Data-Driven Facial Modeling and Animation

1. Research Team

Project Leader:	Prof. Ulrich Neumann, *IMSC and Computer Science*
Post Doc(s):	John P. Lewis
Graduate Students:	Zhigang Deng
Undergraduate Students:	Pamela Fox

2. Statement of Project Goals

Avatars, computer animation, and other forms of human-centric computing all potentially require the generation of novel but realistic facial appearance and motion. Such face modeling and animation has recently been the province of computer artists. While manually animated computer faces can be quite expressive, they usually fall short of complete realism in their appearance and motion (the recent movie *Final Fantasy* is one demonstration of the current state of the art). The manual effort required to model and animate faces also prevents their use in avatars and other contexts where the required motion must be automatically generated.

The Data-Driven Facial Modeling and Animation project explores the use of facial models made directly from motion capture data to address the goals of realism and automation. More specifically, a range of typical face motion data is first captured. This data is then modeled with machine learning techniques. Once a model is available, its output and parameters can be explored to extrapolate or synthesize novel motions. While this general data-driven approach is a recent theme across several research groups, IMSC already has some unique results, as shown below.

3. Project Role in Support of IMSC Strategic Plan

Data-Driven Facial Modeling and Animation is part of the general IMSC effort towards expressive human interaction in virtual and augmented reality environments. The complementary Model-Based Face Computation project is directed toward exploiting prior knowledge of human face structure in various automated face processing tasks.

4. Discussion of Methodology Used

The human face is of paramount importance to other humans, and we can effortlessly make refined judgments about the attentional and emotional qualities conveyed by a speaker. The eyes and mouth are the parts of the face that are most directly used to convey this information, and are the areas that are most attended to by an observer. We have developed a three-dimensional

avatar whose animation is driven from novel audio (speech) using a data-driven approach to synthesizing the crucial mouth and eye movements.

To obtain the mouth motion model, facial motion capture data is analyzed to obtain a movement subspace, then operating in this subspace a smooth mouth movement path for any given utterance is derived from the data using machine learning techniques (Figure 1). The synthesized motions closely resemble ground-truth data obtained from motion capture (Figure 2).

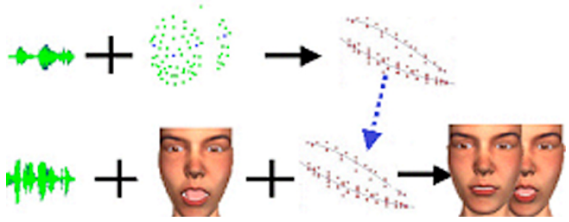

Figure 1. Overview of data-driven speech motion modeling. Motion capture data (top row) is analyzed to produce a movement model. Given novel audio (bottom row), this model is employed to generate audio-synchronized mouth motion.

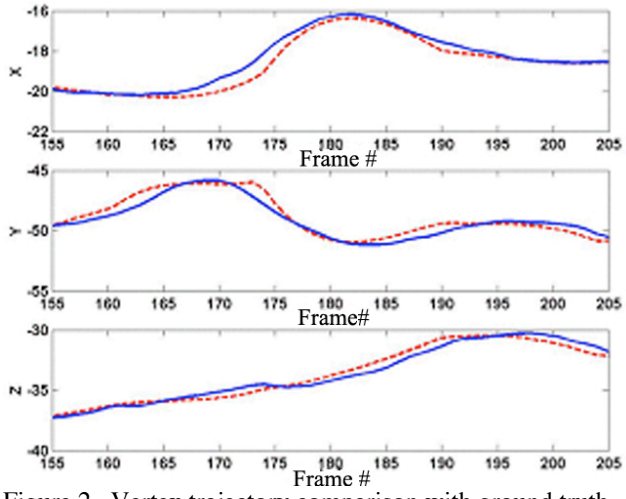

Figure 2. Vertex trajectory comparison with ground truth motion data (not part of the training data). The solid blue line denotes the motion capture; the dashed red line denotes synthetic motion driven by the same audio.

Eye movements and blinks are not strongly correlated with speech sounds, so we adopted a statistical texture synthesis method to capture this independence while still producing characteristic movements. Specifically, we employ the popular non-parameteric sampling method [1], operating on the one-dimensional eye-blink signal (see Figure 3) and the quaternion eye orientation signal (data obtained from motion capture and hand labeling). This class of texture synthesis methods is powerful, and we have shown that the general approach transfers to

movement synthesis as well. Most observers cannot distinguish synthesized motion from the original data played back on the same model.

This technique has the potential of giving the impression of an avatar having different mental states or "moods".

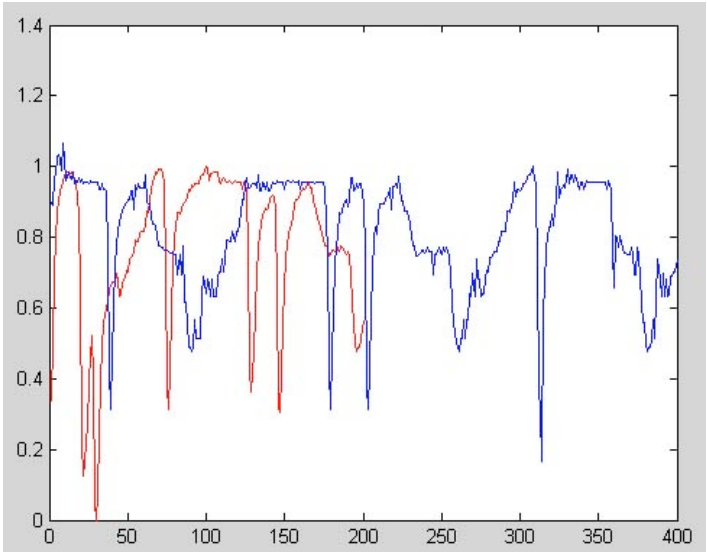

Figure 3. Motion capture data for an eye blinks over time (red graph), and novel synthesized eye blink motion (blue graph).

5. Short Description of Achievements in Previous Years

This project was initiated in 2003. Prototypes of the data-driven speech and eye movement models were completed, and an extended abstract on the eye-movement model was published [2].

5a. Detail of Accomplishments During the Past Year

Both the speech and the eye-movement models have been refined, and the eye movement model has been competed and published [3].

6. Other Relevant Work Being Conducted and How this Project is Different

"Data driven" techniques are a current trend in computer graphics research, and they have been applied to speech modeling as well as other areas such as body animation and texture synthesis. Realistic facial animation demands attention to details such as speech and eye movement. We are the first group applying data-driven modeling to these phenomena.

7. Plan for the Next Year

The data-driven approach is being extended to the animation of facial expressions, and head movement. The result will be a complete data-driven animated head model.

8. Expected Milestones and Deliverables

The current deliverables include roughly 15,000 lines of C++. Our future milestones are indicated in the following table:

- Data-driven facial expression synthesis (continuing)
- Data-driven head movement
- Integration of all data-driven facial animation sub-projects to produce an expressive avatar prototype.

9. Member Company Benefits

Avatars have been proposed for a variety of purposes, and there is experimental evidence that avatars help in some interface scenarios [4]. Current avatars are crude and unexpressive, however; their strange eye movement has caused some observers to describe them as being "drugged" or "schizophrenic".

The mouth and eye movement models developed in this project, derived from real human data, can help member companies realize realistic, pleasant, and expressive avatars.

10. References

[1] A. A. Efros and T.K. Leung, Texture Synthesis by Non-parametric Sampling, *ICCV'99*, 1999, 1033—1038.
[2] Z. Deng, JP Lewis, and U. Neumann, Practical Eye Movement Model using Texture Synthesis, *ACM SIGGRAPH* technical sketch, San Diego, 2003.
[3] Z. Deng, J. P. Lewis, and U. Neumann, Automated Eye Motion using Texture Synthesis, *IEEE Computer Graphics and Applications,* accepted for publication.
[4] I. S. Pandzic, J. Ostermann, and D. Millen, User Evaluation: Synthetic Talking Faces for Interactive Services, *The Visual Computer*, 15,1999, 330—340.