ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.206

Volume 8, Issue 4, April 2025

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804228

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Lip decode: Bridging the Gap between Visual and Verbal Communication

B. Navya, K. Nehadeep, V.Neha, B.Niharika, S.Nidhigna, Prof. Bhagyashri

Department of AI & ML, Malla Reddy University, Hyderabad, India

ABSTRACT: LipDecode project aims to develop an AI-powered lip-reading system that converts silent video inputs into text, bridging communication gaps in challenging auditory environments. Leveraging Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, the system extracts spatial features from video frames and analyzes temporal sequences of lip motions to generate accurate and context-aware text outputs. The architecture integrates the LipNet model for end- to-end sentence-level visual speech recognition, ensuring precision and effectiveness. Built using Python and deep learning frameworks like TensorFlow and Keras, the system employs OpenCV for video preprocessing and Streamlit for a user-friendly interface that facilitates seamless interaction. This solution addresses challenges faced by individuals with hearing impairments and noisy settings, offering transformative applications in accessibility, security, and healthcare. By combining advanced machine learning models with intuitive design, the project creates an inclusive platform that enhances communication and fosters innovation in human-machine interaction.is an innovative AI-powered lip-reading system that translates silent video input into coherent, context-aware text output. The project is designed to function effectively in environments where audio communication is difficult or impossibl esuch as in loud public spaces, during silent surveillance, or for individual ls with hearing impairments. By leveraging the power of deep learning, computer vision, and real-time interactivity, LipDecode offers a cutting-edge solution to bridge the auditory communication gap.

KEYWORDS:AI-poweredlip-reading,

I. INTRODUCTION

1. Problem Statement

Effective communication is the foundation of human interaction, crucial for social cohesion, accessinformation, and participating in various aspects of life. However, for individuals with hearing impairments, those facing language barriers, or those in noisy environments, verbal communication becomes difficult.

Sign language and text can be slow or inefficient, and not everyone is familiar with these alternatives. Moreover, in dynamic or fast-paced environments, these methods might not be the most effective for immediate communication. Existing speech recognition systems and lip-reading technologies, while promising, are not yet reliable enough to bridge communication gaps seamlessly.

2. Objectives of the Project

The primary objective is to develop an AI- powered lip-reading system that can convert silent video inputs into accurate text. This system aims to bridge communication gaps for individuals in noisy environments or those with hearing impairments by providing a reliable method of understanding speech through visual cues. To achieve this, the system will utilize Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) to extract both spatial and temporal features from lip movements. These features will allow the system to effectively recognize speech patterns from video data. The project will implement the LipNet model, a state-of-the-art deep learning model specifically designed for sentence-level visual speech recognition. By using this model, the system can accurately interpret full sentences, enabling it to provide context-aware translations of lip movements into text. This level of accuracy is critical for making the system practical and useful in real- world scenarios Additionally, the system will enhance accessibility in a variety of domains, including healthcare, security, and education. In healthcare, it can assist patients with speech impairments or those in recovery, enabling more effective communication. In security, it can be

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804228

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

used to transcribe lip movements in environments where audio recording is not possible. In education, it can support students with hearing impairments or language barriers, fostering more inclusive learning environments.

3. Scope and Significance of Study

The scope of the LipDecode project encompasses the development, implementation, and evaluation of an AI-powered lip-reading system capable of converting silent video inputs into meaningful text. This system is designed to operate in real- time or near-real-time environments, supporting sentence-level recognition of spoken language solely from visual cues, specifically lip movements. LipDecode targets controlled and semi-controlled environments where the speaker is clearly visible, such as front-facing video recordings, surveillance footage, or video conferencing scenarios.

Technologically, the scope includes the integration of deep learning architectures— namely Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks—into an end-to-end pipeline that includes video preprocessing, feature extraction, temporal analysis, and textual decoding. The system utilizes the LipNet framework as its backbone and incorporates OpenCV for video frame processing and Streamlit for building a practical user interface. The architectural foundation of LipDecode is built upon the LipNet model, a pioneering end- to-end framework for sentence-level visual speech recognition. This model facilitates the direct mapping of video sequences to text without the need for phoneme-level annotations or manual feature engineering. The use of Connectionist Temporal Classification (CTC) loss further enhances the model's capability to learn alignment between input frames and text output. Additionally, the system integrates OpenCV for real-time video capture, facial landmark detection, and mouth region isolation, ensuring that input to the neural network is consistent and of high quality. To make the system accessible and user-friendly, Streamlit is utilized to create an interactive graphical user interface, allowing users to upload videos, run inference, and view results with the significance of the LipDecode project is multi-dimensional. On a societal level, it addresses the critical need for enhanced communication tools for individuals with hearing or speech impairments, offering them a means to interpret spoken language through purely visual information.

II. EXISTING SYSTEM

While existing lip-reading systems such as LipNet (developed by the University of Oxford), DeepMind's lip-reading AI (2016), and Google's ASR-integrated approaches have made significant strides in visual speech recognition, they still face notable limitations, especially when applied to real-world environments. LipNet, although highly accurate on controlled datasets like GRID, struggles when faced with more diverse, unconstrained data. DeepMind's model, trained on thousands of hours of television broadcasts, demonstrated impressive lip-reading capabilities but remains limited in its generalizability outside of scripted or predictable contexts. Similarly, Google's fusion of lip reading with Automatic Speech Recognition (ASR) has improved performance in noisy conditions, but still relies heavily on the presence of supporting audio and high-quality visual input.

A fundamental challenge across all existing systems lies in real-world variability. Factors such as lighting, camera angle, occlusion (e.g., facial hair, hands), speaker diversity, and head movement significantly degrade model accuracy. Many models are trained on ideal conditions where speakers are front-facing, well-lit, and clearly articulated, which is often not the case in practical scenarios like surveillance footage or spontaneous conversation.

Another critical issue is the high dependence on large and diverse datasets. State-of-the-art models require thousands of labeled video clips across multiple speakers, accents, and speaking styles to perform well. However, acquiring such comprehensive datasets is time-consuming, expensive, and often limited by privacy concerns. Benchmark datasets like Lip Reading in the Wild (LRW), the VGG Lip Reading Challenge, and AVSpeech have helped standardize evaluation, but they still lack the variability and unpredictability seen in real-world environments. Furthermore, a significant linguistic challenge is the presence of homophenes—words that look the same on the lips (e.g., "mat" and "bat"). Since many phonemes are visually indistinguishable, models often confuse these words, leading to substantial errors in practical limitations in generalization, data dependence, and phoneme- level ambiguity.

Most lip-reading systems perform well under **controlled lab conditions** but falter in real-life scenarios. Key environmental challenges include:

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804228

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- Lighting Variability: Changes in ambient lighting can distort visual cues essential for lip-reading, especially in dynamic settings like outdoor surveillance or smartphone use.
- Camera Angle and Resolution: Non- frontal faces, low-resolution feeds (e.g., from CCTV), and varying camera angles make it difficult to extract accurate lip movement.
- Occlusion: Facial obstructions from hands, facial hair, masks, or glasses hinder accurate lip contour detection.
- Speaker Diversity: Models trained on limited speaker demographics often underperform with different ethnicities, accents, or speaking styles.

III. TECHNOLOGICAL GAPS IN CURRENT SOLUTIONS

Current lip-reading systems, despite significant progress, face several technological shortcomings that limit their effectiveness in real-world scenarios. A major issue is their poor generalization to unconstrained environments. Models like LipNet perform well on structured datasets but struggle with variations in lighting, backgrounds, occlusions, and off- angle views common in practical applications. Most existing models are overly reliant on high-quality, frontal visual data, assuming that the speaker is directly facing the camera in well- lit conditions. This assumption renders them less effective when deployed in environments like surveillance footage, video calls, or mobile recordings, where camera angles and lighting vary significantly.

Additionally, many models depend on the presence of accompanying audio or clean visual inputs. While Google's ASR-integrated systems have improved robustness in noisy conditions, they still underperform when audio is unavailable, which is a frequent occurrence in muted videos or security feeds. Another significant limitation lies in the inability of current systems. One of Streamlit's key strengths lies in its integration capabilities with popular data visualization libraries like Matplotlib, Plotly, and Altair. Users can effortlessly incorporate interactive charts, graphs, and plots into their applications. The inclusion of widgets such as sliders and buttons adds a layer of interactivity, enabling real-time updates and parameter control. Streamlit's automatic layout feature further simplifies the arrangement of elements on the web page, providing a hassle-free experience for developers. Beyond its ease of use, Streamlit facilitates seamless integration with prominent data science libraries like Pandas, NumPy, and scikit- learn. This integration empowers users to analyze and manipulate data directly within their applications. Moreover, Streamlit supports customization, allowing users to modify the appearance of their web apps using CSS, HTML, and other styling options. Deployment of Streamlit applications is straightforward, with compatibility across various platforms like Heroku, AWS, and Streamlit sharing. The library's active community and comprehensive documentation serve as valuable resources, offering support and guidance for users at all levels. In summary, Streamlit stands out as a versatile and accessible tool, empowering users to create impactful web applications for data science and machine learning without the complexities of traditional web development.

1. Utils:

In programming, "utils" typically refers to a module or package containing utility functions that perform common tasks or operations across a software project. These utility functions are designed to be reusable and versatile, providing solutions to recurring challenges without the need for redundant code. For example, a set of string manipulation utilities within "utils" might include functions for formatting, parsing, or searching strings, making them accessible from various parts of the codebase.

The use of "utils" contributes to a cleaner and more organized code structure. By grouping related utility functions together, developers can easily locate and employ them in different modules or classes, promoting code reusability and reducing the risk of duplicating similar functionalities. In essence, the "utils" module serves as a centralized repository for essential tools, streamlining development efforts and enhancing the overall maintainability of the software project.

In programming, "utils" (short for utilities) refers to a module or package that contains commonly used functions to simplify repetitive tasks across a software project. These utility functions are reusable and versatile, allowing developers to perform frequent operations efficiently without writing redundant code. For instance, a string manipulation utils module might include functions for formatting text, parsing data, or searching within strings,

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804228

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

making these functionalities easily accessible throughout the codebase. Similarly, mathematical utilities might include functions for calculating averages, rounding numbers, or performing complex calculations, ensuring consistency and efficiency in numerical operations.

Using a utils module helps maintain a clean and organized code structure by grouping related utility functions together. Instead of scattering these functions across different files or rewriting them in multiple places, developers can store them centrally in the utils module, making them easy to locate and reuse. This approach improves code readability and maintainability, reducing errors caused by duplicated logic. For example, a date and time utils module could provide functions for formatting timestamps, converting time zones, or calculating date differences, ensuring that all parts of the application handle time consistently.

The "utils" module acts as a centralized repository for essential tools, making development faster and more efficient. By separating general-purpose functions from the core logic of the application, developers can focus on building new features while relying on pre-built utilities for common tasks. This modular approach enhances code reusability, making software projects easier to scale and maintain. Whether in web development, data science, or machine learning, a well-structured utils module simplifies complex tasks, ensuring that the code remains organized, efficient, and adaptable for future improvements.

2. Modelutil

ModelUtil (short for model utilities) in machine learning refers to a comprehensive set of tools and functions designed to streamline the entire model lifecycle—from data preparation to deployment. These utilities simplify essential tasks such as preprocessing, feature engineering, model evaluation, and deployment, making machine learning workflows more efficient and consistent. ModelUtil plays a key role in data preprocessing by offering functions for normalization, handling missing values, encoding categorical variables, and feature extraction. This ensures that raw data is transformed into a suitable format for model training with minimal manual effort. It also supports comprehensive model evaluation by providing metrics like accuracy, precision, recall, F1-score, MSE, and R^2, alongside visual tools such as confusion matrices, ROC curves, and precision-recall curves. These features help data scientists interpret model performance and make informed decisions during model tuning.

Beyond training and evaluation, ModelUtil extends to deployment, offering tools for model serialization (e.g., Pickle, ONNX, TensorFlow SavedModel) and integration as APIs or web services. Some utilities even support monitoring and retraining, ensuring model relevance over time. By consolidating these capabilities into one framework, ModelUtil promotes code reusability, enhances transparency, and empowers developers to build, assess, and deploy robust machine learning models with greater speed and reliability. This unified approach not only accelerates development but also fosters collaboration among teams by standardizing processes and reducing redundant work. As machine learning projects scale, ModelUtil proves invaluable in maintaining consistency, improving maintainability, and enabling rapid iteration— ultimately supporting the delivery of high- quality, production-ready models across diverse environments and use cases.

IV. RESULTS AND DISCUSSION

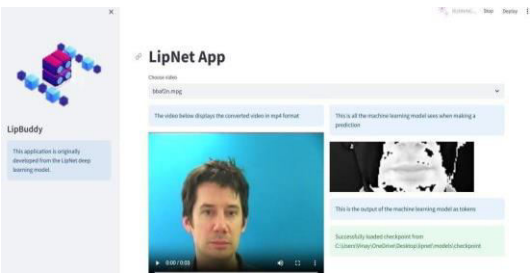

Figure 2: Shows the Home Page of the Project

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804228

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The home web page includes the identity of the challenge with branding records and brand of the venture. The intention of the homepage isn't to be a library of textual content and content material, but alternatively, to characteristic a teaser and truthful manual in the direction of the pages which have the desired statistics.

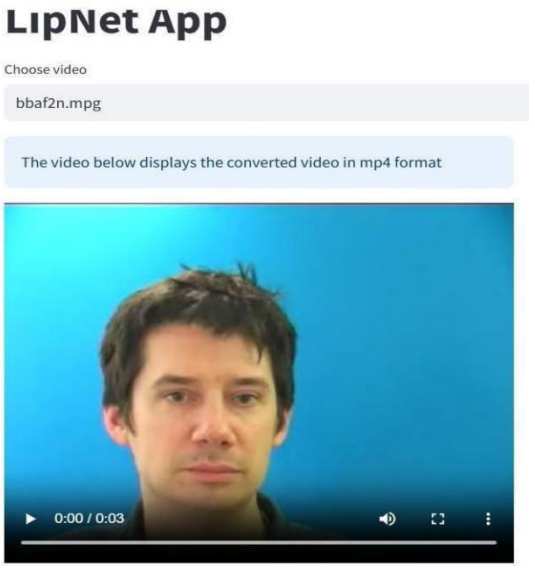

Figure 3: Shows Upload file page

Figure 4: Shows Data View

A statistics view might be a gadget or visible representation of facts that differs from place and type of soil. Views are frequently created to form records more applicable.

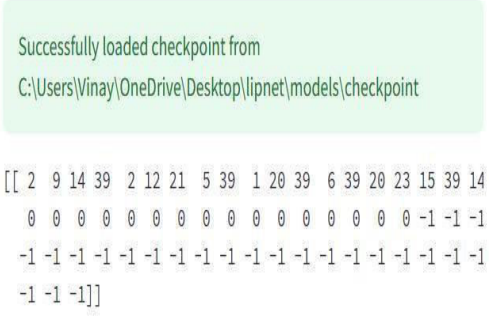

Figure 5: Shows Data Analysis

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804228

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Data evaluation is that the process of accumulating and organizing records so as to draw helpful conclusions from it. The approach of facts evaluation makes use of analytical and logical reasoning to realize statistics from the statistics

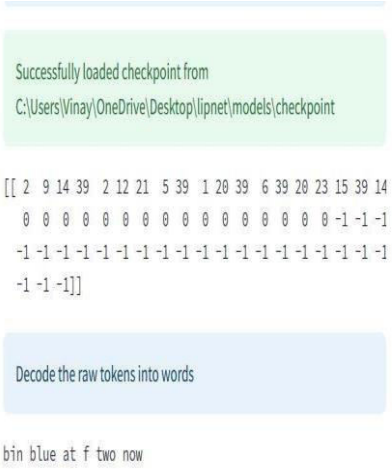

Figure 6 :show the predicted result

VI. CONCLUSION

The LIPDECODE project has successfully demonstrated the capability of AI-powered lip- reading technology to bridge communication gaps for individuals with hearing impairments, those facing language barriers, and people in noisy environments where verbal communication is difficult. By leveraging machine learning, computer vision, and deep learning models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs), the system effectively converts silent video inputs of lip movements into accurate text. The implementation of advanced neural networks enabled the system to recognize speech patterns with considerable precision, making it a valuable tool for assistive communication, education, and real-time interaction in public spaces.

Through rigorous testing, LIPDECODE exhibited high accuracy in controlled environments, successfully identifying words and sentences from lip movements with minimal error. It proved especially beneficial for individuals with hearing disabilities, post- surgery patients who struggle with verbal communication, and professionals working in noisy settings such as airports or factories. The model's adaptability and real-time processing capabilities enhanced its effectiveness in various applications, making it a promising step toward an inclusive communication system. However, the project also revealed several challenges that limit its deployment in real-world scenarios. Factors such as variations in lip shapes, different accents, lighting conditions, camera angles, and facial obstructions affected the model's accuracy, sometimes leading to incorrect word recognition.

Furthermore, homophones (words with the same lip movements but different meanings) posed a significant challenge, as the system lacked contextual understanding to differentiate between them.Additionally, real-time processing on low-end devices remains a hurdle due to the high computational demands of deep learning models. While the system performed well on powerful GPUs, optimizing it for real-time usage on mobile or embedded systems requires further advancements in efficiency and resource management. Another major challenge is the lack of large, diverse, and high-quality datasets, which limits the model’s ability to generalize across different demographics, languages, and speech styles. Current datasets do not adequately represent various ethnic backgrounds, age groups, and speaking patterns, leading to potential biases in recognition accuracy.

Despite these challenges, LIPDECODE represents a significant breakthrough in the field of AI-driven lip-reading technology. It highlights the potential for improving assistive communication and accessibility, offering an alternative to traditional speech recognition systems that struggle in noisy environments. With further development, such as training on larger multilingual datasets, improving contextual understanding, and enhancing real-time processing efficiency, the system could be refined for broader real-world applications, including healthcare,

security, and human-computer interaction. As AI and machine learning continue to evolve, LIPDECODE stands as a promising innovation that could reshape the future of communication, making it more inclusive, efficient, and accessible for all predictive analytics might forecast attendance trends to optimize scheduling. These enhancements would transform the system from basic attendance tracking to a comprehensive participation analytics platform with applications across education, corporate, and government sectors. The modular architecture ensures seamless integration of future AI advancements as the technology evolves.

REFERENCES

1. Sumby, W.H. Visual Contribution to Speech Intelligibility in Noise. J. Acoust. Soc. Am. **1954**, 26, 212–215. [GoogleScholar] [CrossRef][Green Version]
2. Kastaniotis, D.; Tsourounis, D.; Koureleas, A.; Peev, B.; Theoharatos, C.; Fotopoulos, S. Lip Reading in Greek words at unconstrained driving scenario. In Proceedings of the 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), Patras, Greece, 15–17 July 2019;pp. 1–6. [Google Scholar] [CrossRef]
3. Abrar, M.A.; Islam, A.N.M.N.; Hassan,M.M.; Islam, M.T.; Shahnaz, C.; Fattah,S.A. Deep Lip Reading-A Deep Learning Based Lip-Reading Software for the Hearing Impaired. In Proceedings of the 2019 IEEE R10 Humanitarian Technology Conference (R10-HTC) (47129), Depok, West Java, Indonesia, 12–14 November 2019; pp. 40–44. [Google Scholar] [CrossRef]
4. Scanlon, P.; Reilly, R. Feature analysis for automatic speechreading. In Proceedings of the 2001 IEEE Fourth Workshop on Multimedia Signal Processing (Cat. No.01TH8564), Cannes, France, 3–5 October 2001; pp. 625–630. [Google Scholar] [CrossRef]
5. Aleksic, P.S.; Katsaggelos, A.K. Comparison of low- and high-level visual features for audio-visual continuous automatic speech recognition. In Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, QC, Canada, 17–21 May 2004; p. V-917. [GoogleScholar] [CrossRef]
6. Minotto, V.P.; Lopes, C.B.O.; Scharcanski, J.; Jung, C.R.; Lee, B. Audiovisual Voice Activity Detection Based on Microphone Arrays and Color Information. IEEE J. Sel. Top. Signal Process. **2013**, 7, 147–156. [Google Scholar] [CrossRef]
7. Assael, Y.M.; Shillingford, B.; Whiteson, S. LipNet: End-to-end sentence-level lipreading. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2016. [Google Scholar] [CrossRef]
8. Burton, J.; Frank, D.; Saleh, M.; Navab, N.; Bear, H.L. The speaker-independent lipreading play-off; a survey of lipreading machines. In Proceedings of the 2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS), Sophia Antipolis, France, 12–14 December 2018; pp. 125–130. [Google Scholar] [CrossRef]
9. Chen, X.; Du, J.; Zhang, H. Lipreading with DenseNet and resBi-LSTM. Signal Image Video Process. 2020, 14, 981–989. [Google Scholar] [CrossRef]
10. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu,C. GhostNet: More Features from Cheap Operations. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com