

Image & Video Analysis

Helping the fight against crime: computerised lip-reading

Business Challenge

Lip-reading has been used in evidence in court but its use has been controversial because human lip-readers are prone to human errors which cannot easily be quantified. Nevertheless there are situations, in particularly noisy environments for example, when lip motions might be the only evidence available so it is important to understand how speakers move their lips, whether they do so consistently, and whether artificially intelligent computers can be built to recognise visual speech.

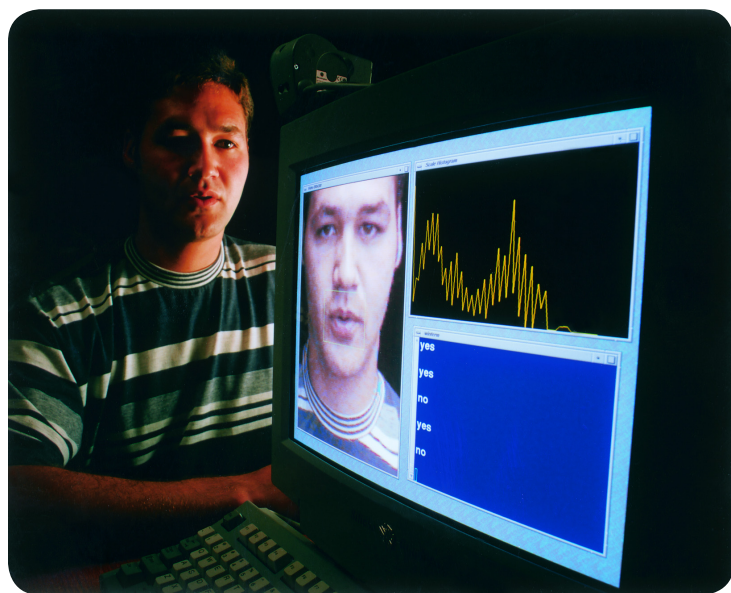
Our Solution & Expertise

A group of our computing experts, led by Dr Richard Harvey, are currently developing a computer lip-reading system which will convert videos of conversations into written transcripts.

Researchers have identified that each speech sound has a particular facial and mouth position, which can be mapped on video and then processed by software that associates the various orientations with the sounds they produce. The technology focuses on mapping the facial movements, called 'features', and converting them into signals which the software can process and 'translate' into text.

The team has collected data, such as videos and photographs, which are being used to determine the exact lip movements and facial expressions associated with specific letter combinations in the most accurate way. There are also plans to extend the speech-recognition system to additional languages.

The University are teaming up with Surrey University, who have built accurate and reliable face and lip trackers. Combining the two technologies will create an innovative and ground-breaking piece of equipment that will be extremely valuable in the fight against crime.



Business Benefits

When this technology is finalised, it will be extremely beneficial to law enforcers collecting evidence to convict criminals and terrorists.

Dr Richard Harvey said: "There are often situations where the police cannot get close enough to hear what a suspected criminal is saying. In such situations, a camera could see the lips and a computer might be able to lip-read. To be effective the systems must accurately track the head over a variety of poses, extract numbers, or features, that describe the lips and then learn what features correspond to what text. To tackle the problem we will need to use information collected from audio speech. So this project will also investigate how to use the extensive information known about audio speech to recognise visual speech."