Miketalk: A Talking Facial Display Based on Morphing Visemes
Technology #8102

Applications

MikeTalk is a text-to-audiovisual speech (TTVS) synthesizer system that converts typed text into audio and produces an accompanying visual stream composed of a talking face enunciating the text. TTVS systems have applications as visual desktop agents, digital actors and virtual avatars. They can also be used for special effects, and may be of interest to psychologists who wish to study visual speech production and perception.

Problem Addressed

Previous efforts to construct a human facial model for TTVS have proved difficult due to their reliance on 3D graphics. 3D image modeling cannot capture the subtleties of texture, color and lighting required to create a model that is photorealistic enough for the purposes of a TTVS system. The inventors have developed a method to construct a facial model from images alone, without any underlying 3D graphics. This model displays increased photorealism and is a major component of MikeTalk.

Technology

Video recordings of a subject enunciating a set of key words are obtained to generate a visual collection of each phoneme in the English language. The recorded frames are manually searched to identify and extract a single image per phoneme. The image set is subjectively reduced to a final set of 16 images, which is termed the viseme set. A viseme is the visual counterpart of a phoneme.

A database of optical flow vectors is built to define the transitions from each viseme image to every other viseme image, for a total of 256 optical flow vectors. Given a pair of viseme images and the optical flow vectors between them, intermediate images between the two endpoints are morphed to enable a smooth transition between the viseme images. The final visual sequence comprises the appropriate viseme transitions, played in synchrony with the audio speech signal generated by the TTVS system.

Advantages

- Image-based technique produces higher photorealism in facial model than 3D modeling
- Initial collection of phonemes that need to be recorded is small
- Optical flow method computes correspondence between visemes without need for tedious morphing methods
- Representing visemes as optical flow vectors allows the morphing of as many intermediate images as necessary to maintain synchrony with audio
Categories For This Invention:
Computer Sciences & Information Technology

Intellectual Property:
Talking facial display method and apparatus
Issued US Patent
6,250,928

Inventors:
Tomaso Poggio
Tony Ezzat

Publications:
Miketalk: A Talking Facial Display Based on Morphing Visemes
Computer Animation
98 Proceedings. pp. 96-102, 06 August 2002
Visual Speech Synthesis By Morphing Visemes
International Journal of Computer Vision
38(1), 45-57

External Links:
Poggio Lab
http://poggio-lab.mit.edu/

Image Gallery: