IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, March 2022
Impact Factor: 6.252

# Visual Speech Recognition

**Supriya Patil[1], Vaibhav Dhoble[2], Saatvik Gawade[3], Pratiksha Jagdale[4], Rohan Jinde[5]**
Assistant Professor, Department of Information Technology[1]
Student, Department of Information Technology[2,3,4,5]
Zeal College of Engineering and Research, Pune, Maharashtra, India

**Abstract:** *The audio-visual speech recognition approach attempts to boost noise-robustness in mobile situations by extracting lip movement from side-face images. Although earlier bimodal speech recognition algorithms used frontal face (lip) images, these approaches are difficult for consumers to utilize because they need them to talk while holding a device with a camera in front of their face. Our proposed solution, which uses a small camera put in a handset to capture lip movement, is more natural, simple, and convenient. This approach also effectively avoids a reduction in the input speech's signal-to-noise ratio (SNR). Optical-flow analysis extracts visual features, which are then coupled with audio features in the context of CNN-based recognition.*

**Keywords:** Convolutional Neural Network, Deep Learning, Image

## REFERENCES

**[1]** Zhang, Xingxuan, Feng Cheng, and Shilin Wang. "Spatio-temporal fusion based convolutional sequence learning for lip reading." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
**[2]** Kurniawan, Adriana, and Suyanto Susanto. "Syllable-Based Indonesian Lip Reading Model." 2020 8th International Conference on Information and Communication Technology (ICoICT). IEEE, 2020.
**[3]** Michelsanti, Daniel, et al. "An overview of deep-learning-based audio-visual speech enhancement and separation." IEEE/ACM Transactions on Audio, Speech, and Language Processing (2021).
**[4]** Desai, Dhairya, et al. "Visual Speech Recognition." International Journal of Engineering Research Technology (IJERT) 9.04 (2020).
**[5]** Fenghour, Souheil, et al. "Deep Learning-based Automated Lip-Reading: A Survey." IEEE Access (2021).
**[6]** Afouras, Triantafyllos, et al. "Deep audio-visual speech recognition." IEEE transactions on pattern analysis and machine intelligence (2018).
**[7]** Zhang, Yuanhang, et al. "Can we read speech beyond the lips? rethinking roi selection for deep visual speech recognition." 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020). IEEE, 2020.
**[8]** Petridis, Stavros, et al. "End-to-end visual speech recognition for small-scale datasets." Pattern Recognition Letters 131 (2020): 421-427.
**[9]** Lee, Wookey, et al. "Biosignal Sensors and Deep Learning-Based Speech Recognition: A Review." Sensors 21.4 (2021): 1399.
**[10]** Lee, Yong-Hyeok, et al. "Audio–visual speech recognition based on dual crossmodality attentions with the transformer model." Applied Sciences 10.20 (2020): 7263.