SILENT SPEECH INTERPRETATION USING AI

[1]Dr.M. Hemalatha, [2]M. Akshayaa

[1]Associate Professor, [2]UG Student
PG & Research Department of Computer Science
Sri Ramakrishna College Arts &Science (Autonomous)
Coimbatore, India.

Abstract- **This project develops a silent speech recognition system that enables people with speech or hearing impairments to communicate by understanding spoken words based on video lips. The system first detects the face and mouth using high-speed object detection technology, then detects visual features such as lip shape and wrinkles. A machine learning classifier trained on videos where lip movements match spoken words then predicts the most likely word based on the extracted features. Promisingly, it has limitations. Due to personal facial features and speech styles, the system may not work well for everyone. Current versions only recognize a limited vocabulary. Background noise, lighting, and head orientation also affect accuracy. Researchers are actively working on improvements. We are exploring techniques to make the system work for more people and to withstand environmental changes. Deep learning models are being explored to process larger vocabularies and capture details. Additionally, it is considered to combine visual information with audio (if available) to achieve even greater accuracy. This technology can transform communication for people with speech or hearing difficulties. It can also enable silent human-computer communication and has applications in education, language learning, and entertainment.**

Keywords: **silent speech recognition, lip reading, speech impairment, hearing impairment, facial features, machine learning, deep learning.**

I.INTRODUCTION

This section covers the field of silent speech recognition, which is a subfield of speech recognition. Silent speech recognition aims to interpret spoken words using only visual cues such as lip movements and facial expressions, eliminating the need for any audio input. This technology is particularly promising for speech-impaired people and offers an alternative way of communication. The document presents a project focused on the development of a silent speech recognition system. The approach is based on the analysis of visual data recorded in videos. The first step involves recognizing the speaker's face and then locating the mouth region in each video frame. This limits the analysis area to facial features relevant to speech recognition. To refine the analysis, the system identifies a certain region in the mouth area, called a region of interest (ROI), where the main visual indicators of speech are located. Once the ROI is defined, the system extracts several features from the video. . from this area during a silent call. These characteristics describe the dynamics of the speaker's lip movements and changes in the shape of the mouth when they silently form words. The extracted features are, for example, those that represent lip contours (edge features), describe the general shape of the mouth, analyze lip color variations (color features) and describe lip surface features (textural features).). By analyzing this combination of visual features, the system tries to recognize spoken words without audio input. This technology can change how people with speech disabilities communicate.

II. RELATED WORK

Xing Wei An, Erwei Yin, and Dong Ming(2022) proposed a novel deep learning architecture called Parallel Inception Convolutional Neural Network (PICNN) for silent speech recognition using surface electromyography (sEMG) signals. Their architecture achieves a high recognition accuracy of 90.76% on a 100-class dataset they designed. Focused on facial sEMG-based silent speech recognition, the PICNN architecture effectively extracts spatial features from sEMG signals, showing promise for real-life applications. Future research will address unsatisfactory recognition rates for specific demands and enhance system robustness through additional experiments and consideration of physiological principles of speech activity.

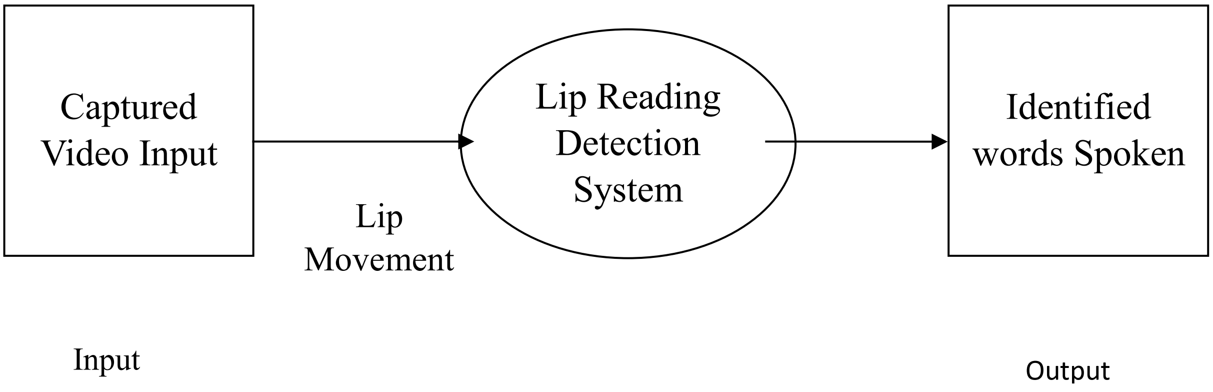
Naga Jyothi and Siddaiah (2018) introduce an ASR-based Airport Enquiry System designed specifically for the Telugu language. Utilizing Convolutional Neural Networks (CNNs), the system capitalizes on features like weight connectivity and local pooling for heightened performance. Through thorough training, the CNN model shows significant improvements, especially with wideband speech signals. The study highlights CNN's effectiveness in handling complex speech recognition tasks, outperforming traditional neural network techniques.

Menshikov Ivan's (2022) study focuses on silent speech recognition using EEG data to advance Brain-Computer Interface (BCI) development for individuals with neurodegenerative disorders and communication difficulties. With EEG recordings from 270 healthy subjects performing nine different commands, including directed movements and a pseudoword, we achieved high accuracy rates through individualized training. Our findings support three key hypotheses regarding the consistency of electrical activity patterns, the effectiveness of smaller individual datasets, and the feasibility of distinguishing silent words with similar brain activity patterns. These insights are pivotal for tailoring BCIs to diverse user groups, especially individuals with disabilities.

III.PROPOSED METHODOLOGY

Inspired by detective work, the silent speech recognition system first uses the Viola-Jones algorithm to quickly recognize faces in a video image. It then zooms in on the mouth area and detects visual cues such as lip shape and wrinkles. These features are provided in a machine learning model trained on a large dataset of silent speech, allowing it to predict the most likely spoken word, much like a detective solving a case based on the evidence collected. This deep learning outperforms traditional methods by automatically identifying key features of raw video data.

IV.DATA FLOW DIAGRAM



Silent Speech Recognition: A Promising Avenue for Communication and Interaction

Interpreting silent speech with artificial intelligence has enormous potential for various applications. This technology allows people with speech or hearing impairments to promote natural communication without voice or sign language. In addition to assisted communication, it improves communication between humans and computers by enabling silent commands that are ideal for noisy or private environments. The potential extends to education and entertainment with language learning programs and an immersive virtual experience where emotions and commands can be transmitted silently. Security and privacy are also enhanced with silent authentication, while silent speech recognition can complement traditional voice recognition by providing an alternative way to understand speech in noisy environments. Deep learning algorithms feed the ability of this system to detect complex patterns of lip movements, resulting in a very accurate interpretation that improves with further study. As research progresses, the ability to recognize a wider range of words and even emotions from silent speech holds enormous promise for the future.

IV. RESULTS AND DISCUSSIONS

In the absence of precise details about the datasets used in silent speech recognition (SSI) research, we can infer some basic features. The data will likely consist of videos of people silently speaking single words or short sentences from a limited vocabulary. These recordings would involve split speakers speaking clearly and with minimal head movement, all recorded in controlled environments with minimal noise and uniform lighting. To train machine learning models, videos are supposed to be tagged with corresponding spoken words or phrases. The dataset size can be smaller compared to general speech recognition because it focuses on controlled elements.

V. CONCLUSION

The automation of the human ability to perform the .lip-reading process is referred to as visual speech recognition which is used in human-computer interaction (HCI), audio-visual speech recognition along disability management. This silent speech recognition technique is based on facial muscle activity and video, without evaluating any voice signals. In this work, a voiceless speech recognition technique is developed that utilizes dynamic visual features to represent facial movements during phonation. The dynamic features extracted from the mouth video are used to classify utterances without using the acoustic data. The proposed technique yields a high recognition rate by applying the face and mouth detection processing using the Viola-jones algorithm. The spatial-related superpixel algorithm is used to identify the

ROI and possible combination of objects present in the video frame and the SVM-QP classifier is used to perform the word classification process. The proposed algorithm achieved significant performance compared to the surface electromyogram.

REFERENCES:

1. CAMPBELL, R., LANDIS, T. and REGARD, M., 1986. Face recognition and lipreading: A neurological dissociation. Brain, 109(3), pp.509-521.
2. Wand, M., Koutník, J. and Schmidhuber, J., 2016, March. Lipreading with long short-term memory. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6115-6119). IEEE.
3. Murase, H. and Sakai, R., 1996. Moving object recognition in eigenspace representation: gait analysis and lip reading. Pattern recognition letters, 17(2), pp.155-162.
4. Stillittano, S., Girondel, V. and Caplier, A., 2013. Lip contour segmentation and tracking compliant with lip-reading application constraints. Machine vision and applications, 24(1), pp.1-18.
5. Wilson, R., Steinfurth, A., Ropert-Coudert, Y., Kato, A. and Kurita, M., 2002. Lip-reading in remote subjects: an attempt to quantify and separate ingestion, breathing and vocalization in free-living animals using penguins as a model. Marine Biology, 140(1), pp.17-27.
6. Zheng, G.L., Zhu, M. and Feng, L., 2014, December. Review of lip-reading recognition. In 2014 Seventh International Symposium on Computational Intelligence and Design (Vol. 1, pp. 293-298). IEEE.
7. Kerr, A.C., White, R.V. and Saunders, A.D., 2000. LIP reading: recognizing oceanic plateaux in the geological record. Journal of Petrology, 41(7), pp.1041-1056.
8. Summerfield, Q., 1983. Audio-visual speech perception, lipreading, and artificial stimulation. In Hearing Science and Hearing Disorders (pp. 131-182). Academic press.
9. Ma, W.J., Zhou, X., Ross, L.A., Foxe, J.J. and Parra, L.C., 2009. Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space. PloS one, 4(3), p.e4638.
10. Alegria, J., Charlier, B.L. and Mattys, S., 1999. The role of lip-reading and cued speech in the processing of phonological information in French-educated deaf children. European Journal of Cognitive Psychology, 11(4), pp.451-472.
11. Lott, B.E. and Levy, J., 1960. The influence of certain communicator characteristics on lip reading efficiency. The Journal of Social Psychology, 51(2), pp.419-425.
12. Fernandez-Lopez, A., Martinez, O. and Sukno, F.M., 2017, May. Towards estimating the upper bound of visual-speech recognition: The visual lip-reading feasibility database. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 208-215). IEEE.
13. Salik, K.M., Aggarwal, S., Kumar, Y., Shah, R.R., Jain, R. and Zimmermann, R., 2019, July. Lipper: Speaker-independent speech synthesis using multi-view lipreading. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 10023-10024).
14. Ludman, C.N., Summerfield, A.Q., Hall, D., Elliott, M., Foster, J., Hykin, J.L., Bowtell, R. and Morris, P.G., 2000. Lip-reading ability and patterns of cortical activation studied using fMRI. British Journal of Audiology, 34(4), pp.225-230.
15. Campbell, R., 1992. The neuropsychology of lipreading. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 335(1273), pp.39-45.
16. Bourguignon, M., Baart, M., Kapnoula, E.C. and Molinaro, N., 2020. Lip-reading enables the brain to synthesize auditory features of unknown silent speech. Journal of Neuroscience, 40(5), pp.1053-1065.
17. O'Sullivan, A.E., Crosse, M.J., Di Liberto, G.M. and Lalor, E.C., 2017. Visual cortical entrainment to motion and categorical speech features during silent lipreading. Frontiers in human neuroscience, 10, p.679.
18. Xu, K., Li, D., Cassimatis, N. and Wang, X., 2018, May. LCANet: End-to-end lipreading with cascaded attention-CTC. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (pp. 548-555). IEEE.
19. Aparicio, M., Peigneux, P., Charlier, B., Balériaux, D., Kavec, M. and Leybaert, J., 2017. The neural basis of speech perception through lipreading and manual cues: Evidence from deaf native users of cued speech. Frontiers in Psychology, 8, p.426.
20. Fung, I. and Mak, B., 2018, April. End-to-end low-resource lip-reading with max-out CNN and LSTM. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2511-2515). IEEE.