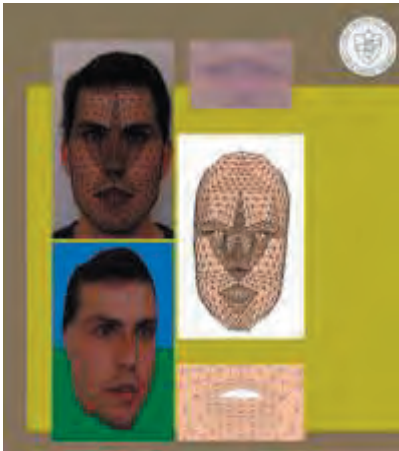


ETRO develops tools for audio-visual face expression communication

ETRO, IMEC's associated lab at the University of Brussels (VUB) has developed tools to integrate both audio and visual components in communication systems. Research currently focuses on the creation of synthetic mouth shapes of a virtual face and emotion recognition in verbal communication. This will be used for a case study with a social robot meant to interact with hospitalized children.

The gestures of a face play an important role in the human-to-human communication. Several recent media applications such as computer games, virtual newsreaders on the internet, and interactive television, make use of a virtual speaking face to support the spoken content. The new trend in the robot industry is to make the robots appear more natural by giving them reactions to observed human emotions and moods. ETRO integrates the audio and visual analysis of speech and gestures to support these applications.

One of ETRO's achievements, in collaboration with the Northwestern University of X'ian (China), is the creation of synthetic mouth shapes of a virtual face corresponding to a spoken sentence. This was done by



Analysis of the motions of a mouth.

exploiting the temporal alignment of the visual lip-shape features and the audio-signal characteristics of the smallest audiovisual speech-units, the visemes. The image analysis used in the training phase of this audio-visual system proceeds as follows: the gesture parameter that represents the amount of opening of the mouth and eyes is a combination of the filtered image window details. The deformations of the finer facial expressions are found by exploiting the correspondence between the motion in the image and the motion in the 3D world of a face virtual model and also by imposing mechanical constraints on the face. With this audio-visual analysis scheme a more plausible talking virtual face is achieved than with a (pure audio) phoneme recognizer. The scheme will be further developed as a lip-reading-aid for the training of a correct pronunciation of hearing-impaired children.

Further, a system for the emotion and intent recognition and synthesis in verbal communication with similar integration of speech and image processing is being developed. Machine learning techniques have already been applied to recognize the emotional intent in a voice. By connecting the visual emotion features to the audio characteristics of numerous examples of expressive speech, it is expected that a better separation of the different emotional states can be detected. ETRO is currently expanding the system with the synthesis of syllables of a synthetic voice carrying emotions. For this purpose the features of pitch, rhythm and loudness of expressive speech examples are transferred to a voice. The plan is to attach a synchronized synthesis of the visual counterpart of the expression, based upon the previously mentioned emotion recognition results. This study focuses on following case-studies in both verbal and nonverbal interaction, among them the 'Anty' social robot which is meant to interact with hospitalized children based on their mood. Anty is being developed in collaboration with the department of robotics and multibody mechanics of the University of Brussels (VUB). With the audio-visual emotion system, ETRO is designing a voice for Anty to speak to the children with comfort, compassion, encouragement,... or to sing them a song. A secondary application of the 'prosody-transplantation' part of the system

Contacts and further information:

- Hichem Sahli, Werner Verhelst, Ilse Ravyse, J. Cornelis
VUB-Department of Electronics and Informatics.
- Audio-Visual Signal Processing, Machine Vision and Digital Speech and Signal processing Groups
www.etro.vub.ac.be
- 'Facial Analysis and Synthesis', I. Ravyse, PhD. Thesis, May 2006.
- 'Context dependent viseme models for voice driven animation', X. Lei, J. Dongmei, I. Ravyse, W. Verhelst, H. Sahli, V. Slavova, Z. Rongchun, 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications, EC-VIP-MC 2003, pp. 649-654.
- Audio-Visual Systems FWO WOG: www.av.s.vub.ac.be
- Anty: <http://anty.vub.ac.be/>