

BREVET DE TECHNICIEN SUPÉRIEUR ELECTRONIQUE / ELECTROTECHNIQUE

SESSION 2000

ANGLAIS

Durée : 2 heures

Coefficient : 2

*Dictionnaire bilingue autorisé
Calculatrices, lexiques et autres outils personnels interdits.*

TRAVAIL À FAIRE :

I) COMPTE-RENDU EN FRANÇAIS DESTINÉ À UN GROUPE DE TECHNICIENS (12 points)

Vous vous efforcerez, dans un compte-rendu comportant entre 150 et 200 mots, de mettre en évidence les informations les plus importantes contenues dans le document.

II) TRADUCTION (8 points)

Traduire en français page 3/3 le passage entre crochets :

de : "Gary Strong, project manager..." jusqu'à "...lip-reading software".

Read my lips

Voice-recognition software that won't be distracted by noise

JUST like us, computers find it tough to hear what's being said in a noisy room. So computer scientists at Carnegie Mellon University in Pittsburgh are teaching them to lip-read.

Whether or not you realise it, you're pretty good at lip-reading, according to Alex Waibel, a computer scientist at CMU. "When people are in a noisy environment they pay more attention to the lips," he says. Lip-reading dramatically improves our understanding of what people are saying.

Waibel's new software, called NLips, is designed to reduce the error rate of speech-recognition software in noisy environments. For software that is, say, 92 per cent successful when the surroundings are quiet, the lip-reading only helps marginally, says Waibel, improving successful recognition to about 93 per cent. But when there is lot of background noise, the success rate of a typical package drops to around 60 per cent – and NLips can bump this up to about 85 per cent.

Like most speech-recognition systems, NLips breaks down speech into discrete sound chunks, called phonemes, but crucially it also combines information from lip movements. Computer-mounted cameras record lip sequences, using tracking software to compensate for any slight movements of the head.

A neural network, which learns as it goes along, constantly monitors lips in the video sequences looking for the 50 visual equivalents of phonemes (*), or "visemes" as Waibel calls them. The software cross-checks the output from the speech recognition program against the visemes. NLips works so well because it combines different sorts of perceptual information, both visual and audio, says Waibel. He admits that the lip-reading software is hopeless on its own. Waibel says his lab is "looking at all these signals and capturing the perceptual world in its entirety", just as humans do.

So far, Waibel and his colleagues have only tested NLips for spelling out words, letter by letter. But he is confident that moving onto continuous speech should be straightforward, because most speech recognition software finds this less of a challenge than spelling. With so many letters sounding similar, ambiguity causes a lot of spelling problems.

Waibel is now working on incorporating NLips into a video conferencing system that can automatically create transcripts of what is said and by whom.

[Gary Strong, project manager for several speech-recognition projects at the National Science Foundation in Arlington, Virginia, believes that it's only a matter of time before speech-recognition software companies follow CMU's two-pronged approach. The next goal, he says, is to put voice recognition inside noisy vehicles – allowing you to give voice commands to your car, for example – but this has in the past been dogged by the unpredictable nature of background vehicle noise. Recognition under these conditions will be almost impossible unless the error rate can be reduced – perhaps by using a tiny camera to feed images to lip-reading software.]

Duncan Graham-Rowe, 14 August 1999, *New Scientist*

(*) phonemes - phonèmes, terme de linguistique relatif aux sons.