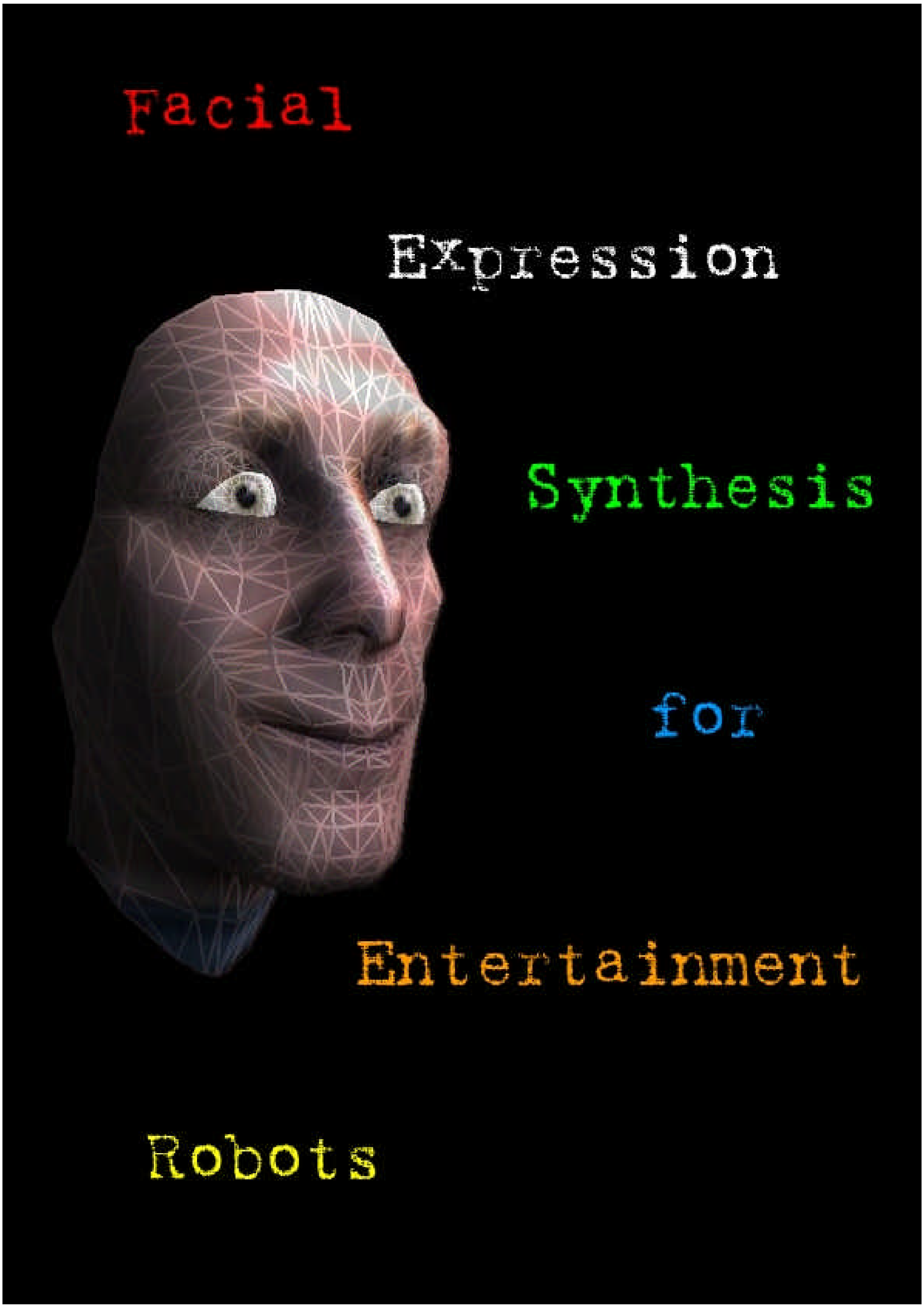
Facial
Expression
Synthesis
for
Entertainment
Robots

Certificate

This is to certify that

Mr.Shalin Shodhan

Roll No. 6775 of Computer Engineering Semester VII class

has successfully presented a Technical White Paper on

“Facial Expression Synthesis for Entertainment Robots”

at

L.D. College of Engineering,

Ahmedabad 3800015.

Date of Presentation	19th September 2002
Staff in Charge	R.J. Doshi
Advisor	Prof. B.V. Buddhadev
Head of Department	Dr. J.S. Shah

Facial Expression Synthesis For Entertainment Robots

by
Shalin Shodhan
Roll No. 6775
Computer Engineering Semester VII
Computer Department
L.D. College of Engineering

Abstract

In the first part, the paper reviews the current state of research in facial expression synthesis for entertainment robots. In the second part, a framework for autonomous facial expression synthesis is presented. A three dimensional simulation of an android head is presented as an implementation of the framework. Artificial Intelligence Markup Language (A.I.M.L.) based text stimulus and response are combined with OpenGL technology to create a simulated android head which synthesizes facial expressions according to the stimuli provided. Various applications and objectives of future work are enlisted in the third part and conclusions are drawn.

Acknowledgements

I sincerely thank

- L.D. College of Engineering Computer Department for providing the required facilities and a good working environment for doing my research.
- Dr. J.S. Shah, Head of Computer Engineering Department, for his support.
- Prof. B.V. Buddhadev for his invaluable suggestions and thoughts.
- Dr. Michael Lyons at the ATR Media Information Science Labs for the JAFFE Image Database and other Facial Expression related information.
- Keith Waters for his wonderful Facial Animation code.
- Dr. Richard Wallace and the A.L.I.C.E. AI Foundation.
- David Calinski for "Hippie", AI Chatbot.
- Dr. Ronak Shodhan for inspiring me to pursue my interests in the fascinating field of Robotics and Artificial Intelligence.
- All the institutions and people who made a vast array of information available to me through the internet.
- My Friends and Family for their love and support.

Contents

List of Figures

1. Introduction

The purpose of this paper is to

1) Review current developments in the field of facial expression synthesis for entertainment robots.
2) Create and implement a framework for autonomous facial expression synthesis.

The Face: A structure of bones driven by muscles and overlaid with skin that forms the most expressive part of the human body. Sensory information from our environment affects our state of mind which is manifested on our face. Thus the face is an indicator and a means of reflecting the complex mental conditions that we experience.

The Expression: The expression is the language of the face. The basic elements of the face collectively orient themselves in various combinations to form what we call an expression. Synthesis of an expression thus involves creating the various combined orientations of bones, muscles and skin. The subtlety with which the human face can manipulate these parameters leads to myriad expressions. The multitude of expressions that we can create due to the complexity of facial structure helps in turn to represent the vast array of mental states that we can have.

Entertaining Robotic Creatures: Entertainment Robotics is the use of robotic technology to create entities with the purpose of providing entertainment.
Example: The robotic pet.

The Implementation: Building a real life animatronic head is the ultimate objective but at the current stage a 3D simulation must suffice. The implementation is titled **“Bezz”** because the very first simulation was a line drawing made of Bezier Curves. In its current form Bezz is an OpenGL based rendering of a 3D human face with a wireframe acting as the bones, muscles actuating the expressions and a texture map for skin. The face can be easily manipulated by using a mouse to create an expression. Storage/retrieval of expressions has been implemented. Behind the face is a brain that receives input from the world such as text and processes a response in form of an expression. The brain uses Artificial Intelligence Markup Language (A.I.M.L.) an XML derivative used to store intelligence in templates. An engine processes text input and uses pattern based text matching system to generate a text response. The standard language and interpreter of AIML have been modified to add expression data to the text response which drives the OpenGL based face program.

2. Review

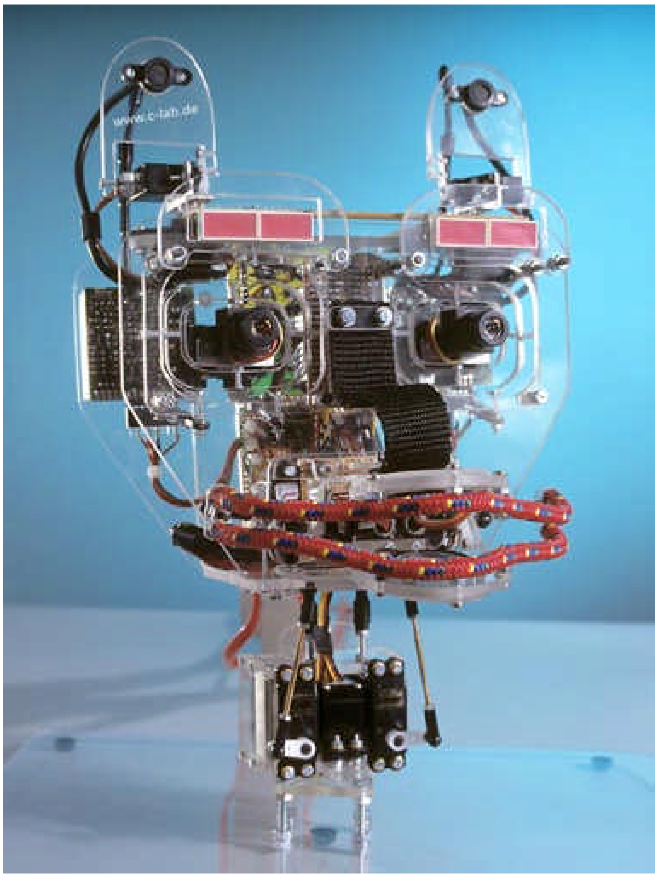

Fig.2.1 MEXI- Machine with Emotionally Extended Intelligence
University of Paderborn (Germany)

Numerous Android Head projects [13] have been undertaken at various research centers around the world. A lot of interesting research on face recognition [1] [2] and facial animation [9] has been conducted. I started my work by studying the Japanese Female Facial Expression Database (JAFFE) maintained and kindly provided to me by Michael J. Lyons [14]. This led to a basic understanding of how an expression can be quantified in terms of orientation of bones and contraction/expansion of muscles. Reviewing various projects of a similar nature I decided to implement a simulated 3D face which can be taught facial expressions and then give it intelligence to synthesize the same in response to stimulus.

I present brief reviews of two of the most inspiring projects that I have come across.

2.1 Doc Beardsley

Fig. 2.1.1 Doc Beardsley (picture from the IAI website [12])

Developed By: Tod Camill and Tim Eck
At: Interactive Animatronics Initiative, Carnegie Mellon University, USA
Objective: Create a convincing illusion of intelligence
Website: http://www.etc.cmu.edu/projects/iai

Philosophy and Goals

- Create conversationally interactive and environment sensitive animatronic characters
- Progress towards the next generation of animatronic character technology

Hardware

- The animatronic design has fourteen degrees of freedom
- Multiplex mc/V2 servo motors
- Functions such as neck rotate, nod and shift.
- Movable Eyelids
- Showsmith hardware for show control
- Cameras for visual perception
- A complete environment and stage designed for interactive shows.

Software

- SolidWorks mechanical design software
- StoryWeb technology is used to encode a non-linear interactive narrative
- Showsmith software for show control
- Sphinx: Open source voice recognition software
- OpenCV visual perception system

Creators' Thoughts

Our goal is to simulate life, or provide the illusion of life. To that end, we are doing just that by trying to simulate a human (or anthropormorphic) entertainment experience through a character. [12]

What we are really doing is akin to designing a clever magic trick -- if the audience gives in, they think they are talking to a living, breathing character. In other terms: the suspension of disbelief. And, if we are successful in suspending the audience's disbelief with technology, that's a cool thing! [12]

I once superglued 37 pieces of toast to the belly of a cat.

- Horatio "Doc" Beardsley

What impresses me most about this project is its wonderfully creative personality development. I also like the clever trick where Doc feigns forgetfulness when he cannot come up with a response.

2.2 Kismet

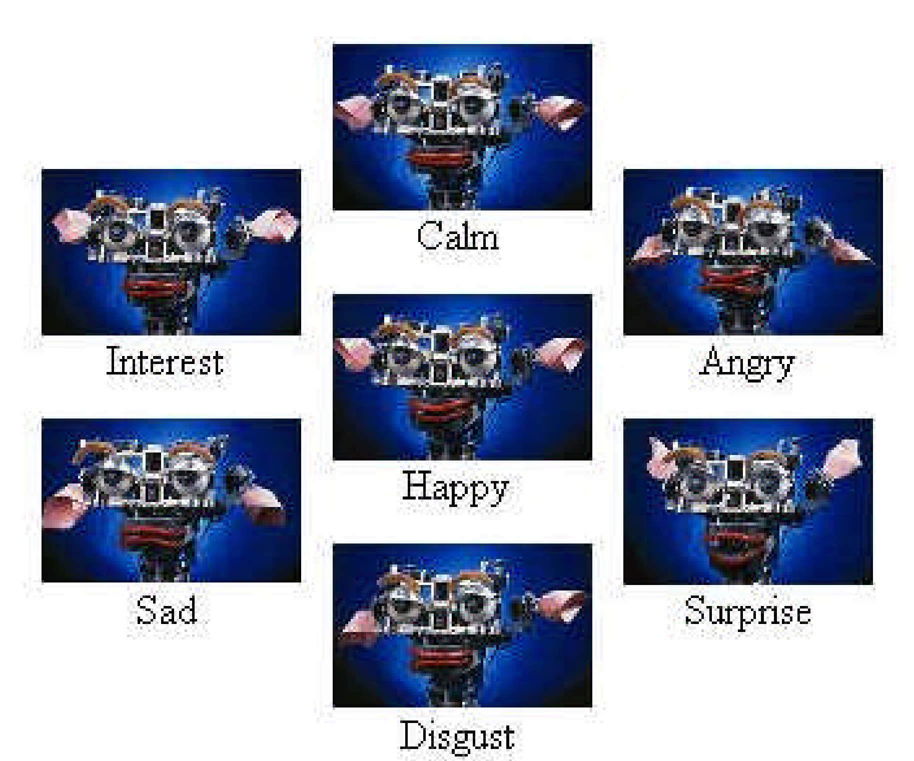

Fig 2.2.1 Kismet (pictures from the Kismet website [11])

Developed by: Cynthia Breazeal
At: Humanoid Robotics Group, Massachusetts Institute of Technology, USA
Objective: To create an intelligent expressive robotic creature
Website: http://www.ai.mit.edu/projects/

Philosophy and Goals

To create sociable humanoid robots that can interact and cooperate with people, and play a part in their daily lives.

Inputs: Visual, auditory, and proprioceptive sensory inputs

Outputs: Vocalizations, facial expressions, communicative cues

Hardware

- **Processors:**
  4 Motorola 68332 microprocessors : LISP
  9 networked 400 MHz PCs :QNX
  Dual 450 MHz PC : NT
  500 MHz PC : Linux

- **Cameras:**
  4 color CCD
  - 2 wide field of view cameras
  - 2 cameras in eye pupils

- **Motors:**
  Maxon DC servo motors
  Numerous motors to give face 15 Degrees of Freedom

Operating Systems

- Linux: Speech Recognition
- Windows NT: Speech Synthesis
- QNX (a real time Unix OS): Vision Processing, Eye and Neck control

Software

- Speech recognition and processing software by Spoken Language Systems Group, MIT
- DECtalk v4.5 speech synthesis

Creator's Thoughts

> The Sociable Machines Project develops an expressive anthropomorphic robot called Kismet that engages people in natural and expressive face-to-face interaction. Inspired by infant social development, psychology, ethnology, and evolution, this work integrates theories and concepts from these diverse viewpoints to enable Kismet to enter into natural and intuitive social interaction with a human caregiver and to learn from them, reminiscent of parent-infant exchanges. To do this, Kismet perceives a variety of natural social cues from visual and auditory channels, and delivers social signals to the human caregiver through gaze direction, facial expression, body posture, and vocal babbles. The robot has been designed to support several social cues and skills that could ultimately play an important role in socially situated learning with a human instructor. These capabilities are evaluated with respect to the ability of naive subjects to read and interpret the robot's social cues, the robot's ability to perceive and appropriately respond to human social cues, the human's willingness to provide scaffolding to facilitate the robot's learning, and how this produces a rich, flexible, dynamic interaction that is physical, affective, social, and affords a rich opportunity for learning. [11]

What impresses me most about Kismet apart from its being a marvel of technology is its learning process and social interactions based on the caretaker-infant model.

Research

The Bezz Project

A Tribute

The Bezz project salutes the engineering of the human face.

I marvel at the accurate and detailed manifestation of complex mental conditions that the face can achieve.

3.1 Motivation

The objective is to create a framework for autonomous expression synthesis and then implement it in the form of a simulated robot or ‘soft-bot’. A long term goal is to develop a real life animatronic robot head.

With advancements in robot technology it will not be long before intelligent machines pervade all facets of our lives. When we begin interacting with automated devices in extensive ways it becomes extremely important that the medium through which they respond is accurate, understandable and pleasing. The face is by far the most expressive part of the human body. As a biological interface it is unparalleled in terms of the effectiveness with which it conveys incredibly complex information. Humans would prefer talking to a face than to a computer screen.

The language of the face is the expression. Hence to have facial interfaces to automated systems we must endow them with expressive capabilities. The focus of this research is to develop a system to create and store facial expressions to create a complete database of expressions. The robot evaluates stimulus from the environment and chooses an expression from this database as response.

A particularly interesting application of such research would be in creating virtual audiences and synthetic listening characters. The character can process an audio signal and extract tonal information to identify notes. The face then synthesizes an expression in response to melody or lack thereof.

3.2 Framework

Autonomous Facial Expression Synthesis:

Bezz is based on the classical robot framework.

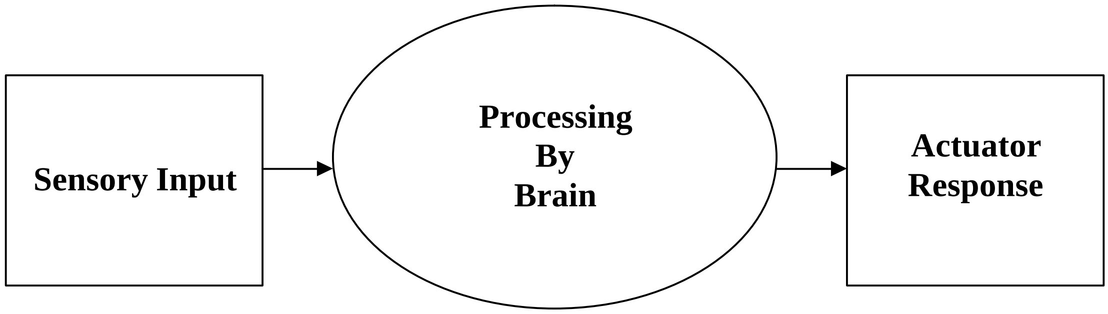

Fig 3.2.1 Classical Robot Framework

- The sensory input is any kind of stimulus provided to the robot.
- The brain is where the input is processed and a response is synthesized.
- The actuator is responsible for manifestation of the response.

The model translates into the following:

Sensory Input: Text, Speech, Music, Video, Still Images, Touch.

Processing: Recognizing Input and synthesizing appropriate expression.

Actuator Response: Showing selected expression on simulated face model.

3.2.1 Stimulus

Types of Stimuli:

Text Input: This is the currently implemented form of input for Bezz. Details regarding implementation are presented in 3.3.2

Voice: This would require speech recognition capabilities and can be implemented using one of the many speech recognition systems available today.

Music: A future development for Bezz will be the ability to distinguish between different notes played consecutively. A robot that appreciates music and develops taste to particular types of music is subject for interesting research. Tonality and Timbre recognition can be used to create a sense of music appreciation in the robot.

Video: “Mirror Mirror on the Wall”
The robot can use a camera to analyze features of various entities and provide an expression in response.

Touch: Consider a robot used to train massage therapists. Touch and pressure sensors monitor the technique of the trainee and provide real time feedback. Touch sensing is a vital system for a robot whose physical contact with its environment has bearing on its response.

Smell: Chemical sensing has many applications for robots looking to gain input from the substances in their environment. A perfume testing robot could be an application.

3.2.2 Brain

The function of Bezz's brain is to

- Recognize and match input with patterns stored in the memory.
- Synthesize response based on match found
- Send signals to actuators to manifest the selected expression

The technology behind the brain is described in depth in section 3.3.2

Training and Storage:
- The brain is trained and stores a large number of input patterns.
- Input patterns will be different corresponding to different stimuli.

For example text input patterns will be language elements where as music input would require sampled notes in various keys or frequency information.

- All such patterns are then mapped to responses in the form of expressions.
- Expressions are quantified and stored in the robot's memory.
- Training is thus provided in terms of pattern creation for various stimuli, and expression creation and storage for response.

3.2.3 Response

Quantification of a Facial Expression

The face is a structure made of bones, muscles and skin.

A graphical simulation uses the following replacements:

Bones: A wireframe model to define facial structure
Muscles: Contraction and Expansion motion provided to the wireframe
Skin: Texture map on top of the wireframe

Any expression is created by a particular orientation of bones and manipulation of skin as driven by muscles.

Thus, the actuators are the facial muscles.
Hence a facial expression can be quantified by storing the amount of contraction or expansion of each of the muscles of the face.

The proposed data structure is as follows:

Structure: struc_Muscle
Name/ Number : String/Integer
Contraction/Expansion: Integer or Float
End

Structure: struc_Face
Expression Name: String
muscles [ ] : struc_Muscle
End

The current implementation uses 18 such muscles to synthesize a large number of expressions.

3.3 Implementation

Why the name Bezz?

The very first implementation of the framework described was made in Visual Basic using Bezier Curves.

Features of the first implementation included:

- Mouth Editing
  - Smile
  - Frown
  - Mouth 'O'

- Key-framing and animation support for mouth.

- Saving and loading animation.

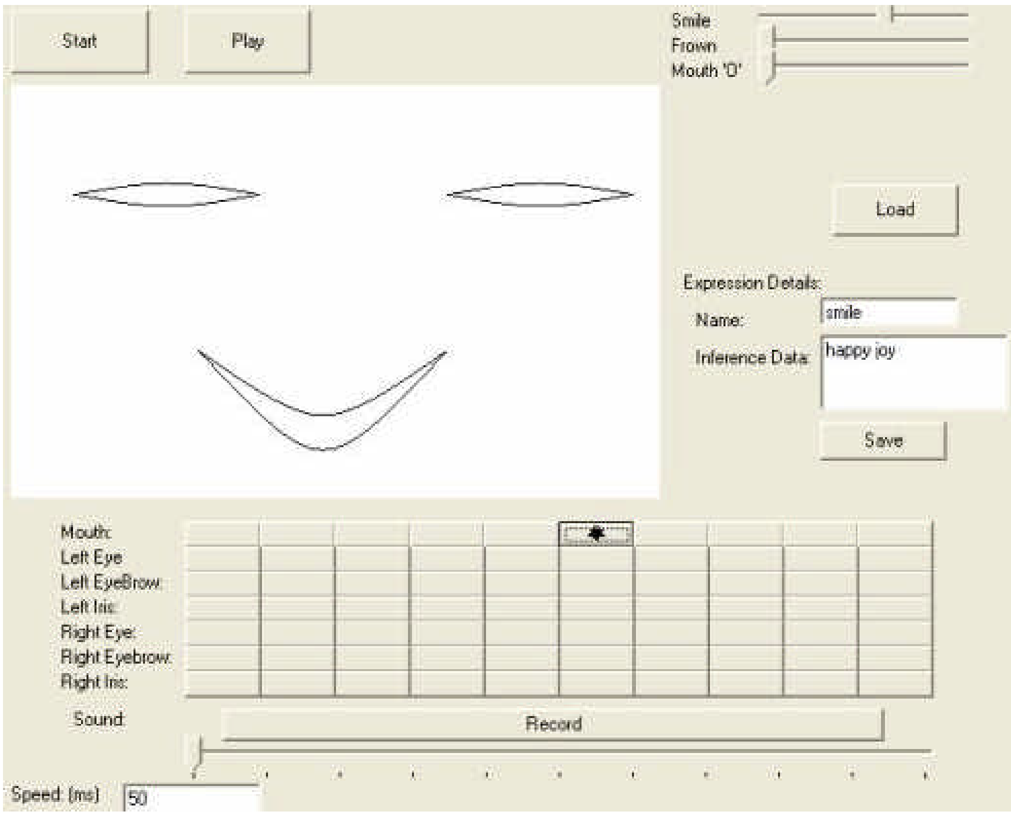

Fig 3.3.1 Screenshot of the first version of Bezz

3.3.1 Android Head Simulation in 3D using OpenGL

The Visual Basic implementation of Bezz was replaced by an OpenGL based three dimensional simulation of an android head written in C/C++. The current implementation is based on Keith Waters' Facial Animation Code [9].

OpenGL is my personal favorite when it comes to graphics APIs. Being comfortable with its functions I could easily understand Keith Waters' excellent code and could add many features to it to suit my own requirements. To learn more about OpenGL please refer to [15].

Features of the Facial Animation Code:
-3D wireframe and various render modes including texturing
-Facial Muscles to manipulate facial expression
Additions/Modifications:
-Code ported to Graphics Library Utility Toolkit (GLUT)
-Mouse based creation of expression added
-Custom storage and retrieval of expressions

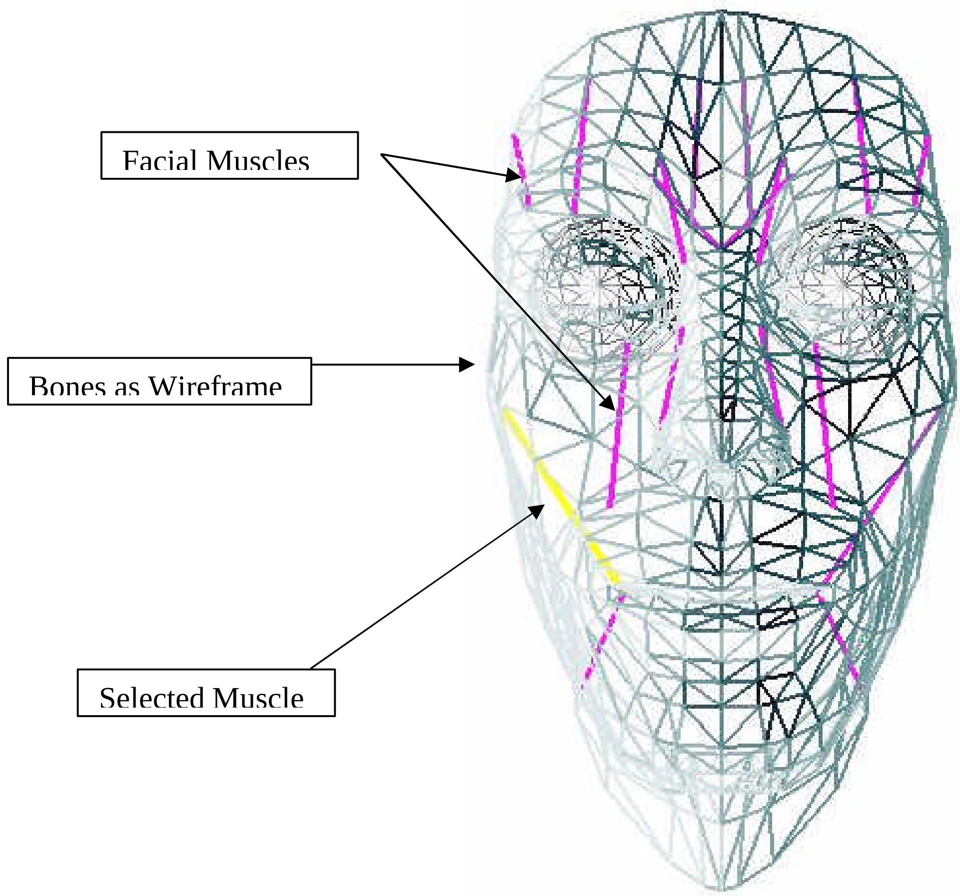

Fig 3.3.2 Android Head Simulation in 3D using OpenGL

3.3.1.1 Expression Creation

Creating facial expressions for Bezz is as easy as clicking a mouse.

Instructions

- Position Mouse over the muscle of you wish to modify
- Click Left Mouse to Contract Muscle
- Click Right Mouse to Expand Muscle
- Keep button pressed for extra contraction/expansion

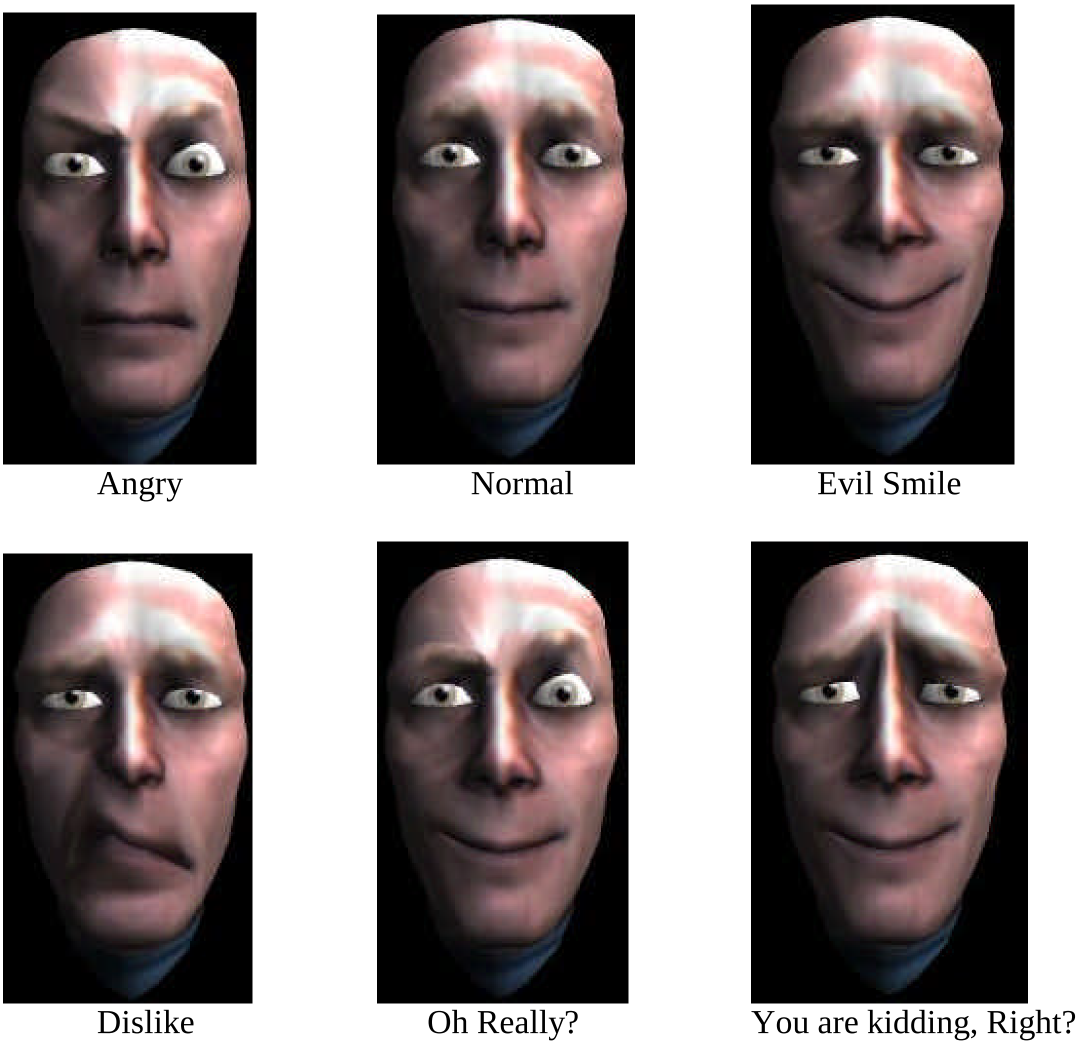

Fig 3.3.3 Example expressions that can be created and subsequently evoked

3.3.1.2 Expression Storage and Retrieval

The most important aspect of the Bezz project is the creation of an expression database that the brain of the robot can refer to when synthesizing response to any stimulus. For efficient implementation of such a database simple data and file structures have been used to quantify an expression in terms of the contraction and expansion of muscles.

A sample expression file: "Evil Smile"
Muscle 0 17
Muscle 1 18
Muscle 2 0
Muscle 3 0
Muscle 4 -12
Muscle 5 -5
Muscle 6 -4
Muscle 7 -3
Muscle 8 0
Muscle 9 0
Muscle 10 0
Muscle 11 0
Muscle 12 0
Muscle 13 0
Muscle 14 0
Muscle 15 0
Muscle 16 0
Muscle 17 0

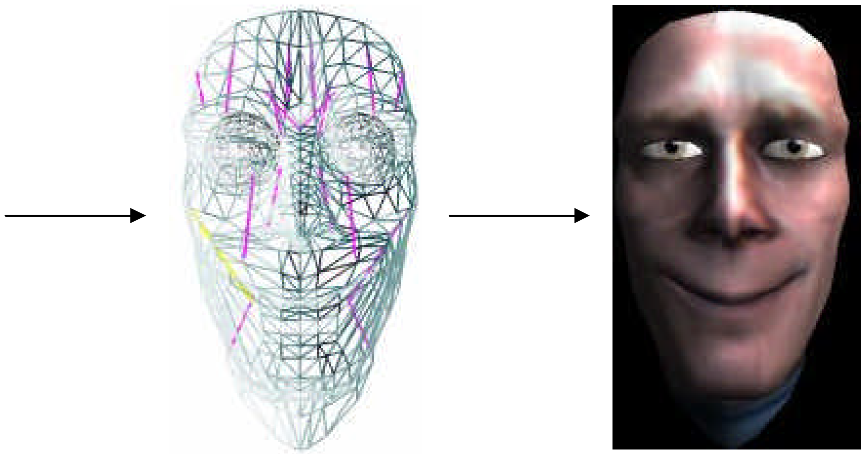

Fig 3.3.4 Expression File Listing and the expression "Evil Smile"

3.3.2 Text Based Stimulus using A.I.M.L.

3.3.2.1 Introduction to A.I.M.L.

In its current implementation Bezz uses a modified version of Artificial Intelligence Markup Language for text input recognition, processing and response synthesis.
A.I.M.L. lies at the center of the robot’s brain. All intelligence is quantified, stored and processed using A.I.M.L. as the common format for intelligence data exchange.
Here is a block diagram of the functioning of A.I.M.L.

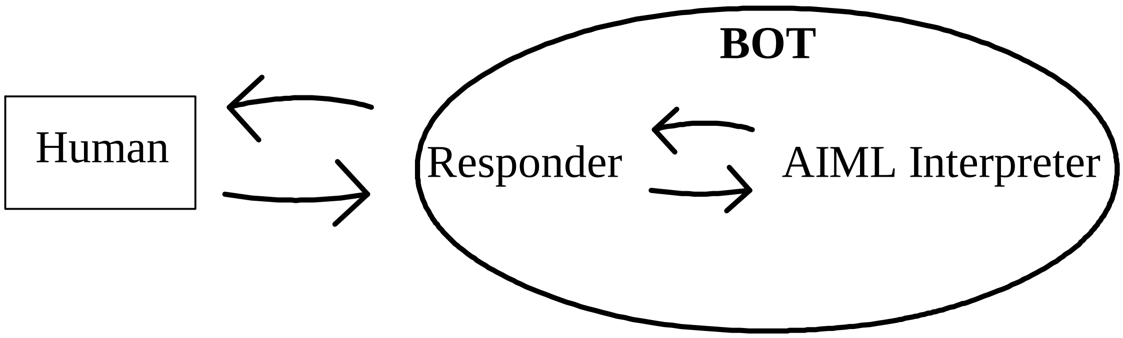

Fig 3.3.5 Block Diagram of functioning of A.I.M.L.

Responder acts as an interface between human and the A.I.M.L. engine whose function is to interpret the data from responder and generate a response which can be sent back to the human via the responder.

Bezz uses a modified version of A.I.M.L. which has been enhanced to include expression data along with text response. The expression data is parsed by the 3D simulation and appropriate expression is manifested.
The A.I.M.L. program and the facial simulation program are thus connected by a pipe.

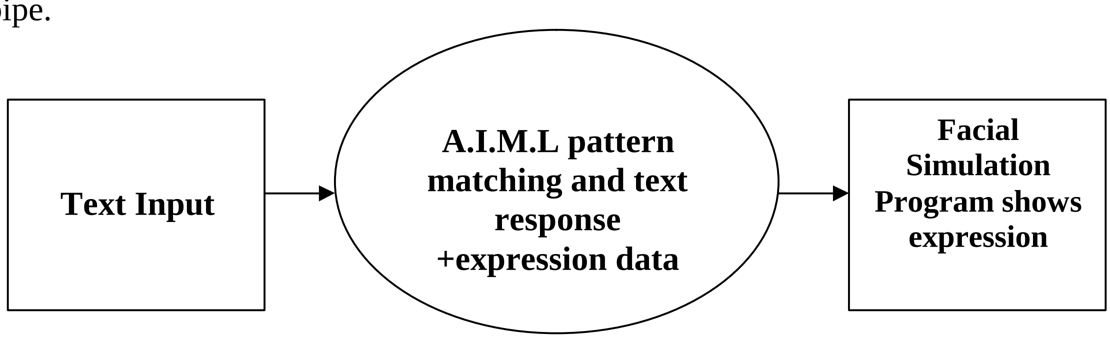

Fig. 3.3.6 Mapping the Classical Robot Framework onto Bezz

A hello world example in A.I.M.L.

```
<category>
<pattern>How are you today</pattern>
<template>
I'm fine, Thank you.
</template>
</category>
```

<category> tag encloses all nodes. There can be many <category> tags in an A.I.M.L. file.

<pattern> tag indicates what the user is expected to type.

<template> is the corresponding response.

3.3.2.2 Extension of A.I.M.L for Expressions

A new custom tag <xp> is used to encode expression data into an A.I.M.L. file.
Example:

```
<category>
<pattern>How are you today</pattern>
<template>
I'm fine, Thank you. <xp> smile
</template>
</category>
```

The A.I.M.L. engine code was also modified to parse out expression data and send to the facial simulation program.

Thus an extensive A.I.M.L. set with corresponding facial expressions is created as the knowledge of the robot.
The A.I.M.L. based pattern matching and response is the robot's intelligence.
The output is a facial expression corresponding to text response both of which are driven by the initial text input

A.I.M.L. related information can be found at the A.L.I.C.E. AI Foundation website [8].

4. Applications

Android production is forecasted to rival automobile production in a couple of decades. Autonomous Facial Expression is the single most important quality required to create a convincingly real interactive character. Facial Expression Synthesis finds the following applications in the field of entertainment robotics and other areas.

Entertainment

- Theme Parks: tour guide robots, Robotic mascots
- Museums: Interactive historical figures
- Game Shows: Robotic characters in televised game programs
- Movies: Animatronic characters in movies Eg. Spielberg's "Jurassic Park" and "AI"
- Robotic Pets: After the phenomenal commercial success of AIBO, the pet robot dog, SONY plans to introduce a human like robot companion.
- Virtual or robotic audiences

Other Fields

- Giving human characteristics to any Automated Response System:
  Any form of automated response, a kiosk, an information desk etc can have a robotic character at the front end to make access to information easy, fast and enjoyable.
- Nurse Bots:
  Care taker robots must have congenial facial expressions to serve their purpose well.
- Study and Quantify Human Behavior and Expression:
  Psychological studies can benefit from research in trying to recreate human expressions and behavior in robots.
- Front End to Expert Systems: A robot teacher
  It would be very useful to characterize good human teachers in form of a robot and back them up with the teacher's knowledge as an expert system. In this way students can benefit from the teaching skills long after the teacher is gone.

5. Conclusions and Future Goals

A review was presented in the first part showcasing some inspirational projects in the field of facial expression synthesis for robots. The second part developed and described a framework for autonomous facial expression synthesis. An implementation titled "Project Bezz" was presented. At the end applications were listed.

My job is far from done. But the work completed so far is very satisfying and motivates me to delve further into the fantastic world of robotics and artificial intelligence.

Bezz is my baby. I intend to train it and inculcate a complete character in the bot. Another objective of the project is to deploy Bezz on the internet so many enthusiasts may add to the robot's expression base as well as train input response.

I also aim to explore applications of expressive robots in music. The goal is to build a robot that can listen to music, appreciate it, learn from hearing and then synthesize its own music. Facial expressions will serve to add completeness to such an artistic entity.

In closing I would like to quote from the Bhagvad Gita:

"Asaktaha Acharan Karma Paramapnoti Purushaha"
"One who works without desire (for fruit) achieves the supreme state"

I have truly enjoyed working on Bezz and my reward lies in carrying out research.

References:

[1] Coding Facial Expressions with Gabor Wavelets
Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi & Jiro Gyoba
Proceedings,
Third IEEE International Conference on Automatic Face and
Gesture Recognition,
April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 200-205.

[2] A Survey of Facial Animation and Modeling Techniques
Jun-yong Noh
Integrated Media Systems Center, University of Southern California
http://csuri.usc.edu/~noh
Ulrich Neumann
Integrated Media Systems Center, University of Southern California
http://www.usc.edu/dept/CGIT/un.html

[3] Fully Embodied Conversational Avatars:
Making Communicative Behaviors
Autonomous
J. CASSELL AND H. VILHJÁLMSSON
MIT Media Laboratory, 20 Ames Street, Cambridge, MA 02139, USA

[4] 3D Models of the Lips for Realistic Speech
Animation
Thierry Guiard-Marigny , Nicolas Tsingos , Ali Adjoudani,
Christian Benoit , Marie-Paule Gascuel ,
Institut de la Communication Parlee, France

[5] Designing PETS:
A Personal Electronic Teller of Stories
Allison Druin, Jaime Montemayor, Jim Hendler, Britt McAlister,
Angela Boltman, Eric Fiterman, Aurelie Plaisant, Alex Kruskal, Hanne Olsen,
Isabella Revett, Thomas Plaisant Schwenn, Lauren Sumida, & Rebecca Wagner
Human-Computer Interaction Lab, Institute for Advanced Computer Studies
University of Maryland, College Park, MD 20742
http://www.umiacs.umd.edu/~allisond/kidteam/robot-index.html

[6] Facing the Music
A Facial Action Controlled Musical Interface
Michael J. Lyons and Nobuji Tetsutani
ATR Media Integration and Communication Research Laboratories

2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan

[7] Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models
Demetri Terzopoulos, Member, IEEE, and Keith Waters, Member, IEEE

[8] A.L.I.C.E AI Foundation
http://www.alicebot.org

[9] Keith Waters Facial Animation Code
http://www.crl.research.digital.com/projects/facial/facial.html

[10] Android World
http://www.androidworld.com

[11] Kismet
http://www.ai.mit.edu/projects/

[12] Doc Beardsley
http://www.etc.cmu.edu/projects/iai

[13] Android Head Projects
http://www.androidworld.com/prod04.htm

[14] JAFFE Database
http://www.mis.atr.co.jp/~mlyons/jaffe.html

[15] OpenGL
http://www.opengl.org