<table>
<thead>
<tr>
<th>Title</th>
<th>Creation and animation of a talking head with lip sync and expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Chikersal, Prerna</td>
</tr>
<tr>
<td>Citation</td>
<td>Chikersal, P. (2013, March). Creation and animation of a talking head with lip sync and expressions. Presented at Discover URECA @ NTU poster exhibition and competition, Nanyang Technological University, Singapore.</td>
</tr>
<tr>
<td>Date</td>
<td>2013</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/11313">http://hdl.handle.net/10220/11313</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2013 The Author(s).</td>
</tr>
</tbody>
</table>
Creation and Animation of a Talking Head with Lip Sync and Expressions

Introduction and Motivation

A Talking Head is an animated model of a human head with synchronized lip movements and expressions. Back in the 1970s, Talking Heads were a major breakthrough in Human Computer Interaction. Ever since, they’ve been used as conversational agents in offline and web-based applications and as a teaching tool to enhance learning in students. Animated talking guides, like the ones used in MS Office, demonstrate a friendly and effective way to assist inexperienced users.

Though, this topic has already been explored, there are certain linguistic challenges which impede the creation of an accurate, realistic and expressive Talking Head. With further advancements, it may be possible to generate accurate Talking Heads which can be lipread by the hearing impaired. Moreover, recent research shows that emotionally expressive avatars facilitate learning in people with autism spectrum disorder.

Phonemes and Visemes

Phonemes are the basic distinctive units of speech sound, which are different for every language. Visemes are generic facial images that are used to describe a particular sound or phoneme. For example:

Viseme “F” = “F” OR “V” in words like “Four” (F AO R) and “Van” (V AE N).

Objectives

The scope of this project is:
1. To create a custom Talking Head, given a person’s photograph and to synchronize the visemes according to the input text.
2. To allow the user to add in expressions as tags into the input text and to produce these expressions in the Talking Head.
3. To solve the three challenges described below.

Challenges

1. Problem of Coarticulation

Coarticulation refers to the changes in the articulation of a phoneme depending on preceding (backward coarticulation) and upcoming segments (forward coarticulation). For example:

- Saying the word “Soon” without Coarticulation. (Different from human beings; hence, inaccurate)
- Saying the word “Soon” with Coarticulation. (Similar to human beings; hence, accurate and what we want to achieve)

2. Variation in intensity of emotions

Intensity of emotions greatly affect the facial movements. So, knowing which expression to enact is not enough; we also need to know the degree of emotion behind it. For example:

User’s expression is “smile”.
But, how happy is he? How much should his avatar smile?

3. Overlap of expressions and visemes

What if you are very angry and pronouncing the phoneme “P” (eg. “Pay”) simultaneously?

Methodology

For continuous 3D facial animation, we use a blendshape approach. Given a set of n facial expressions or visemes and corresponding polygonal meshes, B = {B_1, B_2, B_3, ..., B_n} called blendshapes, we can create new expressions by blending different amounts of the original meshes B_i, using:

B_mixed = \sum w_i \times (B_i - B_0),

where w_i are arbitrary weights and B_0 corresponds to a neutral expression.

Project Title: Talking Head (Lip Sync) Application
Supervisor: A/P Chng Eng Siong