

Talking Faces - Technologies and Applications

Jörn Ostermann, Axel Weissenfeld, and Kang Liu

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Germany

Abstract

Facial animation has been combined with text-to-speech synthesis to create innovative multimodal interfaces. In this lecture, we present the technology and architecture in order to use this multimodal interface in a web-based environment to support education, entertainment and e-commerce applications. Modern text to speech synthesizers using concatenative speech synthesis are able to generate high quality speech. Face animation uses the phoneme and timing information provided by such a speech synthesizer in order to animate the mouth. There are 2 basic technologies that are used to render talking faces: 3D face models as described in MPEG-4 may be used to provide the impression of a talking cartoon or human-like character. Sample-based face models generated from recorded video enable the synthesis of a talking head that cannot be distinguished from a real person. Depending on the chosen face animation technology and latency requirements, different architectures for delivering the talking head over the Internet are required for interactive applications.

Keywords: Face animation, visual speech
