2011

Emotion based Facial Animation using Four Contextual Control Modes

Rajkumar Vijayarangan
University of Windsor

Emotion based Facial Animation using Four Contextual Control Modes

by

Rajkumar Vijayarangan

A Thesis
Submitted to the Faculty of Graduate Studies
through Computer Science
in Partial Fulfillment of the Requirements for
the Degree of Master of Science at the
University of Windsor

Windsor, Ontario, Canada

2011

Emotion based Facial Animation using Four Contextual Control Modes

by

Rajkumar Vijayarangan

APPROVED BY:

____

Dr. Guoqing Zhang
Faculty of Engineering

____

Dr. Robin Gras
School of Computer Science

____

Dr. Xiaobu Yuan, Advisor
School of Computer Science

____

Dr. Dan Wu, Chair of Defense
School of Computer Science

January 24, 2011

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

An Embodied Conversational Agent (ECA) is an intelligent agent that interacts with users through verbal and nonverbal expressions. When used as the interface of software applications, the presence of these agents creates a positive impact on user experience. Due to their potential in providing online assistance in areas such as E-Commerce, there is an increasing need to make ECAs more believable for the user, which has been achieved mainly by using realistic facial animation and emotions.

This thesis presents a new approach of ECA modeling that empowers intelligent agents with synthesized emotions. This approach applies the Contextual Control Model for the construction of an emotion generator that uses information obtained from dialogue to select one of the four modes for the emotion, i.e., Scrambled, Opportunistic, Tactical, and Strategic mode. The emotions are produced in format of the Ortony Clore &Collins (OCC) model for emotion expressions.

DEDICATION

To all my Friends and Family members

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my supervisor Dr. Xiaobu Yuan, IEESM, Associate Professor, School of Computer Science, University of Windsor, ON, Canada for his continuous encouragement and support in presenting this thesis. Deep Appreciation and gratitude goes to him for contributing his suggestions and ideas during my research.

I would also like to thank my thesis committee members Dr. Robin Gras, Dr. Guoqing Zhang and Dr. Dan Wu for spending their valuable time providing feedback about thesis throughout my proposal and defence.

Finally, I would like to thank my friends/colleagues Mr. Johnny Kapps and Ms. Sathulla Sabiha for their moral support and valuable suggestions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Introduction to Human Computer Interaction

Human Computer Interaction, also known as HCI, is one of the most renowned areas of research. It defines the rules for interaction between human and the computer by combining different sciences, psychological and cognitive factors. The goal of Human Computer Interaction is to make a positive impact on the user experience when they interact with computers. The success criteria are mainly determined by how well the machines can understand humans and vice versa. This field of research led to the evolution of different software and hardware that we use in our day to day life. Extensive research in this field has bridged the gap between the humans and computers in the recent past. The evolution of hardware such as mice, keyboards, web camera, microphone and software like word processors, CAD, CAM are considered to be milestones in the field of human computer interaction.

Since the early 1990's researchers have laid their focus in developing avatars or embodied conversational agents for interactions with humans in social situations. The introduction of these avatars or human like characters created a ripple effect in several fields such as filming and animation. Creation of interactive graphical user interfaces has been one of the long term goals in this area of research. Nevertheless this became one of the massive branches of Human Computer Interaction research.

1.2 Overview of Embodied Conversational Agents

The embodied conversational agents are being used in different fields such as:

- Interactive e-learning environments for students through artificial tutors
- Healthcare systems for providing assistance to immobile patients
- e-retail applications for providing customer service
- Non Player Characters (NPC) in gaming environments

In the recent past, the Embodied Conversational Agents (ECA) have gained wide acceptance in the field of Human-Computer Interaction. These agents wear a realistic human face and communicate with users in social situations such as an e-learning system, healthcare system, e-retail environments and games. These agents usually have an interface which is backed up by a suitable dialogue manager and knowledge base. The characters of the ECAs are capable of using voice; animated speech with lip and facial expressions; eye, head and body movements to realize gestures; express emotions; and perform actions or display listening or thinking postures [CemKeskin07].The presence of these agents has made a positive impact on user experience as reported from the previous work of [Cassell00] and [Rickenberg00].

The efficacy of an agent banks on a major factor which is credibility. Embodied Conversational Agents have been made more believable by incorporating emotions. It is very important for these agents to have emotions because they will absolutely increase the user understanding and experience. The agents should not be considered as an individual modality but as the synergic contribution of different communication channels that, properly synchronized, generate an overall communication performance [CemKeskin07]. One of the first major embodied conversational agents that hit the market was the Microsoft Agent which was introduced to help the users of Microsoft products. They used simple 2D cartoon characters called as Merlin, Peedy, Genie and Robby. They were simple conversational characters with a Speech API (SAPI) text to speech engine.

1.3 Motivation

Over the years, significant research in Human Computer Interaction has led to the emergence of different techniques for creating Embodied Conversational Agents. The research on ECA’s can be subdivided into the following:

- Research on realistic 3D facial animation techniques
- Research on the construction of effective cognitive models for the generation of emotions
- Research on spoken dialogue systems

After extensive research from the literature we were unable to find a system in existence which combined the three research categories of ECAs into a current and state-of-the-art design. There is definitely a need to create a robust system that combines the above categories together which could be used in any domain, such as in the customer service based pizza ordering domain.

The goal of this system will be to provide customer service by understanding the customer needs and intentions. The creation of such a system with a 3D talking head driven by emotions will make it more engaging to the end user. This motivated us to create a system that combines the current state-of-the-art 3D facial animation and spoken dialogue systems, along with a new cognitive model for generating emotions by merging Four Contextual Control Model with OCC model for emotions.

1.4 Problem Statement

This thesis will demonstrate different methodologies adopted for creating a new cognitive model of emotion generation which will overcome the limitations of existing models. We have created a new cognitive model for emotions by combining the Four Contextual Control Modes and OCC model for emotions. The developed system will have a 3D talking head with realistic facial animation driven by the new cognitive model for emotion generation along with a powerful dialogue manager known as Partially Observable Markov Decision Process (POMDP) for a Pizza-Ordering domain.

1.5 Thesis Structure

The rest of the thesis is organized as follows. Chapter II provides a deep insight of the existing techniques and methodologies available for the creation of a 3D talking head, highlighting its pros and cons. We also review some of the most popular ECAs available in the market. Chapter III explains the proposed framework in detail highlighting the new approach for modelling emotions, by combining the Four Contextual Control Modes and OCC model for emotions. We also explain the design tools involved that enable the creation of an ECA that displays emotions towards the end of Chapter III. Chapter IV reviews the qualitative and quantitative experimental results, and performance analysis of the proposed method. Chapter V will conclude this thesis with some recommendations and future work.

CHAPTER II

REVIEW OF LITERATURE

Embodied Conversational Agents comprise of a 3D talking head, cognitive model for emotion generation and a dialogue management system which is responsible for generating responses to resolve user's queries by understanding the user's intentions. This chapter will review all three components of an ECA from the literature stating its advantages and limitations. Towards the end of this chapter we will also review some of the existing embodied conversational agents and their limitations, which became a huge motivation for us to address those limitations in this thesis.

2.1 Embodied Conversational Agents: Overview of Architecture

The architecture of an ECA is similar to a human. Humans can analyze the outside environment by means of the five senses: sight, hearing, touch, smell and taste. The verbal expressions in humans are conveyed by means of a voice generated from the vocal cords. The dialogue conversations in humans are steered by the brain. The non verbal expressions in humans are conveyed through their facial features. Given a situation, the decision making ability of a human depends on a vital physical organ called the brain. Humans have the ability to forecast some of the outcomes of their decisions.

Similarly, the ECAs are able to analyze their environment by using different peripheral devices such as cameras, text input, and voice. ECAs convey their verbal expressions by means of a synthetic voice which is generated using a Text To Speech

Engine (TTS). The dialogue conversations in ECAs are steered with the help of a dialogue manager and knowledge base. The knowledge base of the virtual agent can be large in size but the knowledge base should be very sensible, extensible and should be the core for the decision taking abilities of the agents. The ECAs convey non verbal expressions through realistic facial animation. Like a human, given a situation the ECA will react and make its decisions based on a knowledge base and artificial intelligence.

The figure shown below describes the generic architecture of an Embodied Conversational Agent. The agent is driven by animation, a TTS, and dialogue manger which are controlled by the knowledge base.

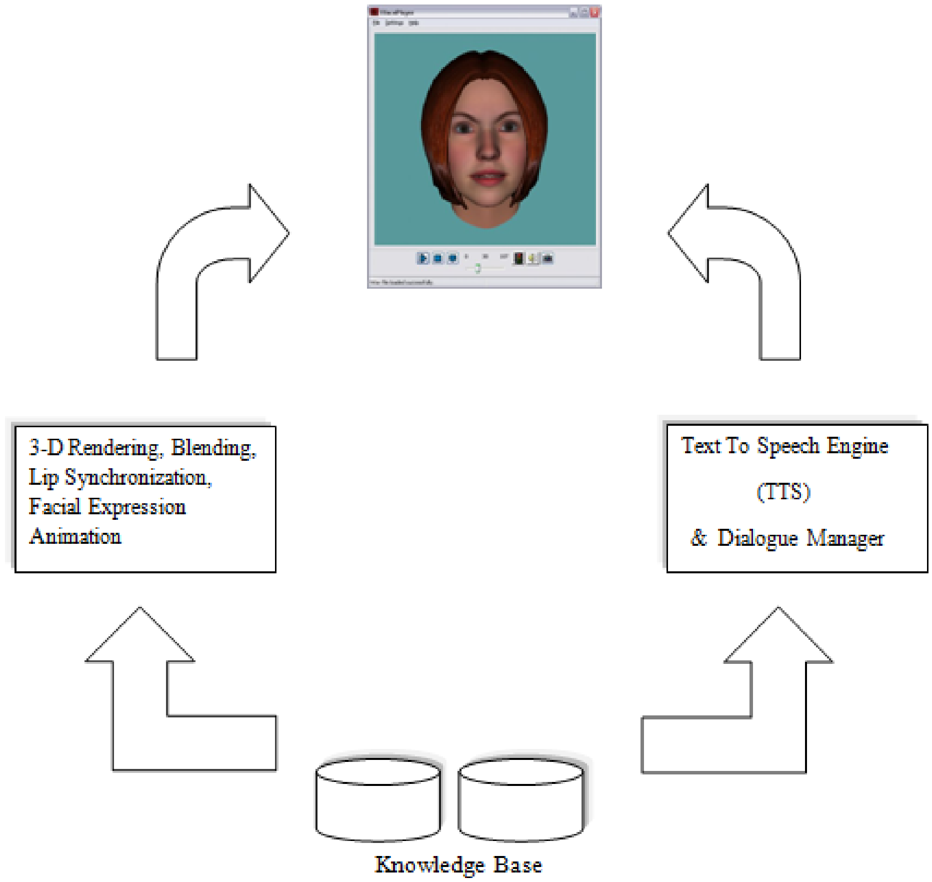


Figure 2.1: General Architecture of an Embodied Conversational Agent

We will now discuss the various components of the Embodied Conversational Agent by introducing various animation techniques available.

2.2 Facial Animation Techniques

Facial animation is a major area of research in Computer Graphics. The introduction of 3D virtual agents and the necessity to create realistic facial animation has enabled this area of research to be a vital part of Human Computer Interaction. Different techniques have emerged addressing the major concerns since Frederic I. Parke [Frederic72] took the lead in 1972. Some of the major concerns in facial animation will be the level of detail added to the facial animation, buffering and rendering.

Facial animation can be classified into two categories [Jun07]:

- Geometric based manipulations
- Image based manipulations

Each category includes a large selection of animation techniques.

The figure shown below provides an outline of the different facial animation techniques available. The subsequent sections will briefly talk about the pros and cons involved with a few of these techniques and finally we will discuss the animation technique adopted for this research.

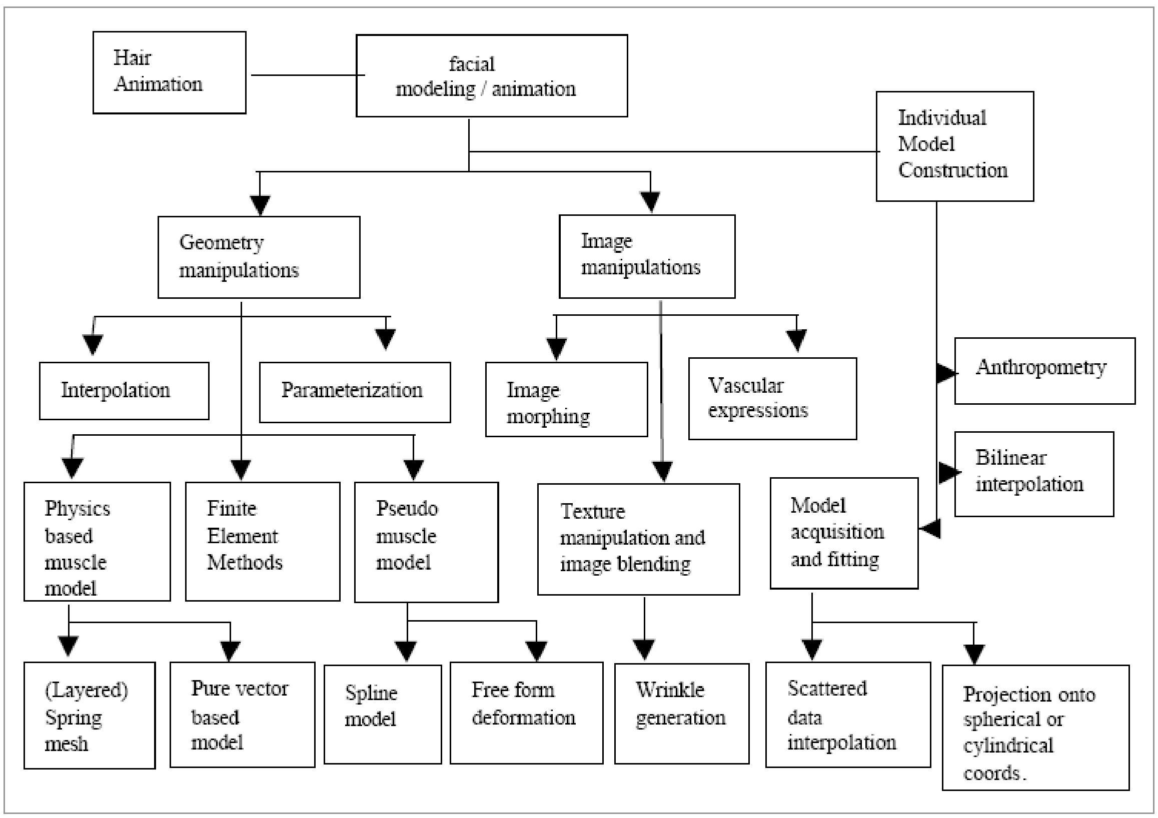


Figure 2.2: Classification of Facial animation techniques

We will begin the discussion on interpolation and parameterization methods, briefly overview image morphing and the physics based muscle models, and finally detail the MPEG4 model of animation.

2.2.1 Interpolation

Interpolation is one of the most common techniques used for facial animation. It is the rendition of images transitioning from one key-frame to another over a specified period of time. The transitioning leads to a smooth facial animation. There are two common types of interpolation techniques used in facial animation:

- Linear Interpolation
- Bi-linear Interpolation

Linear interpolation is the most basic interpolation process. Linear interpolation could be made more realistic by choosing an appropriate interpolation function, such as the cosine interpolation function, which can create smooth transitioning from one key-frame to another. Bi-Linear interpolation can be used for complex facial animation when there are more than just two key-frames.

Interpolations are significantly faster and easier to generate basic animations than most other animation techniques. However, it might be really difficult to create complex facial animation using linear or Bi-Linear interpolation techniques. Interpolation is definitely a good method to produce small set of animations from a few key frames [Jun07]. The interpolation technique is still widely used to generate vivid facial animation. To create a realistic facial animation we have combined the Linear and Cubic interpolation techniques from [Koray04] using XfaceEd, which does not suffer from any limitations as compared to the Linear or Bi-Linear interpolation techniques.

2.2.2 Parameterization

A second facial animation technique, called parameterization, evolved to overcome the limitations of the Linear and Bi-Linear interpolation. Using this technique the creator can specify and control different regions of the face through parameter values [Jun07]. Parameterization provides explicit control of creating realistic facial animations by applying proper deformations to different areas of a facial region. One of the limitations with this technique is that [Jun07] were not able to generate generic facial models with standard parameters to control the deformations. Therefore, they have to specify different parameters for different facial models. For example, a character with a long face has different parameters compared to a character with round face. The complexity of establishing parameters for different facial models is one of the limitations of parameterization.

2.2.3 Image Morphing

A third animation technique called morphing effects a metamorphosis between two target images or models. A 2D image morph consists of a warp between corresponding points in the target images and a simultaneous cross dissolve [Jun07]. Warping is the process of mapping an image onto a regular shape such as a plane or cylinder. Cross dissolving is the process where one image is faded out while another is simultaneously faded in. 2D morphing is one the first morphing techniques that came into existence but suffered from the complexity of requiring manual tuning up the parameters, such as color balancing correspondence selection and tuning of the warp and dissolve parameters.

3D morphing technique evolved to overcome the limitations of the 2D morphing. The 3D morphing created a sense of more realistic facial animation. However, this technique suffered from certain limitations similar to the Linear and Bi-Linear interpolations.

2.2.4 Physics based Muscle Model

Physics based modelling can be classified into three different categories:

- Mass spring systems
- Vector representations
- Layered spring meshes

The work on spring mesh muscle model was first pioneered by Platt and Badler [Platt81] where they proposed that a force is applied to an elastic mesh through muscle arcs will generate realistic facial animation. Platt proposed another model [Platt85] later which consisted of 38 regional muscle blocks interconnected by a spring network.

Alternatively, the work on the vector muscle model was pioneered by Waters [Waters91]. He proposed that a muscle definition consisted of three different elements, such as the vector field direction, an origin and an insertion point. The three vectors are shown in the figure below.

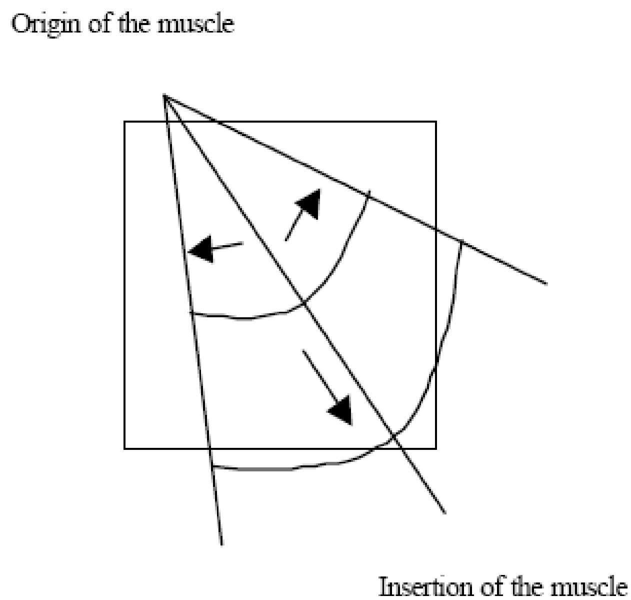


Figure 2.3: Three vectors in a Vector Muscle Model [Waters91]

The vector muscle model was one of the popular models for generating realistic facial animation for different emotions such as anger, fear, joy, surprise, disgust and happiness. The complexity involved in creating such a model was always tedious because we have to position the vector muscles in anatomically correct positions. There are no other ways to automate this process and was always time consuming because we cannot have a generic facial mesh for all models.

The last category of physics based models, the layered spring mesh muscles model, was proposed by Terzopoulos and Waters [Terzopoulos91]. They proposed that the face is made of three layers of deformable meshes that correspond to skin, fatty tissue and muscle tied to bone. The model generated with this technique simulated a true sense of realistic facial animation, but it required extensive computation. The computation time was quite significant for simulating such models.

As discussed, the physics based muscle model can model facial animation well but are often complex and suffered in computation speed. An alternative modeling method, the pseudo muscle model, attempts to lessen the drawbacks of the physics based muscle model.

2.2.6 Pseudo Muscle Model

The pseudo muscle model is classified based on the types of muscle forces.

- Spline Model
- Free Form Deformations

Free form deformation (FFD) deforms volumetric objects by manipulating control points arranged in a three-dimensional cubic lattice [Sederberg96]. There are different types of FFD's such as the Extended Free Form Deformation (EFFD) and Rational Free Form Deformation (RFFD).

RFFD does not provide precise simulation of the actual muscle nor the skin behaviour so it fails to model furrows, bulges, and wrinkles in the skin. Furthermore, since RFFD is based upon surface deformation, volumetric changes occurring in the physical muscle is not accounted for. [Jun07]

On the other hand, the spline based muscle models provided a solution for supporting smooth and flexible deformations. They are defined by a set of specific control points instead of animating the whole mesh itself. The detail of the transformation was impeccable in comparison to the previous techniques discussed above. [Jun07]

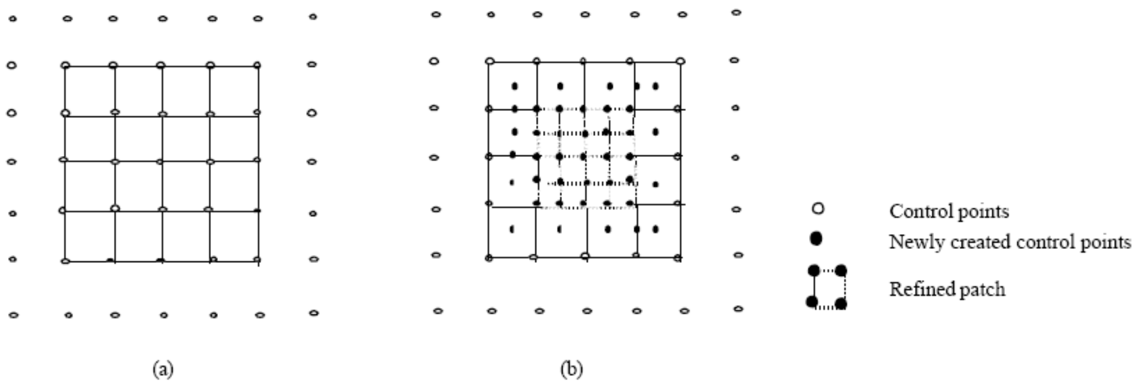


Figure 2.4: Spline based muscle model. (a) Demonstrates a 16 patch surface with 49 control points and (b) demonstrates 4 patches refined to 16 patches [Jun07]

2.2.7 MPEG4 Facial Animation

The last modelling technique that we will discuss is the MPEG4 facial animation technique. MPEG refers to Moving Pictures Experts Group. The MPEG standards focus on integrating natural or synthetic audio with video. The latest release of the MPEG standards is the MPEG4 which focuses specifically on the integration of multimedia content. They have set the standards for creating and driving the animation of an embodied conversational agent. MPEG4 involves the following parameters which help in the creation and animation of an ECA:

- Feature Points (FP)
- Facial Animation Parameters (FAP)
- Facial Animation Parameter Units (FAPU)

Feature points are the standard parameters defined for a standard face by MPEG4. They have to be defined for every single face model so that they can calibrate to work on different MPEG4 players.

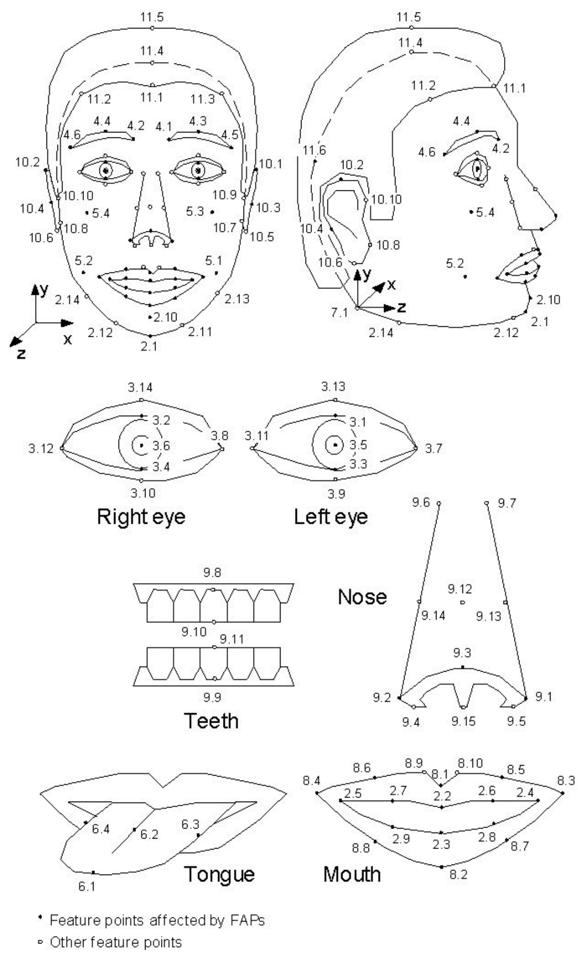


Figure 2.5: Feature points affected by FAPs and other Feature points [Koray04]

According to the MPEG4 standard there are 68 values that define the deformation between two frames of animation. There are two high level parameters representing visemes and emotions. Visemes can be defined as the combination of visuals with speech. We can drive the animation using only the first two FAPs, but the animation will be better if we make use of the low level parameters as well. The standard defines 66 low level parameters that can be used to provide vivid animation. These 66 points refer to specific regions on the face such the right corner lip, chin, and cheeks. Some FAPs correspond to an FP and define translation or rotation of that FP along an axis in three dimensions, while some other FAPs represent rotation of the head and eyes. This is one of the major reasons why the FAPs have to be calibrated prior to use on the face model. [Koray04]

Facial Animation Parameter Units (FAPU) are unit-less values independent of model geometry, which means they can drive different facial models regardless of geometry [Koray04]. FAPU can be defined as the fractions of distances between the key facial features. The figure shows the key facial features that are taken into account.

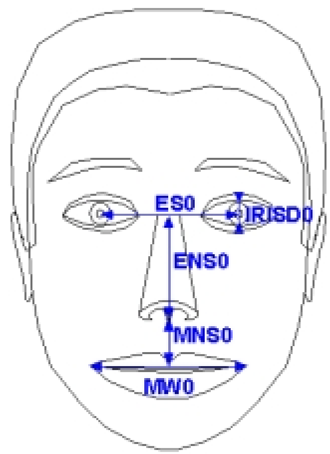


Figure 2.6: MPEG4 FAPU

One of the greatest advantages of coding a face model using FPs and FAPU is that unlike previous animation techniques, we can swap the face models without the problem of calibration and parameterization for animation. Due to the flexibility of MPEG4 coding we have used the MPEG4 facial animation standards to create our 3D talking head.

Having discussed the various animation modeling techniques available we will now discuss the application of emotion modeling on the 3D talking head to simulate realistic emotional conversation.

2.3 Modelling emotions

Emotions can be defined as the internal state of mind that could be affected by any external stimuli. For example: sad emotions arise in a situation when the user is not able to fulfill his/her goals. Emotions are very complex to model. There are few models for emotions that have proven to be effective when dealing with machines. Those emotional models have been created based on the psychological theories of humans. Some of the famous models for emotions are:

- Plutchik's wheel of emotions
- Ortony Clore and Collins Model for emotions (OCC)

2.3.1 Plutchik's wheel of Emotions

Robert Plutchik was an ancient physician who constructed and classified emotions based on certain hypothetical postulates. He established the Plutchik's wheel of emotions in the year 1960. It can be treated as one of the first models of emotions that came into existence, even before the OCC model for emotions.

Plutchik proposed some basic postulates for his emotional theory [Plutchik91]:

- There are a small number of pure or primary emotions.
- All other emotions are mixed that is, they can be synchronized by various combinations of the primary emotions.
- Primary emotions differ from each other with regard to both physiology and behaviour.
- Primary emotions in their pure form are hypothetical constructs or idealized states whose properties can only be inferred from various kinds of evidence.

- Primary emotions may be conceptualized in terms of pairs of polar opposites.
- Each emotion can exist in varying degrees of intensity or levels of arousal.

The figure shown below explains the Plutchik's wheel of emotions. It also shows that how the primary emotions are combined together to get a new emotion.

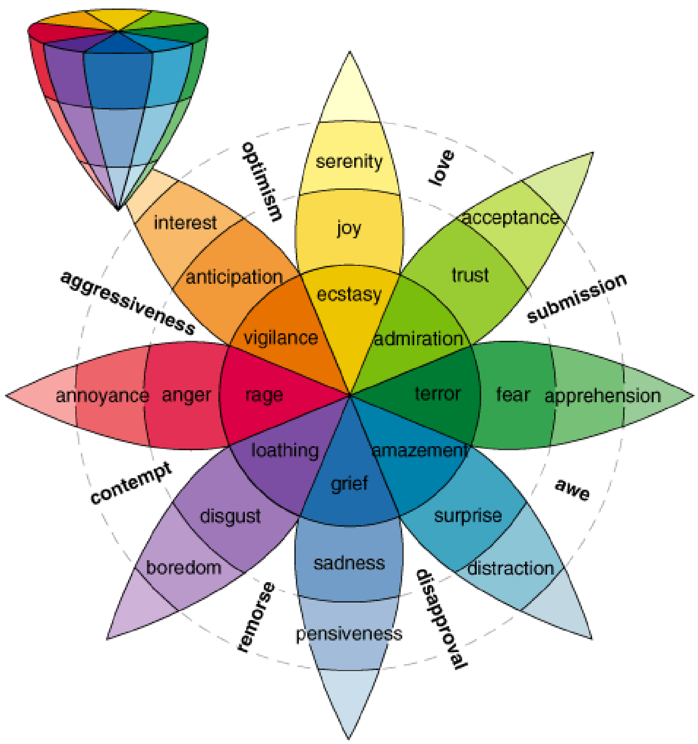


Figure 2.7: Plutchik's Wheel of Emotion [Plutchik91]

Even though Plutchik's emotional theory has been one of the oldest and well known models for emotional theory, it has not found its application in embodied conversational agents. The model classified the emotions as a combination of different primary emotions but never explained how it can be applied to generic computational models that can acquire emotions. Plutchik's emotional theory was a research on the emotions itself rather than a research on how to apply these emotions to computational models. Therefore, it's more convenient to use an emotional model that can be applied to embodied conversational agents. In contrast to Plutchik's model, the OCC model came into existence so that it could be applied to different computational models.

2.3.2 Ortony Clore and Collin's model for Emotions (OCC Model)

OCC model for emotions is one of the most renowned models for emotions. This model was published in 1988 by Ortony, Clore and Collins in their book "The Cognitive Structure of Emotions". The OCC model was one of the first models for emotions that could be applied to machines and embodied conversational agents. It was constructed based on the psychological theories of humans. They classified 22 types of emotions as shown in the figure below. [Ortony88]

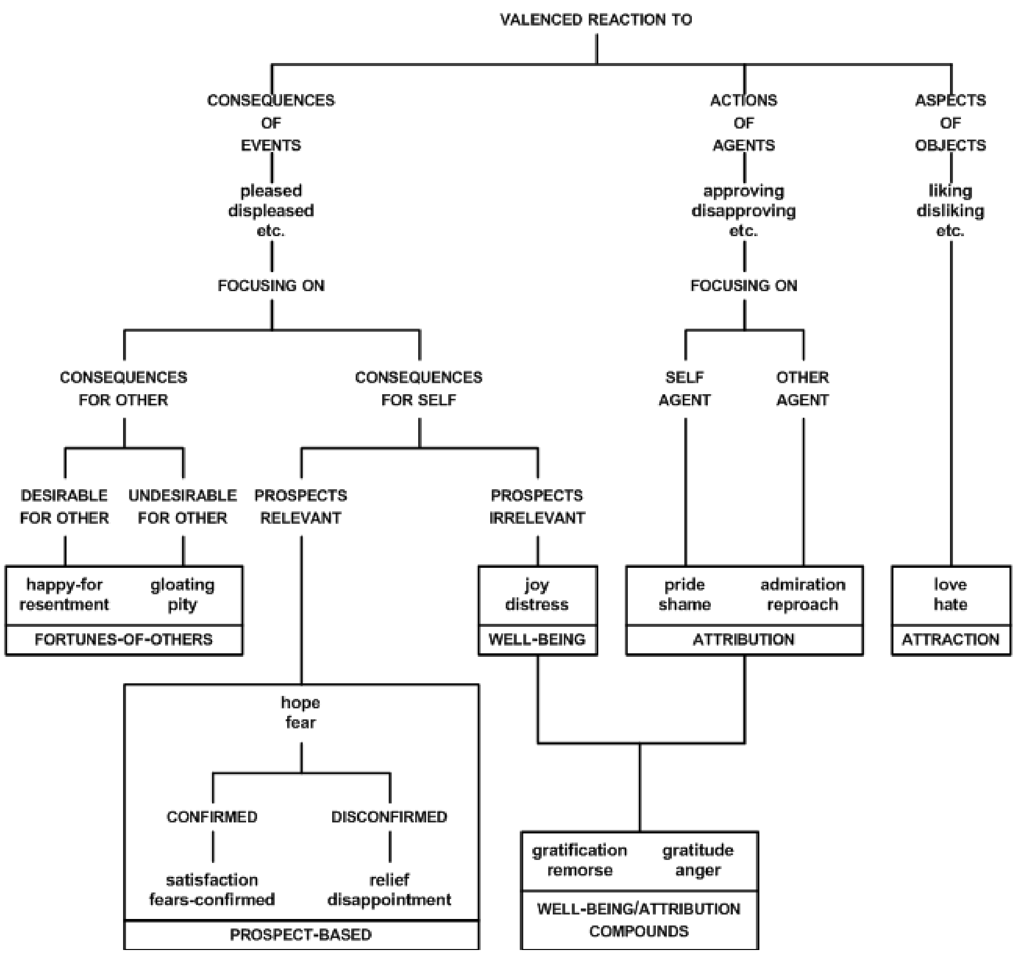


Figure 2.8: The OCC model for emotions [Ortony88]

The 22 types of emotions are:

- Pleased
- Displeased
- Happy-for
- Gloating
- Pity
- Satisfaction
- Fears
- Relief

- Disappointment
- Joy
- Distress
- Pride
- Shame
- Admiration
- Reproach
- Gratification
- Remorse
- Gratitude
- Anger
- Love
- Hate

Ortony, Clore and Collins classified emotion on the following variables:

- Type specification
- Tokens
- Variables affecting intensity

Ortony, Clore and Collins further classified emotions based on their valanced reaction to three factors:

- Consequences of events
- Action of agents
- Aspects of objects

The classification of one of the emotions ‘Fear’ using the above mentioned three factors is shown below:

FEAR EMOTIONS
TYPE SPECIFICATION: (displeased about) the prospect of an undesirable event
TOKENS: apprehensive, anxious, cowering, dread, fear, fright, nervous, petrified, scared, terrified, timid, worried, etc.
VARIABLES AFFECTING INTENSITY:
(1) the degree to which the event is undesirable
(2) the likelihood of the event
EXAMPLE: The employee, suspecting he was no longer needed, feared that he would be fired.

Figure 2.9: The OCC classification of the emotion ‘Fear’ [Ortony88]

The figure shown below represents the emotion type specifictions for all 22 emotions in the OCC model.

Joy:	(pleased about) a desirable event
Distress:	(displeased about) an undesirable event
Happy-for:	(pleased about) an event presumed to be desirable for someone else
Pity:	(displeased about) an event presumed to be undesirable for someone else
Gloating:	(pleased about) an event presumed to be undesirable for someone else
Resentment:	(displeased about) an event presumed to be desirable for someone else
Hope:	(pleased about) the prospect of a desirable event
Fear:	(displeased about) the prospect of an undesirable event
Satisfaction:	(pleased about) the confirmation of the prospect of a desirable event
Fears-confirmed:	(displeased about) the confirmation of the prospect of an undesirable event
Relief:	(pleased about) the disconfirmation of the prospect of an undesirable event
Disappointment:	(displeased about) the disconfirmation of the prospect of a desirable event
Pride:	(approving of) one’s own praiseworthy action
Shame:	(disapproving of) one’s own blameworthy action
Admiration:	(approving of) someone else’s praiseworthy action
Reproach:	(disapproving of) someone else’s blameworthy action
Gratification:	(approving of) one’s own praiseworthy action and (being pleased about) the related desirable event
Remorse:	(disapproving of) one’s own blameworthy action and (being displeased about) the related undesirable event
Gratitude:	(approving of) someone else’s praiseworthy action and (being pleased about) the related desirable event
Anger:	(disapproving of) someone else’s blameworthy action and (being displeased about) the related undesirable event
Love:	(liking) an appealing object
Hate:	(disliking) an unappealing object

Figure 2.10: The OCC type specifications of 22 emotions [Ortony88]

The OCC model has been integrated into several embodied conversational agents. The OCC model defines the rules for generation of emotions. The agents actually obtained the ability to reason and express their emotions based on the above mentioned three factors: consequences of events, action of agents, and aspects of objects. The model has been proved to be successful when applied to embodied conversational agents in the past [Zhen08]. This was one of the main reasons that motivated us to integrate the OCC model to the proposed framework. Even though [Zhen08] established a firm cognitive model for emotions they didn't explain how the emotions can be applied to an embodied conversational agent and how the emotional state of these agents were affected after the emotions were generated. Therefore, the main focus of this thesis is to generate emotions based on the situational information calculating the current emotional state of the agent which is backed up by the OCC model for emotions.

Having discussed the modeling of emotions we will now turn to a discussion of dialogue management systems which drive the Embodied Conversational Agent.

2.4 Dialogue Management Systems: An overview

A dialogue management system is a system that processes the user's queries and gives feedback based on its domain knowledge. Dialogue management systems play a key role in customer service by processing user requests in several domains such as in a flight reservation system. Dialogue management systems also maintain a dialogue history which stores and updates the dialogue states, and finally decides the action taken so as to control the flow of dialogue conversation between user and the machine. According to [Trung06] the dialogue management system is classified into four different categories:

- Finite state and frame based approach
- Information state and probabilistic approach
- Plan based approach
- Collaborative agent based approach

Finite state and frame based approaches are the most basic and simple approaches to handle dialogues. It represents a state transition network with nodes that show the system utterances and the system transition in the network to solve a problem.

The information state and probabilistic approaches were mostly used for constructing multimodal dialogue systems, but unfortunately it was unable to handle to handle high complexity dialogues.

The plan based approaches came into existence to replace the information and probabilistic approaches due to the limitations in handling dialogue complexities.

Finally, the collaborative agent based approach was introduced to replace the plan based approaches because they were able to capture the motivations and the mechanism of the dialogue structure.

Given a specific category of system, various techniques and approaches to the system's dialogue process exist. We will discuss these techniques in the next section.

2.5 Dialogue Management Techniques

Different dialogue management techniques have been introduced since the first one that came into existence in 1999. Some of the popular dialogue management techniques are:

- Handcrafted Dialogue Managers
- Bayesian Network for Dialogue Management
- Supervised Learning and Markov Decision Process
- Semi Markov Decision Process
- Partially Observable Markov Decision Process
- Modified Partially Observable Markov Decision Process

The handcrafted dialogue management systems were one of the oldest techniques that were introduced to handle dialogue conversations between a human and machine. It was not the most efficient technique for creating dialogue management systems. They were time consuming and failed to work properly in complex situations. The greatest advantage of this technique was its simplicity in construction and handling. [Sabiha10]

The Bayesian network was another variant of the of the handcrafted dialogue management systems. It was introduced to replace the legacy handcrafted dialogue managers, but it suffered from processing the sub goals for a system based on different multimodal information sources. [Prodanov04]

The Supervised Learning and Markov Decision Process proved to be better than the existing handcrafted dialogue management systems such as the Bayesian network. It maintained a single dialogue and the actions were learnt from the corpus [Sabiha10]. Even though maintaining a single state was considered to be more effective in comparison to the handcrafted dialogue managers, the Supervised Learning and Markov Decision Process suffered from limitations such as creating a dialogue policy.

The Semi Markov Decision Process was first proposed by [Cuayahuitl09] and the main goal of the system was to choose the best action which resulted in the change of dialogue states. The Semi Markov Decision Process was later replaced by the Partially Observable Markov Decision Process, which we will now focus our discussion on.

2.5.1 Partially Observable Markov Decision Process (POMDP)

According to [Jason05] the POMDP is defined as a tuple, $\{S, A, T, R, O, Z, \lambda, b_0\}$,

where S is a set of states describing the agent's world; A is a set of actions that an agent may take; T defines a transition probability $P(s'|s,a)$; R defines the expected (immediate, real-valued) reward $r(s,a)$; O is a set of observations that the agent can receive about the world; and Z defines an observation probability, $P(o'|s',a)$; λ is a geometric discount factor $0 \leq \lambda \leq 1$ and b_0 is an initial belief state $b_0(s)$. For any given belief state we have the following equation:

$$b'(s') = p(s'|o',a,b) = \frac{p(o'|s',a,b)\; p(s'|a,b)}{p(o'|a,b)}$$

$$= \frac{p(o'|s',a)\sum_{s\in S} p(s'|a,b,s)p(s|a,b)}{p(o'|a,b)}$$

$$= \frac{p(o'|s',a)\sum_{s\in S} p(s'|a,s)b(s)}{p(o'|a,b)}.$$

[Jason05] proved that the POMDP is completely robust and can handle the uncertainty caused by the speech recognition errors by comparing it with existing techniques. One of the first applications of POMDP in dialogue management was the nursing home robot application proposed by Roy, et al. [Roy20]. Since then, POMDP found its applications in several spoken dialogue management systems. The POMDP approach was later replaced by a Modified POMDP model by [Sabiha10].

Contextual dialogue control of the system utilizes the Hollnagel's Contextual Control Model, which we will discuss in the following section.

2.5.2 Hollnagel's Contextual Control Model (COCOM)

In 1993 Eric Hollnagel introduced the contextual control model to assess team behaviour based on the time available. The main contribution was that the system decided the actions based on the context of the situation and available time. He categorized team behaviour into four different modes:

- Scrambled Mode
- Opportunistic Mode
- Tactical Mode
- Strategic Mode

Hollnagel's COCOM was put to test by Stanton, et al. [Stanton01] where they created 24 groups of people with 4 people in each group. There were 74 males and 22 females of the age between 19 and 55. They made six groups of four people in each group work on balancing a simulated gas network system. As an outcome of this experimentation they reliably categorized the four control modes and showed that the progression between control modes conformed to a linear progression.

The figure below shows the COCOM model, detailing the movements available between the different control modes. We have applied the same COCOM to the proposed emotional model for emotion generation.

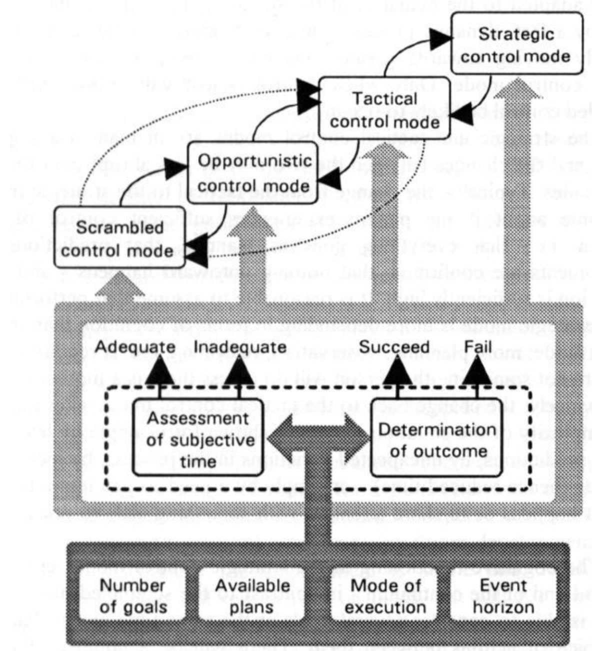


Figure 2.11: Flow of the Four Contextual Control Modes (COCOM) [Stanton01]

2.5.3 Modified Partially Observable Markov Decision Process

The Modified POMDP model was established to overcome the limitations of the existing POMDP approaches. According to [Sabiha10] the existing POMDP approaches considered the world to be static and always made its decisions based on the current belief state. The author also pointed out how the existing POMDP approaches ignored the dialogue history to make decisions effectively. They created a new POMDP approach known as the Modified Partially Observable Markov Decision Process which combined Hollnagel's Contextual Control Model for decision making based on user input [Kim06].

The dialogue manager switches from one mode to another for processing user queries. For example if the dialogue manager is not able to understand the user request it switches to the scrambled mode where panic sets in. The dialogue manager will attempt to move to tactical mode by giving the user a set of options whereas other dialogue managers try to repeat the same question again and again until the user comes up with the right answer.

The dialogue manager achieves the goal by transitioning from mode to another mode using forward planning. The lowest level of control will be the scrambled control mode where the system does not have a proper understanding of the user's queries and the most desirable level of control mode will be strategic mode where the system has a clear understanding of the user's queries. The system chooses the best action based on the context of the dialogue conversation and transitions between the modes depending on the dialogue states, current action, context of the situation and available time. The system proved to be effective in handling the uncertainty caused by speech recognition errors and performed much better at handling conflicts in comparison to the existing POMDP approaches.

Having discussed the dialogue management system we will now turn our focus on the realization of the Embodied Conversational Agent. We start with an overview of a scripting language that is used for animation and speech synthesis of our agent.

2.6 SMIL Agent Scripting Language: An Overview

The Synchronized Multimedia Integration Language script, or SMIL, is used to create interactive audio/visual presentations. SMIL 1.0 is the new standard for driving the facial animation of MPEG4 3D talking heads. Some of the other languages used for driving MPEG4 3D talking heads include:

- APML
- MPML
- VHML

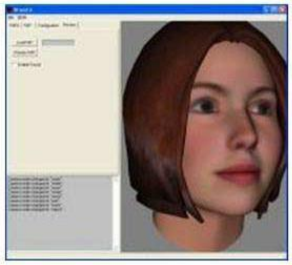
- CML

According to Elena, et al. [Elena05] the above mentioned languages lack independence from communication context, independence from specific character, channel independence, adaptability, standardization, and extendibility. These scripting languages are strictly coupled with the applications and cannot be reused for another application. They are also coupled to the performance parameters of a specific agent. The previous scripting languages lacked the flexibility to be reused with other agents.

On the other hand, SMIL focuses on conveying message to the users through different communication channels such as speech animation, voice, and facial animation. Since it handles different channels for communication at once the users experience a much better interactive environment than before. The script does not carry any information that is specific to the domain and hence, it is very easy to reuse the talking heads and the scripts for different purposes.

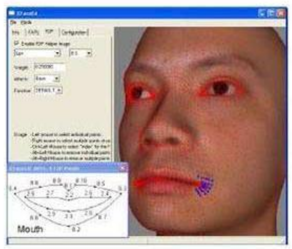
The 3D talking heads driven by the SMIL Agent script could be reused across different domains because the script does not carry any information that is strictly coupled to the domain itself. For example: The 3D talking head used in a healthcare domain can also be used for a pizza ordering domain. The 3D talking heads driven by the SMIL Agent script are completely flexible and adaptable.

Figure 2.12 illustrates two different 3D agents which could have different performance parameters defined. The SMIL scripting language is capable of handling both agents even though different parameters are present. Either agent can then be used a variety of systems and domains.



Alice synthetic agent

Channels	Performance Abilities	Affected by
Voice1	speech	
Voice2	speech	emotion
Face	speech-animation	emotion
	expression	emotion
Head	pointing	
	turning	
Eyes	pointing	



John synthetic agent

Channels	Performance Abilities	Affected by
Voice	speech	
Face	speech-animation	
Head	turning	

Figure 2.12: Alice and John are two different synthetic agents with different performance parameters [Elena05]

2.7 Review of Existing Embodied Conversational Agents

We will now critically review the state-of-the-art of embodied conversational agents in this section. We will be reviewing some of the popular ECA's such as the Microsoft agent, Chinese talking head, LUCIA and chatter bots.

2.7.1 Microsoft Agents

Microsoft Agents were one of the most primitive types of embodied conversational agents. These agents created a new wave in the field of Human Computer Interaction. Microsoft Agents 2.0 was developed to provide assistance to people browsing web pages or even using specific programs on a computer such as Microsoft Office. The characters were able to communicate with users through a text to speech engine or using some pre-recorded sounds. Microsoft Agents found their applications in providing assistance to the users by reading their email aloud, or helping the users navigate a web page.

Some of the famous characters that included in the package were:

- Clipit (the paper clip)
- The Dot (a shape-shifting and color-shifting smiley-faced red ball)
- Hoverbot (a robot)
- The Genius (a caricature of Albert Einstein)

Figure 2.13: Example of a Microsoft Agent - Peedy

With the recent developments in Human Computer Interaction the ECA's of today have reached a different level of technology which is much better than the first set of Microsoft Agents. The ECA's these days have much better interfaces; they have Artificial Intelligence to a certain extent and also have emotions to improve the entire communication experience for users. The following sections will overview some of the state-of-the-art ECAs in research and commercial applications.

2.7.2 Emotional Chinese Talking Head

The virtual agents of present day appear more realistic and natural with the integration of audio and visual display. The user experience and believability increases when the agents are able to express their emotions accompanied by the audio/video displays. Tao, et al. [Tao04] created a similar emotional Chinese talking head system. The talking head recognized the emotions from user text and attempted to produce an emotional audio/video output.

Though the system is effective in terms of recognizing the emotions from text, it lacked the visual display of emotions. The talking head was not realistic comparatively because it did not use multiple channels to express emotions. Every single channel; such as the voice, face, eyes, eyebrows, mouth, body, arms, and head; can realize the emotions individually. The results confirmed that happiness and sadness were not distinguished very effectively using merely facial features while their corresponding acoustic features turned out well.

Happiness and sadness are the two extremes of emotions and it is very important for the user to be able to differentiate between them from the agent. Figure 2.14 shows an example of the talking head responding to user input.

As it turns out the paper laid more emphasis on emotion recognition than the display of emotions from the agent.

Figure 2.14: Emotional Chinese Talking Head [Tao04]

2.7.3 LUCIA

The LUCIA is an Italian emotional 3D talking head. They used INTERFACE to create the 3D model and used Festival for Text to Speech conversion. LUCIA was able to mimic human emotions by observing the human in front of the system as shown in Figure 2.15. It used an optoelectronic device that placed markers on the user's face. The displacement of the facial muscles from the markers denoted a movement and the system looked up a database to find the respective emotion of the user. LUCIA is still one of the best emotional 3D talking heads that can track the human emotions and mimic them. LUCIA basically mimics the human and does not engage in any type of interaction or problem solving. [Piero08]

As a result, although LUCIA is exceptional at understanding the user and generating various emotional responses, it fails at creating emotions based on its own internal goals. The focus of this thesis is to create an emotional talking head that is aimed at problem solving situations where the 3D talking heads can interact with the users and process their queries in an interactive environment.

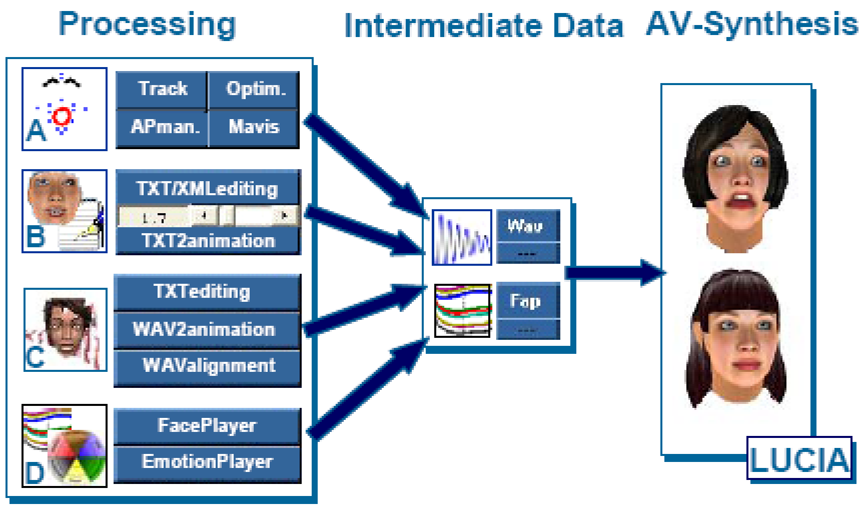


Figure 2.15: Architecture of LUCIA [Piero08]

2.7.4 Emotion based Tutor

Adina, et al. [Adina05] embedded emotions in an artificial tutor based on a BDE architecture. The tutor's emotions were also based on the OCC model and classified as "Affective level" and "Cognitive level". The emotions and attitude of the student played a key role in the planning and decision making process of the tutor. The system aided in stimulating the presence of an empathic learning companion, which encouraged the student towards a faster progress to knowledge acquisition. However, the authors did not discuss anything about an interface (facial animation) that could realize emotions and details regarding the transition from one emotional state to another were unclear. The system also failed to include the factor of emotional history which is important in creating a continuous and believable conversation between user and agent.

2.7.5 GRETA

GRETA is a 3D embodied conversational agent that can interact with people by means of verbal and non-verbal behaviours. GRETA still plays a major role in some of the most significant human computer interaction projects such as:

- HUMAINE – a community that carried out their research on Embodied Conversational Agents. GRETA was an active part of the HUMAINE community .
- SEMAINE – aimed at developing a multimodal architecture with SAL (Sensitive Artificial Listener). It is evident from the video hosted on their website that GRETA is able to communicate with an anonymous user on the other end, using verbal and non verbal behaviours. GRETA is able to display continuous head nods while she is listening to the user and can also gaze when needed. [Semaineu]
- CALLAS – standing for Conveying Affectiveness in Leading-edge Living Adaptive Systems, is an integrated project funded by the European Commission under FP6. The project has been named in the honour of Maria Callas who was a great singer. The project is aimed at developing a system or framework that can bridge the gap between conveying emotional aspects and multimodal inputs. [Callas06]

The 3D model for GRETA was created under the MPEG4 Facial Animation standard. Even though GRETA has found its applications in different projects it is nothing more than a virtual entity that can display emotions. GRETA does not have an emotional model for generating those emotions. In other words GRETA can display complex emotions when it is hardcoded in the system using the APML.

Figure 2.16: GRETA [Stefano08]

2.7.6 Chatter bots

Chatter bots are virtual entities that can communicate with users via text, facial animation, voice and emotional displays. There is not much difference between chatter bots and embodied conversational agents. Chatter bots can be thought of as the primitive, or basic, version of the embodied conversational agents. In general, ECAs have more capabilities compared to chatter bots. There are several chatter bots that are still being used over thousands of websites. We will review a few chatter bots that can be comparable to our proposed ECA:

Verbots [Verbots04] – can be added to your existing web pages for providing virtual assistance to the user. They have an interface that comes with a variety of talking heads which are designed using Flash scripts.

Most of their characters are 2 Dimensional agents and absolutely lack the realism due to that factor. They come with basic emotional capabilities such as smiling or frowning. Therefore, they also lack in displaying true emotions. The Verbots lack proper lip synchronization with the audio and assist the user based on their knowledge base and there is little Artificial Intelligence involved.

The Verbots communicate with the users via text and voice and serve as basic ECA's for any website. The figure shown below is an actual Verbot initiating a conversation with the user. The Verbot's knowledge base did not have any information about greeting the user and henceforth it was not able to recognize the user input. Another drawback of the knowledge based design is that developers have to create a large knowledge base if they wanted the 3D talking head to recognize every single user utterance properly.

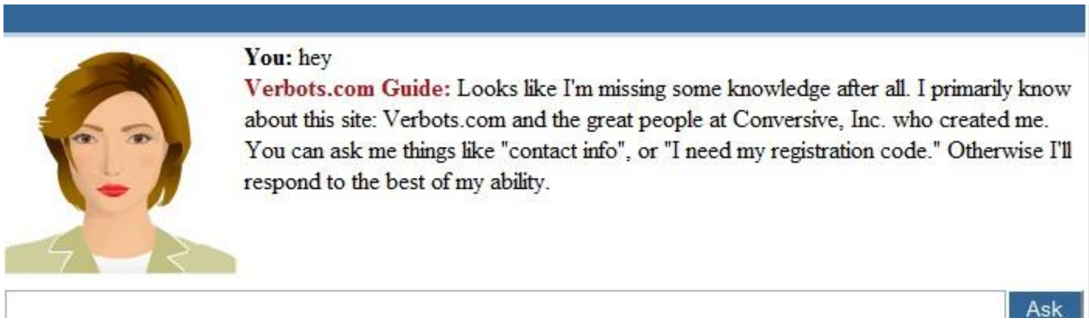


Figure 2.17: An example of a Verbot conversation with a human [Verbots04]

Cleverbot – is a virtual entity that has been hosted online on a server. Users can interact with the Cleverbot by asking questions. The intelligence of the Cleverbot is quite significant in comparison to other chatter bots available. The Cleverbot learns from the users and also from the internet [Cleverbot10]. They communicate with users through text and have no interface. The figure shown below is an actual snapshot taken from the Cleverbot web page. The Cleverbot maintains a knowledge base like any other chatter bot. It is obvious from the figure shown below that their knowledge is definitely limited.

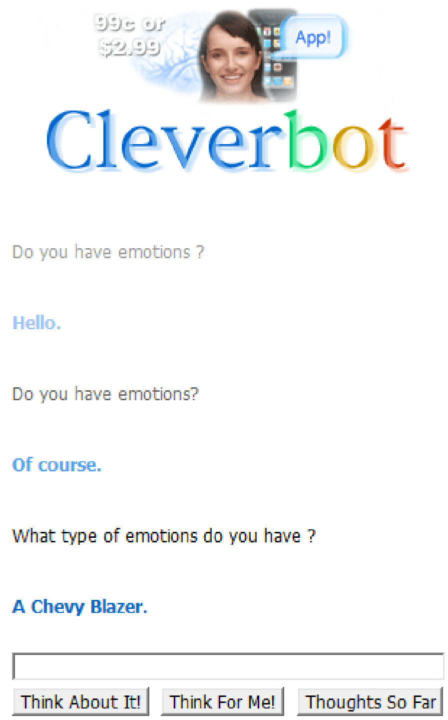


Figure 2.18: An example of a Cleverbot conversation with a human [Cleverbot10]

Aaron- is a virtual agent for Futures Shop, Canada [Futureshop11]. He tries to answer users' queries by performing keyword based searches on the Future Shop website. He has a photorealistic face but does not have any real-time emotions.

The interface basically plays an animation from a repository of pre-recorded animations based on the keywords typed by the user. There are several agents that are being used today in hundreds of websites that just play back a pre-recorded animation from a database of videos like Aaron. They might look realistic in the beginning but as the user continues to use the website he/she will find out that the agent is a pre-recorded animation. The agent can be annoying sometimes because it becomes repetitive with responses. The system cannot handle uncertainty and lacks intelligence.

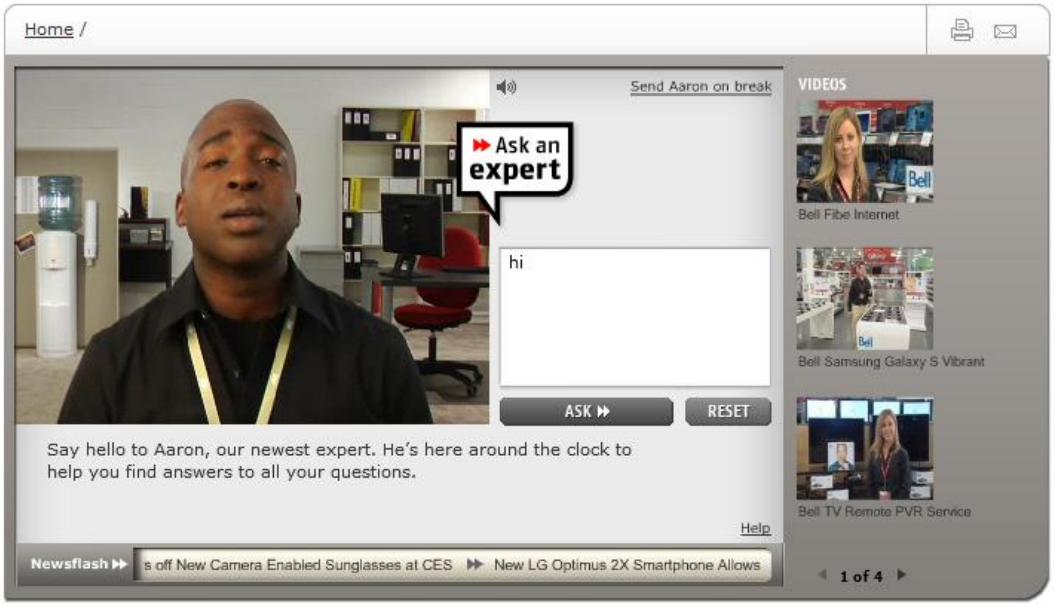


Figure 2.19: Aaron from Future Shop's website [Futureshop11]

Anna – is a representative for the IKEA products on IKEA's web page [Ikea99]. She is a 2-Dimensional cartoon character who provides assistance for the online shoppers. She communicates with the users through text. Like Aaron, she does not have any emotions and has limited user interaction. Her knowledgebase is limited and cannot handle uncertainty. She retrieves the information about the products using the keyword based content retrieval technique. The agent could be annoying at situations where the user mistyped a word or if the user asked something about the website that was not a part of her knowledge base.

Figure 2.20: IKEA's Anna [Ikea99]

2.8 Conclusion

We have reviewed different areas of research such as facial animation, emotion modelling, POMDP dialogue management system, scripting languages for driving facial animation, and some existing embodied conversational agents. Most of the existing 3D talking heads laid more emphasis on individual areas of research, such as capturing user emotion or creating realistic looking emotional output, rather than combining everything together to create a system that can process user requests and interact with them through different communication channels such as voice and facial expressions.

Integration of a powerful dialogue manager such as the Modified POMDP dialogue manager with an effective emotional ECA was never accomplished in the past. The absence of an existing 3D talking head that combines the state-of-the-art MPEG4 facial animation; a dialogue manager, especially the Modified POMDP dialogue manager, backed up by a responsive emotional model; that combines the OCC model and Four Contextual Control Modes are evident from the literature review.

In the next chapter we discuss the methodology and design specifics which realise an integrated dialogue management system which will enhance the user interaction experience.

CHAPTER III

DESIGN AND METHODOLOGY

This chapter will provide a deep insight about the proposed architecture. The beginning of this chapter will discuss the techniques involved in creating an emotional model using OCC model for emotions and Four Contextual Control Modes. Towards the end of this chapter we will also explained the steps involved in creating a 3D talking head with different types of emotions, how the SMIL Agent scripting language drives facial animation via control of all communication channels individually, and the integration of Modified POMDP dialogue manager to the entire system. This chapter will also overview the tools involved in the creation of an embodied conversational agent, explaining its significance and henceforth justifying the reason for choosing those tools. Before we proceed further, we would like to highlight our main contributions towards this thesis work:

- **Creation of a 3-Dimensional Emotional Embodied Conversational Agent that expresses it emotions through different animation channels, such as eyes and lips, using MPEG4 Facial Animation**
- **Creation of SMIL Script Generator that drives the MPEG4 facial animation with speech**
- **Proposal of a new model for Emotion Generation combining the OCC model for emotions and the Four Contextual Control Model (COCOM)**
- **The proposed framework will adhere to the Four Contextual Control Modes and the changes in emotions are “dynamic” and “smooth”**

- **Integration of Modified POMDP Dialogue Manager with the proposed framework and test the functionality of the system in a Pizza Ordering Domain**
- **Extending the Modified POMDP Dialogue Manager to handle special user requests which were not implemented previously**
- **Modifying the existing model to fit the proposed emotional model; generating better confidence scores and rewards**

3.1 Proposed Architecture: An Overview

The figure shown below is the proposed architecture of our system. This section will give an overview of the new architecture and the consecutive sections will review all the components in detail. The core components of the new framework are:

- Modified POMDP dialogue Manager
- Emotion Generator
- SMIL Agent to FAP converter
- Text to speech engine
- MPEG4 Compliant Facial Animation Player

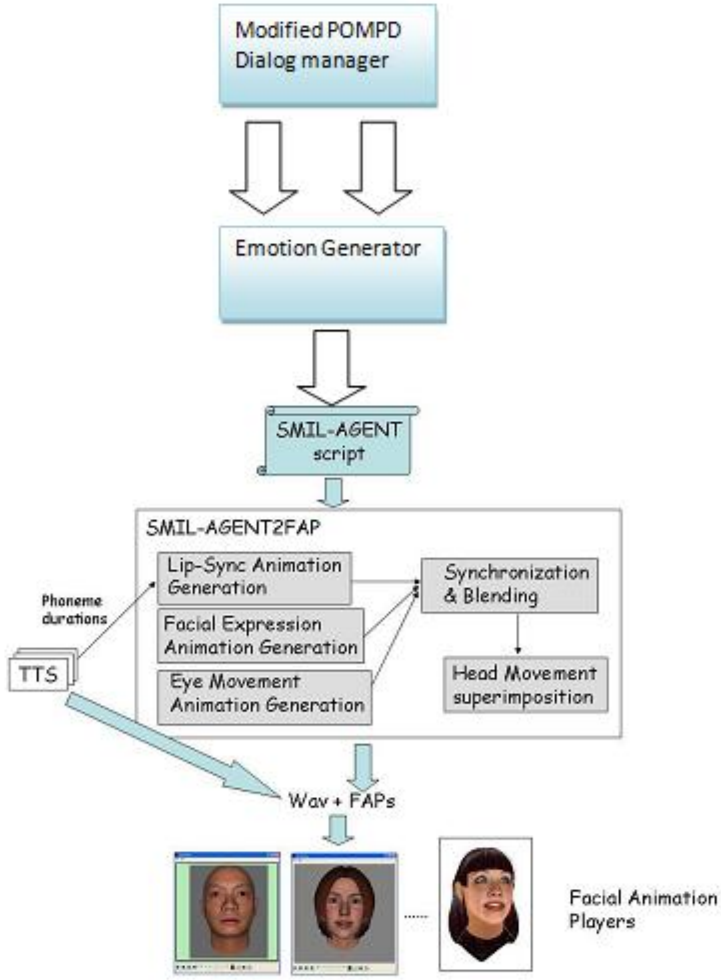


Figure 21: The proposed architecture of our system

The Modified POMDP dialogue manager is responsible of processing the user requests. The dialogue manager processes the queries by switching between the four modes such as Scrambled mode, Opportunistic Mode, Tactical Mode and Strategic Mode. The dialogue manager switches between the modes based on the confidence score obtained for the dialogue state.

The confidence score is a numerical representation of how much the system understood from the current conversation. If the confidence score is low then it means the system did not understand all the user utterances. If the confidence score is high then it means that the system was able to understand most of the user utterances.

For example, if the confidence score is low for the current conversation then the system is in Scrambled Mode where panic sets in because the system has lost its control over the dialogue. The system tries get out of the Scrambled control mode by making a random decision. The random decision might help the system to proceed to the next mode or even fall back to Scrambled Mode. In other words, the random decision might be something that the user actually wanted. In that case the system proceeds further and steps up to the next control mode.

On the other hand, the random decision might not be something that the user actually wanted. In that case, the system falls back to the Scrambled Mode and tries to get more information about the situation. If the user did not provide any information or enough information, then the system steps up to the Opportunistic Mode and tries to give a set of options to the user instead of throwing errors or repeating the same query.

The emotion generator keeps track of the current mode and previous mode, and will generate the appropriate emotion. The rules for generating the emotion is based on a combination of OCC model for emotions and the Contextual Control Modes used by the Modified POMDP dialogue manager. The emotion generated will be transformed into SMIL Agent Script with appropriate parameters for eye movement, blinking, and text to be animated. The text to be animated will be generated by the modified POMDP dialogue manager.

Once the SMIL Agent Script is generated it is processed at the second level which implements a client-server architecture. The SMIL Agent Script to FAP converter acts as the client for processing for SMIL Agent Scripts for the Xface player, which is an MPEG4 Facial Animation player. The Xface player acts as the server and plays the animation. The Client processes the script from time to time and sends it to the server as a “Task” for playing back the animation with appropriate emotions. The text to be animated is converted to speech using MSSAPI 1.5 text to speech engine. The Xface player will playback the animation (FAP) synchronized together with the audio (WAV) file.

We will now explain every individual components of our system in more detail in the following sections.

3.2 Modified POMDP Dialogue Manager

The Modified POMDP dialogue manager evolved to replace the POMDP dialogue manager. The Modified POMDP dialogue manager was obtained from modifying the POMDP model by incorporating the Four Contextual Control Modes. Unlike MDP and POMDP the Modified POMDP model is able to handle uncertainty. The Modified POMDP also uses dialogue history to make dynamic decisions and handle conflicts. The Modified POMDP model is the best dialogue manager available till date that makes use of the situational information to make decisions. The proposed emotional model requires situational information and henceforth the Modified POMDP model is found to be suitable for that purpose. The Modified POMDP model has its own limitations such as handling text and speech errors. If the dialogue manager was not able to understand any of the user utterances it moves to scrambled control mode and make a random decision or move to tactical mode and give the user a set of options to choose from. We extended the Modified POMDP model for better handling of text and speech errors. The proposed emotional model always tries to cope up with its positive emotions and stay happy. The emotion switches to negative emotions when it doesn't understand the user. But the agent heavily tries to avoid the negative emotions by communicating with the Modified POMDP dialogue manager and the dialogue manager tries to clarify anything that it did not understand during the conversation. Therefore, when the agent does not understand any or some of the user utterances then the emotional model instructs the Modified POMDP dialogue manager to compare the strings input by the user to the keywords from the knowledgebase. It is done using the 'Jaro Winkler' algorithm proposed by Winkler in the year of 1990. Java allows you to use the Jaro Winkler library directly. According the Jaro Winkler algorithm calculates the similarity between two given strings and normalizes the value to 0 or 1. If the algorithm generates a '0' then it's not a match and if it generates a '1' then it's a perfect hit. We have used the algorithm to help the Modified POMDP dialogue manager handle the text, speech errors and generate better confidence scores.

3.3 OCC model and Four Contextual Control Modes for Emotion Generation

In this section we will elaborate about the methodology adopted for generating emotions. The OCC model for emotions has been combined with Hollnagel's Four Contextual Control Modes to generate emotions [Stanton01]. The OCC model clearly reviewed a cognitive model for classifying 22 different types of emotions. They have classified emotions based on three factors:

- Consequences of events
- Action of agents
- Aspects of objects

As mentioned in Chapter II even though this model served as a baseline for classifying emotions the authors did not mention specific details on how to implement the emotional model in a software setting. The OCC model also did not specify possible causes for triggering such an emotion. We answered those questions by integrating Hollnagel's Four Contextual Control Modes to the OCC model. According to our new proposed framework, the emotions are definitely an outcome of external stimuli.

The emotions of this 3D talking head are triggered based on the four modes which updates it with situational information. For example, if the agent is in Scramble Mode it means that the agent did not understand some or most of the user utterances. The agent, a goal based or task oriented agent, is in a panic situation where it is not able to process the user requests. The agent's emotional stability tends to vary based on whether it is progressing or regressing from the goal.

A real life situation will draw a clear picture about our proposed emotional framework. For instance, let us assume a situation where a student has to score 75% or above to pass the course and proceed to the final year of his school. Any situations such as the student falling sick during the study holidays, family emergency, etc will definitely affect his emotional stability because it might hold him back from achieving his goal, which is pass the course with 75% or above, and proceed to the final year of his school.

The same logic can be applied to our agent who processes user requests in a pizza ordering domain. The agent switches between the four modes to achieve an optimal goal. The transition between the four modes clearly states the agent's situation and updates the emotions accordingly.

The previous emotional models generated emotions based on fuzzy theories; multimodal input such as camera, microphone, or text; and most of these models were focussed on emotion detection rather than emotion generation. We have established an emotional model for a goal based agent whose emotional stability varies based on whether the agent progresses or regresses from the agent's goal.

3.4 Emotion Generator

The flowchart shown below illustrates the OCC model. The highlighted area denotes that we have used only those emotions from the OCC model that will be appropriate for a goal based agent. The goal based agent is completely different from an empathic agent which focuses on both Consequences for Other (the user) and Consequences for Self (the agent). Instead it will focus only on the Consequences for Self, which will indirectly impact the consequences of others as well.

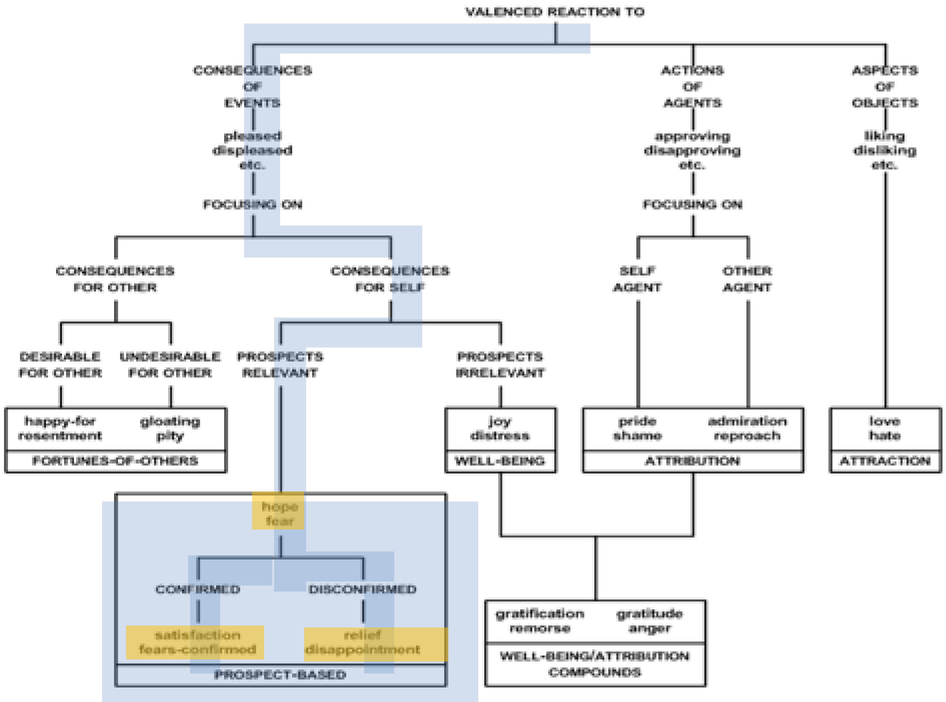


Figure 22: A subset of the emotions from the OCC model used for our thesis [Ortony88]

As we mentioned earlier, Hollnagel's Contextual Control Model for studying team behaviour is based on the Four Contextual Control Modes. Stanton explained the modes as follows [Stanton01]:

- Scrambled Control - is characterised by a completely unpredictable situation where the operator has no control and has to act in an unplanned manner, as a matter of urgency. An example of this may be where there is a sudden accident or emergency, where the operator is unfamiliar with the situation and/or lacks experience in what to do - the engineer's behaviour may be impulsive or even panicky. Consequently this is the mode where most errors occur.

- Opportunistic Control - is characterised by a chance action taken due to time, constraints and again lack of knowledge or expertise and an abnormal environmental state. An example of this may be in a situation where operators are driven by the perceptual dominance of system interface (alarms, lights, noise), and will revert to habitual heuristics, (Reason, 1990). In certain situations, opportunistic control may be used as a way of exploring a problem or situation and testing out alternative solutions, because of an unusual occurrence. This is referred to as 'Explorative Control'.
- Tactical Control - is more characteristic of a pre-planned action, where the operator will use known rules and procedures to plan and carry out short term actions. Consequently, fewer errors will be made than in the previous modes; however the operator is still heavily driven by the immediacy of the situation, and therefore will still be influenced by the system interface.
- Strategic Control - is defined as the 'global view', where the operator concentrates on long term planning and higher level goals. In this mode, the operator will have evaluated the outcome more precisely, and considered the relationship between action and its pre-conditions; s/he will therefore have more overall control of the whole situation or task.

The figure below illustrates the emotions generated based on the transition between the four modes. It also explains how the OCC model has been integrated to the Four Contextual Control Modes that focuses on forward planning.

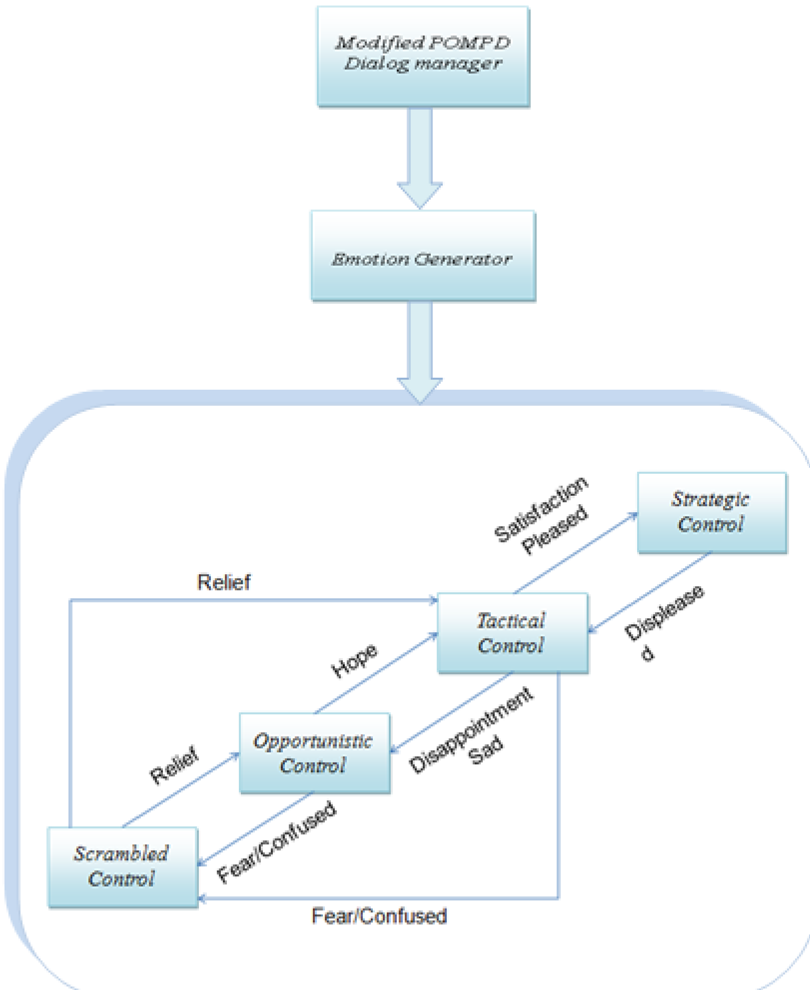


Figure 23: Integration of Four Contextual Modes with OCC Model for Emotions

The COCOM was integrated to the modified POMDP dialogue manager in the previous work and changed modes to process the user requests. However, the previously implemented COCOM model was found to be insufficient when emotions were integrated because mode transitions were discontinuous and rapid, which led to transition between ‘extremes of emotions’. The transitions of modes slightly deviated from the Hollnagel’s Contextual Control Model to process user requests and achieve goals. Even though the emotion generation was dynamic it was not be a smooth transition. Therefore, we modified the previous work to make sure that emotion transitions are dynamic, smooth and at the same time making sure that it does not deviate from Hollnagel’s Contextual Control Model or the OCC model for emotions.

3.5 Algorithm for the Proposed Method

The following table lists the algorithm that implements the OCC model of emotions to the COCOM modeled dialogue manager. We can note that the emotions are not specific to the dialogue itself, rather, relying on the four control modes only. In this way the emotional model is domain independent as the dialogue manager can be used in any application and only the agent’s internal goals are required to be known.

In the first method, the various emotions are calculated based on the current and previous control mode of the agent. These modes are calculated by the dialogue manager based on the agent’s understanding of user input. In the second method, the proper response, speech, and animations are generated and queued to be displayed to the user.

Table 3.1: Algorithm for the proposed method

```
STORED: previousMode;

INPUT : currentMode;

METHOD: get_emotion

 test case(previousMode => currentMode):

     scrambled => opportunistic ?  emotion: relief

     scrambled => tactical ?       emotion: relief
```

```
        opportunistic => tactical ?   emotion: hope
        tactical => strategic ?       emotion: pleased
        strategic => tactical ?       emotion: displeased
        tactical => opportunistic ?   emotion: sad
        tactical => scrambled ?       emotion: fear
        opportunistic => scrambled ?  emotion: fear
   STORED: previousMode; currentMode;
   INPUT: emotion; displayText;
   METHOD: display_emotion
        generate_smil_agent_script:
             convert: display_Text => speech;
             modify: emotion => speech;
             modify: emotion => animation;
        animate_agent:
             synchronize animation;
             synchronize speech;
   save: currentMode => previousMode;
```

3.6 Creating a 3D Emotional Talking Head

The creation of a realistic talking head that could be perceived as engaging is one of the greatest challenges in Human Computer Interaction. There are certain drawbacks in creating such a talking head:

- Lack of Open source tools available that enable researchers to conduct extensive research in this field.
- Some of the Open source tools available for researches have very less or no documentation and do not provide proper support to deal with bugs.
- The techniques or methods involved in creating realistic facial animations are often ambiguous and require exhaustive amount of research.

It is obvious from the literature review in Chapter II that MPEG4 Facial Animation is the new standard for creating realistic facial animation. There aren't many Open source tools available that enable researchers to create more engaging 3D talking heads. Xface is an Open source toolkit that focuses on the researchers pursuing their research on creating Embodied Conversational Agents using MPEG4 Facial Animation.

The toolkit comes with three pieces of software such as:

- Xface Editor
- Xface Player
- Xface Client

We have used all the three pieces of software for creating a 3Dimensional Embodied Conversational Agent.

3.6.1 Creating a 3D model

We have used FaceGen software for creating a 3Dimensional Facial Model. FaceGen is very well renowned for creating 3Dimensional models. It provides the user with an exquisite variety of options to create a 3Dimensional face model of their preference. It allows the users to modify the skin tone, colour, shape of the nose, lips, chin and cheeks. We created a 3Dimensional Facial Model using FaceGen as shown below

Figure 24 Proposed new 3D model

We had to create different morph targets for the 3D model. The morph targets can be classified as:

- Emotion refers to the different types of emotions such as Anger, Sadness, and Fear etc.
- Viseme refers to the visual speech generation
- Modifier refers to the animation such blinking of an eye, frowning, etc.

3.6.2 Emotion

We had to model all the emotions that the 3D talking head will exhibit in action. The agent will display dynamic emotions based on the transitions between the four modes. Therefore, we had to generate a set of morph targets for each emotion using the FaceGen software. The emotions we created for the proposed framework are:

- Relief
- Hope
- Pleased
- Displeased
- Disappointment or Sad
- Fear

Modelling the emotions was another challenge since it involves the change in different parts of the face such as eyes, nose, eyebrows, lips, chin and cheeks. Extra care should be taken when modelling each of these emotions because even the smallest mistake might lead to the wrong understanding of the agent's emotion. For example, if the agent's emotion is surprise we have to make sure that the eyebrows are in a raised position, mouth opened up a bit and eyes are wide open as well. We had to consider every single part of a human face and how it is affected when displaying emotions. The snapshot shown below shows the different emotions displayed by our agent. It is evident that all the six emotions are distinct from each other.

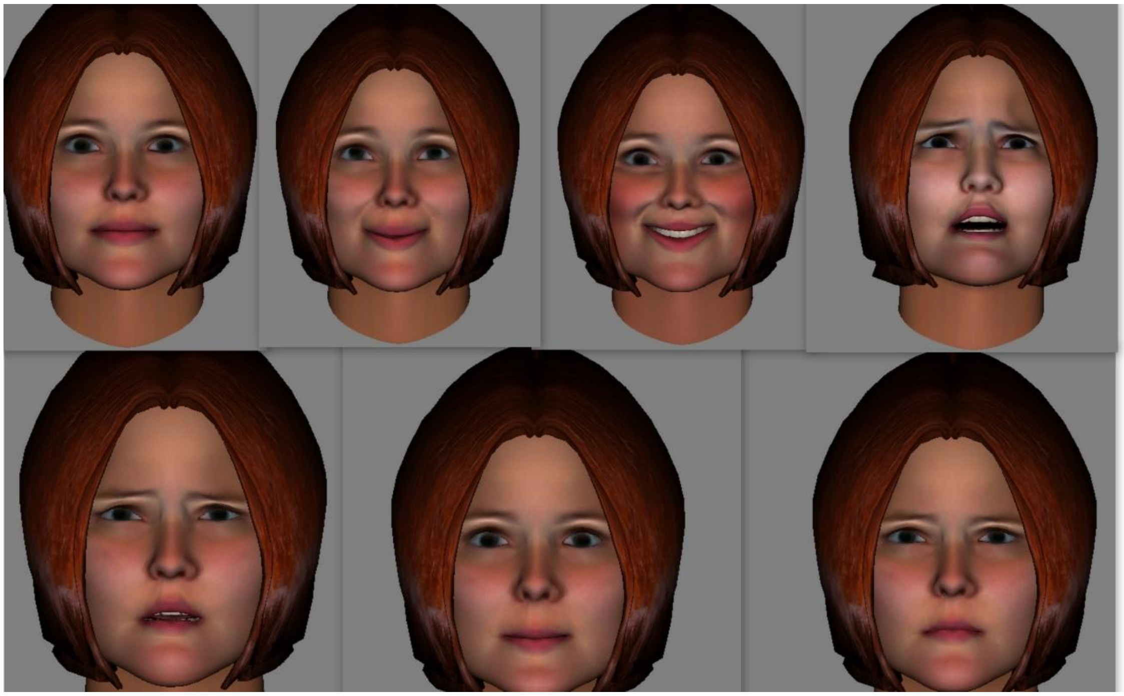


Figure 25 : The emotions from the top left hand corner are: Neutral, Relief, Pleased, and Fear. The emotions from the bottom left hand corner are: Displeased, Hope and Sad.

3.6.3 Visemes

Visemes are the visual counterpart of speech animation generation. We had to model the morph targets for a set of phonemes in such a way that the agent will have proper lip synchronization with the audio. There are 22 phonemes in general that the TTS will use for generating the audio.

Therefore, we had to create the subsequent morph targets for our 3D model for all the phonemes such as:

- aah
- oh

- oh
- ee
- r
- k
- w
- oh
- oh
- oh
- i
- eh
- r
- n
- d
- ch
- th
- f
- d
- q
- b

The figure shown below depicts the morph targets created for some of the phonemes.

Figure 26 : Visemes for the created 3D model

3.6.4 Modifier

Modifiers are really important to make the conversation with a virtual agent more realistic. This includes the blinking of eyes continuously at regular time intervals. The modifiers are the category of morph targets that mimic the humans. Therefore, we modelled the appropriate modifiers that will drive the facial animation in a more realistic fashion. Some of the modifiers we created for our 3D model are shown below.

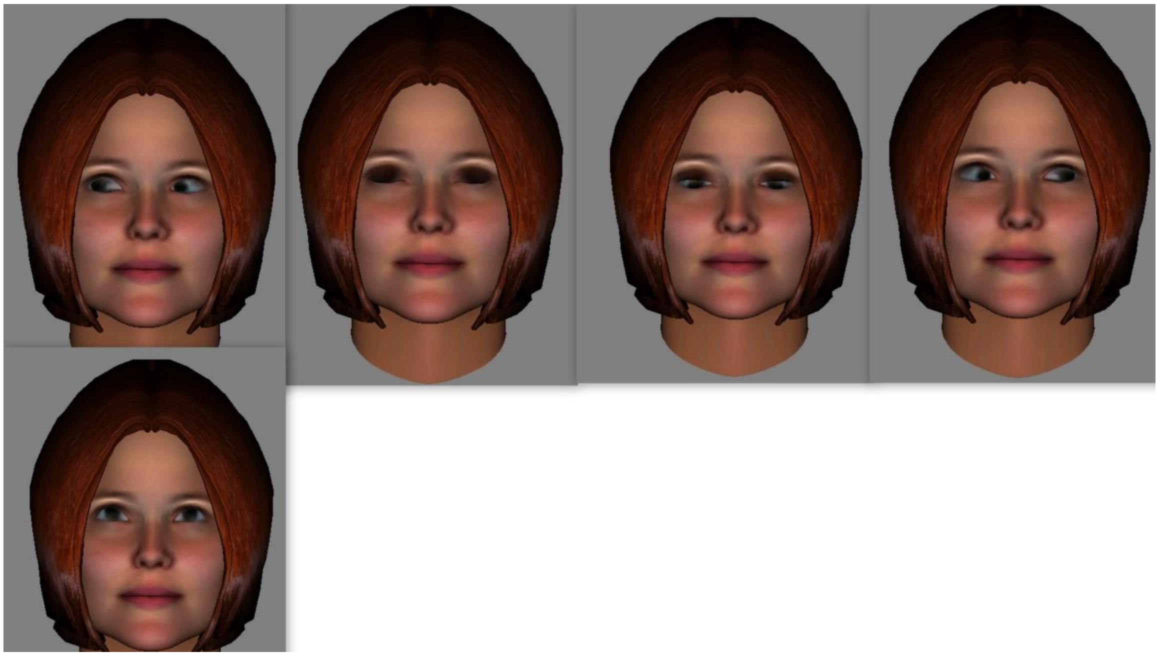


Figure 27: Modifiers for the created 3D model

3.6.5 Applying MPEG4 Facial Animation

The 3D model created in the previous step was imported to XfaceEd along with all the different morph targets. We created the animation of the 3D model according to the MPEG4 standards using the XfaceEd. This step was crucial because we define the weights, deformation functions and the FAP's. A realistic facial animation is created by modifying and fine tuning these parameters. We had already discussed this in detail in Chapter II under MPEG4 Facial Animation.

According to [Koray04] the work flow follows these steps:

- Import on or more static meshes
- Define FAPU
- Define FP regions, weights and deformation functions
- Preview results and fine tune parameters defined in the previous step.

The information about the 3D model is stored under *.fdp extension. The first step was to define to the FAPU. We selected seven points on the face model. It was very important to set the seven points on the face model because the animations are done using the distances defined by these points. The figure shown below explains the first step involved in the creation of a 3D model according to the MPEG4 Facial Animation Standards.

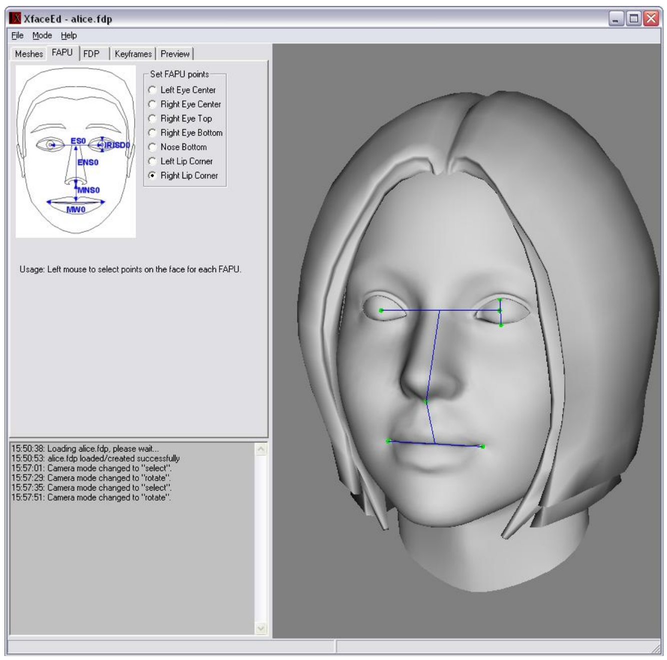


Figure 28: Setting the FAPU for the created 3D model

The second step was to define the Feature Points (FP). There are 84 Feature Points on the head. We don't have to define all the 84 Feature Points since not all FP's are affected by the Facial Animation Parameters (FAP). We have to manually define the weight, deformation function for these FP's. The definitions of these parameters for th cheek area are shown in the figure below. We have used the Raised Cosine deformation function by default that was built with XfaceEd. According to [Koray04] Raised Cosine deformation function has achieved satisfactory results. The Raised Cosine deformation function is defined as follows:

$$\Delta v_p = \left(1 + \cos\left(\pi \times \frac{d_p}{d_{max}}\right)\right) \times w \times fap$$

where Δv_p denotes the Euclidean distance that the point ‘p’ in FP zone should be translated in FAP direction, while d_p is the ‘p’s and d_{max} is the farthest point’s distance in region to the FP.‘w’ is a weight value defined for every deformation function and ‘fap’ is the FAP value.

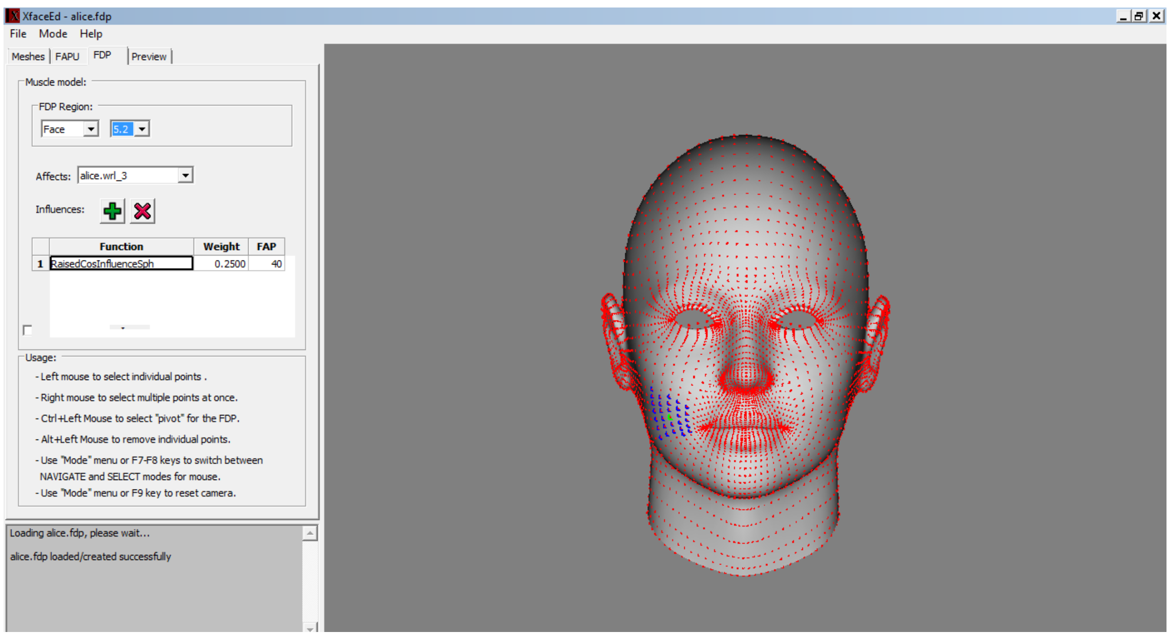


Figure 29: Setting the weights and deformation functions for the created 3D model

Finally, we had to preview the animation and modify the weights of each FAP accordingly. It was a continuous and iterative process in order to achieve realistic facial animation. The preview pane from XfaceEd where we can see the 3D model in action is shown in the figure below.

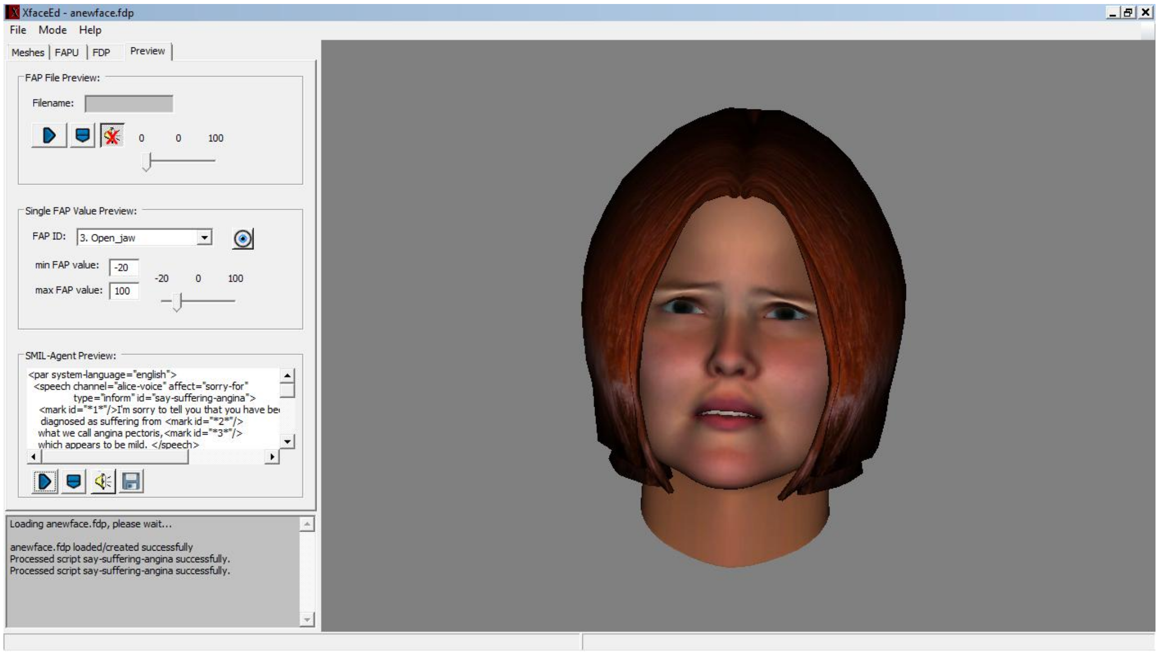


Figure 30 Previewing and Fine tuning the Animation

3.6.6 Blending and Synchronization

The morph targets created in the previous steps are then synchronized and blended together using the interpolation technique. The morph targets from the different categories are chosen accordingly depending upon the type of animation and blended together to produce a vivid and realistic 3D animation.

3.6.7 Interpolation

The technique used for selecting the appropriate morph targets is based on the type of emotion, length of conversation and other parameters which will be supplied by SMIL Script file. We decided to use the interpolation technique for animation out of the different animation techniques available because it doesn’t require any extensive computation and is easier to generate the animation. The linear or bi-linear interpolations are not the best interpolation techniques so we decided to create the animation based on linear and cubic interpolation techniques [Koray04].

$$k(t_u) = k(t_0) \times (1 - u) + k(t_1) \times u$$

$$k(t_u) = k(t_0) \times (2u^3 - 3u^2 + 1) + k(t_1) \times (3u^2 - 2u^3)$$ where u is a value between 0 and 1, $k(t_0)$ and $k(t_1)$ are the vertices of previous and next key frames respectively and $k(t_u)$ designates the morph target interpolated using these two. Interpolation is

3.7 SMIL Scripting Language

Synchronized Multimedia Integration Language (SMIL) is still widely used for the creation of multimedia presentations. We already discussed about the adaptability and flexibility of SMIL in Chapter II under SMIL Agent Scripting Language section. The creation of a realistic 3D talking head relies on the 3D model itself and also the language that drives the animation. SMIL acts as a powerful source for driving the animation by allowing the users to control independent communication channels. We will explain a sample SMIL script from the proposed system as shown below.

The SMIL Agent file has different fields as highlighted below and these different fields will realize different values based on the factors such as the type of emotion to be exhibited by the agent, context of the conversation, the actual text that needs to be animated during run time.

```
<par system-language="english">
<speech channel="new-voice" affect= "Disappointed"
type= "inform" id= "say-wings-drinks">
```

```
Sorry which of the following do you want? We have: Wings, Drink, Pizza, and
Vegetarian Pizza
   </speech>
  <speech-animation channel="new-face" affect= "Disappointed" content-id="say-
wings-drinks"/>
  <action channel="new-eyes" action-type="turning" intensity="0.7"
                content-id="say-wings-drinks">
                 <parameter>LookLeft</parameter>
  </action>
</par>
```

“**Affect**”: The affect attribute can have various values such as “Fear”, “Relief”, “Hope”, “Disappointed”, and “Displeased”, “Pleased”. One of each values will be filled up in “affect=” ” tag depending upon the Control mode the agent is grounded.

“Type”: Tells the type of speech act that is represented by the message to be verbalized. The type attribute will be will filled in with the values as “Sorry-for”, “Greet”, “Question”, “Inform”, “Paraphrase” and “Suggest”.. For ex: If the text to be animated is “I would like to suggest you something that’s better than the current configuration of your computer” then the “type” attribute will take a value “Suggest”.

“Content-id”: Its value is the id attribute value of a speech element. It is used to refer to the <speech>...</speech> element that specifies the content to be animated. The text is converted to speech using a Text to Speech engine during the run time.

"Action Channel": The action channel specifies the channel that should realize the animation. For example, if the agent needs to display a fear emotion then we can have the channel "alice-eyes" to realize that emotion by modifying the values of the attributes such as "action-type" which can be set to "turning" and the "intensity" could be set to "high". The intensity attribute always accepts only numerical values.

We have managed to generate the above the script using our SMIL generator. The SMIL script plays a major role in driving a realistic facial animation. The facial animation does not only rely on the 3D authoring tools but also the scripting language that drives the animation. The next section will elaborate about the playback of the animation using Xface Player.

3.8 Animation Playback.

Xface comes with an MPEG4 Compliant player known as Xface player. The Xface player will be able to process the SMIL script and playback the animation with effortless ease as long as the player has been loaded with the *.fdp file that we created in the previous section which is the configuration file containing all the Fap's and other parameters. The challenge was to establish communication between Xface player and our proposed framework because our framework heavily relies on the player for animation. Xface implements client-server architecture and that is one of the major advantages of Xface. The communication can take place via TCP/IP. We have made that work in our favour and had the client which is our proposed framework send SMIL scripts to the Xface player which acts as the server. It is amazing that Xface allows you to queue all the tasks and the player can start playing the animation one by one. The acknowledgement of sending and receiving the tasks between Client-Server is done by sending out notification messages for each task.

3.9 Conclusion

We have reviewed the techniques involved in the generation of emotions based on the Four Contextual Control Modes and OCC model for emotions. The proposed framework is a new approach for modelling and generating emotions similar to humans. We have also reviewed the design methodology and tools involved in the creation of a realistic 3D talking head. The next chapter will present a complete experimental analysis of the proposed framework to prove the efficiency and effectiveness of the new approach.

CHAPTER IV

ANALYSIS OF RESULTS

This Chapter will review the effectiveness and efficiency of the proposed system. The experimental analysis has been subdivided in two sections. The beginning of this chapter will discuss about the qualitative analysis in detail carried out between the proposed system and some of the most popular Embodied Conversational Agents. As we move towards the end of this chapter we have discussed about the quantitative analysis carried out between the proposed system and the previously developed Modified POMDP dialogue manager. The chapter will also demonstrate the dynamicity of agent's emotions as the mode switches between Scrambled Mode, Tactical Mode, Opportunistic Mode and Strategic Mode adhering to the new approach for modelling emotions by combining Four Contextual Control Modes and OCC model for emotions.

4.1 Theoretical Analysis

We have already reviewed some of the most renowned Embodied Conversational Agents in Chapter II. We will compare those agents with the proposed system against their performance parameters. The comparison will highlight the limitations of the previous work which turned out to be the advantages of the new approach. The table shown below does the comparison between different embodied conversational agents such as GRETA, LUCIA, Emotional Chinese Talking Head from the literature, some existing commercial ECA applications such as Microsoft Agents, Aaron from Future Shop, Anna from IKEA and few significant chatter bots.

	Voice	Animation	Technique	Emotional Model	Emotions	Dialogue Manager
Proposed System	Yes	Yes	MPEG4	Yes	Yes	Modified POMDP
GRETA	Yes	Yes	MPEG4	No	Yes	No
LUCIA	Yes	Yes	MPEG4	No	No	No
Chinese Talking head	Yes	Yes	MPEG4	No	Yes	No
Microsoft Agents	Yes	Yes, but cartoon characters		No	No	No
Aaron from Future Shop	Yes, But not all the time	Plays Pre-recorded video from a database	Flash Script	No	No	No
Anna from IKEA	No	Yes, but 2D animation with no lip movements	Flash Script	No	No	No
Cleverbots	No	No	Flash Script	No	No	No
Verbots	Yes	Yes	Flash Script	No	No	No

Table 4.1 Comparison between existing ECA's with the new system

The table 4.1 compares the various ECA's against our proposed system. The different performance parameters taken into account were:

- Voice: Some of the ECA's do come with voice capabilities with the help of Text to Speech engines and some of them don't have the voice capability which makes their ECA non-realistic.
- Animation: Most of the ECA's have an interface that displays animations for users which creates a positive impact and some of them do not

include a realistic interface such as the 3D facial model implemented in our new system.

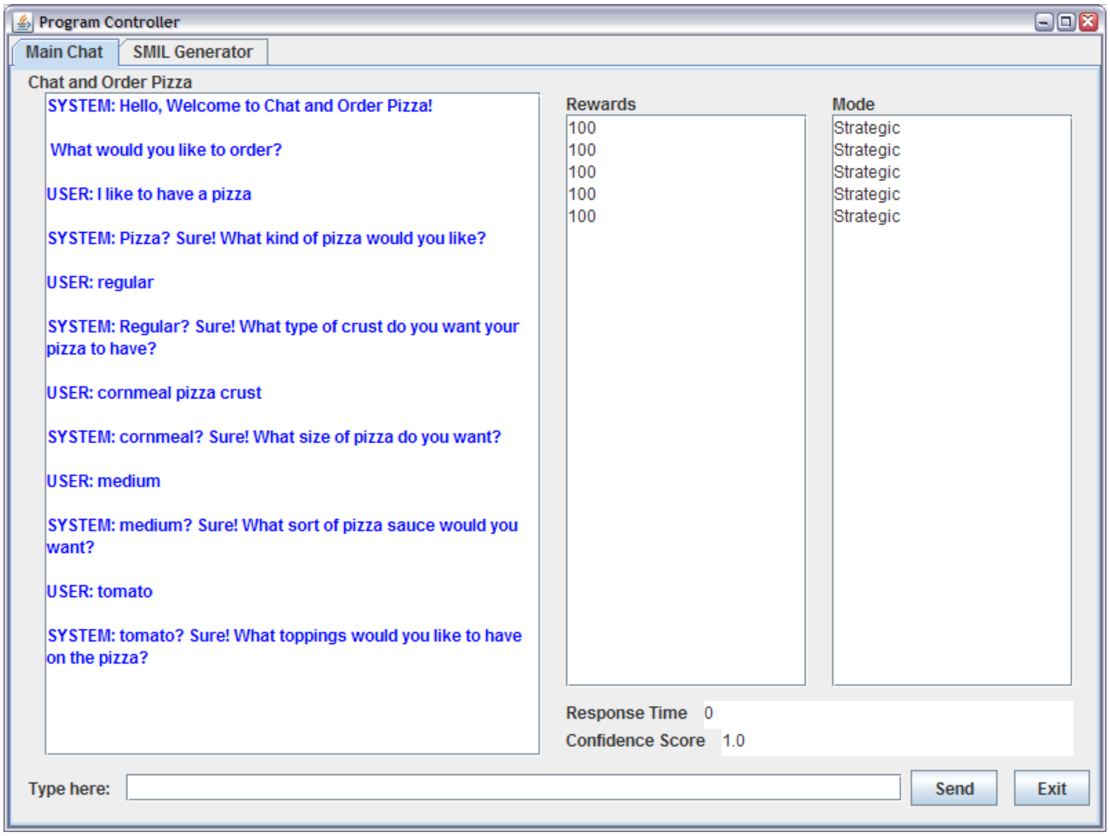
- Technique: The techniques are quite unique for every single agent. The most basic ECA's are still using Flash Script for animation while advanced ECA's use MPEG4 Facial Animation which is driven by powerful languages such as APML, SMIL,CML and VHML.
- Emotional model: It plays an important role in generating emotions on the fly rather than having static emotions coupled to specific context. The emotional models assist in generating dynamic emotions.
- Emotions: It is obvious from Chapter II that only very few ECA's have emotions. Even some of the advanced ECA's lack emotions.
- Dialogue Manager: Dialogue managers are very important when we consider using ECA's in social situations because they are capable of problem solving by understanding the user's intentions. No one has integrated the Modified POMDP dialogue manager to an ECA so far and that is another major contribution in this thesis.

4.2 Quantitative Analysis

In this section we will review the quantitative analysis performed on the new proposed system. We have simulated five test cases out of which the first three test cases shall demonstrate the efficiency of the new proposed system by comparing their confidence scores with the previous work done on Modified POMDP Dialogue Manager [Sabiha10] under different scenarios. We have added a fourth test case that shall demonstrate the efficiency of the new system by comparing the confidence scores generated even when there are too many user's text, speech errors. All the five cases shall demonstrate the dynamicity in the change of emotions on the fly according to the new proposed approach. The five test cases are:

- Normal Dialogue between the user and system
- Dialogue conversation between the user and system with speech and text errors.
- Dialogue conversation between the user and system handling conflicts
- Dialogue Conversation between the user and system where system generates higher confidence scores by recognizing user's text, speech errors
- Dialogue conversation between the user and system to simulate the dynamic change of emotions along with the change in modes.

4.2.1 Test Case: 1



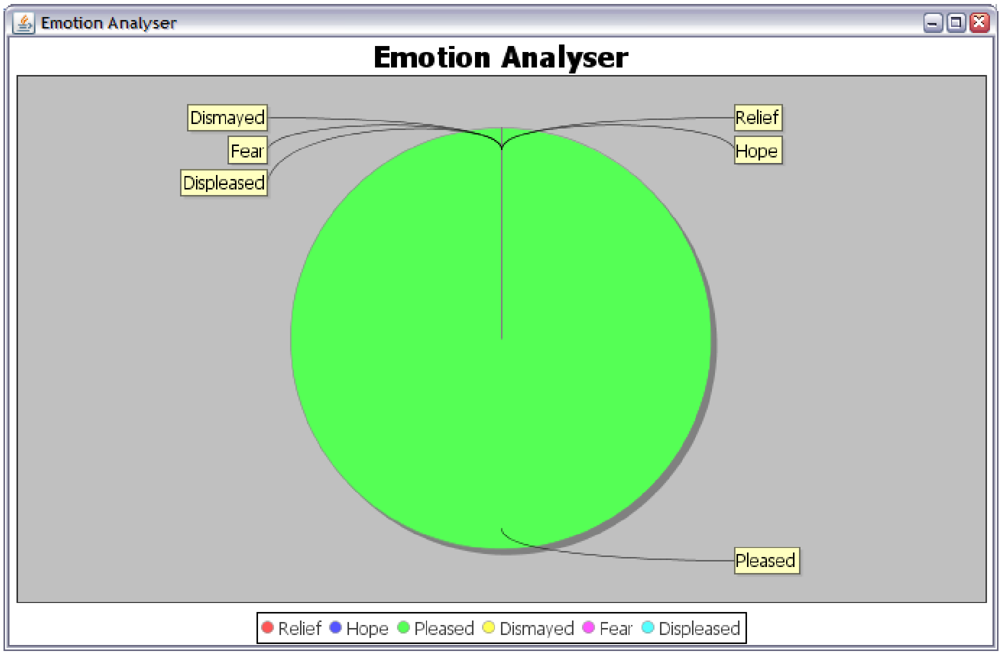


Figure 31: Test Case 1

Confidence Scores of the Old Modified POMDP Dialogue Manager	Confidence Scores of the New Improved POMDP Dialogue Manager
0	100
71.42857	100
86.95652	100
53.846157	100
100	100
N/A	100
N/A	100
N/A	100
N/A	100

Table 4.2: Comparison of Confidence scores for Test Case1

SYSTEM: Hello, Welcome to Chat and Order Pizza!

What would you like to order?

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: I like to have a pizza

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: Pizza? Sure! What kind of pizza would you like?

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: regular

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

++

Emotional Results:

RELIEF: 0/9

HOPE: 0/9

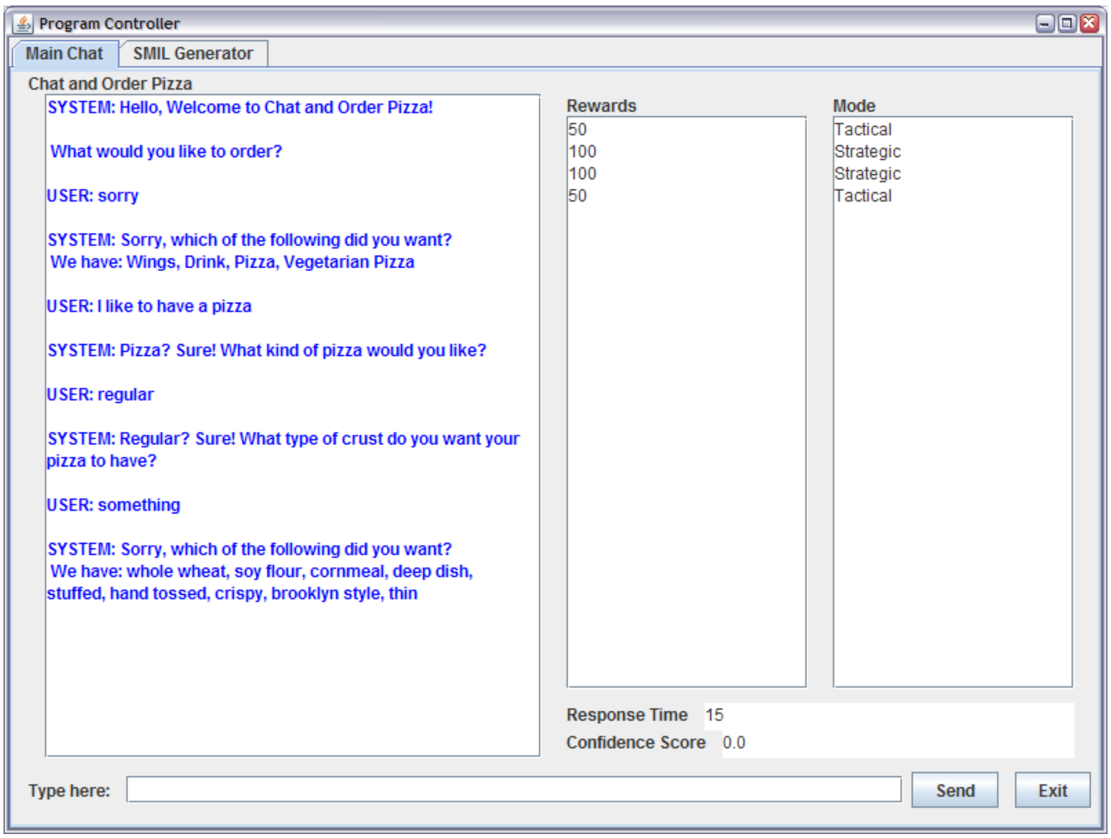
PLEASED: 9/9

DISMAYED: 0/9

FEAR: 0/9

DISPLEASED: 0/9

4.2.2 Test Case 2



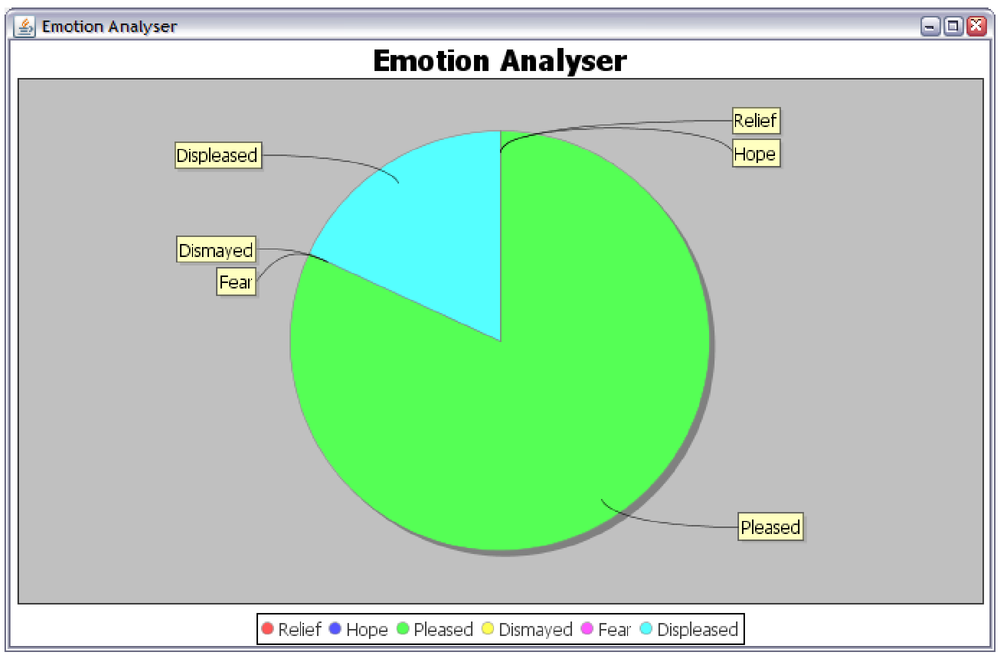


Figure 4.2: Test Case 2

Confidence Scores of the Old Modified POMDP Dialogue Manager	Confidence Scores of the New Improved POMDP Dialogue Manager
0	0
63.636364	100
85.71429	100
53.846157	0
25.0	100
100.0	100
N/A	100
N/A	100
N/A	100
N/A	100
N/A	100

Table 4.3 : Comparison of Confidence scores for Test Case 2

SYSTEM: Hello, Welcome to Chat and Order Pizza!

What would you like to order?

++++++++++++++++++++++++++++++

Emotion: Displeased

USER: sorry

Rewards: 50

Mode: TACTICAL

Confidence: 0.0

SYSTEM: Sorry, which of the following did you want?

We have: Wings, Drink, Pizza, Vegetarian Pizza

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: I like to have a pizza

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

+++

Emotional Results:

RELIEF: 0/11

HOPE: 0/11

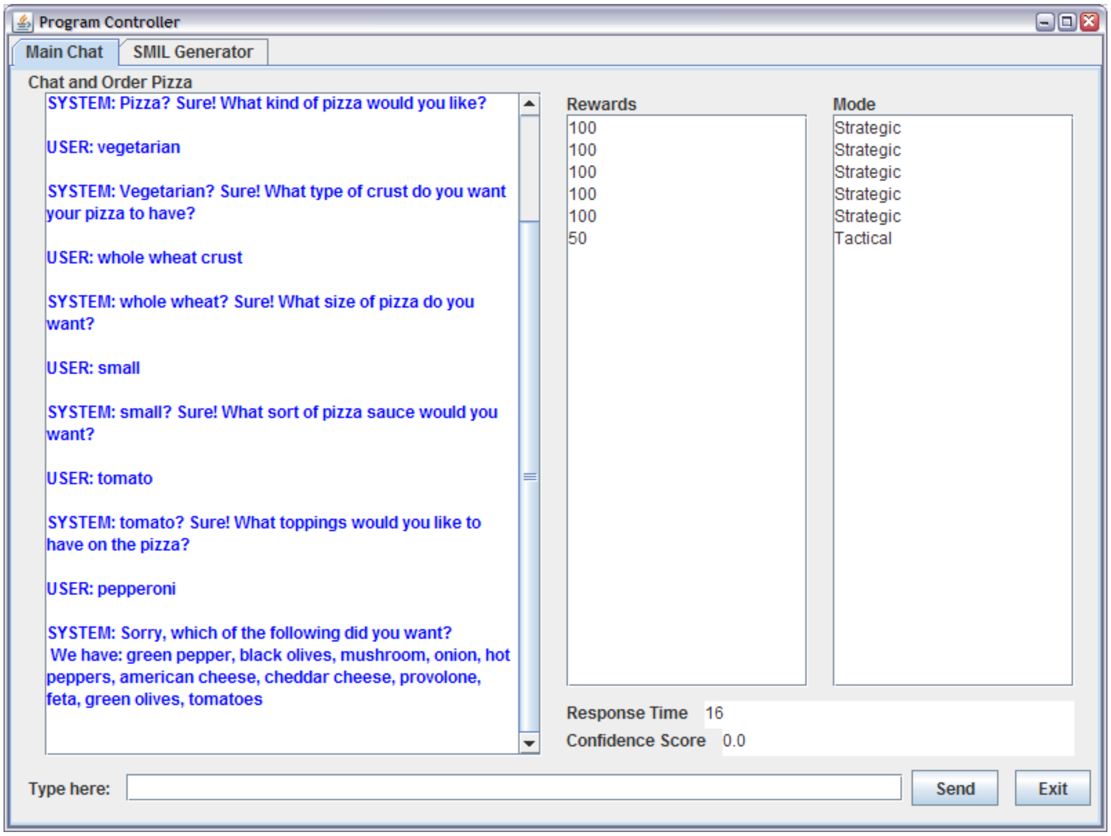
PLEASED: 9/11

DISMAYED: 0/11

FEAR: 0/11

DISPLEASED: 2/11

4.2.3 Test Case 3



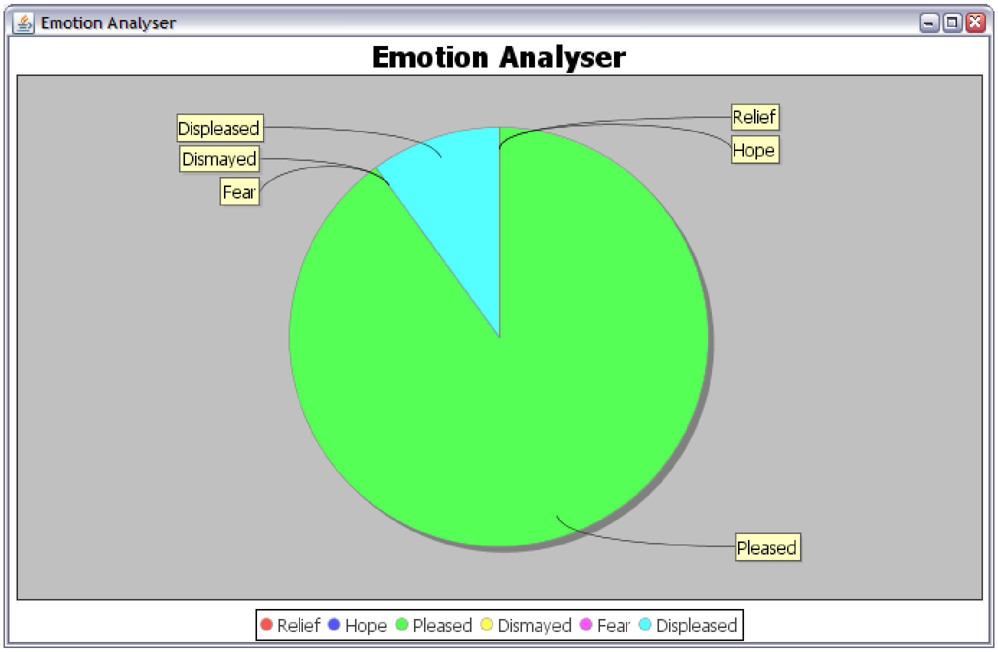


Figure 4.3: Test Case 3

Confidence Scores of the Old Modified POMDP Dialogue Manager	Confidence Scores of the New Improved POMDP Dialogue Manager
0	100
66.66667	100
65.27139	100
100	100
100	100
100	0
N/A	91.66
N/A	100
N/A	100
N/A	100

Table 4.4: Comparison of Confidence scores for Test Case 3

SYSTEM: Hello, Welcome to Chat and Order Pizza!

What would you like to order?

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: I like to order a pizza

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: Pizza? Sure! What kind of pizza would you like?

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: vegetarian

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

++

Emotional Results:

RELIEF: 0/10

HOPE: 0/10

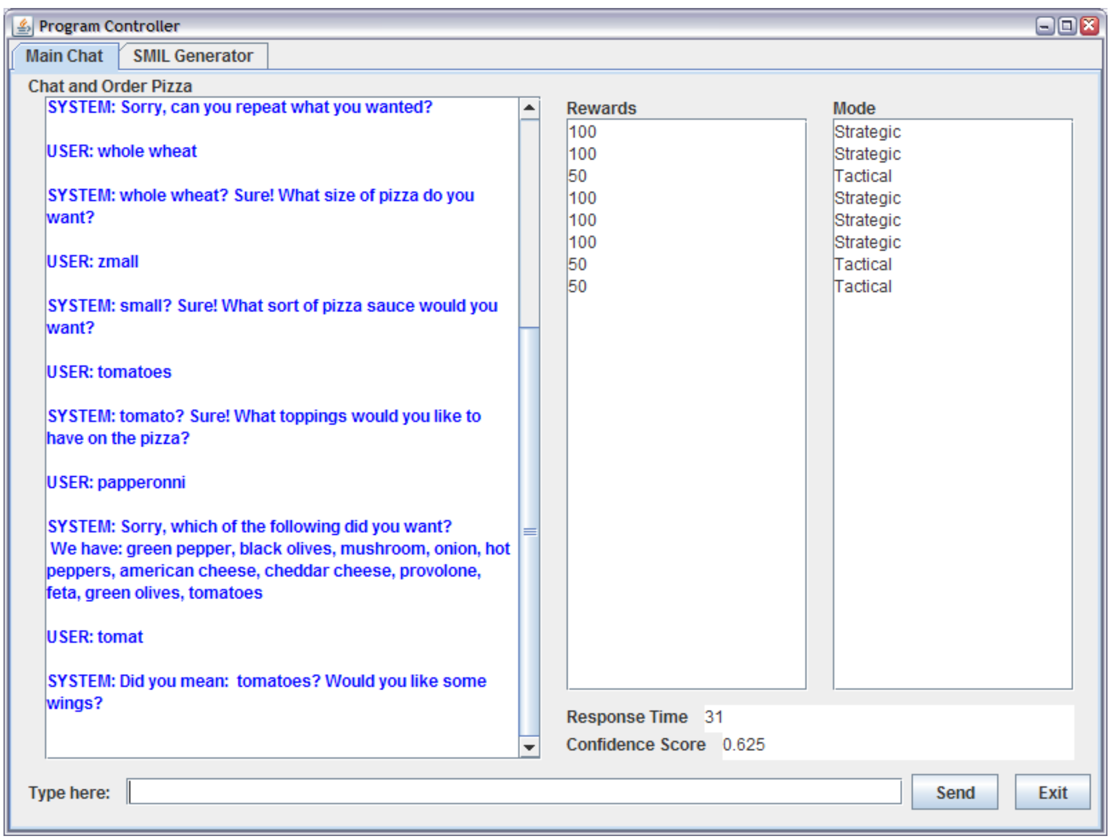
PLEASED: 9/10

DISMAYED: 0/10

FEAR: 0/10

DISPLEASED: 1/10

4.2.4 Test Case 4



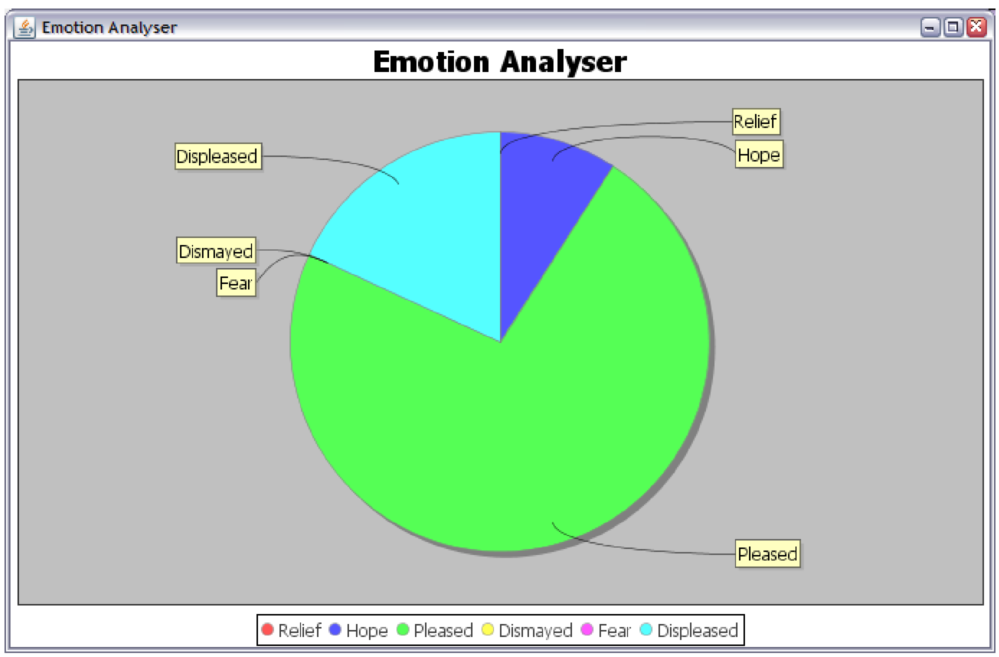


Figure 4.4: Test Case 4

Confidence Scores of the New Improved POMDP Dialogue Manager
94.44444
96.9697
45.454547
90.909094
86.66667
91.66667
0
62.5
100
100

Table 4.5: Comparison of Confidence scores for Test Case 4

SYSTEM: Hello, Welcome to Chat and Order Pizza!

What would you like to order?

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: pizzza

Rewards: 100

Mode: STRATEGIC

Confidence: 0.9444444

SYSTEM: Pizza? Sure! What kind of pizza would you like?

+++++++++++++++++++++++++++++

Emotion: Pleased

USER: vegetariean

Rewards: 100

Mode: STRATEGIC

Confidence: 0.969697

++

Emotional Results:

RELIEF: 0/11

HOPE: 1/11

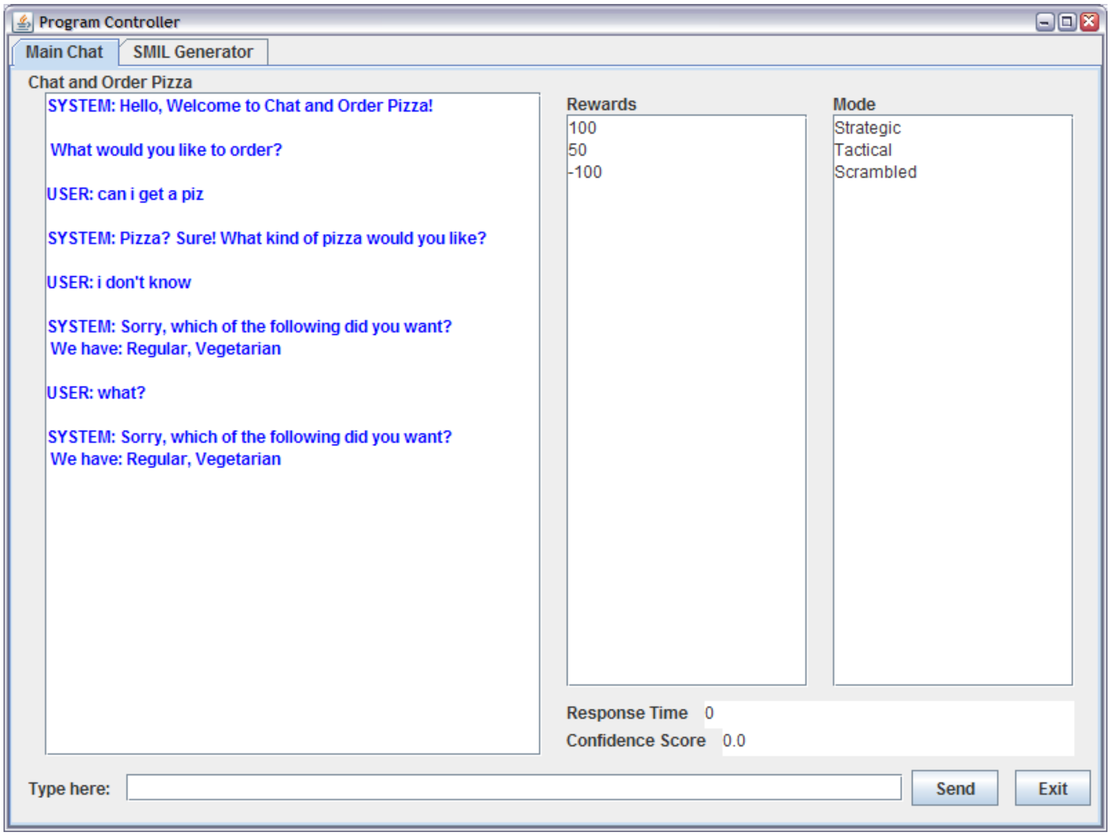
PLEASED: 8/11

DISMAYED: 0/11

FEAR: 0/11

DISPLEASED: 2/11

4.2.5 Test Case 5



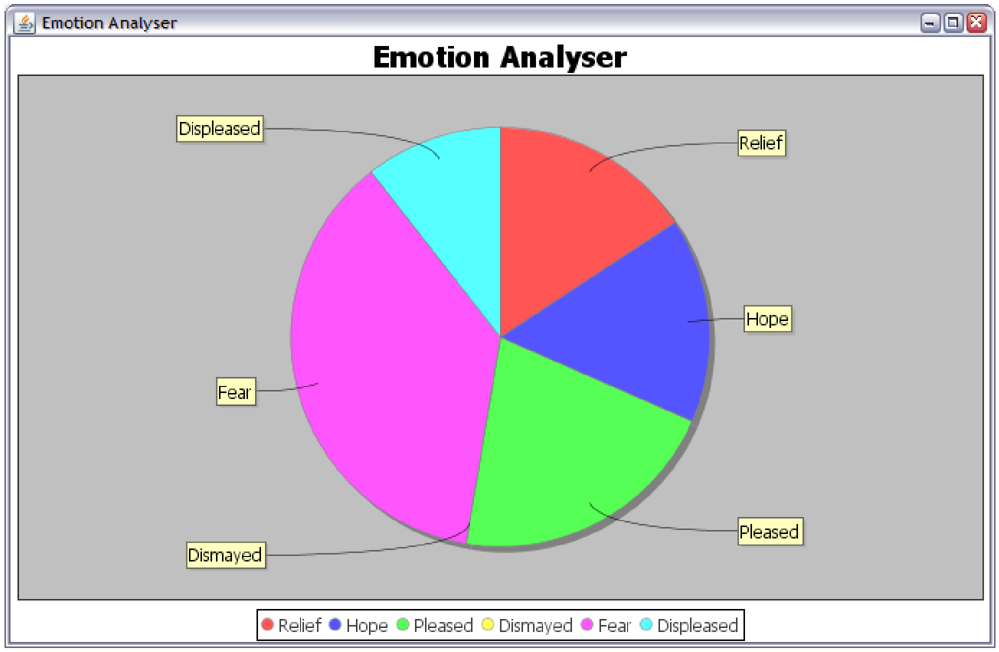


Figure 4.5: Test case 5

Confidence Scores of the New Improved Modified POMDP Dialogue Manager
80
0
0
42.857143
100
0
28.57143
35.714287
50
66.66667

66.66667
100
100
100
33.3333334
66.66667
60

Table 4.6: Confidence scores for Test Case 5

SYSTEM: Hello, Welcome to Chat and Order Pizza!

What would you like to order?

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: can i get a piz

Rewards: 100

Mode: STRATEGIC

Confidence: 0.8

SYSTEM: Pizza? Sure! What kind of pizza would you like?

+++++++++++++++++++++++++++++

Emotion: Displeased

USER: i don't know

Rewards: 50

Mode: TACTICAL

Confidence: 0.0

Emotional Results:

RELIEF: 3/19

HOPE: 3/19

PLEASED: 4/19

DISMAYED: 0/19

FEAR: 7/19

DISPLEASED: 2/19

4.3 Discussion

We simulated five different test cases to analyse the performance of our new approach. It is evident from all the five test cases that the emotions are dynamic and they are conveyed by means of different communication channels such as eye, lip, face and head movements. The animations of the muscle parameters are also realistic during run time and we have managed to control all muscle parameters on the agent's face. Test case 1, 2 and 3 demonstrate the efficiency of the new improved Modified POMDP in comparison to the previous work. The confidence scores generated by the new improved Modified POMDP dialogue manager are compared with the previous work. It is obvious from tables 4.2, 4.3, 4.4, 4.5 that confidence scores generated by the new improved Modified POMDP dialogue manager are higher than previous work. If the system achieves higher confidence scores it means that it can understand the user utterances better in a regular or noisy environment. The test case 4 proved that the agent can handle noisy environments properly in comparison to the previous work. In test case 4, the agent was able understand most of the typos and generate high confidence scores even in a noisy situation which is a significant contribution to this thesis. The previous dialogue manager was not able to handle typos and other errors. It was quite repetitive and could be annoying sometimes from the user's perspective. The emotion analyser is a tool that we created which will output a pie chart of the agent's emotions at the end of every conversation. We created the emotion analyser to find out how well the agent was able to cope up with the current conversation. The goal of the emotional model is to maintain the emotional stability of the virtual agent. It maintains proper stability by communicating with the Modified POMDP dialogue manager. The emotional model thrives to be happy and takes appropriate actions that will keep it happy with the help of the Modified POMDP Dialogue manager. The agent will be happy as long as it understands all the user queries and is able to process them properly. For example, when user types “I want a piza” then the agent will not be able to understand and process the user request. The agent displays negative emotions such as Displeased, Sad, and Fear/Confused and when the agent starts showing negative emotions the emotional model instructs the Modified POMDP Dialogue manager to clarify the errors with users. The Modified POMDP starts clarifying the errors with the user such as “Did you want a pizza?” The user might or might not give a proper answer. But the agent keeps switching modes based on the user’s answers and as the agent switches modes the emotions also change rapidly. It is evident from the pie charts generated by the emotional analyser for the four test cases the agent was pleased most of the time throughout the conversation. The test case 5 was simulated to prove the dynamicity of the agent’s emotions. All six emotions were generated in one conversation to prove that the agent can switch modes rapidly and also display different emotions when the mode changes. We have also attached a part of the conversation log along with its emotions and confidence scores for all the test cases. The complete outputs for all the five test cases have been added to appendix.

4.4 Conclusions

Chapter IV reviewed the efficiency and effectiveness of the new proposed system. The qualitative analysis proved that the proposed system is robust and powerful with more capabilities compared to some of the most renowned current Embodied Conversational Agents. The quantitative analysis produced positive results and was proved by the simulation of five different test cases. We inferred that the confidence scores generated by the new improved Modified POMDP Dialogue manager were much better than the previous work. We have also proved that the emotions generated were dynamic and they tend to change with the transitioning between the four modes. The pie charts generated by the emotion analyser proved that the agent tries to cope up with negative emotions by taking appropriate actions in conjunction with the Modified POMDP Dialogue manager to maintain its emotional stability. This thesis has laid the baseline for the creation of realistic 3D embodied conversational agents accompanied by a powerful dialogue manager such as the Modified POMDP dialogue manager. Some of the future work and recommendations directed towards this area of Human Computer Interaction can be found in the next upcoming Chapter V.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The Embodied Conversational Agents are becoming the current buzzword for development and advancement in e-commerce. The need for ECAs in commercial applications has been increasing since the World Wide Web came into existence. The ECAs wear a human face and provide assistance to people in various fields. The state of art of these agents has changed from the world of 2D to 3D. The Facial Animation associated with these Virtual Agents has opened the doors of research in e-commerce. There are over thousands of websites that use ECAs to assist the online users. But, none of them have an ECA that can be perceived as engaging from the user's perspective. An ECA can be perceived as engaging when it displays emotions and uses different communication channels to convey the same. Therefore, we have created a system that consists of a realistic 3D talking head that can engage with users when used in any social situations or e-commerce applications. The ECAs in the past were coupled to their performance parameters and domains. The new system could be used across any domains because SMIL does not carry information that is related to the domain. The entire system can be re-used across various domains such as healthcare, flight reservation systems, e-retail environments. The experimental results have proved that this new system has the ability to change emotions dynamically on the fly based on the transitioning between the Four Modes and can also process user requests effectively by understanding their intentions even in a noisy environment. The system also defines the rules for emotion generation by combining the Four Contextual Control Modes and OCC Model of Emotions.

5.2 Future Work

The proposed approach will definitely be benchmark for measuring ECAs in the future. But, having said that we can definitely enhance the system to make it even more significant. The agent can be further extended by adding the ability to accept multimodal inputs from users through devices such as camera, microphone etc. There are existing systems that use multimodal input with ECA's but they don't have an ECA's that can display emotions. There are some ECA's being used for mobile applications. But, there aren't any powerful ECA's like ours which can display emotions and handle user request intelligently using a dialogue manager. Therefore, extending the current system to mobile applications will be significant. The thesis could be further extended by calculating the moods of the current agent by analysing the emotional of the agent at the end of each conversation. The emotional model can be further extended by adding personality describing how personality might make an impact on agent's emotions. Significant research has been done in past regarding generating emotions with different emotional intensities. Even though they provided exhaustive techniques for calculating emotional intensities none of them had an interface that can display emotions with different emotional intensities.

APPENDICES

APPENDIX A

Test Case 1

Default Strings loaded
System Property messages loaded
System properties loaded
Loading default properties from trained tagger C:\Documents and Settings\YellowDude\workspace\DialogManager\models\bidirectional-distsim-wsj-0-18.tagger
Reading POS tagger model from C:\Documents and Settings\YellowDude\workspace\DialogManager\models\bidirectional-distsim-wsj-0-18.tagger ... done [3.0 sec].
++++++++++++++++++++++++++++++
Emotion: Pleased

SYSTEM: Hello, Welcome to Chat and Order Pizza!

What would you like to order?
++++++++++++++++++++++++++++++
Emotion: Pleased

USER: I like to have a pizza

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: Pizza? Sure! What kind of pizza would you like?

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: regular

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: Regular? Sure! What type of crust do you want your pizza to have?

++++++++++++++++++++++++++++++
Emotion: Pleased
USER: cornmeal pizza crust

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: cornmeal? Sure! What size of pizza do you want?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: medium

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: medium? Sure! What sort of pizza sauce would you want?

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: tomato

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: tomato? Sure! What toppings would you like to have on the pizza?

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: Chicken

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: chicken? Sure! Would you like some wings?
+++++++++++++++++++++++++++++++
Emotion: Pleased
USER: no

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: No? Sure! Would you like a drink?
+++++++++++++++++++++++++++++++
Emotion: Pleased
USER: no

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: No? Sure! So your final order is:
Pizzas: (1)
0: Pizza --> sauce=tomato, crust=cornmeal, size=medium, toppings=[chicken]
Wings: (none)
Drinks: (none)
Thank you for ordering with us, your order will be delivered in 1 Hour.

Hello, Welcome to Chat and Order Pizza!

What would you like to order?

++

Emotional Results:

RELIEF: 0/9
HOPE: 0/9

PLEASED: 9/9
DISMAYED: 0/9
FEAR: 0/9
DISPLEASED: 0/9

APPENDIX B

Test Case 2

Default Strings loaded
System Property messages loaded
System properties loaded
Loading default properties from trained tagger C:\Documents and Settings\YellowDude\workspace\DialogManager\models\bidirectional-distsim-wsj-0-18.tagger
Reading POS tagger model from C:\Documents and Settings\YellowDude\workspace\DialogManager\models\bidirectional-distsim-wsj-0-18.tagger ... done [3.0 sec].
++++++++++++++++++++++++++++++
Emotion: Pleased

SYSTEM: Hello, Welcome to Chat and Order Pizza!

What would you like to order?
++++++++++++++++++++++++++++++
Emotion: Displeased
USER: sorry

Rewards: 50

Mode: TACTICAL

Confidence: 0.0

SYSTEM: Sorry, which of the following did you want?
We have: Wings, Drink, Pizza, Vegetarian Pizza
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: I like to have a pizza

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: Pizza? Sure! What kind of pizza would you like?

++++++++++++++++++++++++++++++
Emotion: Pleased
USER: regular

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: Regular? Sure! What type of crust do you want your pizza to have?
++++++++++++++++++++++++++++++
Emotion: Displeased
USER: something

Rewards: 50

Mode: TACTICAL

Confidence: 0.0

SYSTEM: Sorry, which of the following did you want?
We have: whole wheat, soy flour, cornmeal, deep dish, stuffed, hand tossed, crispy, brooklyn style, thin
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: whole wheat pizza crust

Rewards: 100

Mode: STRATEGIC

Confidence: 0.90909094

SYSTEM: whole wheat? Sure! What size of pizza do you want?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: medium

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: medium? Sure! What sort of pizza sauce would you want?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: tomato

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: tomato? Sure! What toppings would you like to have on the pizza?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: pepperoni

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: pepperoni? Sure! Would you like some wings?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: no

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: No? Sure! Would you like a drink?
++++++++++++++++++++++++++++++
Emotion: Pleased

USER: no

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: No? Sure! So your final order is:
Pizzas: (1)
0: Pizza --> sauce=tomato, crust=whole wheat, size=medium, toppings=[pepperoni]
Wings: (none)
Drinks: (none)
Thank you for ordering with us, your order will be delivered in 1 Hour.

Hello, Welcome to Chat and Order Pizza!

What would you like to order?

++
Emotional Results:

RELIEF: 0/11
HOPE: 0/11
PLEASED: 9/11
DISMAYED: 0/11
FEAR: 0/11
DISPLEASED: 2/11

Test Case 3

Default Strings loaded
System Property messages loaded
System properties loaded
Loading default properties from trained tagger C:\Documents and Settings\YellowDude\workspace\DialogManager\models\bidirectional-distsim-wsj-0-18.tagger
Reading POS tagger model from C:\Documents and Settings\YellowDude\workspace\DialogManager\models\bidirectional-distsim-wsj-0-18.tagger ... done [3.1 sec].
++++++++++++++++++++++++++++++
Emotion: Pleased

SYSTEM: Hello, Welcome to Chat and Order Pizza!

What would you like to order?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: I like to order a pizza

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: Pizza? Sure! What kind of pizza would you like?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: vegetarian

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: Vegetarian? Sure! What type of crust do you want your pizza to have?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: whole wheat crust

Rewards: 100

Mode: STRATEGIC

Confidence: 0.90909094

SYSTEM: whole wheat? Sure! What size of pizza do you want?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: small

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: small? Sure! What sort of pizza sauce would you want?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: tomato

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: tomato? Sure! What toppings would you like to have on the pizza?
++++++++++++++++++++++++++++++

Emotion: Displeased

USER: pepperoni

Rewards: 50

Mode: TACTICAL

Confidence: 0.0

SYSTEM: Sorry, which of the following did you want?

We have: green pepper, black olives, mushroom, onion, hot peppers, american cheese, cheddar cheese, provolone, feta, green olives, tomatoes

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: green pepper

Rewards: 100

Mode: STRATEGIC

Confidence: 0.9166666

SYSTEM: green pepper? Sure! Would you like some wings?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: no

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: No? Sure! Would you like a drink?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: no

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: No? Sure! So your final order is:
Pizzas: (1)
0: Pizza --> sauce=tomato, crust=whole wheat, size=small, toppings=[green pepper]
Wings: (none)
Drinks: (none)
Thank you for ordering with us, your order will be delivered in 1 Hour.

Hello, Welcome to Chat and Order Pizza!

What would you like to order?

++
Emotional Results:

RELIEF: 0/10
HOPE: 0/10
PLEASED: 9/10

DISMAYED: 0/10
FEAR: 0/10
DISPLEASED: 1/10

Test Case 4

Default Strings loaded
System Property messages loaded
System properties loaded
Loading default properties from trained tagger C:\Documents and Settings\YellowDude\workspace\DialogManager\models\bidirectional-distsim-wsj-0-18.tagger
Reading POS tagger model from C:\Documents and Settings\YellowDude\workspace\DialogManager\models\bidirectional-distsim-wsj-0-18.tagger ... done [3.0 sec].
++++++++++++++++++++++++++++++
Emotion: Pleased

SYSTEM: Hello, Welcome to Chat and Order Pizza!

What would you like to order?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: pizzza

Rewards: 100

Mode: STRATEGIC

Confidence: 0.9444444

SYSTEM: Pizza? Sure! What kind of pizza would you like?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: vegetariean

Rewards: 100

Mode: STRATEGIC

Confidence: 0.969697

SYSTEM: Vegetarian? Sure! What type of crust do you want your pizza to have?
++++++++++++++++++++++++++++++
Emotion: Displeased
USER: while wheat

Rewards: 50

Mode: TACTICAL

Confidence: 0.45454547

SYSTEM: Sorry, can you repeat what you wanted?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: whole wheat

Rewards: 100

Mode: STRATEGIC

Confidence: 0.90909094

SYSTEM: whole wheat? Sure! What size of pizza do you want?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: zmall

Rewards: 100

Mode: STRATEGIC

Confidence: 0.8666667

SYSTEM: small? Sure! What sort of pizza sauce would you want?
++++++++++++++++++++++++++++++

Emotion: Pleased
USER: tomatoes

Rewards: 100

Mode: STRATEGIC

Confidence: 0.9166667

SYSTEM: tomato? Sure! What toppings would you like to have on the pizza?
+++++++++++++++++++++++++++++++
Emotion: Displeased
USER: papperonni

Rewards: 50

Mode: TACTICAL

Confidence: 0.0

SYSTEM: Sorry, which of the following did you want?
We have: green pepper, black olives, mushroom, onion, hot peppers, american cheese, cheddar cheese, provolone, feta, green olives, tomatoes
++++++++++++++++++++++++++++++
Emotion: Hope
USER: tomat

Rewards: 50

Mode: TACTICAL

Confidence: 0.625

SYSTEM: Did you mean: tomatoes? Would you like some wings?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: no

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: No? Sure! Would you like a drink?
++++++++++++++++++++++++++++++
Emotion: Pleased
USER: no

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: No? Sure! So your final order is:
Pizzas: (1)
0: Pizza --> sauce=tomato, crust=whole wheat, size=small, toppings=[tomatoes]
Wings: (none)
Drinks: (none)
Thank you for ordering with us, your order will be delivered in 1 Hour.

Hello, Welcome to Chat and Order Pizza!

What would you like to order?

++
Emotional Results:

RELIEF: 0/11
HOPE: 1/11
PLEASED: 8/11
DISMAYED: 0/11
FEAR: 0/11
DISPLEASED: 2/11

Test Case 5

Default Strings loaded

System Property messages loaded

System properties loaded

Loading default properties from trained tagger C:\Documents and Settings\YellowDude\workspace\DialogManager\models\bidirectional-distsim-wsj-0-18.tagger

Reading POS tagger model from C:\Documents and Settings\YellowDude\workspace\DialogManager\models\bidirectional-distsim-wsj-0-18.tagger ... done [3.0 sec].

++++++++++++++++++++++++++++++

Emotion: Pleased

SYSTEM: Hello, Welcome to Chat and Order Pizza!

What would you like to order?

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: can i get a piz

Rewards: 100

Mode: STRATEGIC

Confidence: 0.8

SYSTEM: Pizza? Sure! What kind of pizza would you like?

+++++++++++++++++++++++++++++++

Emotion: Displeased

USER: i don't know

Rewards: 50

Mode: TACTICAL

Confidence: 0.0

SYSTEM: Sorry, which of the following did you want?

We have: Regular, Vegetarian

+++++++++++++++++++++++++++++++

Emotion: Fear

USER: what?

Rewards: -100

Mode: SCRAMBLED

Confidence: 0.0

SYSTEM: Sorry, which of the following did you want?

We have: Regular, Vegetarian

++++++++++++++++++++++++++++++

Emotion: Fear

USER: reg

Rewards: -100

Mode: SCRAMBLED

Confidence: 0.42857143

SYSTEM: Sorry, can you repeat what you wanted?

++++++++++++++++++++++++++++++

Emotion: Relief

USER: regular

Rewards: 50

Mode: TACTICAL

Confidence: 1.0

SYSTEM: Regular? Sure! What type of crust do you want your pizza to have?

++++++++++++++++++++++++++++++

Emotion: Fear

USER: mm

Rewards: -100

Mode: SCRAMBLED

Confidence: 0.0

SYSTEM: Sorry, which of the following did you want?

We have: whole wheat, soy flour, cornmeal, deep dish, stuffed, hand tossed, crispy, brooklyn style, thin

++++++++++++++++++++++++++++++

Emotion: Fear

USER: broo

Rewards: -100

Mode: SCRAMBLED

Confidence: 0.2857143

SYSTEM: Sorry, can you repeat what you wanted?

++++++++++++++++++++++++++++++

Emotion: Fear

USER: brook

Rewards: -100

Mode: SCRAMBLED

Confidence: 0.35714287

SYSTEM: Sorry, can you repeat what you wanted?

++++++++++++++++++++++++++++++

Emotion: Fear

USER: brookl

Rewards: -100

Mode: SCRAMBLED

Confidence: 0.42857143

SYSTEM: Sorry, can you repeat what you wanted?

++++++++++++++++++++++++++++++

Emotion: Fear

USER: brookly

Rewards: -100

Mode: SCRAMBLED

Confidence: 0.5

SYSTEM: Did you mean: brooklyn style? What size of pizza do you want?

+++++++++++++++++++++++++++++++

Emotion: Relief

USER: medi

Rewards: -50

Mode: OPPORTUNISTIC

Confidence: 0.6666667

SYSTEM: Did you mean: medium? What sort of pizza sauce would you want?

++++++++++++++++++++++++++++++

Emotion: Relief

USER: toma

Rewards: -50

Mode: OPPORTUNISTIC

Confidence: 0.6666667

SYSTEM: Did you mean: tomato? What toppings would you like to have on the pizza?

++++++++++++++++++++++++++++++

Emotion: Hope

USER: pepperoni

Rewards: 50

Mode: TACTICAL

Confidence: 1.0

SYSTEM: pepperoni? Sure! Would you like some wings?

++++++++++++++++++++++++++++++

Emotion: Pleased

USER: yes

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: Yes? Sure! What type of sauce do you want with your wings?

+++++++++++++++++++++++++++++++

Emotion: Pleased

USER: buffalo

Rewards: 100

Mode: STRATEGIC

Confidence: 1.0

SYSTEM: buffalo? Sure! Would you like a drink?

++++++++++++++++++++++++++++++

Emotion: Displeased

USER: y

Rewards: 50

Mode: TACTICAL

Confidence: 0.33333334

SYSTEM: Sorry, can you repeat what you wanted?

+++++++++++++++++++++++++++++++

Emotion: Hope

USER: ye

Rewards: 50

Mode: TACTICAL

Confidence: 0.6666667

SYSTEM: Did you mean: Yes? What type of drink would you like?

++++++++++++++++++++++++++++++

Emotion: Hope

USER: pep

Rewards: 50

Mode: TACTICAL

Confidence: 0.6

SYSTEM: Did you mean: pepsi? So your final order is:

Pizzas: (1)

0: Pizza --> sauce=tomato, crust=brooklyn style, size=medium, toppings=[pepperoni]

Wings: (1)

0: Wings --> buffalo

Drinks: (1)

0: Drink --> pepsi

Thank you for ordering with us, your order will be delivered in 1 Hour.

Hello, Welcome to Chat and Order Pizza!

What would you like to order?

++

Emotional Results:

RELIEF: 3/19

HOPE: 3/19

PLEASED: 4/19

DISMAYED: 0/19

FEAR: 7/19

DISPLEASED: 2/19

REFERENCES

[Adina05] Adina Florea, Eugenia Kalisz, Embedding Emotions in an Artificial Tutor, Proceedings of the Seventh International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC'05), 2005.

[Callas06] http://www.callas-newmedia.eu/

[Cassell20] J. Cassell, T. Bickmore, H. Vilhjálmsson and H. Yan, "More than just a pretty face: affordances of embodiment," In Proceedings of the 5th ICIUI, New Orleans, 2000.

[CemKeskin07] Cem Keskin, Koray Balci, Oya Aran, Lale Akarun: A Multimodal 3D Healthcare Communication System. In 3DTV conference special session: 3D Multimodal Interfaces:

Applications on Edutainment, Medicine and Disabled, Kos Island, Greece, May 2007.

[Cleverbot10] http://www.cleverbot.com/

[Cuayahuitl09] Heriberto Cuayáhuitl, Steve Renals, Oliver Lemon, Hiroshi Shimodaira. Evaluation of a hierarchical reinforcement learning spoken dialogue System. Computer Speech & Language 24(2): 2010, pp. 395-429.

[Elena05] Elena Not, Koray Balci, Fabio Pianesi and Massimo Zancanaro "Synthetic Characters as Multichannel Interfaces" In Proceedings of ICMI05, Seventh International Conference on Multimodal Interfaces, Trento, October 3-7, 2005

[Frederic72] F. I. Parke, Computer Generated Animation of Faces. Proc. ACM annual conference, 1972.

[Futureshop11] http://www.futureshop.ca/en-CA/home.aspx

[Ikea99] http://www.ikea.com/ca/en/

[Jason05] Jason D. Williams, Pascal Poupart, and Steve Young. Factored Partially Observable Markov Decision Processes for Dialogue Management. Proc Workshop on Knowledge and Reasoning in Practical Dialogue Systems, International Joint Conference on Artificial Intelligence (IJCAI), Edinburgh, 2005.

[Jun07] "A Survey of Facial Modeling and Animation Techniques".

[Kim06] MC Kim, PH Seong, E Hollnagel. A probabilistic approach for determining the control mode in CREAM. Reliability Engineering & System Safety, Volume 91, Issue 2, 2006, pp. 191-199.

[Koray04] Koray Balci, Xface: MPEG-4 based open source toolkit for 3d facial animation in AVI04, Working Conference on

Advanced Visual Interfaces, Gallipoli, Italy, May 2004, pp. 25-28.

[Ortony88] Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, UK 1988.

[Piero08] Piero Cosi, Emanuela Caldognetto, Graziano Tisato, Emotional Talking Head: The Development of LUCIA, 2008

[Platt85] S. M. Platt, A Structural Model of the Human Face, Ph.D. Thesis, University of Pennsylvania, 1985.

[Platt81] S. Platt, N. Badler, Animating facial expression. Computer Graphics, 1981, vol. 15(3) pp. 245-252

[Plutchik91] From the book, The Emotions, Robert Plutchik, 1991, pp. 42-45.

[Prodanov04] P .Prodanov, A Drygajlo. Bayesian Networks for Error Handling Through Multimodality Fusion in Spoken Dialogues with Mobile Robots. ISCA Tutorial and Research Workshop (ITRW), 2004.

[Rickenberg20] R. Rickenberg, B. Reeves. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. In Proceedings of CHI, 2000.

[Roy20] Roy, N., Pineau, J., and Thrun, S. Spoken dialogue management using probabilistic reasoning. Proceedingsof the 38th Annual Meeting of the Association for ComputationalLinguistics (ACL-2000), Hong Kong, 2000.

[Sabiha10] Four Mode Based Dialogue Management with Modified POMDP Model, MSc Thesis, School of Computer Science University of Windsor, 2011, pp. 24-29.

[Sederberg96] T. W. Sederberg, S. R. Parry, Free-Form deformation of solid geometry models, Computer Graphics (Siggraph 1996), vol. 20(4), pp. 151 – 160.

[Semaineeu] http://www.semaine-project.eu/

[Stanton01] N A Stanton, M J Ashleigh, A D Roberts, F Xu. Testing Hollnagel's Contextual Control Model: Assessing team behavior in a human Supervisory control task. International Journal of Cognitive Ergonomics, 2001.

[Stefano08] Stefano Pasquariello, Catherine Pelachaud, Greta: A Simple Facial Animation Engine, 2008

[Tao04] Jianhua Tao, Tieniu Tan, Emotional Chinese talking head system, ICMI'04: Proceedings of the 6th international conference on Multimodal interfaces, 2004.

[Terzopoulos91] D. Terzopouls, K. Waters, Techniques for Realistic Facial Modeling and Animation, Proc. Computer Animation 1991, Geneva, Switzerland, Springer–Verlad, Tokyo, pp. 59–74.

[Trung06] Trung.H. BUI, Multimodal Dialogue Management- State of the art.

[Verbots04] http://www.verbots.com/

[Waters91] K. Waters, S. Terzopoulos, Modeling and Animating Faces using Scanned Data, Journal of Visualization and Computer Animation, 1991, Vol. 2, No. 4, pp. 123–128

[Zhen08] Zhen Liu ShaoHua He Wenjian Xiong: A Fuzzy Logic Based Emotion Model for Virtual Human, Cyberworlds International Conference, September 2008, pp. 284-288.

VITA AUCTORIS

Rajkumar Vijayarangan was born in 1985 in Chennai, India. He graduated from A.V.M High school in 2002. He went on to the Anna University where he obtained a B.E in Information Technology in 2007. He is currently a candidate for the Master's degree in Computer Science at the University of Windsor and hopes to graduate in Spring 2011.