AUDIO–VISUAL SPEECH RECOGNITION WITH A HYBRID SVM–HMM SYSTEM (ThuAmPO1)

Author(s) :

Mihai Gurban   (Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland)
Jean–Philippe Thiran   (Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland)

Abstract :

Traditional speech recognition systems use Gaussian mixture models to obtain the likelihoods of individual phonemes, which are then used as state emission probabilities in hidden Markov models representing the words. In hybrid systems, the Gaussian mixtures are replaced by more discriminant classifiers, leading to an improved performance. Most of the time the classifiers used in such systems are neural networks. Support vector machines have also been used in one–modality audio or visual speech recognition, but never in a multimodal audio–visual system. We propose such a hybrid SVM–HMM speech recognizer, and we show how the multimodal approach leads to better performance than that obtained with any of the two modalities individually.

Menu