C.50.615
4

**Final Report for Period:** 07/2000 - 06/2005 **Submitted on:** 08/17/2006

**Principal Investigator:** Essa, Irfan . **Award ID:** 9984847

**Organization:** GA Tech Res Corp - GIT

**Title:**

CAREER: Developing and Evaluating a Spatio-temporal Representation for Analysis, Modeling, Recognition and Synthesis of Facial Expressions

## Project Participants

### Senior Personnel

**Name:** Essa, Irfan
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

### Post-doc

**Name:** Reveret, Lionel
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

### Graduate Student

**Name:** Chen, Alan
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

**Name:** Haro, Antonio
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

**Name:** Xiao, Jun
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

**Name:** Sukel, Kayt
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

**Name:** Ruddarraju, Ravikrishna
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

Started working as a UG researcher. Continued as a PhD student in Electrical and Computer Engineering. Worked on Eye and gaze Tracking

**Name:** Ramanarayanan, Ramji
**Worked for more than 160 Hours:** Yes
**Contribution to Project:**

**Name:** Yin, Pei
**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

Worked as a GRA on this project and then was moved to the NSF ITR project on Audio Visual Fusion. Worked on feature selection and co-training for lip reading from video and audio

**Name:** Huang, Yan

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

Worked on facial tracking as an MS student. In part funded by this project.

**Name:** Brubaker, Stephanie

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

Worked as a GRA on deformable modeling, applied to faces. Won a NSF Graduate Fellowship

**Undergraduate Student**

**Name:** Carter, Scott

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

**Name:** Keenan, Timothy

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

**Name:** Hays, James

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

**Name:** Narayanan, Divya

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

**Name:** Hable, John

**Worked for more than 160 Hours:** Yes

**Contribution to Project:**

**Technician, Programmer**

**Other Participant**

**Research Experience for Undergraduates**

## Organizational Partners

**IBM Almaden Research Center**

Have had interactions with Myron Flickner and David Koons at IBM Almaden Research Center

**Microsoft Corporation**

## Other Collaborators or Contacts

Myron Flickner and David Koons, IBM Almaden Research Center
Brian Guenter, Microsoft Research
Catherine Pelachaud, Universita di Roma 'La Sapienza'
Lionel Reveret, INRIA, France

At Georgia Tech:
Richard Catrambone, John Stasko, Gregory Abowd, Elizabeth Mynatt, James Rehg

## Activities and Findings

**Research and Education Activities: (See PDF version submitted by PI at the end of the report)**

YEAR THREE:
The research funded by this CAREER grant has progressed on four primary areas in the past year. These areas and the specific advances are briefly listed here.

1) Spatio-temporal analysis of speech action from video
One main focus of our effort is to analyze spatio-temporal relationship between how lips move in generation of speech, and the audio signal itself. By analysis of the video and the related audio channel, we seek to build a compact representation for visual speech actions that is suitable for synthesis and recognition of lip movements with speech. We have also added the ability to track these lip shape changes with facial expressions.

2) Eye and Head Tracking
One main constraint faced by existing facial expression (and lip motion) tracking systems is that we need to align the face to appear to have no rigid translations and rotations. This allows for tracking of the nonrigid motion separately. Towards this end, we have focused on several eye and head-pose tracking systems. In this year of our effort, we have built a multi-camera system to track eyes robustly and then by using triangulation extract head orientation. The multi-camera aspect of our approach also allows for a larger area for tracking faces. We have also added Fischer-Discriminats and replaced PDAs which allows for tracking in varying lighting conditions.

3) Toolkits for Facial Modeling and Animation
We are also building various XML-based facial animation toolkits that merge audio and facial action synthesis to generate agents with faces that can talk and make expressions. Our approach is NOT limited to just graphical face models and includes simple robotic systems like the PONG robot provided by IBM Research. In addition, we have also built methods for image-based generation of realistic face models that can be animated.

4) Evaluation of faces in interfaces
One significant aspect of our work was to study how humans also perceive expressions, especially within the context of face-to-face interactions. In this work, we took a slightly different approach this year and chose to collaborate with Professors Catrambone and Stasko, who were interested in studying the importance of faces in an interface. By collaborating with this larger effort by the above-mentioned researchers, we are able to pursue our interest in the modeling and animation of faces in an interface and have successfully shown both (a) the importance of faces in different settings and (b) need for better toolkits and methods discussed in avenues 1 and 2 above.

YEAR TWO:
The primary thrust of our research is the detailed modeling and analysis of a spatio-temporal representation of facial movements. Within this effort, during the last year, we have made some significant progress in the areas of modeling for synthesis skin, analysis and synthesis of speech actions, and evaluations of faces in interfaces, primarily for the purposes of building conversational agents. Our work last year has benefitted extensively from other funding sources, primarily industrial grants from Microsoft, equipment support from IBM, an infrastructure grant from NSF and funding from DARPA on Human Identification.

Modeling of Skin:
Skin is noticeably bumpy in character, which is clearly visible in close-up shots in a film or game. Methods that rely on simple texture-mapping of faces lack such high frequency shape detail, which makes them look non-realistic. More specifically, this detail is usually ignored in real-time applications, or is drawn in manually by an artist. We have developed techniques for capturing and rendering the fine scale structure of human skin. We introduce a method for creating normal maps of skin with a high detree of accuracy from physical data. We use

techniques inspired by texture synthesis to 'grow' skin normal maps to cover the face. Finally, we demonstrate how such skin models can be rendered in real-time on consumer-end graphics hardware.

Analysis and Synthesis of Spatio-temporally coherent, Facial-Speech Actions:
We have developed an Animated speakers Kit that allows for automatic generation of 2D and 3D facial animation directly from the analysis of any speaker's video sequence. It results in a perfectly lip-synched animation as the facial motion is continuously captured and coded in terms of four phonetically-oriented parameters, or Facial Speech Parameters (FSP). The system is based on an analysis-by-synthesis method. An accurate 3D model is learned off-line from an expert phonetician subject to statistically learn the fundamental bio-mechanical degrees of freedom in speech production expressed by the FSP. Using a morphological adaptation, these degrees of freedom are re-mapped on any new subject to analyze his facial motion. The video analysis consists in an optimization procedure which aligns a textured version of the 3D model and the incoming video footage. The facial motion coded by the FSP is re-targetted to 2D and 3D animation. The hypothesis explored in this work is to test if the coding of facial morphology and the coding of facial motion can be separated. Currently, we explored this hypothesis in the domain of speech production with this set of FSP parameters (articulatory hypothesis), which have shown some speakers independent properties.

Evaluation of Anthropomorphic Interfaces:
In this work, we propose a framework for studying anthropomorphic agents that can systematize the research and address the limitation caused by insufficient consideration of key factors that influence the perception and effectiveness of agent-based interfaces. The framework emphasizes features of the agent, the user, and the task the user is performing. Our initial experiment within this framework manipulated the agent's appearance (lifelike versus iconic) and the nature of the user's task (carrying out procedures versus providing opinions). We have found that the perception of the agent was strongly influenced by the task while features of the agent that we manipulated had little effect.

In the coming year of this project, one intention is to combine the above three (3) important phases of our onging effort. The animated speakers toolkit will be used to support easy generation of anthropomorphic agents, which will appear both realistic (using the skin models) and also support non-photo-realistic facial models. We are also at present collaborating with IBM research to get access to their Pong face robots. These robots include cameras that allow for eye tracking. We are now incorporating our earlier work on eye-tracking with these robots to pursue more avenues of face-to-face interaction between humans and robots.

A recent acquisition of a 12 Camera VICON Motion Capture System with functionality to capture both whole-body and facial actions will further aid in the development and validation of detailed facial movements in the upcoming years.

YEAR ONE:
*DATA COLLECTION
1) Collect data of people speaking in natural settings.
2) We are video taping actors who are being engaged in a dialogue to get emotive responses during an interaction.
3) Acquire expression and facial gestures associated with speech.
4) So far, about 10 subjects recorded. This will continue in the next year.
5) Catherine Pelachaud, Universita di Roma 'La Sapienza' is helping with developing protocols for capture of relevant actions.

*MODELING
1) A detailed 3D model of faces that allows for robust modeling of lips.
2) Learn the topology of facial actions associated with speech by analyzing an expert phonetician (Lionel Reveret) to generate a parameter set for speech actions.
3) Radial Basis function approach to warp the generic model to any face in a photograph.
4) An easy method for reliable animation of faces from video data.

*EDUCATION
1) Undergrads undertook projects to take the modeling mentioned above and incorporate it into Alias/Wavefront's MAYA for interactive animation.

**Findings: (See PDF version submitted by PI at the end of the report)**

YEAR THREE
* A compact representation that combines facial movements associated with speech and speech can be extraced from video and can be used for recognition and synthesis.

* Using multiple vision-based eye-trackers, with robust light-independent templates, allows for robust head pose (and eye-gaze) tracking over a larger area. We have also tested this method over extended periods as a part of a simple pilot study.

* A unified and perhaps XML based toolkit is needed to generate realistic facial motions, which can be rendered using realistic and nonphotorealistic graphics models and physical robots with faces.

* Faces are important not just in face-to-face communications, ut also in human-machine interactions, using an anthrophometric agent. We are working on our collaborators to test this and are providing toolkits and other expertise as needed.

YEAR TWO
See year two under research activities

YEAR ONE
* A set of parameters that is learned from data for representing (and therefore recognizing and synthesizing) facial actions associated with speech. These parameters are 'universal' and appear to correlate (given time-alignment) between subjects. We are at present working with 4 such parameters suitable for recognizing lip shapes (and eventually lip-reading) from frontal images.

* A very reliable method facial action tracking from video. This method is robust to lighting varying during an expression and works for subjects of a variety of races.

* An easy extension to reliable animation from video data, with coherent lip shapes.

**Training and Development:**

i. This project is quite interdisciplinary in its nature, providing education and direct research experience in imaging, graphics, and HCI.

ii. Last year, Dr. Lionel Reveret, with Divya Narayanan and Tim Keenan, worked on this effort. Dr Lionel was a post-doctoral researcher and then joined as Assistant Professor/Researcher at INRIA in Grenoble France and still continues to collaborate with PI.

iii. Divya was an UG student who has since graduated and is now a PhD student in bio-imaging at Johns-Hopkins.

iv. Tim Keenan, who worked on the animation part of the project is now a Technical Director at Dreamworks Animation, with credits in movies like Shrek II, Over the Hedge, Madagasscar.

v. Antonio Haro has completing his PhD this year and is working for Nokia Research in Dallas Texas. One of his projects involves faces on a mobile phone.

vi. Ravi Ruddarraju finished his BS in EE and is now continuing his PhD dissertation under the PI at GA Tech.

vii. John Hable finished his BS and MS is CS and is now working on facial capture for Games at EA.

viii. Jun Xiao has just finished his PhD.

ix. Yan Huang has finished her MS and now works for Google.

x. Pei Yin is about to propose for his dissertation and has had two successful internships at Microsoft Research.

**Outreach Activities:**

i. The toolkits discussed earlier are going to be essential in taking this kind of laboratory research to the outside world. We are talking to a few high-school students who could use these toolkits for their own animations.
ii. Several art schools and also production houses have inquired about our software system for use in production. One of the groups in an animation and special effects classes at GA Tech have already used these tools. Due to limitation in supporting such activities, it is unlikely that this will have the impact it could have.

## Journal Publications

Lionel Reveret and Irfan Essa, "Visual Coding and Tracking of Speech Related Facial Motions", Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, 2001, p. , vol. , ( ). Submitted

A. Haro, B. Guenter, I. Essa, "Real-time, Photorealistic, Physically Based Rendering of Human Skin Microstructure", Proceedings of Eurographics Rendering Workshop, London, England, p. 1, vol. 1, (2001). Accepted

A. Haro, I. Essa, M. Flickner, "Detecting and Tracking Eyes by Using their Physiological Properties, Dynamics and Appearance", Proceedings of IEEE Computer Vision and Pattern Recognition Conference 2000, Hilton Head SC, p. 1, vol. 1, (2000). Published

A Haro, I. Essa, M. Flickner, "A Non-Invasive Computer Vision System for Reliable Eye Tracking", Proceedings of ACM CHI Conference, The Hague, Netherlands, p. 1, vol. 1, (2000). Published

A Haro, B. Guenter, I Essa, "Real-time, Photo-realistic, Physically Based Rendering of Fine Scale Human Skin Structure", Proceedings 12th Eurographics Workshop on Rendering, London, England, p. 1, vol. 1, (2001). Published

Jun Xiao, "Understanding the Use and Utility of Anthropomorphic Interface Agents", Proceedings, Extended Abstracts, CHI 2001. Student poster and paper., p. 1, vol. 1, (2001). Published

L. Reveret, I Essa, "Visual Coding and Tracking of Speech Related Facial Motion", Proceedings IEEE CUES in Communications 2001 Workshop, p. 1, vol. 1, (2001). Published

Jun Xiao, John Stasko, Richard Catrambone, "Embodied Conversational Agents as a UI Paradigm: A Framework for Evaluation", AAMAS 2002 proceedings, p. 0, vol. 1, (2002). Accepted

Richard Catrambone, John Stasko, Jun Xiao, "Anthropomorphic Agents as a User Interface Paradigm: Experimental Findings and a Framework for Research", Proceedings of CogSco 2002, p. 1, vol. 1, (2002). Published

Jun Xiao, John Stasko, Richard Catrambone, "Embodied Conversational Agents as a UI Paradigm: A Framework for Evaluation", Proceedings of Embodied Conversationa Agents for AAMAS 2002, p. 1, vol. 1, (2002). Published

Jun Xiao, John Stasko, Richard Catrambone, "Be Quiet? Evaluating Proactive and Reactive User Interface Assistants", Proceedings of INTERACT 2003, p. , vol. , (2003). Accepted

Vivek Kwatra, Arno Schoedl, Irfan Essa, Greg Turk, Aaron Bobick, "GraphCut Texture: Image and Video Synthesis Using Graph Cuts", ACM Transaction on Graphics (issue of the Proceedings of ACM SIGGRAPH 2003 Conference), p. , vol. , (2003). Accepted

Ravi Ruddarraju, Antonio Haro, Irfan Essa, "Fast Multiple Camera Head Post Tracking", Proceedings Vision Interface 2003, Halifax, Canada, p. , vol. , (2003). Accepted

Ravi Ruddarraju, Antonio Haro, Kris Nagel, Irfan Essa, Gregory Abowd, Elizabeth Mynatt, "Perceptual User Interfaces using Vision-based Eye Trackers", International Conference on Multimodal and Perceptual User Interfaces, Vancouver BC, November 2003, p. , vol. , ( ). Submitted

Lionel Reveret, Irfan Essa, Bailley, "Characterization of Speech Actions for Facial Animation", Visual Computer Journal, 2006, p. , vol. , ( ). in preparation

Pei Yin, Irfan Essa, James M. Rehg,, "Asymmetrically Boosted HMM for Speech Reading", IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2004), p. 755, vol. II, (2004). Published

Pei Yin, Irfan Essa, James M. Rehg, "Boosted Audio-Visual HMM for Speech Reading", IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003), p. 68, vol. , (2003). Published

## Books or Other One-time Publications

## Web/Internet Site

**URL(s):**

? http://www.cc.gatech.edu/cpl/projects/animated-speakers/

**Description:**

? http://www.cc.gatech.edu/cpl/projects/graphcuttextures/
? http://www.cc.gatech.edu/cpl/projects/multicameyetracking/
? http://www.cc.gatech.edu/cpl/projects/pupil
? http://www.cc.gatech.edu/cpl/projects/animated-speakers/
? http://www.cc.gatech.edu/gvu/perception/projects/speechreading

## Other Specific Products

**Product Type:**

**Software (or netware)**

**Product Description:**

Animated Speakers TookKit: A toolkit that allows for extraction of facial speech actions and then can be used to generate animations for any 3D model provided with appropriate correspondences.

**Sharing Information:**

making it available via a database of motion capture data

**Product Type:**

**Data or databases**

**Product Description:**

Considerable video data of subjects engaging in various speech tasks that will be made available to the community in the next two years

**Sharing Information:**

through a database library that we are working on with professors at Carnegie Mellon University

## Contributions

**Contributions within Discipline:**

a. To the development of your own discipline(s)?
As noted, this work covers the areas of imaging (analysis and interpretation), graphics, and HCI. Its overall goal is to aid in the analysis and synthesis of facial movements and in that context, our work continues to extend the state of the art in automatic approaches for analysis and synthesis of faces. We do not think we need to emphasize the importance of faces in general. This basic importance makes it essential to continue such studies within the context of computer science research.

**Contributions to Other Disciplines:**

b. To other disciplines of science or engineering?
The primary goal is to study how people make expressions. We hope to pursue more research on using the techniques we have developed here for NIH and other behavioral science type of research. The PI was recently invited to a DARPA planning workshop of virtual face modeling to support facial reconstruction of the injured. He established a team of experts from imaging, and biomedical modeling and put in a proposal to DARPA, which he is expecting to hear back on. It is not clear if DARPA will fund such an effort, however, there are avenues to pursue with NIH.

**Contributions to Human Resource Development:**

The effects of this work will have impact on human resources after some additional study and evaluation.

**Contributions to Resources for Research and Education:**

**Contributions Beyond Science and Engineering:**

The long-term implications to Behavioral Sciences can aid in some important contributions towards medical and mental health issues. Lip-reading work is essential for aiding hearing-challenged individuals.

**Categories for which nothing is reported:**

Any Book

Contributions: To Any Resources for Research and Education

**NSF Project Number: 9984847**
**CAREER: Developing and Evaluating a Spatio-temporal Representation for Analysis, Modeling, Recognition and Synthesis of Facial Expressions**
DATEs: July 1, 2000 - June 30, 2005

1. **Participants: Who has been involved?**
   a. What people have worked on the project?
      i. Irfan Essa, PI, Associate Professor,
         - American Citizen, Male, South Asian
         - PI
      ii. Lionel Reveret, Post-Doctoral Fellow,
         - French, Male, Caucasian.
         - Worked on lip motion modeling from audio/video
      iii. Antonio Haro. GRA, PhD student,
         - American, Male, Hispanic
         - Worked on Head-tracking, Eye-tracking, and skin modeling
      iv. Jun Xiao, GRA, PhD Student,
         - Chinese, Male, Asian
         - Worked on Believable Agents in User Interfaces
      v. Ramji Ramnarayanan, GRA, MS student,
         - Indian, Male, South Asian
         - Worked on using IBM PONG Face Robot
      vi. Ravikrishna Ruddarajju, UROC, GRA, UG & MS student,
         - Indian, South Asian
         - Worked Eyetracking for Visual Interfaces
      vii. Stephanie Wojtkowski (Brubaker), GRA, PhD Student,
         - American, Female. Caucasian
         - Worked on Deformable Modeling
      viii. Timothy Keenan, BS Student
         - American, Caucasian
         - Worked on Lip Modeling
      ix. Yan Huang, GRA, MS Student
         - Chinese, Female, Asian
         - Worked on Tracking of faces
      x. Pei Yin, GRA, PhD Student
         - Chinese, Male, Asian
         - Worked on speech/lip-reading
      xi. Divya Narayanan. UG Student
         - Indian, Female, South-Asian
         - Worked on Audio Analysis
   b. What other organizations have been involved as partners?
      i. IBM Almaden Research (Eye tracking / PONG Robot)
         - Myron Flickner,
         - David Koons
      ii. Microsoft Research Redmond

- Brian Guenter

c. Have you had other collaborators or contacts?

i. Georgia Tech

- Richard Catrombone
- John Stasko
- Gregory Abowd
- Beth Mynatt
- James Rehg

**2. Activities and Findings: What have you done? What have you learned?**

a. What were your major research and education activities?

The research funded by this CAREER grant has progressed on 4 primary areas during the course of the project. These areas and the specific advances within are each are briefly listed here.

Spatio-temporal analysis of speech action from video: One main focus of our effort is to analyze spatio-temporal relationship between how lips move in generation of speech, and the_audio signal itself. By analysis of the video and the related audio_channel, we seek to build a compact representation for visual speech_actions that is suitable for synthesis and recognition of lip movements_with speech. We have also added the ability to track these lip shape_changes with facial expressions. In some recent work, we have added machine learning techniques to undertake feature selection, which in turn helps with merging information for improved speech reading research. This latter work is now being continued under an NSF ITR GRANT with Professor Jim Rehg

Eye and Head Tracking: One main constraint faced by existing facial expression (and lip motion) tracking systems is that we need to align the face to appear to have no rigid translations and rotations. This allows for tracking of the nonrigid motion seperately. Towards this end, we have focused on several eye and head-pose tracking systems. In this year of our effort, we have built a multi-camera system to track eyes robustly and then by using triangulation extract head orientation. The multi-camera aspect of our approach also allows for a larger area for tracking faces. We have also added Fischer-Discriminats and replaced PCAs which allows for tracking in varying lighting conditions. Some of the approaches developed here are in use by other researchers in the fields of HCI.

Toolkits for Facial Modeling and Animation: We are also building various XML-based facial animation toolkits that_merge audio and facial action synthesis to generate agents with faces_that can talk and make expressions. Our approach is NOT limited to_just graphical face models and includes simple robotic systems like the_PONG robot provided by IBM Research. In addition, we have also built_methods for image-based generation of realistic face models that can be_animated.

Evaluation of faces in interfaces: One significant aspect of our work was to study how humans also perceive expressions, especially within the context of face-to-face interactions. In this work, we took a slightly different approach this year and choose to collaborate with Professors Catrambone and Stasko, who were interested in studying the importance of faces in an interface. By collaborating with this larger effort by the above-mentioned researchers, we are able to pursue our interest in the modeling and animation of faces in an interface and have successfully shown both (a) the importance of faces in different settings and (b) need for better toolkits and methods discussed in avenues 1 and 2 above.

b. What are your major research findings?

    i. A compact representation that combines facial movements associated with speech and speech can be extracted from video and can be used for recognition and synthesis.

    ii. Using multiple vision-based eye-trackers, with robust light-independent templates, allows for robust head pose (and eye-gaze) tracking over a larger area. We have also tested this methods over extended periods as a part of a simple pilot study

    iii. A unified and perhaps XML based toolkit is needed to generate realistic facial motions, which can be rendered using realistic and non-photorealistic graphics models and physical robots with faces.

    iv. Faces are important not just in face-to-face communications, but also in human-machine interactions, using an anthropometric agent. We are working on our collaborators to test this and are providing toolkits and other expertise as needed.

    v. Cooperative learning methods are needed to improve recognition accuracies for lip shapes related to speech reading applications. We propose a method that combines boosting methods to select features which are then used with HMMs for improved recognition.

c. What research and teaching skills and experience has the project helped provide to those who worked on the project?

    i. This project is quite interdisciplinary in its nature, providing education and direct research experience in imaging, graphics, and HCI.

    ii. Last year, Dr. Lionel Reveret, with Divya Narayanan and Tim Keenan, worked on this effort. Dr Lionel was a post-doctoral researcher and then joined as Assistant Professor/Researcher at INRIA in Grenoble France and still continues to collaborate with PI.

iii. Divya was an UG student who has since graduated and is now a PhD student in bio-imaging at Johns-Hopkins.

iv. Tim Keenan, who worked on the animation part of the project is now a Technical Director at Dreamworks Animation, with credits in movies like Shrek II, Over the Hedge, Madagasscar.

v. Antonio Haro has completing his PhD this year and is working for Nokia Research in Dallas Texas. One of his projects involves faces on a mobile phone.

vi. Ravi Ruddarraju finished his BS in EE and is now continuing his PhD dissertation under the PI at GA Tech.

vii. John Hable finished his BS and MS is CS and is now working on facial capture for Games at EA.

viii. Jun Xiao has just finished his PhD.

ix. Yan Huang has finished her MS and now works for Google.

x. Pei Yin is about to propose for his dissertation and has had two successful internships at Microsoft Research.

d. What outreach activities have you undertaken to increase public understanding of, and participation in, science and technology?

i. The toolkits discussed earlier are going to be essential in taking this kind of laboratory research to the outside world. We are talking to a few high-school students who could use these toolkits for their own animations.

ii. Several art schools and also production houses have inquired about our software system for use in production. One of the groups in an animation and special effects classes at GA Tech have already used these tools. Due to limitation in supporting such activities, it is unlikely that this will have the impact it could have.

3. **Publications and Products: What have you produced?**

a. What have you published as a result of this work?

i. Major publications (full reference)

- Catrambone, Stasko, Xiao, "Anthropomorphic Agents as a User Interface Paradigm: Experimental Findings and a Framework for Research" Proceedings of CogSci 2002.

- Xiao, Stako, Catrambone, "Embodied Conversational Agents as a UI Paradigm: A Framework for Evaluation", Proceedings of Embodied conversational agents for AAMAS 2002.
- Xiao, Stako, Catrambone, "Be Quiet? Evaluating Proactive andReactive User Interface Assistants", Proceedings of INTERACT 2003.
- Kwatra, Schodl, Essa, Turk, Bobick, "GraphCut Textures: Image and Video Synthesis Using Graph Cuts" in ACM Transaction on Graphics (Issue of the Proceedings of ACM SIGGRAPH 2003 Conference). July 2003.
- Ruddarraju, Haro, Essa, "Fast Multiple Camera Head Pose Tracking," In Proceedings, Vision Interface 2003, Halifax, Canada.
- Ruddarraju, Haro, Nagel, Essa, Abowd, Mynatt "Perceptual User Interfaces using Vision-based Eye Trackers, Submitted to International Conference on Multimodal and Perceptual User Interfaces, Vancouver, BC Nov 2003.
- Pei Yin, Irfan Essa, James M. Rehg, "Asymmetrically Boosted HMM for Speech Reading". in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2004)*, II755-761, Jun. 2004.
- Pei Yin, Irfan Essa, James M. Rehg, "Boosted Audio-Visual HMM for Speech Reading". in *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003)*, pp68-73, Oct. 2003/held in conjunction with *ICCV-2003*.
- Pei Yin, Irfan Essa, James M. Rehg, "Boosted Audio-Visual HMM for Speech Reading". in *Proc. Asilomar Conference on Signals, Systems, and Computers*, pp 2013-2018, Nov. 2003 as an invited paper.
- Brubaker, Essa, Turk, “Deformable Texture”, Submitted for Review 2006.
- Reveret, Essa, Bailley, "Characterization of Speech Actions for Facial Animation," In preparation for The Visual Computer Journal, 2006.

ii. Books and other one-time publications

- None.

iii. What Web sites or other Internet sites reflect this project?

- http://www.cc.gatech.edu/cpl/projects/graphcuttextures/
- http://www.cc.gatech.edu/cpl/projects/multicameyetracking/
- http://www.cc.gatech.edu/cpl/projects/pupil
- http://www.cc.gatech.edu/cpl/projects/animated-speakers/
- http://www.cc.gatech.edu/gvu/perception/projects/speechreading

iv. What other specific products (database, collections, software, inventions, etc.) have you developed?

- N/A

4. **Contributions: Why does it all matter?**

a. To the development of your own discipline(s)?

As noted, this work covers the areas of imaging (analysis and interpretation), graphics, and HCI. Its overall goal is to aid in the analysis and synthesis of facial movements and in that context, our work continues to extend the state of the art in automatic approaches for analysis and synthesis of faces. We do not think we need to emphasize the importance of faces in general. This basic importance makes it essential to continue such studies within the context of computer science research.

b. To other disciplines of science or engineering?

The primary goal is to study how people make expressions. We hope to pursue more research on using the techniques we have developed here for NIH and other behavioral science type of research. The PI was recently invited to a DARPA planning workshop of virtual face modeling to support facial reconstruction of the injured. He established a team of experts from imaging, and biomedical modeling and put in a proposal to DARPA, which he is expecting to hear back on. It is not clear if DARPA will fund such an effort, however, there are avenues to pursue with NIH.

c. To education and development of human resources?

The effects of this work will have impact on human resources after some additional study and evaluation.

d. To physical, institutional, and information resources for science and technology?

N/A at present.

e. To the public welfare beyond science and engineering?

Not presently, but expected on the long run.

5. Special Requirements:

a. A brief summary of the work to be performed during the next year of support if changed from the original proposal

N/A

b. Do special terms and conditions of your award require you to report any specific information that you have not yet reported?

N/A

c. Do you anticipate that more than twenty percent of the funds under your NSF award will remain unobligated at the end of the period for which NSF currently is providing support?

N/A

d. Has there been any significant change in animal care and use, biohazards, or use of human subjects from what was originally approved (or approved later)?

N/A

1. **Activities and Findings: What have you done? What have you learned?**
   a. What were your major research and education activities?

The research funded by this CAREER grant has progressed on FIVE primary areas during the course of the project. These areas and the specific advances within are each area is briefly listed here.

Spatio-temporal analysis of speech action from video: One main focus of our effort is to analyze spatio-temporal relationship between how lips move in generation of speech, and the_audio signal itself. By analysis of the video and the related audio_channel, we seek to build a compact representation for visual speech_actions that is suitable for synthesis and recognition of lip movements_with speech. We have also added the ability to track these lip shape_changes with facial expressions. In some recent work, we have added machine learning techniques to undertake feature selection, which in turn helps with merging information for improved speech reading research. This latter work is now being continued under an NSF ITR GRANT with Professor Jim Rehg

Eye and Head Tracking: One main constraint faced by existing facial expression (and lip motion) tracking systems is that we need to align the face to appear to have no rigid translations and rotations. This allows for tracking of the non-rigid motion separately. Towards this end, we have focused on several eye and head-pose tracking systems. In this year of our effort, we have built a multi-camera system to track eyes robustly and then by using triangulation extract head orientation. The multi-camera aspect of our approach also allows for a larger area for tracking faces. We have also added Fischer-Discriminats and replaced PCAs which allows for tracking in varying lighting conditions. Some of the approaches developed here are in use by other researchers in the fields of HCI.

Toolkits for Facial Modeling and Animation: We are also building various XML-based facial animation toolkits that_merge audio and facial action synthesis to generate agents with faces_that can talk and make expressions. Our approach is NOT limited to_just graphical face models and includes simple robotic systems like the_PONG robot provided by IBM Research. In addition, we have also built_methods for image-based generation of realistic face models that can be_animated.

Evaluation of faces in interfaces: One significant aspect of our work was to study how humans also perceive expressions, especially within the context of face-to-face interactions. In this work, we took a slightly different approach this year and choose to collaborate with Professors Catrambone and Stasko, who were interested in studying the importance of faces in an interface. By collaborating with this larger effort by the above-mentioned researchers, we are able to pursue our interest in the modeling and animation of faces in an interface and have successfully shown both (a) the importance of faces in different settings and (b) need for better toolkits and methods discussed in avenues 1 and 2 above.

Feature Selection for Speech-reading: As an initial step towards modeling lip motions associated with speech and its impact on facial motion, we developed a method to track lips from video. We used this tracking to build representations of lip motion, which were then improved on by high-sped data captured from a VICON Motion Capture System. Using this data, we developed an approach to combine Boosting Methods with HMM based learning methods to select which feature in lip motion are appropriate for speech reading approaches.

1. **Activities and Findings: What have you done? What have you learned?**
   b. What are your major research findings?
      i. A compact representation that combines facial movements associated with speech and speech can be extracted from video and can be used for recognition and synthesis.

      ii. Using multiple vision-based eye-trackers, with robust light-independent templates, allows for robust head pose (and eye-gaze) tracking over a larger area. We have also tested this methods over extended periods as a part of a simple pilot study

      iii. A unified and perhaps XML based toolkit is needed to generate realistic facial motions, which can be rendered using realistic and non-photorealistic graphics models and physical robots with faces.

      iv. Faces are important not just in face-to-face communications, but also in human-machine interactions, using an anthropometric agent. We are working on our collaborators to test this and are providing toolkits and other expertise as needed.

      v. Cooperative learning methods are needed to improve recognition accuracies for lip shapes related to speech reading applications. We propose a method that combines boosting methods to select features which are then used with HMMs for improved recognition.