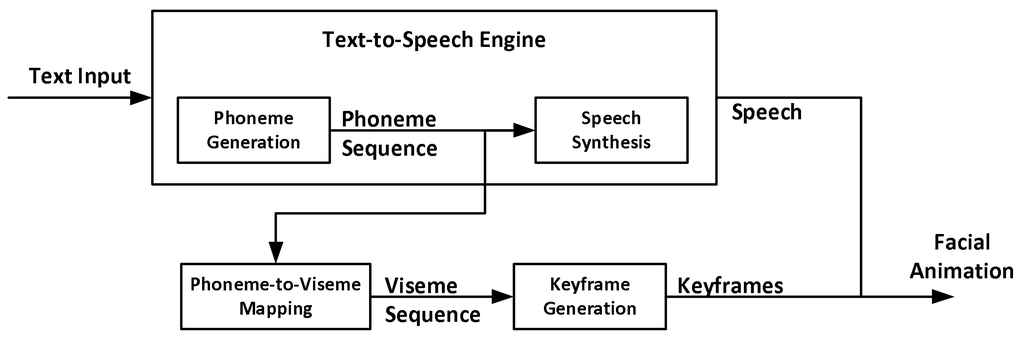
Facial Animation Framework for Web and Mobile Platforms

Engin Mendi, and Coskun Bayrak
Computer Science Department, University of Arkansas at Little Rock

Introduction

- visual interaction improves:
  - speech intelligibility
  - speech perception
- new generation applications
  - intelligent communication systems
  - human-machine interfaces
  - interfaces for handicapped people
- proposed system
  - conversion of the text into 3D face animation
  - synthetic voice
  - synchronization of the head and eye movements with
    - emotions
    - word flow of sentences



Facial Animation

- keyframe based animation
  - keyframe: deformed version of a face shape
  - mapping of visemes (mouth shape) and phonemes (smallest part of a spoken word)
  - interpolation to produce smooth motion
- MPEG-4 animation
  - ISO standard developed by MPEG (Moving Picture Experts Group)
  - 84 feature points (FP)
  - 68 facial animation parameters (FAPs)
  - the face motion is controlled by FAPs driving the animation on the FPs

Text-To-Speech Conversion

- text-to-speech engine
  - Microsoft speech API
  - Java speech API
- speech synthesizing
  - synthetic speech data
  - phonemes with their durations

Viseme Generation

- mapping from the set of phonemes
- blending process
  - interpolation and synchronization of the timing and phonetic parameters
- ω is an arbitrary weight such that $\omega \in [0, 1]$, $\nu(t_0)$ and $\nu(t_1)$ are the vertices of previous and next visemes respectively, and $\nu(t_r)$ designates resultant viseme interpolated using these two.
- interpolation:

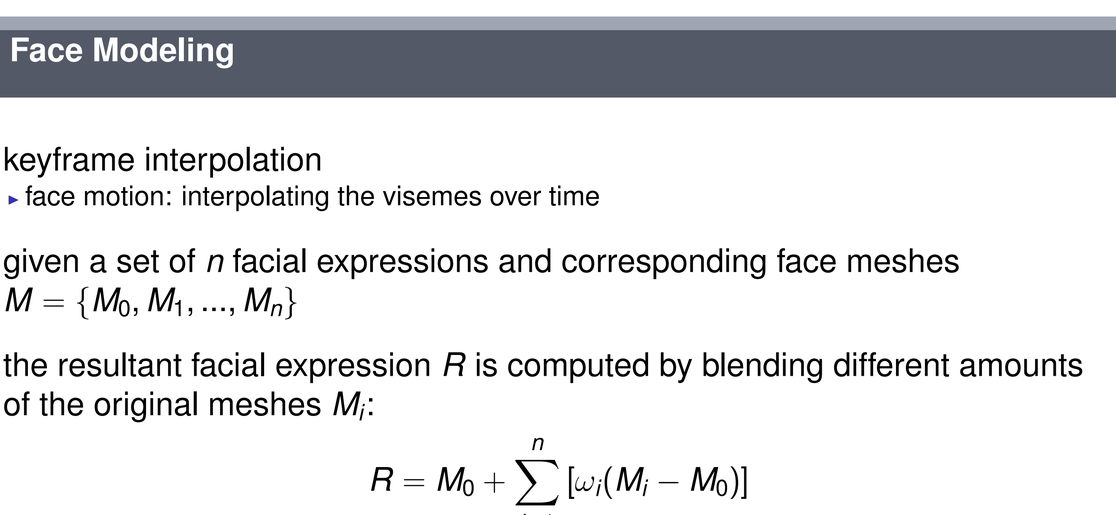
$$\nu(t_r) = \nu(t_0).(1 - \omega) + \nu(t_1).\omega$$

Face Modeling

- keyframe interpolation
  - face motion: interpolating the visemes over time
- given a set of n facial expressions and corresponding face meshes $M = \{M_0, M_1, ..., M_n\}$
- the resultant facial expression R is computed by blending different amounts of the original meshes M_i:

$$R = M_0 + \sum_{i=1}^{n} [\omega_i(M_i - M_0)]$$

- ω_i: arbitrary weights and M_0: neutral expression
- M_i become more exaggerated in R when ω_i gets larger
  - weights and their sum are between $[0, 1]$ avoid exaggeration
- FaceGen editor
  - realistic 3D faces
  - facial expressions
- 3D face models:



Facial Expressions

- tagged text input
  - expression tags (i.e. <joy>Hello<\joy>)
  - turn into given emotion on the face while the face model is speaking
- emotional expressions of Obama face model
  - anger, disgust, joy, fear, surprise, sadness

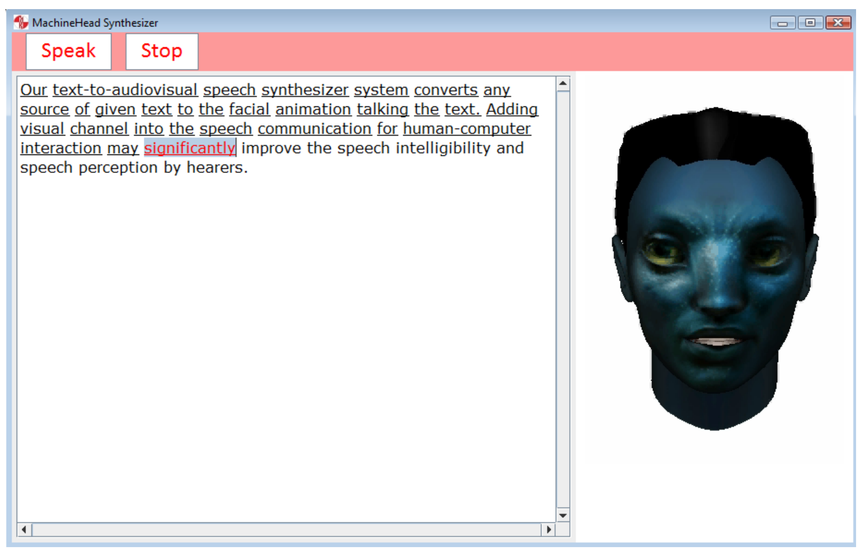
Virtual Face Synthesis

- generation of lip movements with emotional expressions corresponding to speech
- speech-visual synchronization
  - conversion of the input sentence into phonemes and speech wave
  - phonemes are mapped into visemes and sent to the face model to realize lip movements

Virtual Talking Face Environments

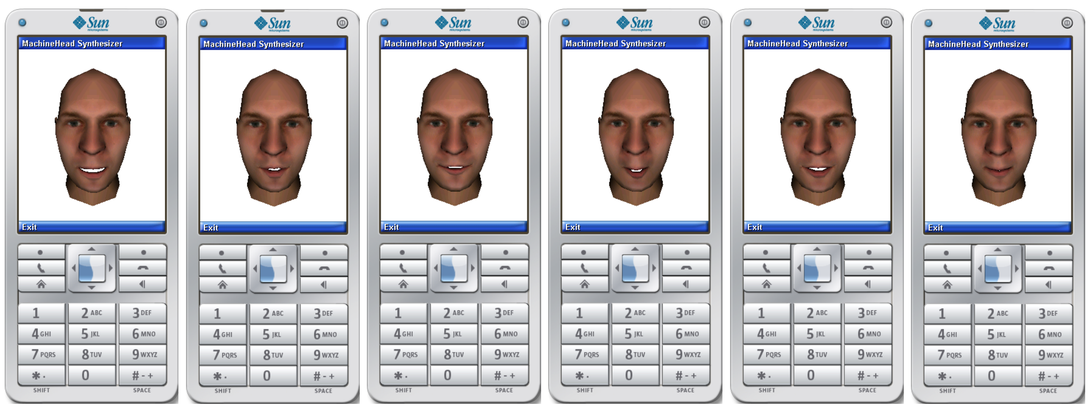
- web environment
  - speaks free text typed, text files and internet sites
  - speed and pitch control
  - language support

- animation pad
  - speaks text files and highlight words currently being spoken.



- mobile environment
  - OpenGL ES (OpenGL for embedded systems)
  - mobile emulator
  - emotional expressions: anger, disgust, joy, fear, surprise, sadness:

- fragments during speech:



Computer Science Department, University of Arkansas at Little Rock, Little Rock, AR, USA
Mail: {esmendi, cxbayrak}@ualr.edu