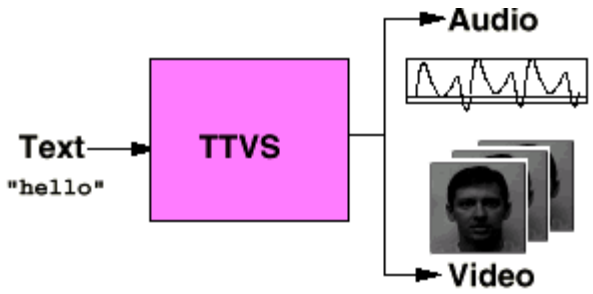
MIT9904-15
Adaptive Man-Machine Interfaces

Tomaso Poggio

Project Overview

We proposed two significant extensions of our recent work on developing a text-to-visual-speech (TTVS) system (Ezzat, 1998). The existing *synthesis* module may be trained to generate image sequences of a real human face synchronized to a text-to-speech system, starting from just a few real images of the person to be simulated. We proposed 1) to extend the system to use morphing of 3D models of faces — rather than face images — and to output a 3D model of a speaking face and 2) to enrich the context of each viseme to deal with coarticulation issues. The main applications of this work are for virtual actors and for very-low-bandwidth video communication. In addition, the project may contribute to the development of a new generation of computer interfaces more user-friendly than today's interfaces. Our text-to-audiovisual speech synthesizer is called MikeTalk. MikeTalk is similar to a standard text-to-speech synthesizer in that it converts text into an audio speech stream. MikeTalk also produces an accompanying visual stream composed of a talking face enunciating that text. An overview of our system is shown in the figure.



Proposal for an extension

We have been successful to attract, as we planned to try to do, Volker Blanz as a part-time postdoc. However, our progress and rates of expenditures have been slower than expected on the 3D subproject because Volker, who will be mainly responsible for it working with Tony Ezzat, has been delayed in joining CBCL and he will join us only part time. He spent two weeks recently at MIT and will come back for two more months during the summer.

Research Plan (July 2000- June 2001)

We plan to:

1) finish development of our new morphable model, which we have already tested with very promising results. With it we will be able to deal with coarticulation. We will demonstrate several experiments with the system and implement a close to real-time version of it.

2) develop and finish the extension of our system to 3D models of faces and produce as output a 3D face, complete with texture. The work will be done in collaboration with Thomas Vetter and Volker Blanz. We will record a 3-dimensional face as it dynamically utters the same visual corpus we have designed and extract 3D visemes. We will do experiments to evaluate the quality of the data and our capability to synthesize new 3D visemes In August and September with Volker Blanz at MIT we will be able to drive a 3D model with the parameters (of the morphable model) supplied by Tony's technique. We have tested the basic approach and produced good quality animations between closed mouth and open mouth (with a 3D face).

3) We will add control of expression in collaboration with Thomas Vetter and Volker Blanz working from University of Freiburg with Tony Ezzat (at MIT).

4) We hope now to be able to animate a synthetic 3D face with high photorealism starting from a single 2D image of the face of a person. It would be interesting to evaluate how faithful the visual speech will be to the specific speaking pattern of an individual.

5) Assess the realism of the talking face. We plan to perform several psychophysical tests to evaluate the realism of our system.

External Collaborators

Thomas Vetter – now Professor in Freiburg – and Dr. Volker Blanz will work with us, partly at MIT but also in Freiburg.