

# Visual Speech Recognition

Tejasweeni Hampe<sup>1</sup>, Vaibhav Dhoble<sup>2</sup>, Saatvik Gawade<sup>3</sup>, Pratiksha Jagdale<sup>4</sup>, Rohan Jinde<sup>5</sup>

Assistant Professor, Department of IT, Zeal College of Engineering and Research, Pune<sup>1</sup>

Student, Department of IT, Zeal College of Engineering and Research, Pune<sup>2,3,4,5</sup>

**Abstract:** *The general media discourse acknowledgment strategy utilizing lip development extricated from side-face pictures to endeavour to increment commotion strength in versatile conditions. Albeit most past bimodal discourse acknowledgment techniques utilize front facing face (lip) pictures, these techniques are difficult for clients since they need to hold a gadget with a camera before their face while talking. Our proposed strategy catching lip development utilizing a little camera introduced in a handset is more regular, simple and helpful. This technique likewise successfully evades a diminishing of sign to-commotion proportion (SNR) of information discourse. Visual elements are separated by optical-stream examination and joined with sound elements in the system of CNN-based acknowledgment.*

**Keywords:** Convolutional Neural Network, Deep Learning, Image Processing, etc.

## 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. "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. Lee, Wookey, et al. "Biosignal Sensors and Deep Learning-Based Speech Recognition: A Review." Sensors 21.4 (2021): 1399.
- [9] Lee, Yong-Hyeok, et al. "Audio-visual speech recognition based on dual cross modality attentions with the transformer model." Applied Sciences 10.20 (2020): 7263.