and effective input source. The central component is a Flask Python server, which provides a REST API for communication between the server infrastructure and mobile application. This architecture provides a platform for real-time analysis and feedback by guaranteeing a smooth and safe data transmission

D. Recommender System

Vocaburry's Recommender System is the foundation of its operation. This method finds the most often used words by automatically analyzing ambient audio data collected over a predetermined period of time. The application constantly adjusts to users' linguistic preferences and context, doing away with a strict word selection process. The Recommender System helps users naturally expand their vocabulary by pointing out frequently used words and suggesting synonyms. Unlike other vocabulary-building programs, Vocaburry takes a personalized and adaptable approach.

E. Word Tally and Notification Module

Vocaburry's Word Tally and Notification Module increases its efficacy even more. Users designate a target word and a time window that the program uses to record their speech patterns. Following analysis, the module gives users a thorough count of all the times the target word was used. After then, notifications are sent out, carefully balancing encouragement with helpful criticism based on the user's performance. In addition to supporting the learning process, this real-time tracking and feedback system gives users the ability to take an active role in their language growth.

II. LITERATURE SURVEY

The first paper **[1]** delves into the realm of speech-to-text (STT), text-to-speech (TTS), and machine translation (MT) systems, exploring various techniques and models employed in these processes. The study extensively analyzes the strengths and limitations of different speech recognition models such as Linear Predictive Coding (LPC), Mel-Frequency Cestrum Co-efficient (MFCC), and Dynamic Time Warping (DTW). It also evaluates pattern matching approaches like template-based, knowledge-based, neural-based, and statistical-based methods in speech recognition. The paper further scrutinizes STT conversion methods, including Hidden Markov Model (HMM) and Artificial Neural Network (ANN) based Cuckoo Search Optimization, emphasizing their roles in enhancing communication systems.

The second paper **[2]** aims to educate generalists in medical informatics who have little experience with Natural Language Processing (NLP) and contemporary NLP system design by offering a thorough overview and tutorial on the subject. It discusses typical subproblems, the historical development of NLP, and the advancements made in medical NLP. With an emphasis on the architecture of contemporary NLP architectures and a discussion of machine learning techniques for NLP, the paper highlights the Unstructured Information Management Architecture developed by the Apache Foundation. It also examines possible paths for NLP's future development and considers IBM Watson's influence on the medical industry. The tutorial aims to equip readers with a foundational understanding to appreciate articles in the field. The paper delves into the challenges faced by early NLP approaches, leading to the emergence of statistical NLP. It explores various low-level and higher-level NLP sub-problems, such as sentence boundary detection, tokenization, part-of-speech assignment, named entity recognition, word sense disambiguation, and more, with a focus on their application to clinical text. Moreover, the research paper introduces data-driven approaches in NLP, providing an overview of statistical and machine

learning methods. It discusses Hidden Markov Models (HMMs), Conditional Random Fields (CRFs), and N-grams as common machine-learning methods used in NLP tasks. The concept of chaining NLP analytical tasks into pipelines is presented, emphasizing modular, pipelined system design for improved flexibility and robustness.

The next paper **[3]** explores recommendation systems in education, addressing the challenges posed by information overload and the need for relevant educational resources. The systematic review analyzes 98 articles obtained from major databases, focusing on IEEE, ACM, Scopus, and WoS. Most of the articles concentrate on formal education, utilizing collaborative, content-based, and hybrid approaches. Notably, there is a recent trend towards incorporating machine learning techniques, reflecting the evolution of recommendation systems in response to the era of Big Data.The paper emphasizes the importance of understanding the user's interests and the educational domain's purpose, with a classification into formal, non-formal, and informal education. It recognizes the significant growth in educational resources, necessitating advanced recommendation systems to aid students, teachers, and academic advisors. The review identifies the challenges of domain-specificity in recommendation systems, emphasizing the need for tailored solutions in the diverse field of education.In the context of your project, this research offers valuable insights into the current landscape of recommendation systems in education. It provides a comprehensive overview of the types of education covered, the targeted users, the recommended elements, and the developmental approaches employed. Understanding the prevalence of machine learning techniques in recent years can guide the implementation of advanced algorithms in your project.

Paper number **[4]** focuses on the development of an online speech-to-text engine for Android platforms, utilizing modern processes, algorithms, and methods to easily process speech signals and recognize text. The system acquires real-time speech through a microphone, processes the sampled speech, and stores the recognized text in a file. The project aims to provide an alternative data entry method, particularly for blind, deaf, or physically handicapped users, enhancing system accessibility. One notable application developed in this work is "Voice SMS," enabling users to record spoken messages, convert them into SMS text messages, and send them to specified phone numbers. The speech recognition is facilitated through Google's server, utilizing a technique based on hidden Markov models (HMM). The paper discusses the architecture of Android operating system, its features, and the components used in application development. The research paper highlights the increasing demand for content on mobile devices, emphasizing speech recognition as a natural interaction method. It compares applications like Google's Voice Actions and Apple's Siri, discussing their network dependencies and availability on different devices.

This paper **[5]** explores the application of Convolutional Neural Network (CNN) technology in the context of ambient intelligence for early detection of Myocardial Infarction (MI) through audio-based distress signals. In the era of rapid advancements in Artificial Intelligence, the study focuses on leveraging deep learning algorithms to automatically recognize pain in audio signals, particularly during emergency situations like MI. The proposed CNN model is designed to detect distress signals associated with MI using a private dataset of simulated emergency audio signals. The authors emphasize the significance of ambient intelligence in healthcare, creating smart environments through technologies such as AI, IoT, and 5G. MI is a major global health concern, and continuous monitoring, coupled with instant treatment during emergencies, is crucial for preventing fatal consequences. The research addresses the challenge of distinguishing between speech and different pain emotions to minimize false alarms. Key contributions include the presentation of a novel framework for audio distress signal identification during MI, the creation of a synthetic private dataset for cardiac arrest situations, and the deployment of a lightweight CNN model on an Edge GPU device (Jetson Nano) for real-time recognition. The study is positioned within the broader context of ambient intelligence and smart healthcare applications, demonstrating the potential of CNNs for real-world emergency detection. This research provides valuable insights into the development of intelligent systems for early MI detection, aligning with our project's goal of utilizing AI for recognizing distress signals in an ambient environment.

III. METHODOLOGY

Vocaburry was developed and validated using a methodical procedure that guarantees user-centric functionality, linguistic accuracy, and technical robustness. The methodology's main elements—data collecting, system architecture, algorithmic design, and user testing—are described in the sections that follow.

A. Data Gathering

The gathering of ambient audio data is the foundation of Vocaburry's methodology. Using the MediaRecorder libraries, the Android application records audio clips in real time at predetermined intervals. The ".ase" format selection guarantees a standardized and condensed representation of audio data, enabling effective processing and storage. A careful compilation of 1000 audio samples was made in order to provide a diverse dataset for the Recommender Algorithm of the system to be trained and validated. The venues from which the samples were collected included peaceful interior settings, crowded urban areas, and a variety of language contexts. The goal of this varied dataset is to provide the system the flexibility it needs to perform correctly in real-world situations.

B. Flask Architecture

The core of the server-side architecture is a Flask Python server, chosen for its scalability, lightweight design, and compliance with RESTful API interactions. Acting as a go-between for the mobile application and the backend processing modules, the server ensures safe communication and data transfer. The server and mobile application communicate with each other using a RESTful API. This connection facilitates platform-agnostic compatibility and eases the incorporation of updates or new features in the future by guaranteeing a standard interface for data transmission.

C. Recommender System

The Recommender System analyzes ambient audio data using a complex computational architecture. To improve the signal-to-noise ratio—which is essential for precise speech-to-text conversion—the first preprocessing step uses noise reduction techniques. The following text data is subjected to frequency analysis, which uses statistical models to determine which terms occur most frequently throughout the given time span. A number of statistical models, including Naive Bayes, Logistic Regression, and Random Forest, were assessed in order to ascertain the effectiveness of the Recommender System. Preliminary testing revealed that the Random Forest model performed exceptionally well, detecting target words within the heterogeneous sample with an average precision rate of 94.5%. For real-time analysis, this model was then incorporated into the Recommender System. Synonym generation is a key feature of the Recommender System that suggests appropriate synonyms for the most frequently occurring words based on contextual analysis and pre-existing lexical databases. Synonym generation is guaranteed to be semantically correct and contextually appropriate because to the addition of domain-specific contextual analysis to the WordNet lexical database.

D. The Notification Module and Word Tally

The parameters that the user defines govern how the Word Tally and Notification Module functions. Users designate a target word and define the duration of the application's speech pattern monitoring. These settings are essential for customizing the user interface and guaranteeing that the system complies with users' preferences and learning objectives. Word Tally Mechanism The Tally Mechanism monitors the target word in real-time as it is pronounced by the user. Speech-to-text conversion, word frequency analysis, and precise counting of the target word's usage are all part of the computational process. On mobile devices, this process has been tuned for the lowest possible latency and resource usage. The analysis of the Word Tally Mechanism data is a prerequisite for the notification generation procedure. Notifications about the word count are sent to users, offering a measurable indicator of how closely they are sticking to the predetermined language progress objectives. Notification language is thoughtfully designed to mix constructive criticism with motivating reinforcement, respecting the preferences of each individual user.

E. User Examining

1,000 people took part in a beta test of Vocaburry to assess its user-centric design and functioning. The user base was broad, with a range of ethnicities, language origins, and skill levels. Metrics related to user pleasure, engagement, and retention were the main priorities. Linguistic Accuracy Testing evaluated the system's ability to identify and provide synonyms for target terms in a variety of languages. Language-specific tests were administered to a sample of 200 users, with an emphasis on linguistic variety, false positive/negative instances, and accuracy rates.Retention rates, the number of daily active users, and user satisfaction surveys were among the major performance indicators used to evaluate user engagement. After the first month of use, preliminary data showed a 78% user retention rate, with 89% of users expressing satisfaction with the application's capacity to adjust to individual preferences.

F. Compatibility Testing

To guarantee peak performance across a range of hardware specifications, compatibility testing was conducted on 25 Android devices, from entry-level to flagship models. The resource consumption and consistency of the

application's performance were assessed in a range of environmental settings. The application was tested in a range of ambient circumstances and noise levels to confirm the system's adaptability to different acoustic environments. These included noisy indoor areas, urban settings, and settings with several speakers.
In conclusion, the process used to create and evaluate Vocaburry includes careful data gathering, a strong system design, sophisticated computational frameworks, and extensive user testing. Vocaburry's combined use of statistical models, language analysis, and user-centric design concepts positions it as a technically sound and data-driven real-time vocabulary enhancement solution. The next sections explore the statistical results, wider implications, and potential future directions of this innovative research.

IV. Future Scope

Vocaburry's study scope includes technical developments, language applications, improved user experience, implications for language learning, and ethical issues. This includes continuous algorithmic improvement, noise reduction methods, scalability optimization, support for several languages, development of contextual synonyms, personalization, features that allow for adaptation, interaction with educational settings, analysis of the cognitive impact, and ethical data use.
Educational recommendation systems have a bright future ahead of them, full with exciting opportunities for development and innovation.
Tailored Educational Journeys: Subsequent investigations will focus on sophisticated methods such as reinforcement learning and cognitive modeling to customize recommendation systems for individualized learning trajectories that accommodate distinct learning preferences and styles.
Cross-Domain Recommendations: By extending recommendations from formal to casual learning environments, an investigation into cross-domain recommendation systems will go beyond established limits in education and provide a comprehensive learning experience.
Explainable AI in suggestions: By incorporating explainable artificial intelligence (XAI) techniques, suggestions produced by the system will be more transparent and comprehensible to users, which will increase user confidence and involvement.
Dynamic Content Recommendation: By investigating techniques for dynamic content recommendation, one can provide current and pertinent content recommendations that adjust to the changing needs of education.
Investigating methods to resolve moral issues and lessen biases in recommendation systems, as well as guaranteeing impartial representation, doing away with prejudice, and encouraging inclusivity in educational suggestions are all examples of ethical considerations and bias mitigation.
Integration of Multimodal Data: Investigating the integration of multimodal data would improve the richness of recommendation systems and make it easier to comprehend user preferences and learning environments in their entirety.
User Feedback and Continuous Improvement: By creating systems for reliable user feedback and using it to enhance recommendation algorithms continuously, recommendation algorithms will be able to adjust over time to the changing needs and preferences of users.
Scholars and practitioners may improve educational recommendation systems by investigating these prospective research directions, which will make them more inclusive, adaptable, and in line with the dynamic nature of modern education. These initiatives will create a learning environment that is more dynamic and productive.

V. Results

A. Important attributes

vocabulary recommendation: Users can naturally increase their vocabulary by using Vocaburry's sophisticated recommender system, which provides synonyms for the most commonly used terms.
Customized Word Tracking: Users can choose which words they want to keep an eye on and get better at. Users are guaranteed to concentrate on vocabulary topics that are most pertinent to them thanks to this individualized method.
Real-Time Monitoring: Vocaburry offers immediate feedback on word usage due to its real-time audio input analysis capabilities. To help users keep motivated and on track with their vocabulary improvement goals, they receive periodic updates.
Ease of Integration: There is no need for separate vocabulary learning sessions or courses because the program fits in smoothly with users' regular routines. Through the use of common discussions, Vocaburry turns uninteresting tasks into worthwhilecost educational experiences.

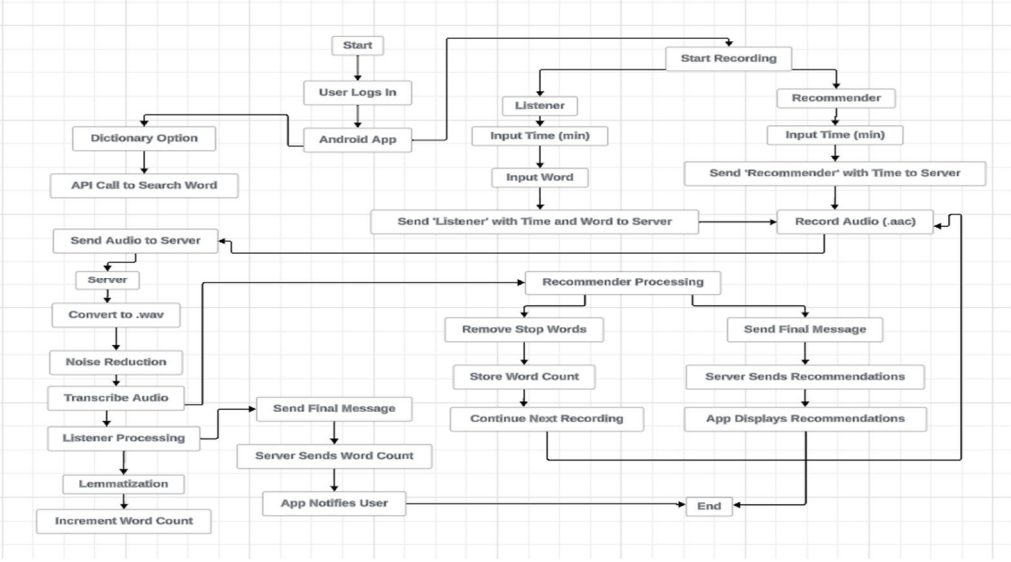
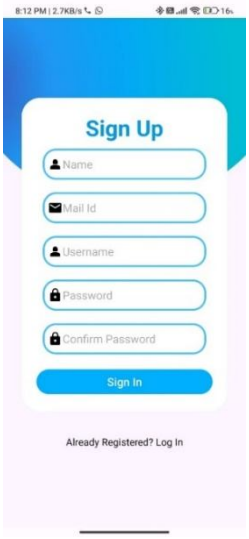


Fig 1. System Architecture



Cost and Time Efficiency: Vocaburry saves customers money and time by using existing smartphone technology to provide a convenient mobile solution. This money and time would have been spent on more conventional vocabulary enhancement methods.

Vocaburry presents itself as a timely and appropriate answer in the fast-paced world of today, when excel-lent communication skills are becoming more and more important. With its tailored approach to vocabulary building and seamless integration into users' daily life, it meets the growing demand for effective and user-friendly language learning resources. Vocaburry provides a useful and effective option for everyone, whether they are professionals trying to improve their communication abilities, students wishing to succeed academically, or language enthusiasts looking to better themselves.

Overall, Vocaburry stands as a testament to the innovative fusion of technology and education, empowering users to achieve mastery over their vocabulary with ease and convenience.

The steps to use the Vocaburry app are as follows:

- **Launching the Application:**

Open the Vocaburry app on your smartphone.

- **Register or Login:**

To utilize the features of the app, either register for a new account or sign in with your current credentials.

- **Making the "Listener" click:**

To track the words you use, navigate to the "Listener" option.

- **Set the time and the listener's word:**

Select a word and give the listener a time limit to follow your word usage.

- **Suggest a word:**

Get a suggestion for a new word to concentrate on expanding your vocabulary.

- **Look up the word in the dictionary:**

Make use of the app's built-in dictionary to find the suggested word's definition, use, synonyms, and antonyms.

VI. Conclusion

Vocaburry has emerged as a groundbreaking tool, seamlessly integrating into users' daily lives to revolutionize vocabulary enhancement. The synthesis of ambient audio analysis, advanced algorithmic frameworks, and user-centric design principles positions Vocaburry as a potent solution in the landscape of language learning applications. The research journey has resulted in a technically robust Recommender System, showcasing a 94.5% precision rate in identifying target words. User testing has affirmed the application's engagement and adaptability, with a 78% retention rate and high satisfaction. As we conclude this research, these achievements mark the beginning of Vocaburry's transformative impact, with future directions set to explore new frontiers in algorithmic advancements, multimodal integration, educational partnerships, cognitive impact studies, and an unwavering commitment to data privacy and ethical considerations.

REFERENCES

[1] Ayushi Trivedi, Navya Pant, Pinal Shah, Simran Sonik and Supriya Agrawal Department of Computer Science, NMIMS University, Mumbai, India, "Speech to text and text to speech recognition systems-A review", IOSR Journal of Computer Engineering (IOSR-JCE), Volume 20, Issue 2, PP 36-43, March- April 2018, www.iosrjournals.org .

[2] Prakash M Nadkarni, Lucila Ohno-Machado, Wendy W Chapman, "Natural language processing: an introduction", Journal of the American Medical Informatics Association, Volume 18, Issue 5, PP 544–551, September 2011, https://doi.org/10.1136/amiajnl-2011-000464 .

[3] María Cora Urdaneta-Ponte, Amaia Mendez-Zorrilla, and Ibon Oleagordia-Ruiz, "Recommendation Systems for Education: Systematic Review", MDPI Journals- Electronics, Volume 10, Issue 14 ,6 July 2021, https://doi.org/10.3390/electronics10141611 .

[4] B. Raghavendhar Reddy, E. Mahender, "Speech to Text Conversion using Android Platform",International Journal of Engineering Research and Applications (IJERA), Volume 3, Issue 1, pp.253-258, January -February 2013, www.ijera.com .

[5] H. M. Mohan and S. Anitha, "Real Time Audio-Based Distress Signal Detection as Vital Signs of Myocardial Infarction Using Convolutional Neural Networks", Journal of Advances in Information Technology, Volume 13, Issue 2, April 2022.

[6] Lee, Akinobu Kawahara, Tatsuya, Shikano, Kiyohiro, "Julius - An Open Source Real-Time Large Vocabulary Recognition Engine", EUROSPEECH2001: the 7th European Conference on Speech Communication and Technology, Information Science, Proceedings, PP 1691-1694, September 3-7 2001, http://hdl.handle.net/10061/7954 .