|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SNO | AUTHORS | TITLE | PUBLISHING | TECHNIQUES AND DATASET | PROS | CONS |
| 1 | Zhang, Shiliang, et al. | Robust audio-visual speech recognition using bimodal DFSMN with multi-condition training and dropout regularization. | *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019. | Deep Neural Network  NTCD-TIMIT corpus | Proposed bimodal-DFSMN with multi-condition training can significantly improve the performance of AVSR system. | In the experiments, the performance of video-only ASR system is not very good. |
| 2 | Bhaskar, Shabina, T. M. Thasleema, and R. Rajesh. | A Survey on Different Visual Speech Recognition Techniques | *Data Analytics and Learning*. Springer, Singapore, 2019. 307-316. | Automatic Lip reading,  Hidden markov model  CUAVE, TULIPS, XM2VTS, AV-TIMIT, AVCAR, MODALITY | combination of both transform and geometric based technique will lead to a better visual feature extraction approach | The review reveals that only the two databases AVCAR and XM2VTS considered about the speaker number above 70 and most of them considered the acoustic noise only rather considering both acoustic and visual noises. |
| 3 | Dhairya Desai  Priyesh Agrawal  Priyansh Parikh  Mr. Piyush Kumar Soni | Visual Speech Recognition | International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 9 Issue 04, April-2020 | LSTMs, CNNs, Decision Tree and  Neural networks.  GRID corpus dataset, Full-lip images and  Patches around tracked lips obtained by face-alignment, two datasets were used for the training and testing of the  method containing front-view video of the speaker and the sideview video of the speaker | Visual speech recognition alludes to the detailed feature-based  analysis on the lips and its surrounding environment. | The accuracy achieved here is approx. 32% which is not great according to the article but a complex model with many environmental variables and features to be considered needs more future developments. |
| 4 | Anonymous authors | LARGE-SCALE VISUAL SPEECH RECOGNITION | Under review as a conference paper at ICLR 2019 | Group normalization  a deep learning-based visual speech recognition model with production-grade word-level decoding techniques.  traditional speech technique  visual speech recognition dataset,  Large-Scale Visual Speech Recognition (LSVSR) dataset,  LRS and MV-LRS datasets | This work presents a scalable solution to continuous visual speech recognition.  This article’s pipeline uses landmark smoothing, a blurriness filter, an improved speaking classifier network and outputs phonemes. |  |
|  |  |  |  |  |  |  |