



LIPREAD BY LIOPA: AUTOMATED LIP-READING ENGINE

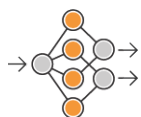
LipRead is a cloud-based Deep Learning engine that deciphers speech from speaker lip movements. LipRead provides enhanced speech recognition in situations where speaker audio is either not available or of poor quality. LipRead powers Liopa's market driven solutions, providing a Visual Speech Recognition (VSR) capability for a wide range of applications.

LIPREAD: How does it work?



Capture

- Speaker positioned in front of device camera (smartphone, tablet, PC, webcam)
- Range of commodity camera types supported
- Range of pose angles.



Analysis

- Video uploaded to LipRead VSR engine & analyzed to detect & track speaker head, facial landmarks & finally lip region
- Lip movement (viseme) features extracted from video frames
- Features analyzed using Deep Neural Network (DNN) techniques.



Results

- Output of feature analysis compared to universal model built from large quorums of training data
- Probabilistic representation of most likely words/phrases created
- Choice of result formats available.

VSR TECHNOLOGY

Liopa's VSR technology is based on the principle of viseme analysis. A viseme is a unique lip movement that can be used to describe a specific sound. It is the visual equivalent of a phoneme or unit of sound in spoken language. Using visemes, the hearing-impaired can view sounds visually via studying a person's lip movement.

Liopa's AI based VSR technology implements this process programmatically by:

- Capturing a video of a subject speaking
- Tracking & extracting the movement of the subject's lips
- Performing lip movement feature extraction
- Using Deep Neural Network (DNN) techniques to analyze features extracted from each frame of the video
- Comparing the results of the analysis (on a viseme by viseme basis) with a universal model to determine what has been spoken.

A deep neural network (DNN) is an artificial neural network (ANN) with multiple hidden layers between the input and output layers giving the potential of modelling complex data with fewer units than a similarly performing shallow network.

Liopa has developed a DNN based VSR system which leverages a proprietary and patent pending combination of leading edge neural network techniques.



LIPREAD IN USE: POWERING MARKET SOLUTIONS

Making 'Voice-first' better

- Voice-driven applications are now ubiquitous, as is evident from the rise in popularity of virtual personal assistants (for e.g., Siri, Cortana, Alexa, etc.) and voice-activated in-car command systems. These applications use Audio Speech Recognition (ASR) technology to decipher speech from the audio waveform. ASR performs very well in 'clean' environments, but word accuracy generally degrades in more real-world, noisy environments.
- LipRead VSR is audio noise agnostic and can combine with the ASR to improve overall system performance in situations where background noise exists and where a camera can be trained on the head of the speaker.
- Liopa has developed a series of patent-pending techniques to dynamically integrate 3rd party ASR solutions with LipRead VSR giving optimal accuracy across varying levels of audio & video noise.
- **Liopa's AVSR solution can truly unlock the potential of the voice-first interface.**

Voiceless Speech Recognition

- There are situations where voice-based communication is not possible – either the person cannot speak in the normal sense or their voice cannot be heard. The LipRead VSR engine can be used in these circumstances to provide a VSR-only communications capability, initially supporting a constrained list of words & phrases.
- Liopa is trialing a communications aid for Tracheostomy patients in UK hospitals. The patient's lip movements are captured by a smartphone camera, uploaded to LipRead, analyzed and the correct phrase selected from a pre-determined vocabulary.
- Voice-driven machinery, a key tenant of today's 'Smart' factories, needs to operate in very noisy industrial environments where the operator can often not be heard. The interface is a constrained command/control vocabulary and thus very suitable for use of LipRead.

Keyword spotting

- LipRead can analyze video footage to determine whether or not certain words/phrases have been spoken. The LipRead VSR engine will produce 'lattices' containing most likely word 'units' - these lattices are subsequently searched to identify patterns analogous to a user-supplied key word or phrase.
- Ideal when audio track is non-existent or of poor quality.
- Can be used to analyze large volumes of CCTV footage to identify segments of interest.

Anti-Spoofing

- Liopa's LipSecure product leverages the LipRead VSR engine to provide a very strong anti-spoofing mechanism for online authentication.
- The best authentication systems are multi-modal and contain at least one Biometric modality (e.g. fingerprint, facial recognition) but these techniques can be 'spoofed' by an artificial representation of the biometric data (for e.g., false fingerprint, photograph, etc.). The ability of these authentication systems to differentiate between a live person and a spoof is critically important.
- LipSecure ...
 - challenges the user to speak/mime a random sequence of digits
 - uses the LipRead VSR engine to validate the user response and confirm whether or not a live subject is present.
- LipSecure is much more robust than alternative software-based anti-spoofing techniques.