Atakan Günal, Emre Ozcelik
2014

Speech to Lip - Abstract

In this project, we have developed a system that generates animated lip motions on a 3D modeled human face that correspond to speech in Turkish. Lip motions are among the most important aspects of human communication. Computerized facial animation has been one of the earliest areas of interest in the field of computer graphics as human facial expression is key to conveying elements of human communication. Lip motions are arguably the most important element of a person's facial expression because speech is the primary mode of communication among humans. It constitutes an important part of how we speak, and it is contains enough of the message during speech that it is possible to understand what someone is saying by lip reading. Therefore, animation of lip movements that correspond to speech is crucial for achieving an effective animation of human face.

In the system we have developed, we have used speech recognition techniques to convert Turkish speech audio to text. The system can use pre-recorded audio files or the user can use the microphone input to create a new recording each time just before use. Voice activation detection is applied to the recording in order to identify when the speech has started and ended. The system later does alignment of the speech text to the recording. This generates the timing information for each phoneme and produces a timed phoneme sequence that will be used in the animation. For the animated 3D face, we have created a 3D model from a 2D portrait image. Each phoneme in the Turkish language maps to one of eight lip shapes called visemes. We have created animation points for each of the eight visemes on the 3D model. Our graphics rendering application reads in the phoneme sequence data and triggers transitions between the animation points according to the mapping between the phonemes and the visemes. The system can be incorporated in many application areas including communication, aiding the hearing-impaired, automatic generation of lip synchronization for animated motion picture and video games.