Read my lips: New technology spells out what's said when audio fails

March 24 2016

New lip-reading technology developed at the University of East Anglia (UEA) could help in solving crimes and provide communication assistance for people with hearing and speech impairments.

The visual speech recognition technology, created by Dr Helen L. Bear and Prof Richard Harvey of UEA's School of Computing Sciences, can be applied "any place where the audio isn't good enough to determine what people are saying," Dr Bear said.

Dr Bear, whose findings will be presented at the International Conference on Acoustics, Speech and Signal Processing (ICASSP) in Shanghai on March 25, said unique problems with determining speech arise when sound isn't available - such as on CCTV footage - or if the audio is inadequate and there aren't clues to give the context of a conversation. The sounds '/p/,' '/b/,' and '/m/' all look similar on the lips,

but now the machine lip-reading classification technology can differentiate between the sounds for a more accurate translation.

Dr Bear said: "We are still learning the science of visual speech and what it is people need to know to create a fool-proof recognition model for lip-reading, but this classification system improves upon previous lip-reading methods by using a novel training method for the classifiers.

"Potentially, a robust lip-reading system could be applied in a number of situations, from criminal investigations to entertainment. Lip-reading has been used to pinpoint words footballers have shouted in heated moments on the pitch, but is likely to be of most practical use in situations where are there are high levels of noise, such as in cars or aircraft cockpits.

"Crucially, whilst there are still improvements to be made, such a system could be adapted for use for a range of purposes - for example, for people with hearing or speech impairments. Alternatively, a good lip-reading machine could be part of an audio-visual recognition system."

Prof Harvey said: "Lip-reading is one of the most challenging problems in artificial intelligence so it's great to make progress on one of the trickier aspects, which is how to train machines to recognise the appearance and shape of human lips."

The research was part of a three-year project and was supported by the Engineering and Physical Sciences Research Council (EPSRC).

The paper, Decoding visemes: Improving machine lip-reading, will be published on March 25, 2016 in the *Proceedings of the IEEE* International Conference on Acoustics, Speech and Signal Processing 2016.