

**FACIAL EXPRESSION ANIMATION BASED ON  
CONVERSATIONAL TEXT**

**HELGA MAZYAR**

**NATIONAL UNIVERSITY OF SINGAPORE**

**2009**

**FACIAL EXPRESSION ANIMATION BASED ON  
CONVERSATIONAL TEXT**

**HELGA MAZYAR**  
*(B.Eng. ISFAHAN UNI. OF TECH.)*

Supervisor: DR. TERENCE SIM

**A THESIS SUBMITTED  
FOR THE DEGREE OF MASTER OF  
COMPUTING**

**DEPARTMENT OF COMPUTER SCIENCE  
SCHOOL OF COMPUTING  
NATIONAL UNIVERSITY OF SINGAPORE**

**MAY 2009**

---

# Acknowledgements

---

This research project would not have been possible without the support of many people. The author wishes to express her gratitude to her supervisor, Dr. Terence Sim who was abundantly helpful and offered invaluable assistance, support and guidance.

The author also like to extend her thanks to Dr. Hwee Tou Ng for offering suggestions and advice, which proved to be of great help in this project. Deepest gratitude are also due to the members of Computer Vision laboratory without whose support and suggestions this study would not have been successful. Special thanks to Ye Ning, for his kind assistance and support.

Finally, the author would also like to convey thanks to the Singapore Agency of Science, Technology and Research (A\*Star) for providing the financial means and opportunity to study and live in Singapore.

---

# Contents

---

<b>Contents</b>	<b>i</b>
<b>List of Tables</b>	<b>vi</b>
<b>List of Figures</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Facial Expressions . . . . .	3
1.2.1 Facial Expression of Emotion . . . . .	4
1.3 Emotion . . . . .	5
1.3.1 Basic Emotions . . . . .	5
1.3.2 Mixed Emotions . . . . .	6
1.4 Statement of Problem . . . . .	6
1.5 Contribution . . . . .	7
1.6 Applications . . . . .	7
1.7 Organization of the Paper . . . . .	8
<b>2 Existing Works</b>	<b>10</b>
2.1 Emotional Classification Through Text . . . . .	10
2.1.1 Lexicon Based Technique(LBT) . . . . .	11
2.1.2 Machine Learning Techniques (MLT) . . . . .	13

2.1.3	Existing emotional Text Classification Systems . . . . .	18
2.2	Facial Expressions Synthesis . . . . .	20
2.2.1	Traditional Methods . . . . .	21
2.2.2	Sample-based Methods . . . . .	22
2.2.3	Parametric Methods . . . . .	22
2.2.4	Parameter Control Model . . . . .	26
2.2.5	Listing of Existing Facial Animation Systems . . . . .	26
<b>3</b>	<b>Experiments–Text Classification with Lexicon-Based Techniques</b>	<b>28</b>
3.1	Overview of Lexicon-Based Text Classifier . . . . .	28
3.2	Emotion Analysis Module . . . . .	29
3.2.1	Affect Database . . . . .	31
3.2.2	Word-level Analysis . . . . .	33
3.2.3	Phrase-level Analysis . . . . .	33
3.3	Experiment . . . . .	34
3.3.1	Corpus . . . . .	34
3.3.2	Results and Discussion . . . . .	37
<b>4</b>	<b>Experiments–Text Classification with Machine Learning</b>	<b>39</b>
4.1	Overview of Text Classification System . . . . .	39
4.2	Data representation . . . . .	42
4.2.1	Bag-of-words (BoW) . . . . .	42
4.3	Feature selection . . . . .	43
4.3.1	Chi-squared (CHI) . . . . .	44
4.4	Evaluation measures . . . . .	44
4.5	Results and Discussion . . . . .	45
<b>5</b>	<b>Experiments–Animation Module</b>	<b>48</b>
5.1	Expression of Mixed emotions . . . . .	48
5.2	Results and Discussion . . . . .	52

<b>6</b>	<b>User study</b>	<b>58</b>
<b>7</b>	<b>Conclusion</b>	<b>61</b>
	<b>Bibliography</b>	<b>63</b>
<b>A</b>	<b>Emoticons and abbreviations database</b>	<b>71</b>
<b>B</b>	<b>List of selected features for text classification</b>	<b>73</b>
<b>C</b>	<b>Facial Action Coding (FAC) System</b>	<b>74</b>
<b>D</b>	<b>User Study</b>	<b>76</b>

---

# Summary

---

Real time expressive communication is important as it provides aspects of the visual clues that are present in face-to-face interaction but not available in text-based communications. In this Master thesis report, we propose a new text to facial expression system (T2FE) which is capable of making real time expressive communication based on short text. This text is in the form of conversational and informal text which is used commonly by user of online messaging systems.

This system contains two main components: The first component is text processing component. The task of this component is to analyze text-based messages used in usual online messaging systems to detect the emotional sentences and specify the type of emotions conveyed by these sentences. Second component is the animation component and its task is to use detected emotional content to render relevant facial expressions. These animated facial expressions are presented on a sample 3D face model as the output of the system.

The proposed system differs from existing T2FE systems by using fuzzy text classification to enable rendering facial expressions for mixed emotions. To find out if the rendered results are interesting and useful from the users point of view, we performed a user study and the results are provided in this report.

In this report, first we study the main works done in the area of text classification and facial expression synthesis. Advantages and disadvantages of different techniques are presented to decide about the most suitable techniques for our

T2FE system. The results of the two main components of this system as well as a discussion on the results are provided separately in this report. Also the results of the user study is presented . This user study is conducted to estimate if the potential users of such system find rendered animations effective and useful.



---

# List of Tables

---

2.1	Existing emotional text classification systems and main techniques used.	19
2.2	Existing emotional text classification systems categorized by text type.	19
2.3	Facial Animation Parameters. . . . .	24
2.4	Existing facial expression animation systems . . . . .	27
3.1	Some examples of records in WordNet Affect database. . . . .	32
3.2	Some examples of records in Emoticons-abbreviations database . . . .	33
3.3	Sentence class distribution. . . . .	35
3.4	Sample sentences of the corpus and their class labels. . . . .	36
3.5	Results of classifying text with lexicon-based text classifier. . . . .	37
4.1	Summary of SVM sentence classification results. . . . .	45
4.2	Results of SVM classifier-Detailed accuracy by class. . . . .	46
6.1	Results of user study . . . . .	60
C.1	FAP groups. . . . .	74

---

# List of Figures

---

1.1	The general idea of the system. . . . .	3
1.2	Main components of our T2FE system. . . . .	4
1.3	Ekman six classes of emotion. . . . .	6
2.1	SVM linear separating hyperplanes. . . . .	16
2.2	SVM kernel concept. . . . .	17
2.3	An example of traditional facial animation system. . . . .	21
2.4	Examples sample-based methods. . . . .	22
2.5	Sample single facial action units . . . . .	23
2.6	Sample FAP stream. . . . .	25
2.7	Shape and grayscale variations for a facial expression. . . . .	26
2.8	Results of the model proposed by Du and Lin. . . . .	26
3.1	Overview of Lexicon-based text classifier . . . . .	29
3.2	Proposed emotion analysis module. . . . .	30
3.3	The interactive interface of our implementation. . . . .	34
4.1	A simple representation of text processing task applied in our system. . . . .	41
5.1	Basic shapes. . . . .	49
5.2	Illustration of linear interpolation used for generating interval frames. . . . .	50
5.3	Static and dynamic parts of 3D face model. . . . .	52

5.4	Neutral face( <i>FACENT</i> ) used as the base face in the experiment. . . . .	53
5.5	Basic shapes used for the experiment. . . . .	53
5.6	Interpolation of Surprise face. . . . .	54
5.7	Interpolation of Disgust face. . . . .	54
5.8	Blending of basic faces. . . . .	56
5.9	Over-animated faces. Some deformed results of animation module. . .	57
6.1	A sample entry of user study. . . . .	59
C.1	Feature points defined in FAC system. . . . .	75

---

# List of Symbols and Abbreviations

---

Abbreviation	Description	Definition
ag	Anger	page 28
AU	Action unit	page 23
BoW	Bag-of-Word	page 42
CHI	CHI-squared	page 44
dg	Disgust	page 28
FAC	Facial action coding	page 23
FAP	Facial animation parameter	page 24
FDP	facial definition parameters	page 24
fp	false positive	page 45
fn	false negative	page 45
fr	Fear	page 28
hp	Happiness	page 28
LBT	Lexicon based technique	page 11
ME	maximum entropy	page 15
MLT	Machine learning technique	page 13
MPL	minimum path-length	page 12
NB	Naive Bayes	page 14
NLP	natural language processing	page 10

Abbreviation	Description	Definition
PMI	Pointwise mutual information	page 12
PMI-IR	Pointwise mutual information-Information retrieval	page 12
sd	Sadness	page 28
sp	Surprise	page 28
SNHC	Synthetic/natural hybrid coding	page 24
SVM	Support Vector Machine	page 15
T2FE	text to facial expression	page iv
tp	true positive	page 45

# Chapter 1

---

## Introduction

---

### 1.1 Motivation

One of the interesting challenges in the community of human-computer interaction today is how to make computers be more human-like for intelligent user interfaces.

Emotion, one of the user affect, has been recognized as an important parameter for the quality of the daily communications. Given the importance of the emotions, affective interfaces using the emotion of the human user are gradually more desirable in intelligent user interfaces such as human-robot interactions. Not only this is a more natural way for people to interact, but it is also believable and friendly in human-machine interaction. In order for such an affective user interface to make use of user emotions, the emotional state of the human user should be recognized or sensed in many ways from diverse modality such as facial expression, speech, and text. Among them, detecting the emotion within an utterance in text is essential and important as the first step in the realization of affective human-computer interfaces using natural language. This stage is defined as perception step[11]. In this study, we mainly focus on short text for perception and try to find out emotion conveyed through this kind of text. Although the

methods provided in this report for perception are applicable to long text, we do not extend our study to long text perception. This is basically because there is a high chance of having variety of emotional words from different groups of emotions in long text (for example having happy and sad emotional words in the same text). This fact might cause different emotions to neutralize the effect of each other which leads to get neutral faces as the output of the animation module which is not exciting for the potential users of this system. Also, using short text reduce the analysis time which is needed for online communication as the main application of this T2FE system.

Another important domain in the area of human-computer interaction is generation step, regarding production of dynamic expressive visual and auditory behaviors . For this research paper, we narrow the visual behaviors down to facial expressions and auditory behaviors are not discussed.

In this report, at first we study the techniques widely used to reason about emotions automatically from short conversational text as well as the methods used in the computer animation area for expressing emotions on a 3D face. We investigate the promising techniques and propose a new technique for our text to facial-expression system. The performance of our system is measured using machine learning measures.

It is important to note that one of the main characteristics our system is the ability to show mixed emotions on face and not only the based emotions (we will cover the definitions of basic and mixed emotions in section 1.3). Also, we present the results of a user study performed to see if users of such system find watching an animated face, which is animated using mixed emotions extracted from text messages, useful and interesting.

As mentioned before, in our proposed system the sentences are analyzed and the appropriate facial expressions are displayed automatically on a 3D head. Figure 1.1 demonstrates the general idea of this system and Figure 1.2 shows

main components of our T2FE system.

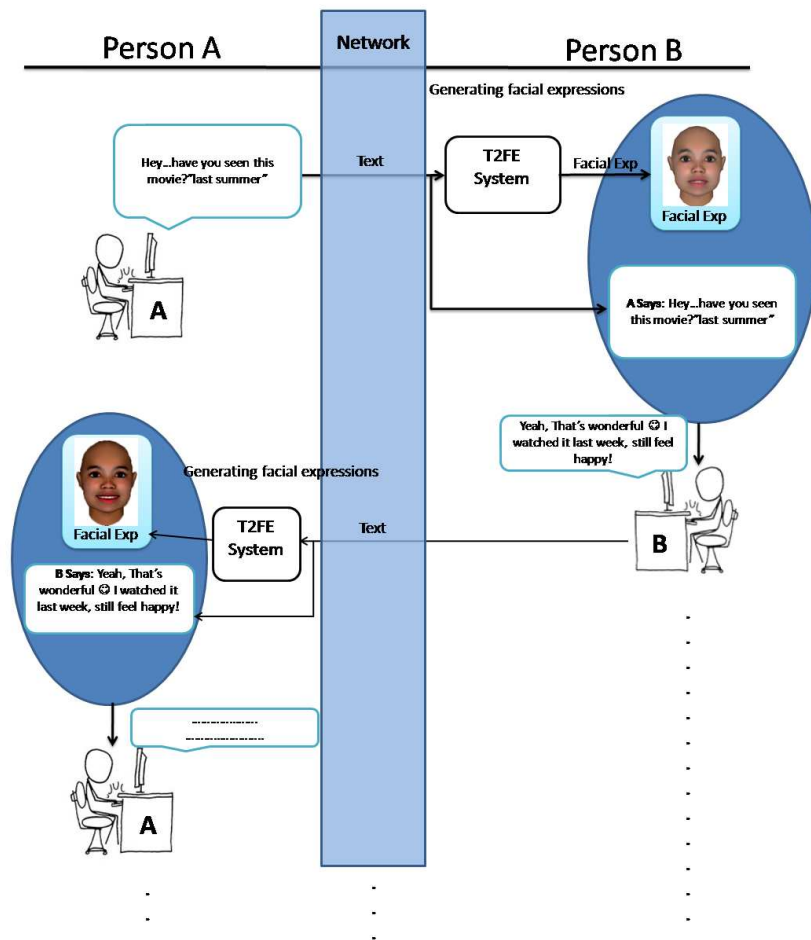


Figure 1.1: The general idea of the system. A chat session between two persons (A and B) is taking place utilizing T2FE system. Users of the system can watch the extracted facial-expression animation as well as the original text message.

## 1.2 Facial Expressions

A facial expression is a visible manifestation of the affective state, cognitive activity, intention, personality, and psychopathology of a person [26]. Facial expressions results from one or more motions or positions of the muscles of the face and play several roles in communication and can be used to modify the meaning of what is being said[69].



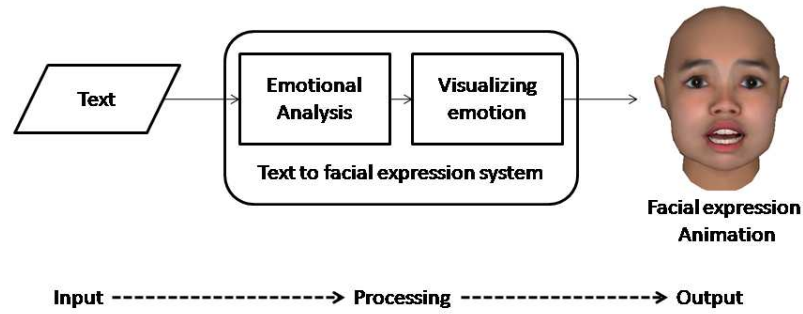


Figure 1.2: Main components of our T2FE system.

Facial expression is also useful in controlling conversational flow. This can be done with simple motions, such as using the direction of eye gaze to determine who is being addressed.

One sub-category of facial expression which is related to non-verbal communication is emotional facial expressions which we will discuss more in the following subsection.

### 1.2.1 Facial Expression of Emotion

Emotions are linked to facial expressions in some undetermined loose manner [41]. Emotional facial expressions are the facial changes in response to a person internal emotional states, intentions, or social communications. Intuitively people look for emotional signs in facial expressions. The face seems to be the most accessible window into the mechanisms which govern our emotional behaviors [29].

Given their nature and function, facial expressions (in general), and emotional facial expressions (in particular), play a central role in a communication context. They are part of non-verbal communication and are strongly connected to daily communications.

## 1.3 Emotion

The most straightforward description of emotions is the use of emotion-denoting words, or category labels [86]. Human languages have proven to be extremely powerful in producing labels for emotional states: Lists of emotion-denoting adjectives were compiled that include at least 107 items [86]. It can be expected that not all of these items are equally central. Therefore, for specific research aims, it seems natural to select a subset fulfilling certain requirements.

In an overview chapter of his book, Robert Plutchik mentions the following approaches to proposing emotion lists: Evolutionary approaches, neural approaches, a psychoanalytic approach, an autonomic approach, facial expression approaches, empirical classification approaches, and developmental approaches [70]. Here, we just focus on the facial expression approach and divide emotions into two main categories, **basic emotions** and **mixed emotions** for more discussion.

### 1.3.1 Basic Emotions

There are different views on the relationship between emotions and facial activity. The most popular one is the basic emotions view. This view assumes that there is a small set of emotions that can be distinguished discretely from one another by facial expressions. For example, when people are happy they smile and when they are angry they frown.

These emotions are expected to be universally found in all humans. In the area of facial expressions, the most accepted list is based on the work by Ekman [28].

Ekman devised a list of basic emotions from cross-cultural research and concluded that some emotions were basic or biologically universal to all humans. His list contains these emotions: **Sadness, Happiness, Anger, Fear, Disgust** and **Surprise**. These basic emotions are widely used for modeling facial expression of emotions ([36, 96, 59, 8]) and are illustrated in Figure 1.3.

Some psychologists have differentiated other emotions and their expressions from those mentioned above. These other emotion or related expressions include contempt, shame, and startle. In this paper, we use the Ekman set of basic emotions because his set is widely accepted in the facial animation community.



Figure 1.3: Ekman six classes of emotion: Anger, Happiness, Disgust, Surprise, Sadness and Fear from left to right.

### 1.3.2 Mixed Emotions

Although there is a small number of basic emotions, there are many other emotions which humans use to convey their feelings. These emotions are mixed or derivative states. It means that they occur as combinations, mixtures, or compounds of the primary emotions. Some examples of this category are: blend of happiness and surprise, blend of disgust and anger and blend of happiness and fear.

Databases of naturally occurring emotions show that humans usually express low-intensity rather than full blown emotions, and complex, mixed emotions rather than mere basic emotions downsized to a low intensity [86]. The fact motivated us to use these category of emotion for animating facial expressions. For some sample illustrations of these category of emotions please refer to Figure 2.4 or the results of our animation system, Figure5.8.

## 1.4 Statement of Problem

We propose a new text to facial expression system which is capable of making real time expressive communication based on short text. This text is in the form

of conversational and informal text which is used commonly by user of online messaging systems.

This system contains two main components: The first component is text processing component. The task of this component is to analyze text-based messages to detect the emotional sentences and specify the type and intensity of emotions conveyed by these sentences. Second component is the animation component and its task is to use detected emotional content to render relevant facial expressions. Mixed classes of emotions are used in this system to provide more realistic results for the user of the system.

The rendered facial expressions are animated on a sample 3D face model as the output of the system.

## **1.5 Contribution**

Existing T2FE systems ([37, 5, 14, 36, 97, 96, 90]) are composed of two main components: The text processing component, to detect emotions from text, and the graphic component which uses detected emotions to show relevant facial expressions on the face. Our studies show that for the graphic part, researchers use basic classes of emotions and other types of emotions are ignored.

Our proposed T2FE system differs from existing T2FE systems by using fuzzy text classification to enable rendering facial expressions for mixed emotions. The user study conducted for this thesis show that most of the users of such systems find the expressions of mixed classes of emotions a better choice for representing the emotions in the text.

## **1.6 Applications**

Synthesis of emotional facial expression based on text can be used in many applications. First of all, such system can add another dimension to understanding

on-line text based communications. Although these days technology has enriched multi-modal communication, still many users prefer text based communication. Detecting emotion from text and visualizing emotion can help in this aspect.

Secondly, this system can be a main component for development of other affective interfaces in human-computer Interaction. For projects such as embodied agents or talking heads, conveying emotional facial expressions are even more important than verbal communication. These projects have important roles in many different areas such as animation industry, affective tutoring on e-learning system, virtual reality and web agents.

## **1.7 Organization of the Paper**

Chapter 2 of this thesis covers the literature review and related works. In this chapter significant works done in the area of text classification and facial animation systems are explained separately: Section 2.1 explains two well-known approaches proposed for automatic emotional classification of text in the Natural Language Processing research community followed by a discussion of the advantages and disadvantages of two approaches. Section 2.2 explains the main approaches proposed for rendering emotional facial expressions.

Chapter 3 and chapter 4 explain our experiments of text classification using two different approaches of text classification. For each experiment, the results are presented followed by a discussion on the accuracy of the implemented text classifier.

Chapter 5 explains the animation module of our T2FE system. This chapter includes explanation of the animation module as well as some frames of rendered animation for different mixed emotions. These results are followed by a discussion on the validity and quality of the rendered facial expressions.

Chapter 6 presents a user survey conducted to find out if users find the results of the implemented system interesting and useful. Finally, chapter 7 concludes

this paper with suggestions for the scope of future work and some concluding remarks.

## Chapter 2

---

# Existing Works

---

In this chapter, we overview significant existing works in the area of emotional text classification and facial expression's animation respectively.

### 2.1 Emotional Classification Through Text

Emotion classification is related to sentiment classification. The goal of sentiment classification is to classify text based on whether it expresses positive or negative sentiment. The way to express positive or negative sentiment are often the same as the one to express emotion. However emotion classification differs from sentiment classification in that the classes are finer and hence it is more difficult to distinguish between them.

In order to analyze and classify emotion communicated through text, researchers in the area of natural language processing(NLP) proposed a variety of approaches, methodologies and techniques. In this section we will see methods of identifying this information in a written text.

Basically, there are two main techniques for sentiment classification: *Lexicon based* techniques(symbolic approach) and *machine learning* techniques. The symbolic approach uses manually crafted rules and lexicons [65][64], where the

machine learning approach uses unsupervised, weakly supervised or fully supervised learning to construct a model from a large training corpus [6][89].

### **2.1.1 Lexicon Based Technique(LBT)**

In lexicon based techniques a text is considered as a collection of words without considering any of the relations between the individual words. The main task in this technique is to determine the sentiment of every word and combine these values with some function (such as average or sum). There are different methods to determine the sentiment of a single word which will be discussed briefly in the following two subsections.

#### **Using Web Search**

Based on Hatzivassiloglou and Wiebe research [39], adjectives are good indicators of subjective, evaluative sentences. Turney[83] applied this fact to propose a context-dependent model for finding the emotional orientation of the word. To clarify this context dependency, we can consider the adjective "unpredictable" which may have a negative orientation in an automotive review, in a phrase such as "unpredictable steering", but it could have a positive orientation in a movie review, in a phrase such as "unpredictable plot".

Therefore he used pairs consisting of adjectives combined with nouns and of adverbs combined with verbs. To calculate the semantic orientation for a pair Turney used the search engine Altavista. For every combination, he issues two queries: one query that returns the number of documents that contain the pair close (defined as "within 10 words distance") to the word "excellent" and one query that returns the number of documents that contain the pair close to the word "poor". Based on this statistical issue, the pair is marked with positive or negative label. The main problem here is the classification of text just into two classes of positive and negative because finer classification requires a lot of



computational resources.

This idea of using pairs of words, can be formulated using Pointwise Mutual information (PMI). PMI is a measure of the degree of association between two terms, and is defined as follow [66]:

$$PMI(t_1, t_2) = \log \frac{p(t_1, t_2)}{p(t_1) \times p(t_2)} \quad (2.1)$$

PMI measure is symmetric ( $PMI(t_1, t_2) = PMI(t_2, t_1)$ ). It is equal to zero if  $t_1$  and  $t_2$  are independent and can take on both negative and positive values.

In text classification, PMI is often used to evaluate and select features from text. It measures the amount of information that the value of a feature in a text (e.g. the presence or absence of a word) gives about the class of the text. Therefore, higher values of PMI present better candidates for features.

PMI-IR [82] is another measure that uses Information Retrieval to estimate the probabilities needed for calculating the PMI using search engine hitcounts from a very large corpus, namely the web. The measure thus becomes as it is shown in the following equation:

$$PMI-IR(t_1, t_2) = \log \frac{hitCounts(t_1, t_2)}{hitCounts(t_1) \times hitCounts(t_2)} \quad (2.2)$$

### **Using WordNet**

Kamps and Marx used WordNet[34] to determine the orientation of a word. In fact, they went beyond the simple positive-negative orientation, and used the dimension of appraisal that gives a more fine-grained description of the emotional content of a word. They developed an automatic method[45] using the lexical database WordNet to determine the emotional content of a word. Kamps and Marx defined a distance metric between the words in WordNet, called minimum path-length (MPL). This distance metric is used to find the emotional weights for

the words. Only a subset of the words in WordNet can be evaluated using MPL technique, because for some words defining the connecting path is not possible.

### **Improving Lexicon Based Techniques**

Lexicon based techniques have some important drawbacks mainly because they do not consider any of the relations between the individual words. They can often be more advantageous if they consider some relations between the words in a sentence. Several methods are proposed to fulfill this need. We mention here briefly Mulder and al.'s article [63], which discusses the successful use of an affective grammar.

Mulder et al. in their paper [63] proposed a technique that uses affective and grammar together to overcome the problem of ignoring relations between words in lexicon based techniques. They noted that simply detecting emotion words can tell whether a sentence is positive or negative oriented, but does not explain towards what topic this sentiment is directed. In other words, what is ignored in lexicon base technique is the relation between attitude and object.

The authors studied how this relation between attitude and object is formalized and combined a lexical and grammatical approach:

- Lexical, because they believe that affect is primarily expressed through affect words
- Grammatical, because affective meaning is intensified and propagated towards a target through grammatical constructs.

### **2.1.2 Machine Learning Techniques (MLT)**

In supervised method a classifier (e.g. Support Vector Machines (SVM), Naive Bayes (NB), Maximum Entropy (ME)) is trained on the training data to learn the sentiment recognition rules in text. By feeding a machine learning algorithm a large training corpus of affectively annotated texts, it is possible for the system to

not only learn the affective value of affect keywords as the job done with Lexicon based techniques, but such a system can also take into account the valence of other arbitrary keywords (like lexical affinity), punctuation, and word co-occurrence frequencies [56].

The method that in the literature often yields the highest accuracy uses Support Vector Machine classifier[83]. The main drawback of these methods is that they require a labeled corpus to learn the classifiers. This is not always available, and it takes time to label a corpus of significant size. In the following subsections we briefly explain some of the most important text classifiers:

### **Naive Bayes Classifier(NB)**

One approach to text classification is to assign to a given document  $d$  the class  $cls$  which is determined by  $cls = \arg \max P(c|d)$ . Here,  $c$  is any possible class considered in the classification problem. Based on Bayes rule:

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)} \quad (2.3)$$

After detecting features ( $f_i$ 's) from document based on the nature of the problem, to estimate the term  $P(c|d)$ , Naive Bayes assumes that  $f_i$ 's are conditionally independent given  $d$ 's. Therefore the training model will act based on the following formula.

$$P(c|d) = \frac{P(c) \prod_{i=1}^k P(f_i|c)^{n_i(d)}}{P(d)} \quad (2.4)$$

Naive Bayes classifier simplifies the job by its conditional independence assumption, which clearly does not hold in real-world situations. However, Naive Bayes-based text categorization still tends to perform surprisingly well [52]. Domingos and Pazzani [25] showed that Naive Bayes is optimal for certain problem classes with highly dependent features.

## Maximum Entropy

Maximum entropy classification (ME) is another machine learning technique which has proved effective in a number of natural language processing applications [12]. ME estimates  $P(c|d)$  based on the following formula:

$$P(c|d) = \frac{1}{Z(d)} \exp\left(\sum_i \lambda_{i,c} \times F_{i,c}(d, c)\right) \quad (2.5)$$

$F_{i,c}$  is a feature/class function for feature  $f_i$  and class  $c$ . The value of  $F_{i,c_1}(d, c_2)$  is equal to 1 when  $n_i(d) > 0$  (meaning that feature  $f_i$  exists in document  $d$ ) and  $c_1 = c_2$ . Otherwise it is set to 0.

$Z(d)$  is a normalization function and is used to ensure a proper probability:

$$Z(d) = \sum_c \exp\left(\sum_i \lambda_{i,c} \times F_{i,c}(d, c)\right) \quad (2.6)$$

The  $\lambda_{i,c}$  s are feature-weight parameters and are the parameters to be estimated. A large  $\lambda_{i,c}$  means that  $f_i$  is considered a strong indicator for class  $c$ . The parameter values are set so as to maximize the entropy of the induced distribution subject to the constraint that the expected values of the feature/class functions with respect to the model are equal to their expected values with respect to the training data: the underlying philosophy is that we should choose the model that makes the fewest assumptions about the data while still remaining consistent with it, which makes intuitive sense [66].

Unlike Naive Bayes, ME makes no assumptions about the relationships between features, and so might potentially perform better when conditional independence assumptions are not met. It is shown that some times, but not always, ME outperforms Naive Bayes at standard text classification [66].

## Support Vector Machines

Support vector machines (SVMs) have been shown to be highly effective at traditional text categorization, generally outperforming NB [43]. They are large-

margin, rather than probabilistic, classifiers, in contrast to NB and ME.

In the two-category case, the basic idea behind the training procedure is to find a hyperplane, represented by vector  $\vec{w}$ , that not only separates the document vectors in one class from those in the other, but for which the separation, or margin, is as large as possible (See Figure 2.1).

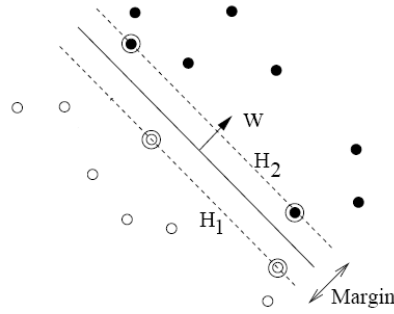


Figure 2.1: Linear separating hyperplanes ( $W$ ,  $H_1$  and  $H_2$ ) for SVM classification. Support vectors are circled.

This search corresponds to a constrained optimization problem. Letting  $c_j \in \{-1, 1\}$  (corresponding to positive and negative) be the correct class of document  $\vec{d}_j$ , the solution can be written as the following equation:

$$\vec{w} = \sum_i \gamma_i c_i \vec{d}_i, \gamma_i > 0 \quad (2.7)$$

where the  $\gamma_i$ s are obtained by solving a dual optimization problem. For more details please refer to Burges tutorial on SVM [18].

Those  $\vec{d}_j$  such that  $\gamma_i$  is greater than zero are called support vectors, since they are the only document vectors contributing to  $\vec{w}$ . Classification of test instances consists simply of determining which side of  $\vec{w}$ 's hyperplane they fall on.

Figure 2.1 is a classic example of a linear classifier, i.e., a classifier that separates a set of documents into their respective classes with a line. Most classification tasks, however, are not that simple, and often more complex structures are needed in order to make an optimal separation. This situation is depicted in

Figure 2.2.(a). Here, it is clear that a full separation of documents would require a curve (which is more complex than a line).

Figure 2.2 shows the basic idea behind SVM.. In Figure 2.2.(b) we see the original documents mapped, i.e., rearranged, using a set of mathematical functions, known as **kernels**. The process of rearranging the objects is known as mapping (transformation). Note that in this new setting, the mapped objects are linearly separable and, thus, instead of constructing the complex curve (left schematic), all we have to do is to find an optimal line that can separate mapped documents.

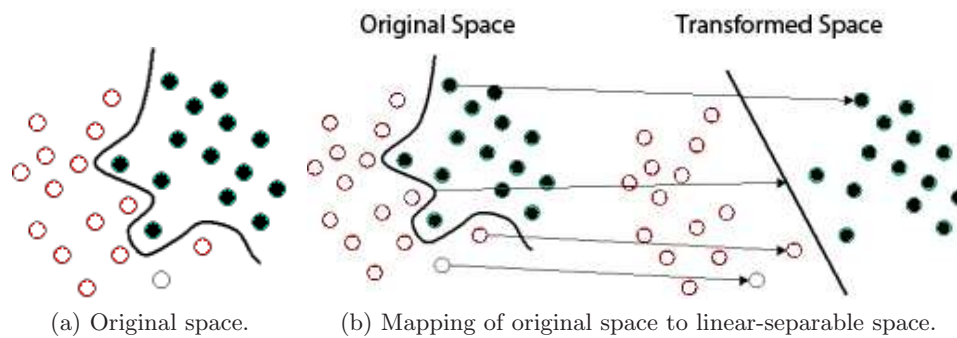


Figure 2.2: SVM kernel concept.

There are non-linear extensions to the SVM, but Yang and Liu [92] found the linear kernel to outperform non-linear kernels in text classification. Hence, we only present linear SVM.

### Multi-classification with SVM

So far, we explained SVM for binary classification but there are more than two classes in the classification task. We call this a multi-classification problem. Regarding SVM classifier, the dominating approach for multi-classification is to reduce the single multiclass problem into multiple binary problems where each of the problems yields a binary classifier. There are two common methods to build such binary classifiers:

1. **One-versus-all:** In this method each classifier distinguishes between one of the labels to the rest. Classification of new instances for one-versus-all case is done by a winner-takes-all strategy, in which the classifier with the highest output function assigns the class.
2. **One-versus-one:** In this method each classifier distinguishes between every pair of classes. For classification of a new instance, every classifier assigns the instance to one of the two classes, then the vote for the assigned class is increased by one vote, and finally the class with most votes determines the instance classification.

### **2.1.3 Existing emotional Text Classification Systems**

To complete the literature survey on the emotional text classification techniques, here we present the list of existing systems proposed for affective text classification (text classification based on the emotional content of the text) as well as the base techniques used in the systems. This list is shown in Table 2.1.

In a different listing of the existing works on emotional text classification, Table 2.2 shows the existing works based on text type (short or long) and the type of emotions considered in the classification. Based on the importance of conversational text in online communication and this table content, conversational text is potentially a good area of research.

System	Technique	System	Technique
[36]	LBT	[67]	ML (SVM,NB,ME)
[62]	LBT (PMI)	[72]	LBT (PMI)
[80]	LBT	[35]	LBT
[61]	LBT	[90]	LBT
[14]	LBT	[74]	LBT
[83]	LBT (PMI)	[54]	ML(SVM)
[37]	LBT	[79]	ML(NB,SVM)
[53]	LBT(PMI)	[23]	ML(NB)
[65]	LBT(with Grammer)	[7]	LBT
[50]	ML(ME)	[15]	LBT
[56]	LBT	[91]	ML(SVM)
[24]	ML		

Table 2.1: Existing emotional text classification systems and main techniques used.

<i>EMOTION TYPE</i>	<i>TEXT TYPE</i>					
	Long(# of Sen >15)		Short (# of Sen <15)			
	Formal <sup>‡</sup>	Informal <sup>§</sup>	Formal		Informal	
			NonDialogue	Dialogue	NonConv <sup>¶</sup>	Conv <sup>  </sup>
Positive or Negative	[83],[67],[13],[15]	[13]	-	-	[40],[61],[62]	[23]
Ekman emotions	[53],[54],[6]	-	[24],[90],[72],[74],[36]	[50],[14],[35],[37]	[7],[79],[56]	[65]
Other basic emotions	[80]	[91]	-	-	-	-

Table 2.2: Existing emotional text classification systems categorized by text type.

<sup>‡</sup>Formal text does not contain informal words/phrases. This group contains News, News headlines and articles.

<sup>§</sup>Informal text contains informal words/phrases or emoticons. This group contains blogs, film reviews, written conversations and stories.

<sup>¶</sup>Non conversational.

<sup>||</sup>Conversational.



## 2.2 Facial Expressions Synthesis

A facial expression is a visible manifestation of the affective state, cognitive activity, intention, personality, and psychopathology of a person [26]; it plays an important non-verbal communicative role in interpersonal relations.

Mehrabian [60] showed that facial expressions of the speaker contributes for 55 percent to the effect of the spoken message, while the verbal part (i.e., spoken words) of a message contribute only for 7 percent to the effect of the message as a whole and the vocal part (e.g., voice intonation) contributes for 38 percent. As a consequence of the information that they carry, facial expressions play an important role in communications.

Since facial expressions can be a very powerful form of communication, they should be used in enhanced Human-Machine interfaces . Unfortunately, the synthesis of proper conversational expressions is extremely challenging. One reason for this is that humans are amazingly good at recognizing facial expressions and can detect very small differences in both motion and meaning.

A second reason can be found in the subject matter itself: The physical differences between an expression that is recognizable and one that is not can be very subtle.

Facial expression generation has attracted many researchers since the early 1970s and many studies are published in this area. To have a more compact presentation, in this study we only review those papers that are more related and suited to T2FE systems.

Approaches to facial animation in general can be studied based on the methods applied to achieve animation: *Traditional methods*, *Sample-based methods*, *Parametric methods* and *Parameter control methods*. Traditional methods are based on image processing algorithms such as image warping and morphing to synthesis facial movements while sample-based methods aim to generate facial animation based on a large dataset of animations.

Parameterized systems assign weighted vertices of the face mesh to every parameter. During animations, the vertices are displaced according to the parameter value. Parameter control methods try to adopt a flexible model for shape and grayscale attributes and animate the face based on the emotion vectors. In the following subsections, we will describe these methods in more details.

In another categorization, as far as the output is concerned, it could be a 2D image [57, 71, 84, 97] or a 3D surface model [68, 69, 95].

### 2.2.1 Traditional Methods

The traditional method of generating facial expressions is geometric warping-based approaches which were proposed by Aizawa et al [4]. This method can be used to animate 2D image or 3D wire-frame face model. In this method, the parameters of face animation are extracted from facial expression image set first, and then mapped to a new person's face by using geometry-controlled image warping [10, 55].

The main drawback of this approach is that it can only capture the facial feature's geometric changes, completely ignore their illumination variations. Some results generated using this approach are shown in Figure 2.3.

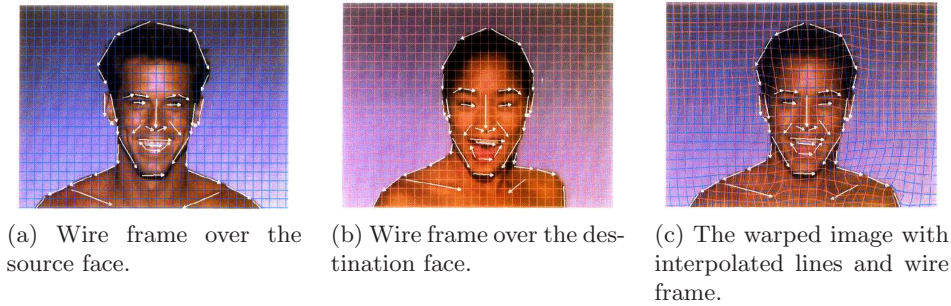


Figure 2.3: An example of traditional facial animation using warping techniques. Results taken from [10]

## 2.2.2 Sample-based Methods

In order to generate photo-realistic animation, many researchers have proposed sample-based methods [69, 33, 16, 78]. In this kind of approaches, a large amount of sample images are collected and stored at the training process, and then used to synthesize expression images by using editing and morphing methods. For example, Pighin et al. in [69] used a combination of several photos to generate expression images. In their system, user can interact with computer to design the expression he/she wants. Based on the input, the system can use different weight to mix the training samples for synthesizing different expressions. Although the sample-based approaches can obtain very realistic expressions, they are hard to generate expressions for a new person.

Some results of this method are shown in Figure 2.4.



(a) A global blend between surprised (left) and sad (center) produces a worried expression (right).

(b) Combining the upper part of a neutral expression (left) with the lower part of a happy expression (center) produces a fake smile (right).

Figure 2.4: Examples of results generated by sample-based methods. The examples are taken from [69]

## 2.2.3 Parametric Methods

Synthesizing facial expression by means of parametric control methodology is also very popular. Actually, there are many researches on this subject in recent years. Here, we briefly explain two main systems designed using parametric approach: Facial action coding system and MPEG4 animation.

## Facial Action Coding System

Ekman and Friesen [31][30] built a system for describing all visually distinguishable facial movements, called the Facial Action Coding System or FACS. It is based on the enumeration of all action units(AUs) of a face that cause facial movements. Some samples of AUs are shown in Figure 2.5.

There are 46 AUs in FACS that account for changes in facial expression. The combination of these action units result in a large set of possible facial expressions. AU combinations may be additive, in which case combination does not change the appearance of the constituents, or non-additive, in which case the appearance of the constituents changes. For example smile expression is considered to be a combination of pulling lip corners (AU 12+13) and/or mouth opening (AU 25+27) with upper lip raiser (AU 10) and bit of furrow deepening (AU 11). However this is only one type of a smile; there are many variations of the above motions, each having a different intensity of actuation.

Although the number of atomic action units is small, more than 7,000 combinations of action units have been observed.

FACS provides the necessary detail with which to describe facial expression. Despite its limitations, this method is the most widely used method for measuring human facial motion for both human and machine perception.

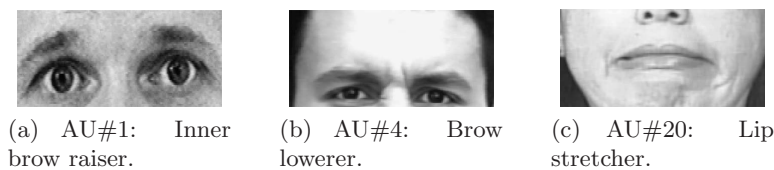


Figure 2.5: Sample single facial action units

## MPEG-4 Animation

Aiming for efficiently representing facial expressions and animations, MPEG4-SNHC (synthetic/natural hybrid coding), a sub-protocol of the MPEG4 standard for video compression, contains two components: synthetic objects and natural objects.

One of the standards within the first component, when combined with a 3D human model, is to provide an efficient description for transferring the related parameter information regarding body motions and facial expressions in a real-time manner, thus increasing the associated compression ratio. These parameters can be divided into two categories: facial expression-related and body motion-related.

The first category, pertaining to this work, consists of two parts: *FDPs* (facial definition parameters) and *FAPs* (facial animation parameters). Please refer to Table 2.3 for some examples of FAPs and their description.

FDPs define the shape of the model while FAPs define the facial actions. Given the shape of the model, the animation is obtained by specifying the FAP-stream that is for each frame the values of FAPs (see Figure 2.6).

#	FAP Name	FAP Description
...	...	...
3	open-jaw	Vertical jaw displacement (does not affect mouth opening)
4	lower-t-midlip	Vertical top middle inner lip displacement
5	raise-b-midlip	Vertical bottom middle inner lip displacement
6	stretch-l-cornerlip	Horizontal displacement of left inner lip corner
7	tretch-r-cornerlip	Horizontal displacement of right inner lip corner
...	...	...

Table 2.3: Facial Animation Parameters.

In a FAP-stream, each frame has two lines of parameters. In the first line the activation of a particular marker is indicated (0, 1) while in the second, the target values, in terms of differences from the previous ones, are stored.

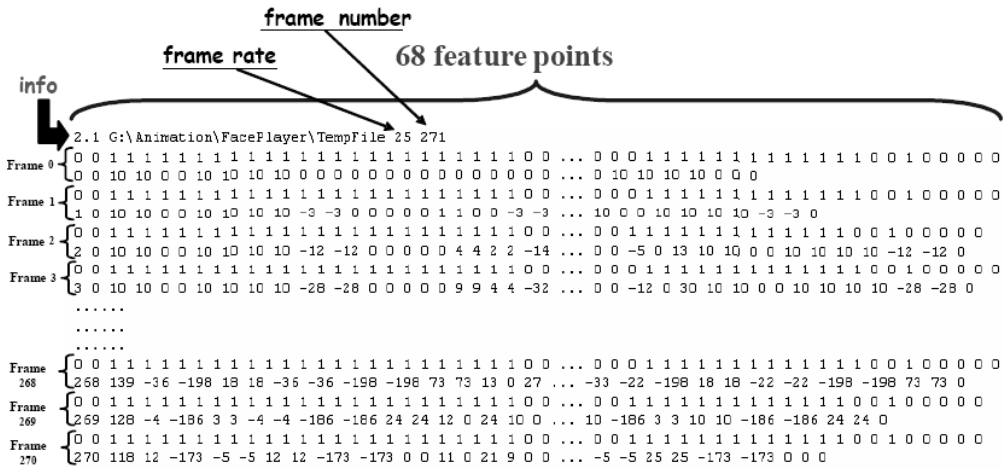


Figure 2.6: Sample FAP stream.

Raouzaïou et al. [71] made use of this scheme for modeling facial expression animations. They defined some control points on the face and used them with FAP information to animate face. These control points are shown in appendix C.

In fact, MPEG-4 standard has defined six basic expressions in its facial animation parameters (FAP), including happiness, sadness, surprise, anger, disgust and fear. The value of each parameter indicates the ingredient of corresponding expression embodied in the image, and can be used by user to simulate the mixed expressions. For a complete illustration of FAPs and FAP groups refer to appendix C.

Following this standard, many research groups have been actively developing compatible facial animation systems with various implementations [51, 93, 17]. The main drawback of this approach is that it can only make cartoon-like animation, while is difficult to generate very natural characteristic facial expression.



Figure 2.7: Shape and grayscale variations for a facial expression. Examples taken from [27].

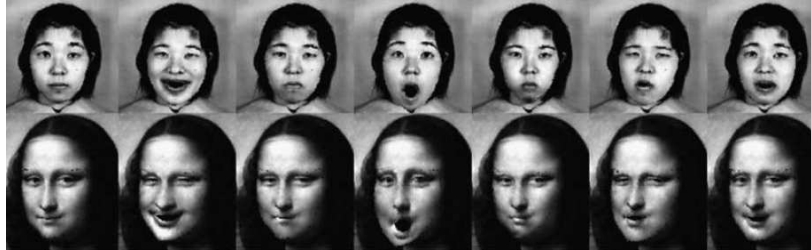


Figure 2.8: Results of the model proposed by Du and Lin on a training set entry(up) and on a new person(down)[27].

## 2.2.4 Parameter Control Model

Du and Lin [27] proposed a parameter control model to synthesize comprehensive facial image. They adopted Flexible Model, proposed by Cootes and his colleague [9], in their method. The model represents both shape and grayscale appearance of an elastic object (Figure 2.7), and is built by performing a statistical analysis over a training set of example images. They used JAFFE [58] as the database and the attached corresponding evaluation scores of each image to train a mapping function so that complicated expression can be manipulated by an emotion vector. Some results of this method are shown in Figure 2.8.

## 2.2.5 Listing of Existing Facial Animation Systems

There are many works in the field of facial expression animation. For this master thesis, we studied the systems which look promising for our T2FE system. Systems that generate natural expressions and are simple enough to be able to use for his master study. These systems are presented in Table 2.4. Because one of

the main attributes of the animation engine needed for our T2Fe system is the real-time ability, we also consider this attribute in this table.

These systems are which look suitable Table 2.4 lists the existing systems and their attributes in the area of facial animation.

System	Control Parameters	Facial Model	Real time
CSLU[1]	system specific	parametric	yes
Chai03[22]	system specific	performance-based	yes
BEAT[21]	expressions	-	yes
Ezzat02[32]	prototype images	sample-based	no
Wojdel'05[88]	FACS	param./sample-based	yes
Albrecht02[5]	expressions	sample-based	yes
Koch98[47]	FACS	sample-based	yes
CharToon[75]	system specific	parametric	yes
Kshirsagar03[49]	MPEG-4	sample-based	yes

Table 2.4: Existing facial expression animation systems



## Chapter 3

---

# Experiments–Text Classification with Lexicon-Based Techniques

---

As explained in earlier sections, our proposed system is composed of two main parts: text processing part and animation part. In this section we focus on text processing part and explain methods used for automatic sentence classification into classes of emotion. We use Ekman classes of emotion in our classification task including happiness(*hp*), sadness(*sd*), fear(*fr*), anger(*ag*), surprise(*sp*), disgust(*dg*). Furthermore, we dedicate another class to neutral sentences to consider sentences that do not convey any emotional content.

It is important to note that the aim of this text classification is not just extracting the dominant emotion of a given sentence. In fact, we are interested in finding probabilities of classifying a given sentence to each of the seven classes.

### 3.1 Overview of Lexicon-Based Text Classifier

In this section, the lexicon-based text classifier proposed for classifying text to emotion classes is explained. Figure 3.1 shows the general idea of this system.

In this system, the Emotion Analysis module receives the input text and uses

Affect database to analyze emotional content of the text and finds the emotional weights (rates) of each class of emotion. These weights will be used in the animation module to render the relevant facial expressions. We will explain the details of this text classification in next subsections.

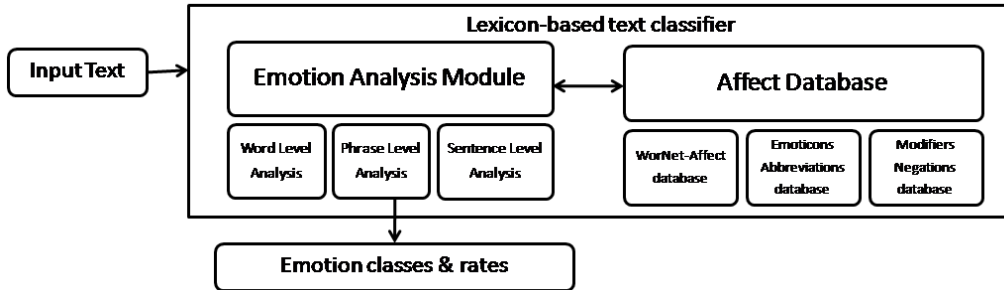


Figure 3.1: Overview of Lexicon-based text classifier

## 3.2 Emotion Analysis Module

In this section we describe our emotion analysis model. This model is shown in Figure 3.2 and is composed of 6 main components:

1. Input preprocessing module. The main job of this module is to split input to sentences and words, remove stopwords and quotations.
2. Affect database. This component builds the affect core of the system. We will explain it in more detail in the next subsection.
3. Word-level analysis module. This module is used to find the emotional weights of each word or sign(in case of emoticons or abbreviations) using affect database and transfer these values to the phrase-level analysis module. For each word we refer to affect database to estimate the emotional weights. In addition, some heuristic rules are employed to adjust these weights which will be discussed in section 3.2.2.

4. Phrase-level analysis module. This module is used to compute the phrase level emotional weights. In this level, some heuristic rules are used to adjust the weights given by word-level analysis which will be discussed in subsection 3.2.3.
5. Sentence-level analysis module. This module is used to compute the sentence level emotional weights.
6. Output module. Used for illustrating the emotional weights for the classes of emotion as shown in Figure 3.3.

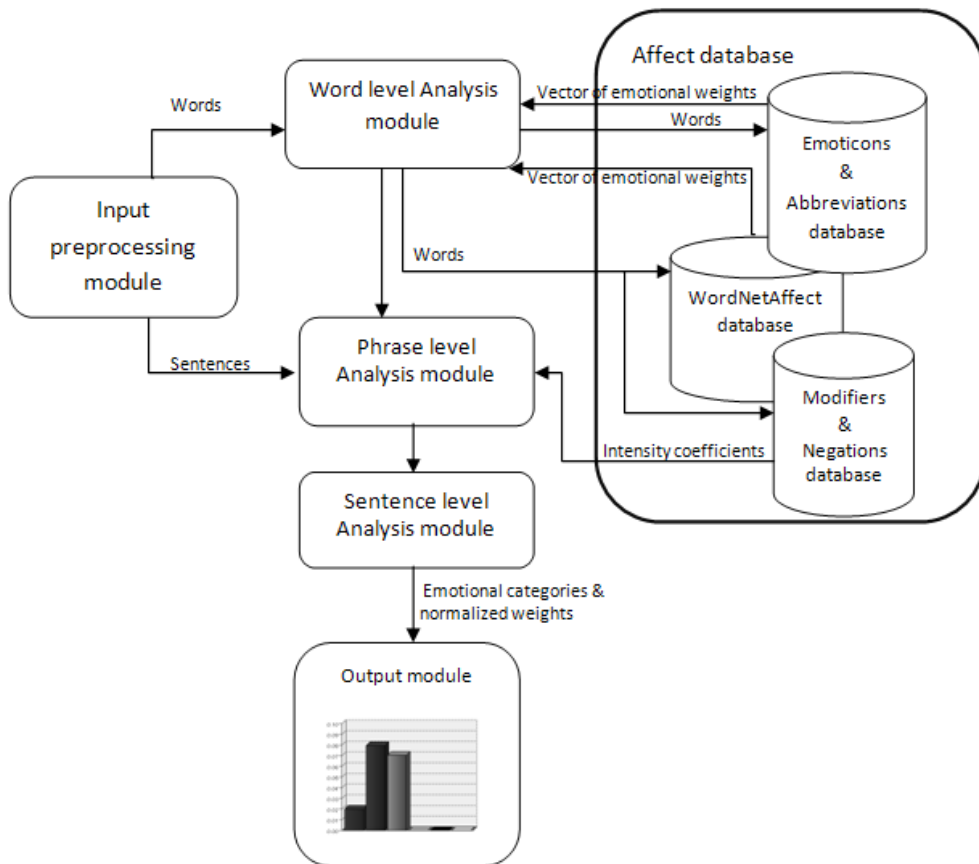


Figure 3.2: Proposed emotion analysis module.

Briefly speaking, the whole procedure of detecting emotion from text and classifying the text into one of the six emotional categories are as follows: For a

given input, the input preprocessing module is used to split the input to sentences and words and send the words one by one to the word-level analysis modules and the regarding sentence to the phrase-level analysis module.

In the word-level, at first we look for exclamation signs, question signs and words written in uppercase since these factors intensify the affect of the whole sentence. In the next step, for each word, we make a query to WordNet-Affect database and emoticons-Abbreviation database to find the affect weights conveyed with each word for all the emotional categories. Also, searching the word in modifiers-negations database returns the coefficients which will be used in the phrase-level to adjust the phrase level emotional weights.

In phrase-level analysis, we use the weights obtained from the word-level and the intensify-negation coefficients to compute the weights of the phrase which compose a sentence. At the next module, sentence-level analysis module, the whole emotional weights of the input are computed and sent to the output module. We show these weights by  $weight_i$  where  $i \in \{sd, hp, anger, fr, dg, sp, ne\}$ . In the output module, these weights will be normalized and written to the output interface. The normalization is based on the following formula:

$$NormWeight_i = \frac{weight_i}{\sum_i \alpha_i \times weight_i} \quad (3.1)$$

where  $i \in \{sd, hp, anger, fr, dg, sp, ne\}$  and  $\alpha_i$  illustrates the overall effect on each emotional category resulted from phrase-level analysis.

### 3.2.1 Affect Database

The affect database composes the main affect of our system and contains three main databases: WordNet-Affect which contains 4698 affective words, emoticons-abbreviations database containing 111 affect records ,intensity modifiers-negations database.

### 1. WordNet-Affect

For the source of affect lexicons, WordNetAffect [76, 19] was used to find the affect meaning of each word. This dataset is a linguistic resource for the lexical representation of affective knowledge which was created with the aim to support applications relying on language recognition and generation. Some examples of this dataset and their assigned emotional weights are listed in Table 3.1. To view the full list please refer to WordNet Domains web site [3].

Text	General Weight	hp	sd	ag	fr	dg	sp
annoyed	1	0	0	0	0	0.266	0
casual	0.355	0	0	0	0	0	0.177
hard	0.085	0	0.038	0	0	0	0

Table 3.1: Some examples of records in WordNet Affect database.

### 2. Emoticons-abbreviations database

In today’s textual communication, people usually use emoticons, such as :-), :-0, :-(, and specific abbreviations, such as LOL, OMG, rofl, to clarify the meanings and emotions of the sentences.

In order to support abbreviated language in our emotion model and interpret the affective features of emoticons, a dataset containing 111 records was created. Each record of this dataset contains the textual form of the emoticon or abbreviation followed by the emotional values for each emotional class. There is also a *general weight* field in this database that shows the intensity of the emoticon/abbreviation. Some records of this database are listed in Table 3.2 while the complete list is provided in appendix A.

### 3. Intensity modifiers and negations

In our daily conversation and written, we use adjectives and adverbs (e.g. very, awfully, dreadfully, eminently, exceedingly) to change the intensity of

Text	General Weight	hp	sd	ag	fr	dg	sp
LOL	1	1.0	0.0	0.0	0.0	0.0	0.0
:-@	1	0.0	0.0	1.0	0.0	0.5	0.0
WOW	1	0.0	0.0	0.0	0.0	0.0	1.0
:-)	1	1.0	0.0	0.0	0.0	0.0	0.0

Table 3.2: Some examples of records in Emoticons-abbreviations database

a specific word. In the proposed model a set of modifiers is collected and ranked to handle the effect of intensity modifiers.

### 3.2.2 Word-level Analysis

After splitting the input to tokens, we refer to affect database to estimate the emotional weights for the word. In addition, We employ some heuristic rules to adjust these weights:

1. Words written in upper case: This is a sign of more intensive word.
2. The effect of the previous word:
  - Previous word is an intensity modifier (e.g. "extremely") => Intensify the emotional weights by multiplying the weight to modifier effect.
  - Previous word is a negation word (e.g. no, not, don't, don't, haven't, weren't, wasn't, didn't) => flip the weights of the affect word by multiplying weights by -1.

### 3.2.3 Phrase-level Analysis

For phrase level analysis, some heuristic rules are used to find the overall emotion of the sentence.

1. Number of exclamation signs in a sentence: the more exclamation signs the higher emotional weights

2. More emoticons with more emotional signs (e.g. :DDDD) intensifies the emotional weights

### 3.3 Experiment

The proposed lexicon-based text classifier is implemented using Java. The program can work in two modes: the interactive mode where user of the program can enter arbitrary text. In this mode, the weights of each emotional category and the dominant weight will be shown in the form of bar charts. A sample output of this mode is shown in Figure 3.3.

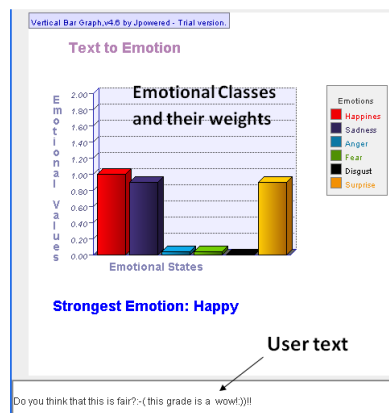


Figure 3.3: The interactive interface of our implementation.

The second mode is the test mode, where we used a well-known publicly available and labeled dataset to test the accuracy of our implementation. This corpus and the test results are more described in the following subsections.

#### 3.3.1 Corpus

For text classification part of our system, we use the a subset of corpus prepared by Szpakowicz in [8]. This database contains of 173 blog posts containing a total of 15205 sentences. The sentences were labeled with emotion category ( one of the 7 categories of happiness, sadness, fear, surprise, disgust, anger and no-emotion)

and emotion intensity (high, medium and low) by four judges. In this paper we just consider the emotion category independent and do not use the emotion intensity.

Furthermore, we just select the sentences for which the annotators agreed on the emotion category. This limitation narrows the number of the sentences down to **4090 sentences**. These sentences include conversation words and emoticons which makes this dataset a good candidate for learning systems based on informal conversations systems such as our system.

Table 3.3 and Table 3.4 show the distribution of the sentences based on the emotion category and some sample sentences of this corpus respectively.

Sentence Class	Frequency	Percentage
No-emotion	2800	0.68
Happiness	536	0.13
Surprise	115	0.02
Disgust	172	0.04
Sad	173	0.04
Angry	179	0.04
Fear	115	0.02

Table 3.3: Sentence class distribution.

We can see from Table 3.3 that in this corpus most of the sentences are labeled with no-emotion label and there is a high distribution skew, where other classes are very small. This means that for these classes, we have a few samples to learn from. In chapter 4, we will see how this skew can effect the classification task.



Sentence	Class
WE WERE TOTALLY AWESOME!!!!	hp
I don't know what happened to my happiness, I woke up feeling down and miserable and in fact it's worse.	sd
Wow, I hardly ever have plans.	sp
First off, it's a hike to get up to this place, and I can't see worth shit in the dark.	dg
Sheldon and I told him to shut up.	ag
The second day I went in and I was so paranoid.	fr
See yaaaa tomarrow.	ne

Table 3.4: Sample sentences of the corpus and their class labels.

Table 3.4 shows some sample sentences from the corpus used for experiments. We can see that the text contains many attributes of a conversational text such as abbreviations (such as “it’s”) and conversational words (such as using “yaaa” instead of “you”). Also, we can see that during text-based messaging people might use words in capital letters to show higher level of emotion (first sentence).

### 3.3.2 Results and Discussion

The results of this test are shown in Table 3.5. In this table, the accuracy of the classification task is provided for each class. The *Accuracy* measure showed in this table, represents the number of sentences correctly classified.

Emotion	Accuracy
ne	0.43
sp	0.32
hp	0.37
dg	0.34
sd	0.26
ag	0.32
fr	0.28

Table 3.5: Results of classifying text with lexicon-based text classifier.

The average accuracy of emotion analysis module implementation is 33.14 percent which is still better than a random classifier that provides an accuracy of about 14 percent for 7 classes.

One reason for this low accuracy of this classifier, is the fact that there are many cases that the emotion of a sentence is hidden in the content of the sentence and not just the words that make a sentence. Therefore, searching emotional words in an affect database might not be the best solution to classify a sentence to one of the classes of emotion. Although, enriching the affect database might help in this regard, we are never sure to be able to store all of the possible emotional words in a database.

Another reason is that in lexicon-based technique, the input text is considered

as a collection of words without considering any of the relations between the individual words.

In the next chapter, we will focus on the machine learning techniques to improve this accuracy.

## Chapter 4

---

# Experiments–Text Classification with Machine Learning

---

### 4.1 Overview of Text Classification System

In the previous chapter, we explained our lexicon-based text classifier and explained that in this classifier, the input text is considered as a collection of words without considering any of the relations between the individual words. In fact, the main task in this technique is to determine the sentiment of every word and combine these values with some function. To cover this drawback, many researchers proposed using machine learning techniques for text classification and reported better results using this techniques (for more details please refer to section 2.1.2). In this chapter we explain our experiments of emotional text classification using machine learning techniques.

It is important to note that the aim of this text classification is not just extracting the dominant emotion of a given sentence. In fact, we are interested in finding probabilities of classifying a given sentence to each of the seven classes and use these probabilities as the blending weights in the graphic module. In other words we are looking for fuzzy classification of text and not the precise

classification. To cover the need of fuzzy classification, we use fuzzy set theory developed by Zadeh[94] that allows concepts that do not have well-defined sharp boundaries.

In contrast to the classical set theory, in which any object should be classified as a member or non-member of a specific set, an object in fuzzy theory can partially belong to a fuzzy set. A *membership function* is used to measure the degree to which an object belongs to a fuzzy set. This value is a number between 0 and 1.

Based on these definitions of fuzzy sets and membership functions, we can define our fuzzy set( $A$ ) and membership functions ( $Mem$ ) as follows:

$$A = corpus = \{s_1, s_2, \dots, s_n\} \quad (4.1)$$

$$Mem_i(s_k) = \begin{cases} prob(s_k|i) & , i \in \{hp, sd, fr, dg, sp, ag\} \\ 1 - \sum_{\sigma} prob(s_k|\sigma) & , i = ne, \sigma \in \{hp, sd, fr, dg, sp, ag\} \end{cases} , 1 \leq k \leq n \quad (4.2)$$

After calculating the values of the member functions, these values will be used to blend 3D face models for six classes of emotion together and generate the new head. We will explain this in more details in chapter 5. In the following sections we will explain our text classification experiment and the results.

An overview of the sentence classification task is shown in Figure 4.1. Briefly speaking, we use a labeled corpus as our learning dataset. For short text classification, many researchers used different classifiers such as Naive Bayse (NB), Decision Trees (DT), and Support Vector Machine (SVM).

In our work, SVM is selected as the classifier as it has been traditionally used for text categorization with great success[42, 44]. SVM is well-suited for text categorization because of the large feature sets involved and SVM's ability to project data into multiple dimensions to find the optimal hyperplane.

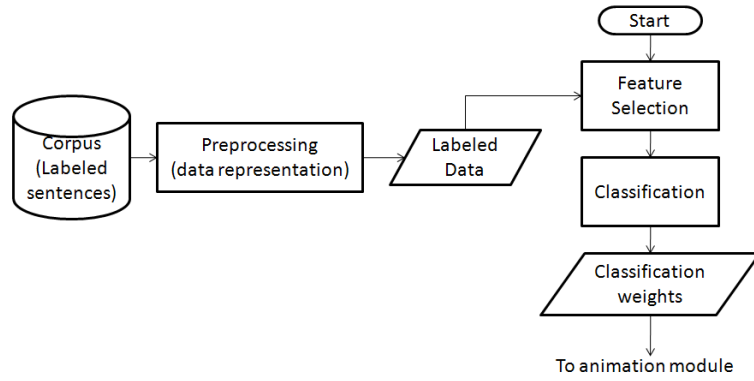


Figure 4.1: A simple representation of text processing task applied in our system.

In case of short-text classification, we refer to the experiments done by Khoo et al. in [46] for short text and sentence classification. Their experiments show that SVM classification algorithm generally outperforms other common algorithms. The authors also analyzed different feature selection algorithm including Chi-squared, Information Gain, Bi-Normal Separation and Sentence Frequency. They evaluated these various feature selection by inspecting the performance of classifiers, and concluded that for sentence classification the results of different feature selection algorithms are almost the same and a there is not a significant difference among the results. They suggest that for sentence classification, a cheap and simple feature selection algorithm is enough and further processing might lead to losing a large portion of features which is basically not useful for short text classification.

Based on this discussion, we use SVM classifier for our text processing part, with a linear kernel and One-versus-all scheme for multi-category classification.

The One-versus-all scheme helps us in fuzzy text classification by investigating the results of classifying one sentence to all of the classes of emotions. We will use these results as the probabilities while calculating membership functions in equation 4.2. There are non-linear extensions to the SVM, but Yang and Liu found the linear kernel to outperform non-linear kernels in text classification

[92]. Hence, we only present linear SVM results.

The platform used for applying the classification algorithms is the machine learning library WEKA [87].

For testing the classifiers, 10-fold cross validation procedure is used. With this procedure all the labeled sentences are randomly divided to 10 sets of equal size and training is done on 9 sets and the classifier is tested on 1 set. This procedure is repeated for 10 times and the average accuracy is considered as the accuracy of the classifier. We will explain the results in the following sections.

## 4.2 Data representation

Before explaining the details of the experiments and the results, we explain techniques used for data representation and features extraction for sentences.

Data representation is a domain specific problem and the technique used for this task should be selected based on the specific aims of the project. For example, the best data representation used for the task of classifying text based on subject (topic selection) might not be the best candidate for detecting emotions from text. However, in this research to pay more attention to main contributions of this paper we do not focus to find the best techniques for data representation and use the well known and widely used techniques. For this experiment, we use Bag-of-words (BoW) representation which is popular for its simplicity and computational efficiency [20].

### 4.2.1 Bag-of-words (BoW)

In this technique, the whole corpus is broken into an ordered set of words and each distinct word corresponds to one feature. If there are  $N$  number of distinct words in the corpus, the bag will contain  $N$  members and each text is transformed to a vector of  $N$  elements as  $\langle a_1, a_2, \dots, a_N \rangle$  where  $a_k$  is the weight of the  $k^{th}$  word of the bag in that text. Different researchers propose different definitions

to calculate these weights such as the frequency of the word in the text. In our work, because we are dealing with short text, we use the binary weights which shows the existence or absent of the specific word in that text.

To explain more about our BoW representation, suppose that our corpus contains two sentences  $S_1$  and  $S_2$ .  $S_1 = \text{“See yaaa tomorrow !!!”}$  and  $S_2 = \text{“I’ll talk to you tomorrow”}$ . Processing  $S_1$  adds four words to BoW: “See”, “yaaa”, “tomorrow” and “!!!” and the size of BoW will increase to four.  $S_2$  is tokenized into five words: “I’ll”, “talk”, “to”, “you” and “tomorrow” and the first four words will be added to BoW (“tomorrow” is already inside BoW). After this step the BoW looks like this ordered set :

$$BoW = \{See, yaaa, tomorrow, I'll, talk, to.you\}$$

With this BoW, the first and the second sentence are converted to the following binary representation respectively :

$$S_1 = \langle 1, 1, 1, 0, 0, 0, 0 \rangle$$

$$S_2 = \langle 0, 0, 1, 1, 1, 1, 1 \rangle$$

After this step, learning algorithms are applied on these representations to build text classifier.

### 4.3 Feature selection

When we are using machine learning techniques, we are usually dealing with large datasets leading to thousands or ten thousands of features for learning aims. These large number of features put a very high load on the learning algorithms, in our case the classification algorithms.

Using feature selection algorithms, we can sort and select the best features and reduce the loads on the classification problem. Here we briefly explain Chi-squared which is widely used as a feature selection algorithms for text classification [46].



### 4.3.1 Chi-squared (CHI)

This algorithm measures the independence of each possible feature from all of the classes and ignores the features that show high independence [38]. For sentence classification experiment, each word is considered as a candidate feature and the independence from each of the classes of emotion is measured, the maximum score is taken as CHI score and is used as selection criteria. Lower score means better candidate.

In our case, CHI measures the independence of word  $w$  and each class  $C_i$  as follow:

$$CHI(w, C_i) = \frac{N \times (\alpha \times \delta - \beta \times \gamma)^2}{(\alpha + \gamma) \times (\alpha + \beta) \times (\beta + \gamma) \times (\delta + \gamma)} \quad (4.3)$$

$$\text{where } \begin{cases} \alpha = \# \text{Occurrence } w \text{ and } C_i \text{ together} \\ \beta = \# \text{Occurrence } w \text{ without } C_i \\ \gamma = \# \text{Occurrence } C_i \text{ without } w \\ \delta = \# \text{Occurrence neither } w \text{ nor } C_i \end{cases} , i \in \{ne, hp, sd, sp, fr, dg, ag\}$$

## 4.4 Evaluation measures

To evaluate and compare the results of our experiments, we use three standard measures used widely in classification algorithms: Precision, recall and Fmeasure[73]. In the task of classifying text into class  $C_i$  these measures are defined as follows :

$$Precision = \frac{tp}{tp + fp} \quad (4.4)$$

$$Recall = \frac{tp}{tp + fn} \quad (4.5)$$

$$Fmeasure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4.6)$$

$$\text{where } \left\{ \begin{array}{l} tp = \# \text{sentences correctly classified into } C \\ fp = \# \text{sentences incorrectly classified into } C \\ fn = \# \text{sentences incorrectly not classified into class } C \\ C \in \{ne, hp, sd, sp, fr, dg, ag\} \end{array} \right.$$

## 4.5 Results and Discussion

For this experiment, we used the same corpus used in the lexicon-based text classification experiment (refer to section 3.3.1). Table 4.1 shows the summary of the text classification results gained using best 200 features selected with Chi-square feature selection methods out of 7970 features. These 7970 features are in fact all the words existing in the corpus, ignoring the duplicates. Selected features are listed in appendix B.

As we can see in table 4.1, the total number of instances used in the experiment is 4090 and 79.58 percent of them are correctly classified into emotion. This accuracy shows good progress compared to overall accuracy of 33.14% gained from the works explained in subsection 3.3.2.

Total number of instances	4090
Number of correctly classified instances	3255
Number of incorrectly classified instances	835
Accuracy (percentage of correctly classified instances)	79.58%

Table 4.1: Summary of SVM sentence classification results.

To investigate the results in more details, we show class-by-class results in Table 4.2. The values in this table show how well the each class was predicted in terms of different measures: True Positive, False Positive, Precision, Recall, and Fmeasure (please refer to section 4.4 for definition of these terms.).

Class	TP Rate	FP Rate	Precision	Recall	F-measure
ne	0.976	0.551	0.794	0.976	0.876
hp	0.487	0.017	0.813	0.487	0.609
sp	0.304	0.004	0.714	0.304	0.427
dg	0.413	0.006	0.755	0.413	0.534
sd	0.266	0.003	0.821	0.266	0.402
ag	0.335	0.003	0.845	0.335	0.480
fr	0.417	0.002	0.889	0.417	0.568
Weighted Avr.	0.796	0.380	0.798	0.796	0.768

Table 4.2: Results of SVM classifier-Detailed accuracy by class.

As shown in this table, Precision values of all of the classes are higher than 0.7, and the weighted average of total Precision is close to 0.80 which is a very good precision value. Also, the False Positive values are very low for all of the classes except *ne* class. This means that there is a high chance that a sentence is classified in class **ne** while it is labeled as an sentence with emotional contents by human judges. On the other hand, the low False Positive values for the other classes show that if sentence is classified to a class, for example *hp*, there is a high chance that this sentence is truly a *hp* sentence.

The analysis of results for True Positive measure show that all of the classes except *ne* have a low True Positive rate. This low rate for classes of emotion (*hp,sp,dg,sd,sg,fr*) and high rate for *ne* convey the fact that many sentences which are labeled with emotional classes by judges, are classified into *ne* class by classifier. In fact, our classifier is a bias classifier and is eager to classify sentences into *ne* emotion class. However, in case of classifying sentence into one of the emotional classes, the result of the classifier is highly accurate and the same as the labels annotated using human judges.

To investigate this problem deeper, we refer to the distribution of data in our training set. As it is shown in Table 3.3, 0.68 percent of the sentences of our training corpus are labeled with *ne* and some classes are very small. This means that for these small classes we have a very few positive examples to learn from.

Researchers in the area of machine learning suggested some methods to overcome the problem of bias classifier affected from highly skewed data [81, 85]. In this experiment we do not focus on methods to solve this problem. Instead, we try to estimate the accuracy of our classifier with better measures and use F-measure, derived from Precision and Recall, to reflect the biased behavior of our classifier. As reported in [46] using F-measure can avoid the misleading of Precision or Recall in classification problems. The values of F-measure are presented in the last column of Table 4.2.

## Chapter 5

---

# Experiments–Animation Module

---

### 5.1 Expression of Mixed emotions

In this section, we present a model for generating facial expression arising from mixed emotions. Here, by mixed emotions we are referring to those emotions which are a blend of two or more basic emotions (refer to section 1.3.2 for more details).

We formulate our model at the level of facial expressions. In other words, we do not build the expressions of mixed emotions from scratch. We use the basic expressions of emotions and blend these expressions together to build new expressions. This idea of blending some basic shaped together and generating new shaped is called Shape Blending in computer animation and has a great practical use [77] and can be categorized as a subset of sample-based approach (section 2.2.2).

To be able to generate the expressions of mixed emotions for each frame, we need two sets of parameters: basic shapes and the weights for blending the basic shapes together.

- Basic shapes

Based on the needs of our system, we choose the facial expressions of basic

classes of emotion as our basic shapes, these shapes are shown in Figure 5.1. We use the notation  $FACE^\sigma$  to refer to these shapes, where  $\sigma \in \{hp, sd, fr, dg, sp, ag\}$ . Each of these  $FACEs$  are made of vertices  $v_1$  to  $v_n$  and can be positioned in the space using their 3D coordinates as shown in equation 5.1, where  $n$  is the number of vertices and  $k$  is the  $k^{th}$  vertex of  $FACE^\sigma$ .

We consider the neutral face as the base face and use  $FACE^{nt}$  to refer to this for our next discussions. The goal of the animation module is to animate this base face into a particular emotional face as specified with weights gained from the text-processing module.

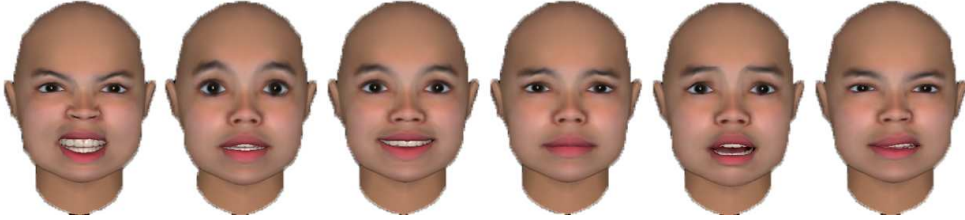


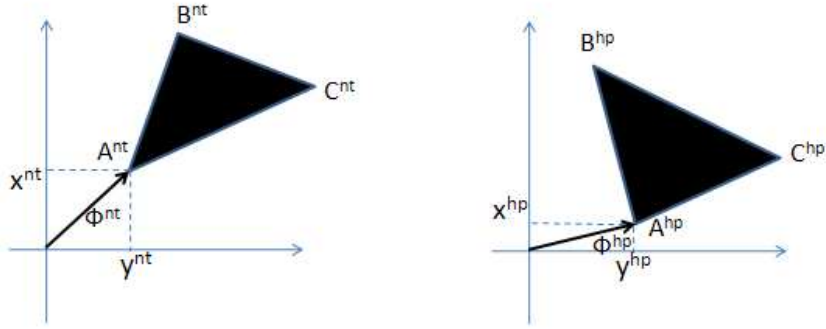
Figure 5.1: Basic shapes: Anger, Surprise, Happiness, Sadness, Fear and Disgust from left to right.

- Weights

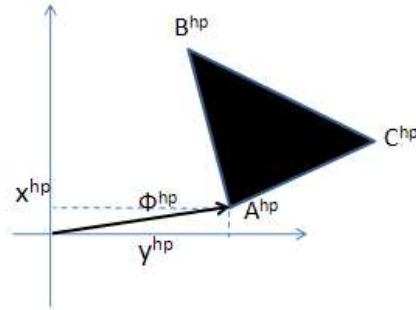
The weights are measured by processing the text to evaluate the classification weights based on the algorithm explained in chapter 4 and more specifically with equation 4.2.

$$FACE^\sigma = \begin{bmatrix} v_1^\sigma \\ \cdot \\ v_k \\ \cdot \\ v_n^\sigma \end{bmatrix} = \begin{bmatrix} xmax_1^\alpha & ymax_1^\alpha & zmax_1^\alpha \\ \cdot & \cdot & \cdot \\ xmax_k^\alpha & ymax_k^\alpha & zmax_k^\alpha \\ \cdot & \cdot & \cdot \\ xmax_n^\alpha & ymax_n^\alpha & zmax_n^\alpha \end{bmatrix}, FACE^{nt} = \begin{bmatrix} x_1^{nt} & y_1^{nt} & z_1^{nt} \\ \cdot & \cdot & \cdot \\ x_k^{nt} & y_k^{nt} & z_k^{nt} \\ \cdot & \cdot & \cdot \\ x_n^{nt} & y_n^{nt} & z_n^{nt} \end{bmatrix} \quad (5.1)$$

where  $\sigma \in \{hp, sd, fr, dg, sp, ag\}$ .



(a) Start frame - A sample triangle from Neu- (b) End frame - The same triangle in a happy  
tral face. face



(c) Prototype Happiness frame - The same triangle in the  $FACE^{hp}$

Figure 5.2: Illustration of linear interpolation used for generating interval frames.

Based on these two parameters (basic shapes and weights), the animation module generates the faces for each frame of the animation.

To better explain this task, let us explain the whole work flow of animation module using a triangle instead of the whole face model. Figure 5.2(a) shows triangle  $ABC$  (representative of  $FACE^{nt}$ ) in the first frame of animation and Figure 5.2(b) shows the same triangle in the last frame of animation. Given the coordinates of these two triangles, we can interpolate the shape of frame in time  $t$  using the following equations:

$$\begin{cases} x(t) = x^{nt} + t \times \frac{x^{hp} - x^{nt}}{f} \\ y(t) = y^{nt} + t \times \frac{y^{hp} - y^{nt}}{f} \end{cases}, 0 \leq t \leq f \quad (5.2)$$

Where  $f$  is the number of frames in the animation.

In our model, we always use the Neutral face as the start frame. In this example we assume that the last frame shows the triangle in the happy face, which is the reason we use  $x^{hp}$  and  $y^{hp}$  to refer to the coordinates of the points in the last frame.

Now let us suppose that the change from neutral to happy face is originated from the happiness weight of sentence  $s$ . Therefore, we can calculate the position of the vertices of the triangle in the last frame by applying the following equation on the positions of the vertices in  $FACE^{hp}$ . In this equation  $Mem_{hp}$  is calculated using equation 4.2 and  $xmax^{hp}$  is the x coordinate of vertex  $A$  in  $FACE_{hp}$ .

$$x^{hp} = x^{nt} + Mem_{hp}(s) \times (xmax^{hp} - x^{nt}) \quad (5.3)$$

In general, the last frame of the animation might be a blend of the all the emotions. To blend all of the emotions together we sum  $Mem_{hp}(s) \times (xmax^{hp} - x^{nt})$  over all of the six classes of emotions. In case of using a face model instead of triangle, we can rewrite equation 5.3 for vertex  $k^{th}$  in the following form:

$$x_k^\sigma = x^{nt} + \sum_{\sigma} [Mem_{\sigma}(s) \times (xmax_k^\sigma - x_k^{nt})], \sigma \in \{hp, sd, fr, dg, sp, ag\} \quad (5.4)$$

Where  $xmax_k^\sigma$  and  $x_k^{nt}$  are the x coordinate of  $k^{th}$  vertex in  $FACE^\sigma$  and Neutral face respectively as shown in equation 5.1. Using the same approach, we can write the following equations for  $y$  and  $z$  coordinates.

$$y_k^\sigma = y^{nt} + \sum_{\sigma} [Mem_{\sigma}(s) \times (ymax_k^\sigma - y_k^{nt})] \quad (5.5)$$



$$z_k^\sigma = z_k^{nt} + \sum_{\sigma} [Mem_{\sigma}(s) \times (zmax_k^\sigma - z_k^{nt})] \quad (5.6)$$

Using equations 5.1, 5.2 and 5.4 to 5.6, we can generate *NEWFace* for  $t^{th}$  frame of the animation with respect to the emotional weights obtained by processing sentence  $s$ .

$$NEWFace(t) = FACE^{nt} + \frac{1}{f} \times t \times \sum_{\sigma} [Mem_{\sigma}(s) \times (FACE^{\sigma} - FACE^{nt})] \quad (5.7)$$

where  $0 \leq t \leq f, \sigma \in \{hp, sd, fr, dg, sp, ag\}, f = \#frames \text{ in animation}$

## 5.2 Results and Discussion

The basic shapes used in our experiments (Figure 5.5, are rendered using FaceGen Modeller software [2]. The neutral head which is used as the base face is shown in figure 5.4.



(a) Skin.

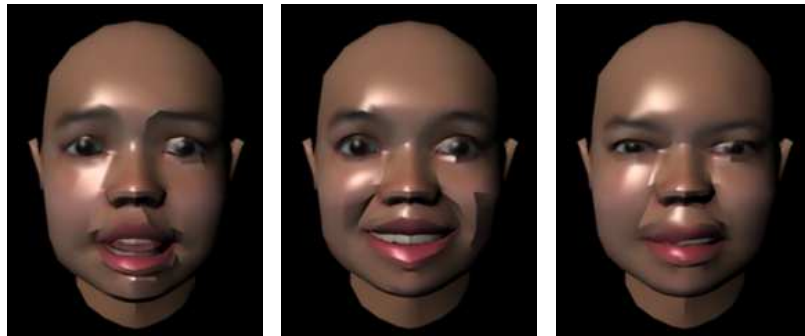


(b) Eyes, teeth, tongue and sock.

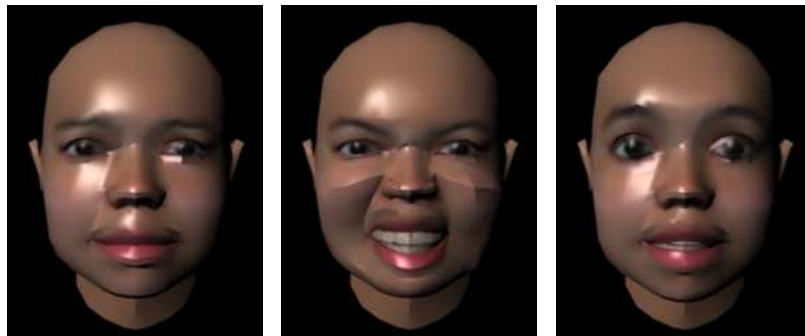
Figure 5.3: Static and dynamic parts of 3D face model.



Figure 5.4: Neutral face( $FACE^{nt}$ ) used as the base face in the experiment.



(a) Fear( $FACE^{fr}$ ). (b) Happiness( $FACE^{hp}$ ). (c) Disgust( $FACE^{dg}$ ).



(d) Sadness( $FACE^{sd}$ ). (e) Anger( $FACE^{ag}$ ). (f) Surprise( $FACE^{sp}$ ).

Figure 5.5: Basic shapes used for the experiment.

The head model is composed of 7 main parts: skin, eyes(left and right), sock, tongue and teeth (upper and lower). The animation parameters (weights) are applied to skin, teeth, sock and tongue whereas eyes are static. The whole model is composed of 1802 triangles and 981 vertices. The model is shown in Figure 5.3.

The interpolation of new faces are done based on equation 5.7. In Figure 5.6 and Figure 5.7 the results of interpolation algorithm are shown for surprise and disgust emotions respectively. In these two figure, the rightmost and leftmost faces are basic shaped and the three interval faces are rendered using our proposed algorithm. The following sets of parameters are used to render these two sets of images:

$$f = 90, t = \{0, 25, 50, 75, 90\}$$

$$\left\{ \begin{array}{l} Mem_{sp} = 1, Mem_{\sigma} = 0, \sigma \in \{hp, sd, fr, dg, ag\} \quad Figure5.6 \\ Mem_{dg} = 1, Mem_{\sigma} = 0, \sigma \in \{hp, sd, fr, sp, ag\} \quad Figure5.7 \end{array} \right.$$

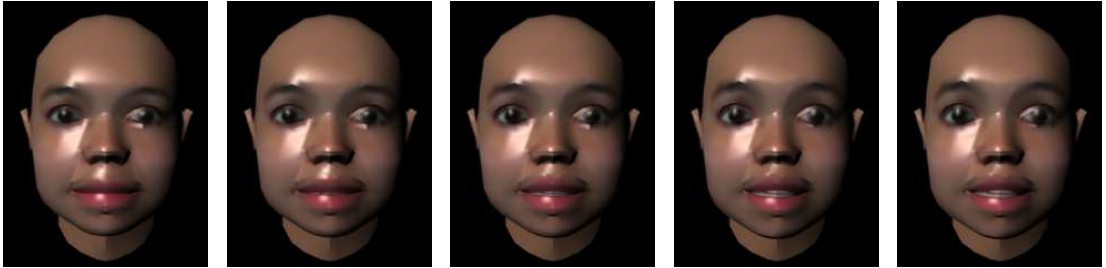


Figure 5.6: Interpolation of Surprise face from neutral face(left) to maximum-surprise-face(right).

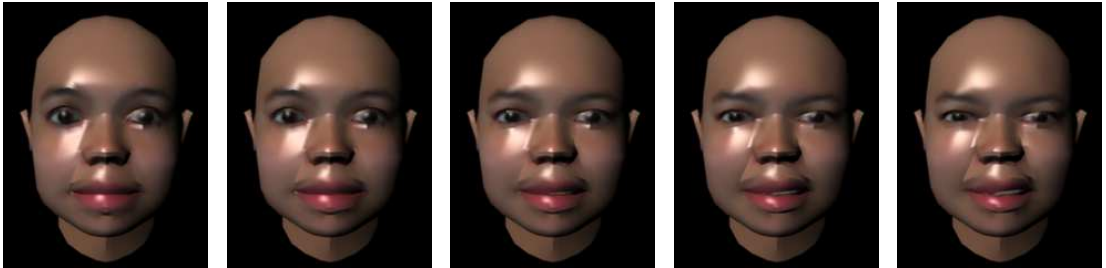


Figure 5.7: Interpolation of Disgust face from neutral face(left) to maximum-disgust-face(right).

Figure 5.8 shows the results of blending two basic shapes. For each set of images, the first two faces show the basic shapes and the third face is the new

face rendered using equation 5.7. For the image sets shown in this figure, the following parameter are used respectively:

$$f = 90, t = 45, \begin{cases} Mem_{fr} = 0.5, Mem_{hp} = 0.5, Mem_{\sigma} = 0, \sigma \in \{sd, dg, sp, ag\} \\ Mem_{dg} = 0.5, Mem_{hp} = 0.5, Mem_{\sigma} = 0, \sigma \in \{fr, sd, sp, ag\} \\ Mem_{sd} = 0.5, Mem_{sp} = 0.5, Mem_{\sigma} = 0, \sigma \in \{fr, dg, hp, ag\} \\ Mem_{hp} = 0.5, Mem_{sp} = 0.5, Mem_{\sigma} = 0, \sigma \in \{sd, dg, fr, ag\} \end{cases}$$

The animation module works well in many cases. However, to investigate more about the quality of the animations generated with this system, we tried different parameters and we found out that this module might render deformed images while using heavy blends. It means that, while rendering the new face, if we blend many basic shapes (usually more than three basic shapes) together to generate the new face, the results may not look very well. This deformation is more obvious when the new face is generated with blending emotions that cause very different effects on the face, for example blending happy, disgust and surprise faces together. We call this problem *over-animated face* and some of the samples are shown in Figure 5.9.

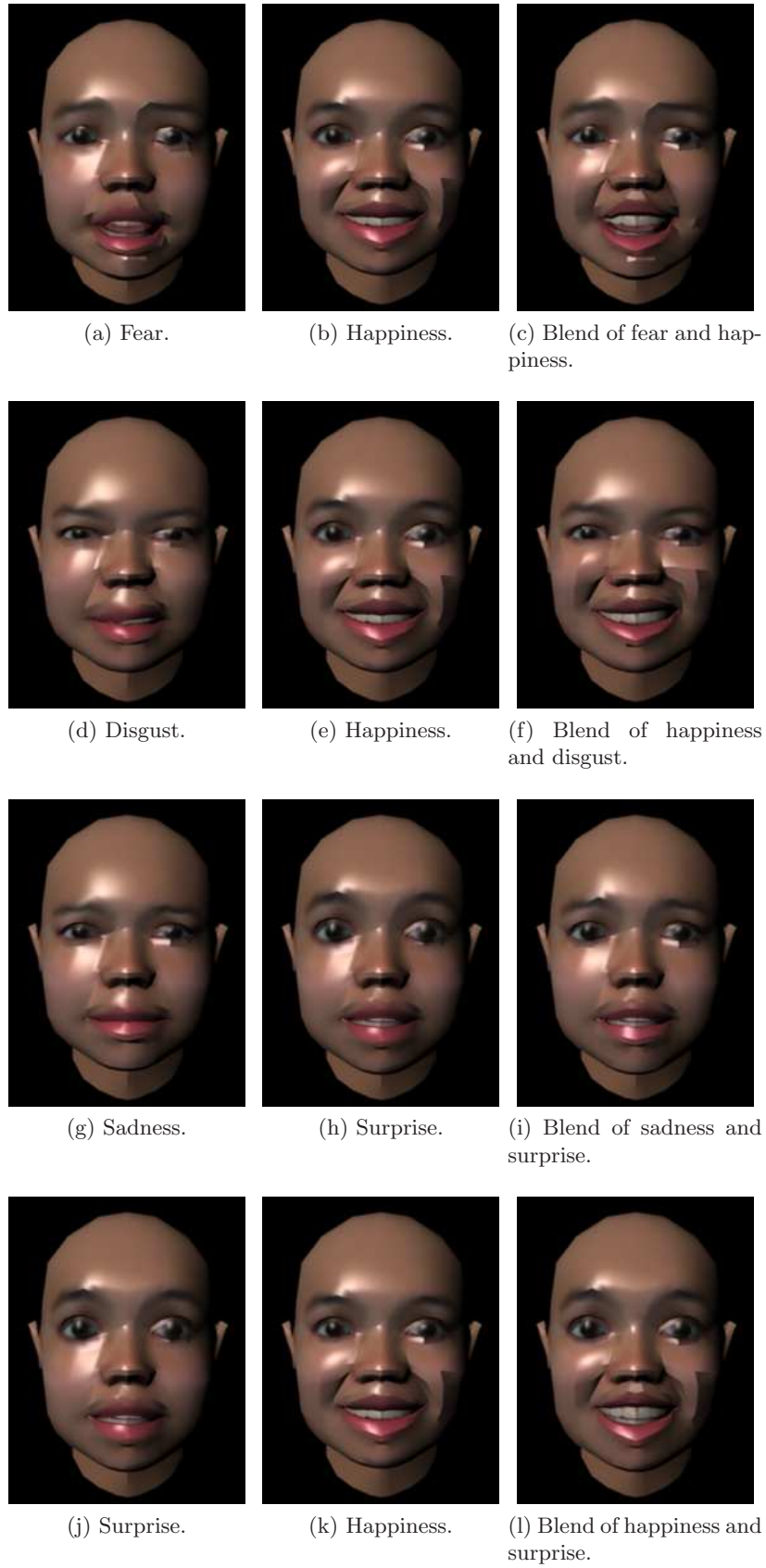
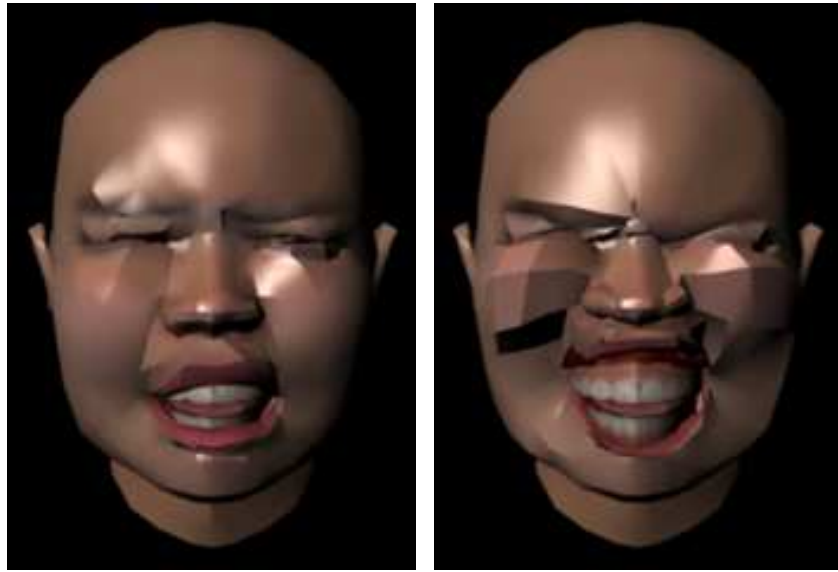


Figure 5.8: Blending of basic faces.



(a) Blending Happiness, Disgust and Surprise  
(b) Blending Happiness, Disgust, Surprise, Fear and Sad

Figure 5.9: Over-animated faces. Some deformed results of animation module.

## Chapter 6

---

# User study

---

In this chapter we explain the on-line user study performed to find out if people find our T2FE interesting and useful. In this user study we examine if the users choose animation of mixed emotion over basic emotion for a given text. Here, we do not perform a test to study if showing facial animation from text can be useful and interesting to potential users. Instead, we refer to Koda's comprehensive analysis of the effects of lifelike characters on computer-mediated communication [48]. Her studies indicate that using life like avatars enhanced with facial expressions, instead of text-only messages, improves user experiences and build enthusiasm toward participation and friendliness in communication.

To find out user's preference about animation of mixed emotions or basic emotions we designed an experiment as follows:

Eight text messages are used for this experiment. All of these messages convey mixed emotions to the reader but they are different in the type and intensity of the emotion which is hidden in the text.

For each text, we rendered two animations to show the emotional meaning of the text. One of the animations, shows the dominant emotion of the text on a sample head while the other one shows the mixed emotions which is a blend of two dominant emotions hidden in that particular text.

The participant in this study are asked to select the animation that better represents the emotion of the text. Therefore, they can choose between the mixed-emotion animation and basic-emotion animation. There is another choice that users can choose if they do not feel any difference between the two animations. In fact, the main goal here is to find out if users prefer to see mixed emotions on the face or not.

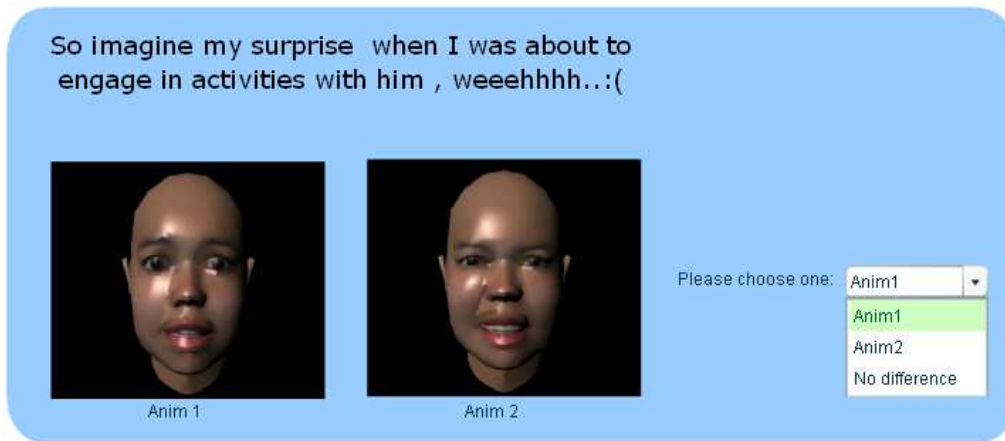


Figure 6.1: A sample entry of user study.

A sample entry of this experiment is shown in Figure 6.1 in which the text has a mixed emotion of Disgust and Surprise. In this figure, *Anim1* illustrates Surprise feeling (dominant emotion) where *Anim2* illustrates mixture of Surprise and Disgust feeling (mixed emotion). In the main user study, we randomize the animations for mixed emotion and dominant emotion between Anim1 and Anim2. This helps to avoid cases where user is eager to consistently select Anim1 (or Anim2) for the whole user study. The complete user study is presented in appendix D.

This user study has 34 participant. Table 6.1 shows the selections of users for each text as well as the emotion type hidden in text.

These results show that 51 percent of the total answers to this study, select mixed emotion while 30 percent select dominant emotion as the better represen-



Text#	emotion	mixed emotion	basic-emotion	No Difference
1	sd-sp	22	10	2
2	dg-hp	12	14	8
3	af-hp	22	8	4
4	sp-dg	10	16	8
5	dg-sd	22	8	4
6	fr-hp	18	8	8
7	sd-sp	18	10	6
8	sd-sp	16	8	10

Table 6.1: Results of user study

tations. The remaining 19 percent do not see any difference between basic and mixed emotions. **This study shows that majority of the participants prefer expressions of the mixed emotions as the representation of emotion in text.**

## Chapter 7

---

# Conclusion

---

This research report introduced the problem of facial expression's animation based on conversation text and its applications in the area of computer human interaction. In this paper, this problem was divided to two main tasks: emotional text classification and facial expression analysis. The significant works done in both areas were discussed and the advantages and disadvantages of main approaches were explained.

For emotional text classification we explored the lexicon based techniques and machine learning techniques. Although lexicon-based techniques benefit from their simplicity and begin free of learning data, their accuracy is not as good as machine learning methods. Among different machine learning methods, the best choice for text classification is proven to be Support Vector Machines (SVM) which guarantees the best performance. Based on these statements and our experiments of text classification techniques, in our T2FE system we used SVM as the core of our text classification task. In our text classification experiment, the fuzzy classification concepts were merged with SVM to build a fuzzy text classifier which is able to classify text into basic classes of emotion. The overall accuracy of our text classifier is 79.58%.

Facial expression animation is also reviewed in this paper and different ap-

proaches based on the traditional methods, sample-based methods, parametric methods and parameter control methods were discussed. Based on this survey, many of the works done in the area of facial animation use standard techniques such as MPEG4 animation and Facial Action Coding system to generate human-like facial movement although the results are not very realistic. While reviewing different facial animation systems, we noticed that most of the existing works focus on animating basic facial animation and mixed-emotions animation are not widely studied and doing experiments in the area of animating face using mixed classes of emotion looks as an interesting and novel work.

We proposed a facial expression animation system for mixed emotions which is able to render animations based on blending expressions of basic classes of emotion. The implementation of this system and the results of this system were fully described in the relevant chapter.

A user study is also conducted to estimate if the potential users of our T2FE system find rendered animations effective and useful. This study was designed with special attention to the difference between animation of basic and mixed emotions. **The results of this study showed that the majority of the participants in this experiment selected the expression of mixed emotions as a better choice for representation of emotion in the text.**

---

# Bibliography

---

- [1] Cslu toolkit. Website. <http://cslu.cse.ogi.edu/toolkit/>. [cited at p. 27]
- [2] Facegen modeller. Website. <http://www.facegen.com/index.htm/>. [cited at p. 52]
- [3] Wordnet domains. Website. <http://wndomains.itc.it/download.html>. [cited at p. 32]
- [4] K. Aizawa, H. Harashima, and T. Saito. Model-based analysis synthesis image coding(MBASIC) system for a person’s face. *Signal Processing: Image Communication*, 1(2):139–152, 1989. [cited at p. 21]
- [5] I. Albrecht, J. Haber, K. Kahler, M. Schroder, and H.P. Seidel. ” May I talk to you?:-)”-facial animation from text. In *Pacific Conference on Computer Graphics and Applications*, pages 77–86, 2002. [cited at p. 7, 27]
- [6] C.O. Alm, D. Roth, and R. Sproat. Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of HLT/EMNLP 2005*, 2005. [cited at p. 11, 19]
- [7] S. Aman and S. Szpakowicz. Using Roget’s thesaurus for fine-grained emotion recognition. [cited at p. 19]
- [8] S. Aman and S. Szpakowicz. Identifying expressions of emotion in text. *Lecture Notes in Computer Science*, 4629:196–205, 2007. [cited at p. 5, 34]
- [9] C.J. Andreas and F. Timothy. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 743–756, 1997. [cited at p. 26]
- [10] T. Beier and S. Neely. Feature-based image metamorphosis. *ACM SIGGRAPH Computer Graphics*, 26(2):35–42, 1992. [cited at p. 21]
- [11] A. Belz. And now with feeling: developments in emotional language generation. Technical report, Technical Report ITRI-03-21, Information Technology Research Institute, University of Brighton, Brighton, 2003. [cited at p. 1]
- [12] A.L. Berger, V.J. Della Pietra, and S.A. Della Pietra. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71, 1996. [cited at p. 15]

- [13] E. Boiy, P. Hens, K. Deschacht, and M.F. Moens. Automatic sentiment analysis in on-line text. In *Proceedings of the 11th International Conference on Electronic Publishing, Openness in Digital Publishing: Awareness, Discovery & Access*, 2007. [cited at p. 19]
- [14] AC Boucouvalas, Z. Xu, and D. John. Expressive image generator for an emotion extraction engine. *People and Computers*, pages 367–382, 2004. [cited at p. 7, 19]
- [15] D.B. Bracewell, J. Minato, F. Ren, and S. Kuroiwa. Determining the emotion of news articles. *Lecture notes in computer science*, 4114:918, 2006. [cited at p. 19]
- [16] C. Bregler, M. Covell, and M. Slaney. Video Rewrite: driving visual speech with audio. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 353–360. ACM Press/Addison-Wesley Publishing Co. New York, NY, USA, 1997. [cited at p. 22]
- [17] G. Breton, C. Bouville, and D. Pelé. FaceEngine a 3D facial animation engine for real time applications. In *Proceedings of the sixth international conference on 3D Web technology*, pages 15–22. ACM New York, NY, USA, 2001. [cited at p. 25]
- [18] C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998. [cited at p. 16]
- [19] Strapparava C. and Valitutti A. WordNet-Affect: an affective extension of WordNet. In *In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pages 1083–1086, May 2004. [cited at p. 32]
- [20] A. Cardoso-Cachopo and A.L. Oliveira. An empirical comparison of text categorization methods. *Lecture Notes in Computer Science*, pages 183–196, 2003. [cited at p. 42]
- [21] J. Cassell, H.H. Vilhjálmsón, and T. Bickmore. BEAT: the Behavior Expression Animation Toolkit. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 477–486. ACM New York, NY, USA, 2001. [cited at p. 27]
- [22] J. Chai, J. Xiao, and J. Hodgins. Vision-based control of 3 D facial animation. In *Symposium on Computer Animation: Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation: San Diego, California*, volume 26, pages 193–206, 2003. [cited at p. 27]
- [23] N. Chambers, J. Tetreault, and J. Allen. Approaches for automatically tagging affect. In *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, 2004. [cited at p. 19]
- [24] T. Danisman and A. Alpkocak. Feeler: Emotion classification of text using vector space model. *AISB 2008 Symposium on Affective Language in Human and Machine*, 2:53–60, 2008. [cited at p. 19]
- [25] P. Domingos and M. Pazzani. On the optimality of the simple bayesian classifier under zero-one loss. *Machine Learning*, 29(2):103–130, 1997. [cited at p. 14]
- [26] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 974–989, 1999. [cited at p. 3, 20]
- [27] Y. Du and X. Lin. Emotional facial expression model building. *Pattern Recognition Letters*, 24(16):2923–2934, 2003. [cited at p. 26]

- [28] P. Ekman and W.V. Friesen. *The repertoire of nonverbal behavior*. Mouton de Gruyter, 1969. [cited at p. 5]
- [29] P. Ekman and W.V. Friesen. *Unmasking the face: a guide to recognizing emotions from facial clues*. Prentice-Hall, 1975. [cited at p. 4]
- [30] P. Ekman and W.V. Friesen. Manual for the facial action coding system. *Consulting Psychologist*, 1977. [cited at p. 23]
- [31] P. Ekman, W.V. Friesen, J.C. Hager, and A.H. Face. Facial action coding system. *Consulting Psychologists Press*, 1978. [cited at p. 23]
- [32] T. Ezzat, G. Geiger, and T. Poggio. Trainable videorealistic speech animation. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 388–398. ACM Press New York, NY, USA, 2002. [cited at p. 27]
- [33] T. Ezzat and T. Poggio. Facial analysis and synthesis using image-based models. In *International Conference on Automatic Face and Gesture Recognition*, pages 116–121, 1996. [cited at p. 22]
- [34] C. Fellbaum and I. NetLibrary. *WordNet: an electronic lexical database*. MIT Press USA, 1998. [cited at p. 12]
- [35] S. Fitrianie and LJM Rothkrantz. My\_Eliza, a multimodal communication system. In *Proceedings of Euromedia2003*, pages 14–22, 2003. [cited at p. 19]
- [36] S. Fitrianie and LJM Rothkrantz. A text-based synthetic face with emotions. In *Proceedings of Euromedia2006*, pages 28–32, 2006. [cited at p. 5, 7, 19]
- [37] Siska Fitrianie and Leon J. Rothkrantz. The generation of emotional expressions for a text-based dialogue agent. In *TSD '08: Proceedings of the 11th international conference on Text, Speech and Dialogue*, pages 569–576, Berlin, Heidelberg, 2008. Springer-Verlag. [cited at p. 7, 19]
- [38] L. Galavotti, F. Sebastiani, and M. Simi. Experiments on the use of feature selection and negative evidence in automated text categorization. *Lecture notes in computer science*, pages 59–68, 2000. [cited at p. 44]
- [39] V. Hatzivassiloglou and J.M. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th conference on Computational linguistics-Volume 1*, pages 299–305. Association for Computational Linguistics Morristown, NJ, USA, 2000. [cited at p. 11]
- [40] Y. Hu, J. Duan, X. Chen, B. Pei, and R. Lu. A new method for sentiment classification in text retrieval. *Lecture Notes In Computer Science*, 3651:1, 2005. [cited at p. 19]
- [41] C.E. Izard. Emotions and facial expressions: a perspective from differential emotions theory. *The Psychology of Facial Expression*, 1997. [cited at p. 4]
- [42] T. Joachims. Learning to classify text using support vector machines: Methods, theory, and algorithms. *Computational Linguistics*, 29(4). [cited at p. 40]
- [43] T. Joachims. *Text categorization with support vector machines: Learning with many relevant features*. Springer, 1997. [cited at p. 15]
- [44] T. Joachims, C. Nedellec, and C. Roveiro. Text categorization with support vector machines: learning with many relevant. In *Machine Learning: ECML-98 10th European Conference on Machine Learning, Chemnitz, Germany*. Springer, 1998. [cited at p. 40]

- [45] J. Kamps, M. Marx, R.J. Mokken, and M. de Rijke. Using WordNet to measure semantic orientation of adjectives. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, volume 4, pages 1115–1118, 2004. [cited at p. 12]
- [46] A. Khoo, Y. Marom, and M. Albert. Experiments with sentence classification. In *ALTW2006 Australian Language Technology Workshop*, pages 18–25, 2006. [cited at p. 41, 43, 47]
- [47] R.M. Koch, M.H. Gross, and A.A. Bosshard. Emotion editing using finite elements. In *Computer Graphics Forum*, volume 17, pages 295–302. Blackwell Synergy, 1998. [cited at p. 27]
- [48] T. Koda. *Analysis of the Effects of Lifelike Characters on Computer-mediated Communication*. PhD thesis, Kyoto University, 2006. [cited at p. 58]
- [49] S. Kshirsagar, S. Garchery, G. Sannier, and N. Magnenat-Thalmann. Synthetic faces: Analysis and applications. *International Journal of Imaging Systems and Technology*, 13(1):65–73, 2003. [cited at p. 27]
- [50] C. Lee. Emotion recognition for affective user interfaces using natural language dialogs. In *Proceedings of the 16th IEEE International Symposium on Robot & Human Interactive Communication*, pages 798–801, 2007. [cited at p. 19]
- [51] W.S. Lee, M. Escher, G. Sannier, and N. Magnenat-Thalmann. Mpeg-4 compatible faces from orthogonal photos. In *Proc. Computer Animation*, volume 99, pages 186–194, 1999. [cited at p. 25]
- [52] D.D. Lewis. Naive (Bayes) at forty: The independence assumption in information retrieval. *Lecture Notes in Computer Science*, pages 4–18, 1998. [cited at p. 14]
- [53] K.H.Y. Lin and H.H. Chen. Ranking reader emotions using pairwise loss minimization and emotional distribution regression. In *EMNLP*, pages 136–144, 2008. [cited at p. 19]
- [54] K.H.Y. Lin, C. Yang, and H.H. Chen. What emotions do news articles trigger in their readers? In *Annual ACM Conference on Research and Development in Information Retrieval: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, volume 23, pages 733–734, 2007. [cited at p. 19]
- [55] P. Litwinowicz and L. Williams. Animating images with drawings. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 409–412. ACM New York, NY, USA, 1994. [cited at p. 21]
- [56] H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 125–132, 2003. [cited at p. 14, 19]
- [57] Z. Liu, Y. Shan, and Z. Zhang. Expressive expression mapping with ratio images. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 271–276. ACM New York, NY, USA, 2001. [cited at p. 21]
- [58] M.J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1357–1362, 1999. [cited at p. 26]

- [59] N. Mana and F. Pianesi. HMM-based synthesis of emotional facial expressions during speech in synthetic talking heads. In *Proceedings of the 8th international conference on Multimodal interfaces*, pages 380–387. ACM New York, NY, USA, 2006. [cited at p. 5]
- [60] A. Mehrabian. Communication without words. *Communication Theory*, pages 193–200, 2007. [cited at p. 20]
- [61] R. Mihalcea and H. Liu. A corpus-based approach to finding happiness. In *Proceedings of the AAAI Spring Symposium on Computational Approaches to Weblogs*, 2006. [cited at p. 19]
- [62] G. Mishne. Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access*, 2005. [cited at p. 19]
- [63] M. Mulder, A. Nijholt, M. Uyl, and P. Terpstra. A lexical grammatical implementation of affect. *Lecture Notes in Computer Science*, pages 171–178, 2004. [cited at p. 13]
- [64] F. Nasoz, K. Alvarez, C.L. Lisetti, and N. Finkelstein. Emotion recognition from physiological signals using wireless sensors for presence technologies. *Cognition, Technology & Work*, 6(1):4–14, 2004. [cited at p. 10]
- [65] A. Neviarouskaya, H. Prendinger, and M. Ishizuka. Textual affect sensing for sociable and expressive Online communication. *Lecture Notes in Computer Science*, 4738:218, 2007. [cited at p. 10, 19]
- [66] K. Nigam, J. Lafferty, and A. McCallum. Using maximum entropy for text classification. In *IJCAI-99 workshop on machine learning for information filtering*, pages 61–67, 1999. [cited at p. 12, 15]
- [67] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics Morristown, NJ, USA, 2002. [cited at p. 19]
- [68] F.I. Parke. Computer generated animation of faces. In *Proceedings of the ACM annual conference-Volume 1*, pages 451–457. ACM New York, NY, USA, 1972. [cited at p. 21]
- [69] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D.H. Salesin. Synthesizing realistic facial expressions from photographs. In *Computer graphics proceedings, annual conference series*, pages 75–84. Association for Computing Machinery SIGGRAPH, 1998. [cited at p. 3, 21, 22]
- [70] R. Plutchik. *The psychology and biology of emotion*. Harpercollins College Div, 1994. [cited at p. 5]
- [71] A. Raouzaïou, N. Tsapatsoulis, K. Karpouzis, and S. Kollias. Parameterized facial expression synthesis based on Mpeg-4. *EURASIP Journal on Applied Signal Processing*, 10:1021–1038, 2002. [cited at p. 21, 25]
- [72] J. Read. Recognizing affect in text using pointwise-mutual information. *Master’s thesis, University of Sussex*, 2004. [cited at p. 19]
- [73] C.J. RIJSBERGEN. *Information retrieval*. University of Glasgow, 1979. [cited at p. 44]



- [74] L.J.M. Rothkrantz and A. Wojdel. A text based talking face. *Lecture notes in computer science*, pages 327–332, 2000. [cited at p. 19]
- [75] Z. Ruttkay and H. Noot. Animated CharToon faces. In *Proceedings of the 1st international symposium on Non-photorealistic animation and rendering*, pages 91–100. ACM Press New York, NY, USA, 2000. [cited at p. 27]
- [76] KR Scherer and HG Wallbott. Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66(2):310–28, 1994. [cited at p. 32]
- [77] T.W. Sederberg and E. Greenwood. A physically based approach to 2–D shape blending. *ACM SIGGRAPH Computer Graphics*, 26(2):25–34, 1992. [cited at p. 48]
- [78] S.M. Seitz and C.R. Dyer. View morphing. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 21–30. ACM New York, NY, USA, 1996. [cited at p. 22]
- [79] E. Spyropoulou, S. Buchholz, and S. Teufel. Sentence-based emotion classification for text-to-speech. *International Workshop on Computational Aspects of Affectual and Emotional Interaction*, 2008. [cited at p. 19]
- [80] P. Subasic and A. Huettner. Affect analysis of text using fuzzy semantic typing. *Fuzzy Systems, IEEE Transactions on*, 9(4):483–496, 2001. [cited at p. 19]
- [81] L. Tang and H. Liu. Bias analysis in text classification for highly skewed data. In *Proceedings of the Fifth IEEE International Conference on Data Mining*, pages 781–784, 2005. [cited at p. 47]
- [82] P.D. Turney. Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. *Lecture Notes in Computer Science*, pages 491–502, 2001. [cited at p. 12]
- [83] P.D. Turney et al. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, 2002. [cited at p. 11, 14, 19]
- [84] H. Wang and N. Ahuja. Facial expression decomposition. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 958–965, 2003. [cited at p. 21]
- [85] G.M. Weiss and F. Provost. Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research*, 19(2):315–354, 2003. [cited at p. 47]
- [86] C.M. Whissell. The dictionary of affect in language. *Robert Plutchik and Henry Kellerman (Ed.), Emotion: Theory, Research, and Experience*, pages 113–131, 1989. [cited at p. 5, 6]
- [87] I.H. Witten and E. Frank. Data mining: practical machine learning tools and techniques with Java implementations. *ACM SIGMOD Record*, 31(1):76–77, 2002. [cited at p. 42]
- [88] A. Wojdel and LJM Rothkrantz. Parametric generation of facial expressions based on FACS. In *Computer Graphics Forum*, volume 24, pages 743–757. Blackwell Synergy, 2005. [cited at p. 27]

- [89] C.H. Wu, Z.J. Chuang, and Y.C. Lin. Emotion recognition from text using semantic labels and separable mixture models. *ACM Transactions on Asian Language Information Processing (TALIP)*, 5(2):165–183, 2006. [cited at p. 11]
- [90] Z. Xu, D. John, and AC Boucouvalas. Expressive image generation: Towards expressive Internet communications. *Journal of Visual Languages and Computing*, 17(5):445–465, 2006. [cited at p. 7, 19]
- [91] Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. Emotion classification using web blog corpora. In *WI '07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 275–278, Washington, DC, USA, 2007. IEEE Computer Society. [cited at p. 19]
- [92] Y. Yang and X. Liu. A re-examination of text categorization methods. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 42–49. ACM New York, NY, USA, 1999. [cited at p. 17, 42]
- [93] L. Yin and A. Basu. Generating realistic facial expressions with wrinkles for model-based coding. *Computer Vision and Image Understanding*, 84(2):201–240, 2001. [cited at p. 25]
- [94] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965. [cited at p. 40]
- [95] Y. Zhang, E.C. Prakash, and E. Sung. Efficient modeling of an anatomy-based face and fast 3d facial expression synthesis. In *Computer Graphics Forum*, volume 22, pages 159–169. Blackwell Synergy, 2003. [cited at p. 21]
- [96] X. Zhe and AC Boucouvalas. Text-to-emotion engine for real time internet communication. In *Proceedings of International Symposium on Communication Systems, Networks and DSPs*, pages 164–168, 2002. [cited at p. 5, 7]
- [97] C. Zhou and X. Lin. Facial expressional image synthesis controlled by emotional parameters. *Pattern Recognition Letters*, 26(16):2611–2627, 2005. [cited at p. 7, 21]

# Appendices

## Appendix A

# Emoticons and abbreviations database

Text	GW*	hp	sd	ag	fr	dg	sp	Text	GW*	hp	sd	ag	fr	dg	sp
:)	1	1	0	0	0	0	0	:-j	1	0	1	0	0	0	0
:-)	1	1	0	0	0	0	0	:P	1	0.5	0	0	0	0	0
=)	1	1	0	0	0	0	0	:-P	1	0.5	0	0	0	0	0
(:	1	1	0	0	0	0	0	:p	1	0.5	0	0	0	0	0
(-:	1	1	0	0	0	0	0	:-p	1	0.5	0	0	0	0	0
:D	1	1	0	0	0	0	0	:d	1	0.5	0	0	0	0	0
:-D	1	1	0	0	0	0	0	:'(	1	0	1	0	0	0	0
;	1	1	0	0	0	0	0	:-(	1	0	1	0	0	0	0
;-)	1	1	0	0	0	0	0	;(	1	0	1	0	0	0	0
;D	1	1	0	0	0	0	0	:(	1	0	1	0	0	0	0
;-D	1	1	0	0	0	0	0	:c	1	0	1	0	0	0	0
>:D<	1	1	0	0	0	0	0	:-c	1	0	1	0	0	0	0
/:)	1	1	0	0	0	0	0	:-S	1	0	0	0	0.5	0.5	0
:)]	1	1	0	0	0	0	0	:S	1	0	0	0	0.5	0.5	0
:]	1	1	0	0	0	0	0	:-s	1	0	0	0	0.5	0.5	0
:-]	1	1	0	0	0	0	0	:s	1	0	0	0	0.5	0.5	0
;;)	1	1	0	0	0	0	0	:-/	1	0	0	0	0.5	0.5	0
:*	1	1	0	0	0	0	0	:/	1	0	0	0	0.5	0.5	0
:-*	1	1	0	0	0	0	0	:	1	0	0	0	0.5	0.5	0
:x	1	1	0	0	0	0	0	x.x	1	0	0	0	1	0.5	0
:-x	1	1	0	0	0	0	0	:O	1	0	0	0	0	0	1
:>	1	0.5	0	0	0	0	0.5	:-O	1	0	0	0	0	0	1
:>	1	0.5	0	0	0	0	0	:0	1	0	0	0	0	0	1
B)	1	1	0	0	0	0	0	:-0	1	0	0	0	0	0	1
B-)	1	1	0	0	0	0	0	!:	1	0	0	0	0	0	1
8)	1	1	0	0	0	0	0	x(	1	0	1	1	0	0	0
8-)	1	1	0	0	0	0	0	xx(	1	0	1	1	0	0	0
8>	1	1	0	0	0	0	0	:-@	1	0	0	1	0	0.5	0
8->	1	1	0	0	0	0	0	:@	1	0	0	1	0	0.5	0
O:-)	1	1	0	0	0	0	0	>:(	1	0	0.5	1	0	0.5	0
>:)	1	1	0	0.5	0	0	0	>:-(	1	0	0.5	1	0	0.5	0
>:-)	1	1	0	0.5	0	0	0	8o	1	0	0	1	0	0.5	0

– Continue on the next page

\*General weight

– Continued from previous page.

Text	GW*	hp	sd	ag	fr	dg	sp	Text	GW*	hp	sd	ag	fr	dg	sp
I-)	1	0.5	0	0	0	0	0	>:0	1	0	0	1	0	0	1
@:-)	1	1	0	0	0	0	0	>:-0	1	0	0	1	0	0	1
@}->-	1	1	0	0	0	0	0	:&	1	0	0	0	0	0.5	0
@>->-	1	1	0	0	0	0	0	:-&	1	0	0	0	0	0.5	0
(-)	1	1	0	0	0	0	0	h8	1	1	0	0	0	0	0
(@)	1	0	0	0	0	0	1	<3	1	1	0	0	0	0	0
(@_@)	1	0	0	0	0	0	1	{}	1	1	0	0	0	0	0
.-)	1	1	0	0	0	0	0	omg	1	0	0	0	0	0	1
:O)	1	1	0	0	0	0	0	oic	1	0	0	0	0	0	1
8-}	1	1	0	0	0	0	0	)	1	1	0	0	0	0	0
<:-P	1	1	0	0	0	0	0	-	1	1	0	0	0	0	0
=D>	1	1	0	0	0	0	0	( )	1	0	1	0	0.5	0.5	0
@-)	1	1	0	0	0	0	0	-( )	1	0	1	0	0.5	0.5	0
\m/	1	1	0	0	0	0	0	]:( )	1	1	0	0	0	0	0
<):)	1	1	0	0	0	0	0	rofl	1	1	0	0	0	0	0
:)	1	1	0	0	0	0	0	lol	1	1	0	0	0	0	0
:(	1	0	1	0	0	0	0	:roff:	1	1	0	0	0	0	0
:-( )	1	0	1	0	0	0	0	:lol:	1	1	0	0	0	0	0
8( )	1	0	1	0	0.5	0.5	0	wow	1	0	0	0	0	0	1
8-( )	1	0	1	0	0.5	0.5	0	:wow:	1	0	0	0	0	0	1
=( )	1	0	1	0	0	0	0	wow!	1	0	0	0	0	0	1
:<	1	0	1	0	0	0	0								

---

\*General weight

## Appendix B

---

# List of selected features for text classification

---

! !! !!! !!!! :( :) :] :D :P :p ;) ;D ?) ?? :^ ^\_^ AAAHHH  
AARRRGHH adored afraid aggravated agitated agonized alarmed  
amazed amused angry anguished annoyed anxious aroused ashamed  
attracted awful awkward BAD bashful bitter bored brave calm  
caring cautious cheerful comfortable compassionate concerned confused  
contempting critical curious cynical damn DAMNNN delighted depressed  
desired disappointed discouraged disgusted disliked dismayed distressed  
disturbed doesn't don't down dreadful eager ecstatic edgy elated  
embarrassed encouraged engagement enthused excited excited exhausted  
exhausted exhilarated fatigued fearful fidgety fond frightened frustrated  
funny furious glad gleeful gloomy greedy griefstricken guilty HAHAHA  
happy hassled hateful helpless hesitant homesick hopeful hopeless  
horrible humiliated hurt hysterical impatient indifferent infatuated  
inferior insecure insulted irate irked irritated isolated jealous jittery  
jolly joyous lazy leery liked loathe LOL lonely loved loved loving mad  
mean melancholic mischievous miserable moody mortified neglected  
nervous nice NOWAY numb optimistic overwhelmed panicky passionate  
patient pessimistic pleased proud puzzled queasy rageful raptured  
regretful rejected relieved reluctant restless ridiculous rushed sad safe  
satisfied scared scornful secure sensitive shaky shocked shy silly sleepy  
smile sorry stressed surprised suspicious tender tense terrified tired  
tired troubled uncomfortable uneasy unhappy unsafe unsettled upset  
victorious warm weary well why woeful wonderful worried wrathful

## Appendix C

---

# Facial Action Coding (FAC) System

---

Group	# of FAPs
Visems and expressions	2
Jaw, Chin, inner lower lip, corner lips, middle lip	16
Eyeballs, pupils, eyelids	12
Eyebrow	8
Cheeks	4
Tongue	5
Head rotation	3
Outer lip positions	10
Nose	4
Ears	4

Table C.1: FAP groups.

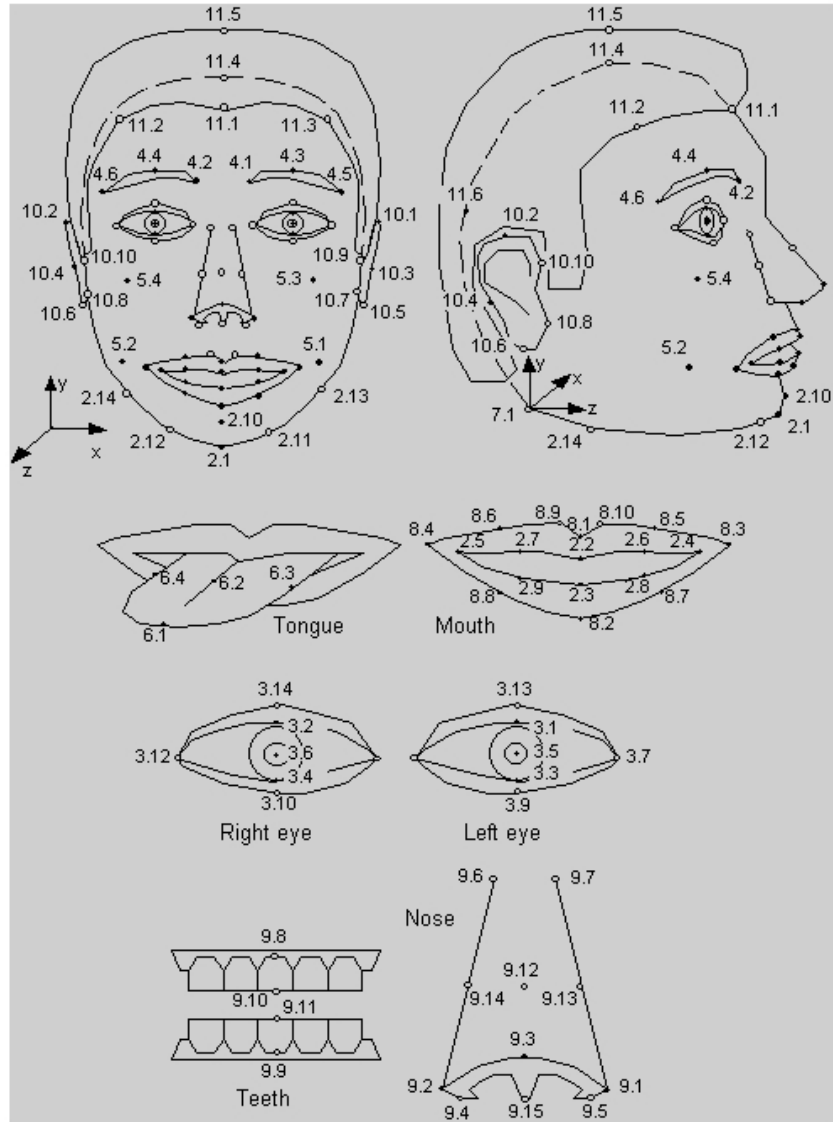



Figure C.1: Feature points defined in FAC system.



# Appendix D


## User Study

1) Hey, I'm shocked by the news, is it true that u had a car accident yesterday?!!




Avatar 1 Avatar 2 Please choose one Avatar1 Avatar2 No difference

2) I hate it when people cant decide, come on! U eager to join us and pay for the tickets?



Avatar 1 Avatar 2 Please choose one Avatar1

3) I told u , I'm afraid of height, but maybe I'll try flyer in the weekend. Mina told me it's so much fun



Avatar 1 Avatar 2 Please choose one Avatar1

4) So imagine my surprise when I was about to engage in activities with him , weeehhhh...:(



Anim 1



Anim 2

Please choose one:

5) I was pretty bummed out, because it meant they put a hold on my cheque, I hate when they do that, what's the bloody point?



Anim 1



Anim 2

Please choose one:

6) So I kept walking and then the other dog that crossed the street came back running at me and I said "OHH MY GOD!!!". funny yeah??? Now u can laugh!



Anim 1



Anim 2

Please choose one:

7) Really, what did make me think that??? :(



Anim 1



Anim 2

Please choose one:

8) All the time I was thinking what a fabulous gift for him on his birthday!! But now, he says he does not like pets at all!! what is wrong with him?



Anim 1



Anim 2

Please choose one: