Emotion Extraction Engine: Expressive Image generator

Xu Zhe, David John and Anthony C. Boucouvalas

Multimedia Communications Research Group,
School of Design, Engineering and Computing,
Bournemouth University,
Fern Barrow, Poole,
Dorset, BH12 5BB, UK.

{zxu, djohn, tboucouv}@bournemouth.ac.uk

Abstract:
In this paper, we present the latest development of an emotion extraction engine used for real-time internet text communication. Real time expressive communication is important as it provides aspects of the visual clues that are present in face-to-face interaction not available in ordinary text-based communications. In former papers, we demonstrated a text-to-emotion engine that can analyse the emotional content present in a real-time chat environment and can deliver the emotional parameters necessary to invoke an appropriate expressive image. In this paper, we present a quick and user-friendly method to generate the necessary expressive images. Images can be generated for six emotion categories from one original neutral image. In each category, three different emotion intensities can be achieved. Users only need to provide a single default image, define six control points and two control shapes to generate all the images. This paper also presents the preliminary findings from a series of experiments that have been carried out to test the efficiency of our expressive image generator.

Keywords: Emotion, Expression, Warping, Morphing.

1 Introduction

Emotion research can be traced back to 500BCE, Heraclitus (the Greek philosopher of the late 6th century BCE) who concluded that the emotional state is characterised by a mixture of expressive body parameters [1].

Communication of emotions through facial expression is an active research area. Darwin and other recent contemporary researchers such as Buck [2] and K. Dautenhahn [3] have shown great attention to this area.

With the rapid development of Computing and Internet, research on of emotions is associated with research on computer networks, for example the "Affective Communication" [16], "Computational Emotions" [17] projects, and our emotion extraction engine.

The emotion extraction engine, can analyse user input in the form of text sentences, has been developed and is presented in detail in [4]. Sentences are analysed in real-time for emotive content and the emotion is represented by an appropriate facial expression displayed automatically. In paper [5], a series of experiments that were carried out to test the performance and the effectiveness of the engine are described.

The experiment results in [5] show that most users prefer an online chatting interface that includes text with expressive images than text alone. The challenge is how to generate expressive images for all users in a fast and friendly way without individuals having to resort in taking many expressive photographs.

To solve this challenge, we report here the development of the expressive image generator discussed in detail in this paper.

Finally, we have carried out tests for assessing the performance of the expressive image generator**.** The first test assesses the correct recognition of the expressive images without any text information; the second test assesses the effective recognition of the expressive images including text information.

This paper is organised as follows. In section 2 the emotion extraction engine is reviewed. In section 3 the background knowledge of image generation is discussed. In section 4 the motion image generator is described in detail. In section 5 the test strategy for the expressive image generator is given. Section 6 illustrates possible applications using the engine. Finally in section 6 conclusions are given.

2 Emotion extraction engine

An emotion extraction engine that can analyse user input in the form of text sentences has been developed. When emotional content is detected the engine will send the parameters needed for selecting the expressive images across the network. When receiving side engine receives the parameters, the corresponding expressive images will be selected and displayed. In this way, the transmissions of images are avoided since the only data transmitted over the network is the text parameters. As a result the bandwidth requirement is extremely low.

The emotion extraction engine includes two sub-systems: the emotion analysis system and the expressive image generator. The emotion analysis system includes three parts: input text analysis, tagging system and parser. A general description is

presented here. For detail information, please refer to [4] and [5].

The working flow of the emotion analysis system is shown in figure 1.

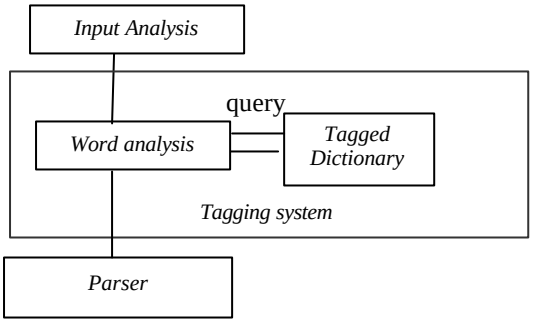

Figure 1: The working flow of the emotion analysis system

- ***Input analysis function***

Users' text input is sent from the user interface to *the input analysis function*. The engine analyses only one sentence at a time. The *input analysis function* will replace all punctuation with pre-defined characters and send the analysed sentences to the *tagging system.*

- ***The tagging system***

The w*ord analysis function* splits the sentences from *input analysis function* into words and searches *the tagged dictionary* to find the corresponding tag category. The outputs include the *word category* and the *emotional tag*.

Daily communications involve about two thousand words [6]. In order to identify the words, a special designed dictionary was set up. In this project, a database containing 16400 words was used. The database includes three fields: *word*, *category* and *emotional tag*. The *category* field contains the corresponding word category (noun, verb, adjective etc) and the *emotional tag* field describes whether that word belongs to one of the six emotion types.

Unlike tagging methods used in some existing systems, e.g., BNC [7] and Brown Corpus [8], the entire word is appropriately tagged in the dictionary in order to keep the response time to a minimum.

- ***Parser***

Receiving the output from the *tagging system*, the *parser* will try to identify the emotion content. The *parser*'s analysis is accomplished through the use of rewrite rules and tree representations [9]. According to pre-defined rules, the *parser* will search for the current emotional words, the person to whom the emotional words refer to and the intensity of the emotional words. The *parser*'s outputs are the emotional parameters and are sent across the network.

- **Expressive images**

The output from the parser will be sent out through the network to related users. When receiving the output, the corresponding expressive images are selected from a database and displayed.

3 Expressive image generation background

- **Universal Expressions**

Research in facial expression has concluded that there are six universal categories of facial expressions that are recognised across cultures [10]. The categories are *happiness*, *sadness*, *anger*, *fear*, *disgust* and *surprise*. Within each of the categories, a wide range of expression intensity and variation of detailed expressions exists. Our engine requires a set of expressive images representing the emotions.

- **Existing Expression image algorithms**

To generate expressive images a number of algorithms and theories have been developed. The most well known algorithms and software include the "Facial Action Coding System" [11], "FaceWorks"[12] and "Synthesising Realistic Facial Expressions" [13].

The major challenges for these algorithms and software are the need to reduce computation time and the development of a user-friendly interface. For example, some programs may require a database to hold different faces and compare the new face with the existing samples. Some programs may require users to create a complex model at the face that covers almost all the edges of the face. It is difficult for users to identify all the points required by the complex model accurately and it is not always possible to generate images to be used in real time communication systems.

Instead of using a complex model and a heavy burden computation algorithm, we developed a simple algorithm based on image warping and image morphing. Users only need to choose six control points and draw two control areas on the original neutral image. Then six images that correspondingly belong to the six expression categories will automatically be generated and stored for future use. In each emotion category, three different emotion intensities are achieved. In total eighteen expressive images are generated from a default image.

- **Image warping**

Image warping is the act of distorting a source image into a destination image according to a transformation between source space (u,v) and destination space (x,y) [14]. The transformation function f() describes the destination (x,y) for every location (u,v) in the source. The function is presented as follows.

$$x=f_x(u,v)$$
$$y=f_y(u,v)$$

The function f() will correspondingly stretch or compress an area defined by source space and destination space.

To apply warping to an image, we may apply the transformation function f() to each pixel. The pseudocode is shown in figure 2.

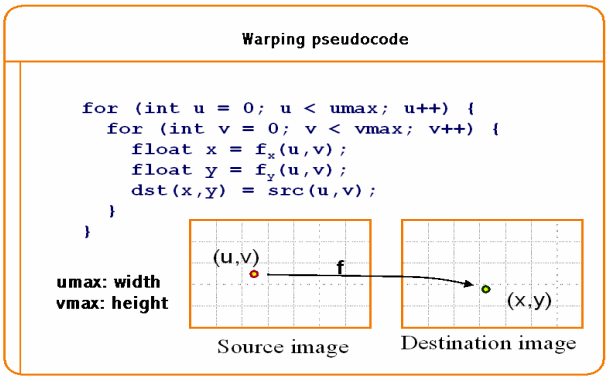

Figure 2: the warping pseudocode

According to Ekman's Facial action Coding System [11] and PARKE's Computer Facial Animation [10], a limited number of muscles on the faces are responsible for the expression generations. For example only four muscles contribute to expressing sadness. It is not necessary to implement warping for facial expression generation across the whole image since most parts of the face remain unchanged.

- **Image morphing**

Image morphing is an image processing technique used for the metamorphosis from one image to another. The usual morphing technique is to generate a sequence of intermediate images. Those images put together with the original images would represent the change from first image to the last [15].

4 expressive image generator

To generate expressive images, users need to upload a neutral face image to the expressive image generator. The generator is based on the local area warping and morphing technology discussed above. The structure of the expressive image generator is shown in figure 3.

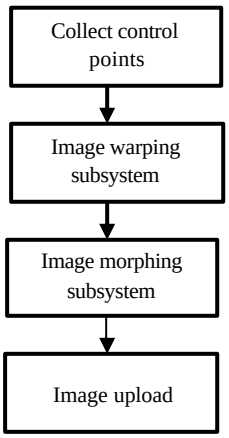

Figure 3: Expressive image generator structure.

- **Emotion categories**

In reference [10] and [11], 64 action units are defined that are responsible for the movement of the face. In this paper, we limit the investigation to the facial action units required to create expressive images and define the following rules to generate the expressive images.

Happiness: The eyebrows are relaxed, the mouth is wide with the corners pulled up toward the ears.
Sadness: The inner portions of the eyebrows are pulled up above the upper eyelid and the mouth is relaxed.
Anger: The eyebrows are pulled downward and together. The mouth is closed with the upper lip slightly compressed or squared off.
Disgust: The middle eyebrows are pulled upward and the mouth is slightly opened with the upper lip squared off.
Fear: Eyebrows are raised and pulled together. The eyebrows are bent upward. The mouth may be dropped slightly open.
Surprise: The eyebrows are raised up, the upper eyelids are opened and the mouth is dropped open [10].
For each expression category, three different emotion intensities are calculated. Based on the above rules, the movement of mouth, eyebrows and lips are quantified. From low intensity to high intensity, the movement is proportionately enlarged also.

- **Control Points**

To generate different expressions, start points and finish points are required. Users only need to select the start points. The finish points are calculated automatically based on the user's selection.

After uploading the neutral image to the system, users will be guided to select six start points and three control areas. The six start points include: left corner of the mouth (LM), right corner of the mouth (RM), outer edge of the left eyebrow (LOE), inner edge of the left eyebrow (LIE), outer edge of the right eyebrow (ROE) and inner edge of the right eyebrow (RIE).

The three control areas include the outer edge of the lips, and the inner edge of the eyelids for the left eye and right eye. These areas will be used in image morphing subsystem.

These parameters will be sent to the image warping subsystem to generate the expressive images.

- **Image Warping subsystem**

The image warping subsystem implements the model based warping. Instead of interactive manipulation, the model based warping subsystem has the advantages of accuracy and speed.

To generate the model, the facial action coding system and several different expressive images from different persons were analysed. We have named the kernel that generates the images the "expression model mask". The mask is constituted by two sets of points. The first set are the points selected by users, which is called the start points. The second set are the control points of the start points, which is called the finish points. The values of the finish points are calculated relative to the start points. By applying the masks to images of individual faces, corresponding expressions can be generated.

The finish points are calculated as the start points plus an integer value that depends on the expression being

generated. The function to calculate finish points is shown below.

$$Finish.x = Start.x + a$$
$$Finish.y = Start.y + b$$

With user-selected points and the calculated finish points, The image warping subsystem can generate intermediate images. These images are sent to the image morphing subsystem.

- **Image morphing subsystem**

The weakness of the image warping subsystem is that it can not generate new pixels, for example to open the mouth and widen the eyes. The emotion *fear* and *surprise* require an opened mouth and widen eyes. The warping subsystem can not achieve this affect. To solve this problem, we implemented an image morphing subsystem.

First several different images of mouths and eyes were chosen from *surprise* and *fear* images. When the morphing subsystem receives the intermediate images from the warping subsystem, it will replace the mouth and eyes for the images belonging to the appropriate emotion categories. The mouth and eyes from pre-chosen images are pasted to the images. To remove the gaps between the pre-prepared images and the intermediate images, a gaussian blur operation is applied to the edges.

Examples of expressive images generated by our expressive image generator are shown in figure 4.

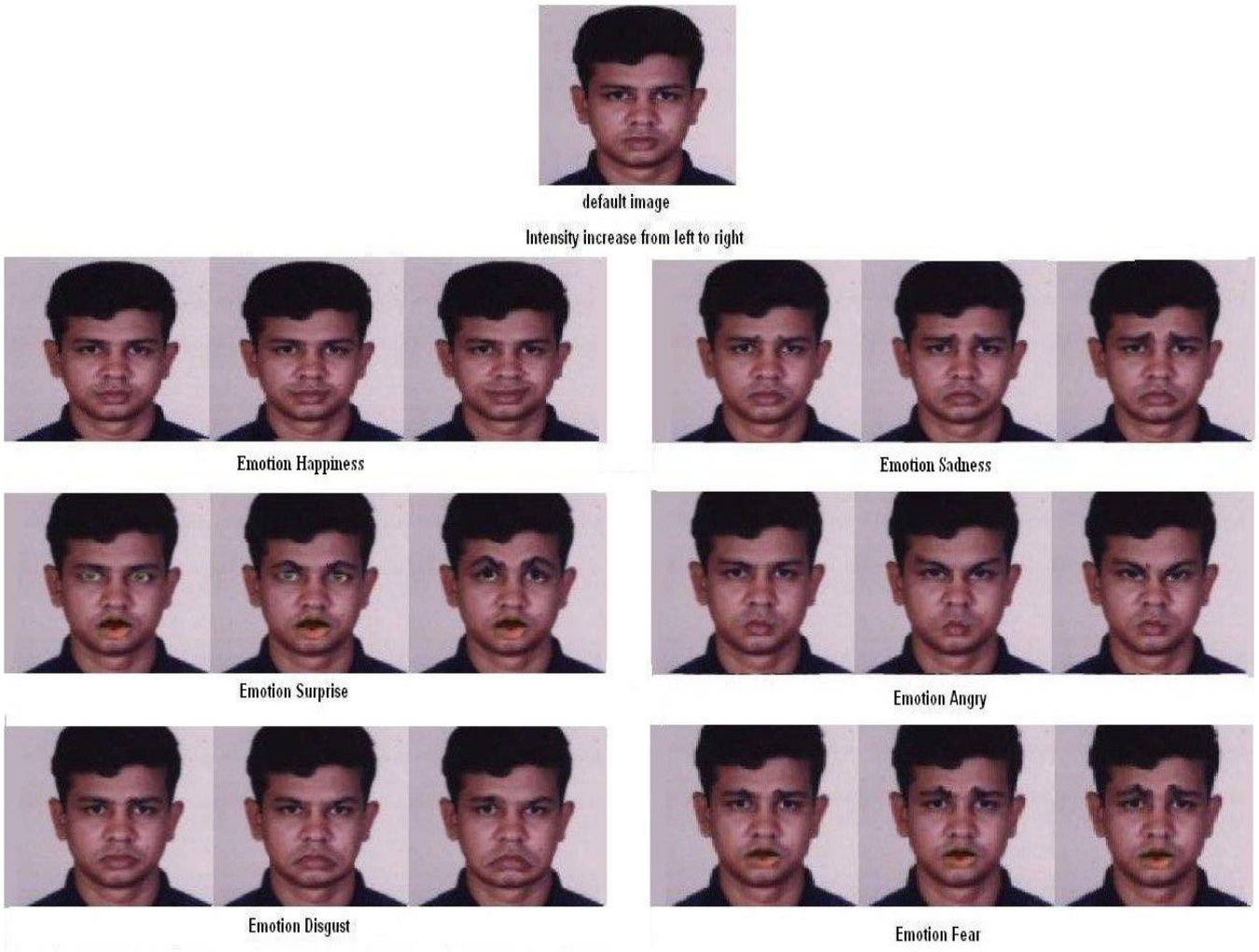

Figure 4: Generated expressive images

5 Experiment strategy

To test the correctness and effectiveness of the expressive image generator, two types of experiments were carried out. The first was an image only test, in which the test subjects view the generated images in isolation. The second type is the image plus text test, in which users view the images with contextual information.

Null hypothesis and chi-square test

The null hypothesis predicted that each of the six expression categories would be recognised only at the random level, at 17%. A recognition rate significantly above this would disprove the null hypothesis and prove its inverse, namely, that the images were recognised at a significantly greater rate than chance, and therefore that expressive image generator could be used successfully with the emotion extraction

engine. The chi-square tests were carried out for both types to analyse the null hypothesis.

Image only experiment

The purpose of this experiment is to test the effectiveness of the expressive images themselves. Four people's facial images were upload to the image generator, and 18 images were generated for each person. The eighteen images belong to the six expression categories, each with three images of different expressive intensities were created.

A total of 35 students and staff from Bournemouth University participated in this experiment. Each subject was told to view the images and identify which category of emotion each image expressed. The answers to be chosen include *happiness, sadness, anger, surprise, fear disgust* and *not sure*.

The results are shown in table 1. (1, 2, 3 in the x axis represents increasing corresponding intensity in that emotion category, 1 is the lowest and 3 is the highest. the y axis shows the percentage of subjects who identified the emotion correctly).

Result analysis

For each category of image, the chi-square test is carried out. The corresponding obtained values for the six emotion categories *Happiness, sadness, surprise, angry, fear and disgust* are 245.3, 456.1, 563.2, 123.4, 245.7 and 156.2. The values mean that they are extremely significant $p < .01$. The results show that users did not classify the images randomly.

The intensity increasing in the expressive image generator means that the characteristic of the corresponding emotion increases. We predicted that users should recognise more images correctly with the intensity increasing. The results proved our prediction. From table 1, we also find that the

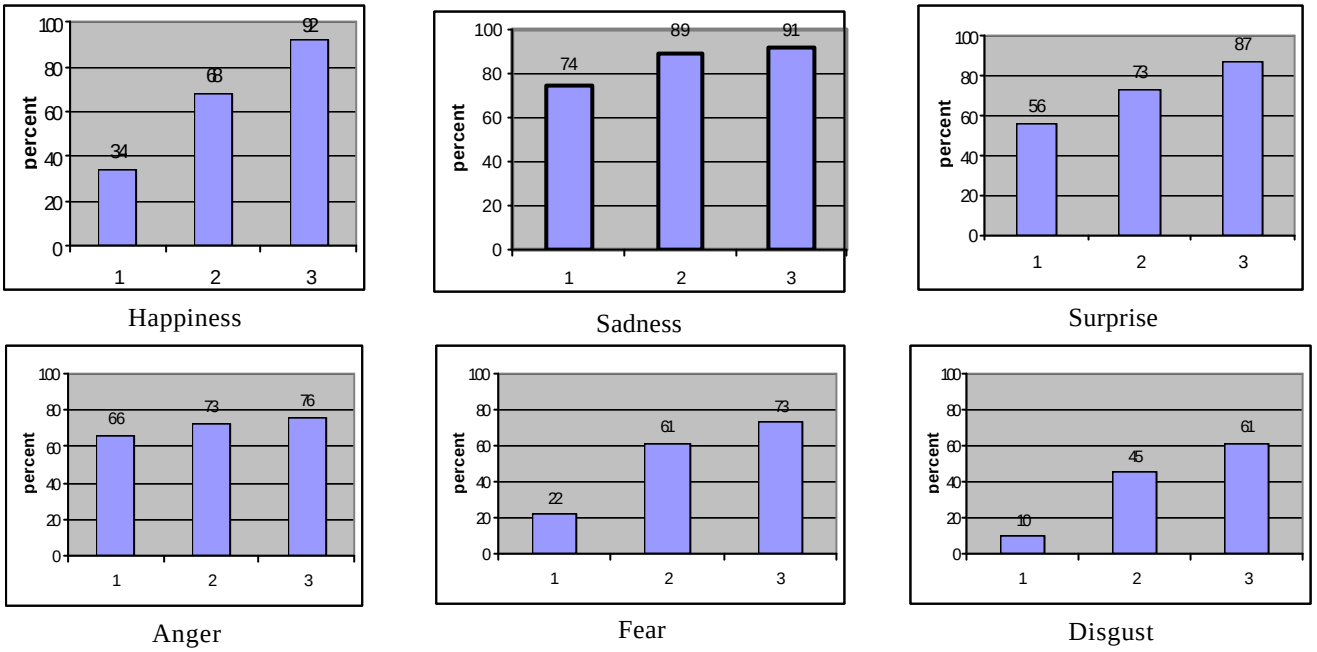

Table 1 Results for different emotion categories

the images in the emotion categories *disgust* and *fear* were not recognised as successfully as others.

Image plus text test

A total of 35 students and staff from the Bournemouth University participated in this experiment. Each subject was told to view a set of 18 pre-selected expressive images, which were generated using our expressive image generator. Below each image a short sentence was written. For example, for a generated "*happy*" image with high intensity, the text is "I am extremely happy". The subjects answer whether the image is appropriate to the text or not.

The results are shown in table 2. (In these figures, 1, 2, 3 in the x-axis represents the increasing corresponding intensity in that emotion category; 1 is the lowest and 3 is the highest. The y-axis shows the percentage of the subjects who agree that the image is appropriate to the text.)

Result analysis

For each category of image, the chi-square test is carried out again. The corresponding obtained values for emotion categories *happiness, sadness, surprise* and *anger* are 8.257, 24.03, 9.91 and 9.94. Those are extremely significant at $p < .01$. For categories *fear* and *disgust*, the obtained values are 3.671 and 4.82, which are marginally significant at $p<0.1$. The results show that users did not choose their answers randomly.

This experiment also proved that by increasing expressive intensity, more subjects will correctly recognise the expressive images.

It is shown that with text context information, the subjects correctly recognised more expressive images than in the first test. For emotion categories *happiness, sadness, surprise, fear*, more than 70% images were correctly recognised. For emotion categories *fear* and *disgust*, on average more than 60% images were correctly recognised. For all images with

medium and high intensity, more than 78% are recognised correctly.

In this case, the expressive image generator can be successfully used together with emotion extraction engine in a chatting environment.

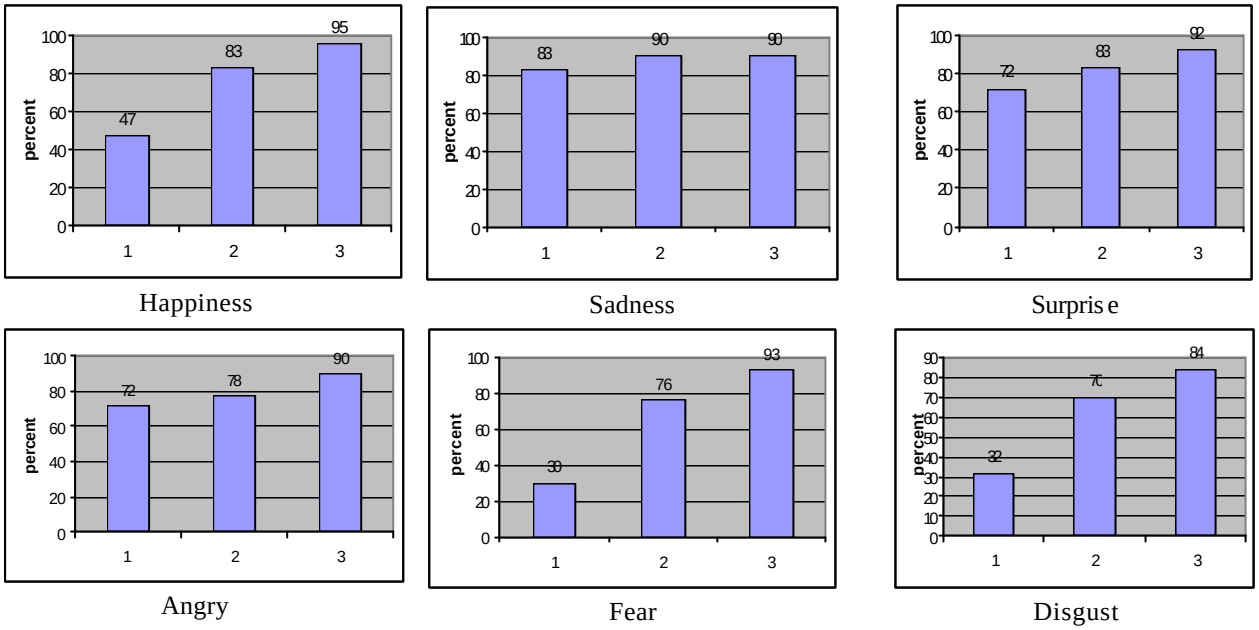

Table 2: Results for image plus text test

6 Conclusions

A text based real-time Internet communication system using an emotion extraction engine was developed and operates successfully. Since only text parameters are transmitted over the network, the bandwidth requirements for real-time communications is extremely low with the text to emotion extraction engine.

From previous experiments in paper [5], we established that people prefer using expressive image plus text in the chatting environments. The challenge in this paper is how to generate the expressive images for each user in real time.

An expressive image generator has been developed successfully. The generator can receive an uploaded neutral facial image and generate eighteen different expressive facial images. The eighteen facial images belong to six universal categories: *happiness*, *sadness*, *disgust*, *anger*, *surprise* and *fear*. For each category, three images with different emotion intensities are archived.

Using the expressive image generator, expressive images for each user can be produced in real time and in a user-friendly way. The generator can be used with the emotion extraction engine in a number of environments that provide text context information e.g. chatting room, story reader and online games. The generator's effectiveness may be enhanced with improvements to the model: expression model mask. Possible refinements include adding more control points and more control areas.

The experiments show that by increasing the emotion intensity, more acceptable expressive images can be obtained.

The test results demonstrate that the generator can create acceptable expressive images for a chatting environment when accompanied by text labels. The expressive image generator solved the image problem described in paper [4] and [5].

7 References

[1] http://emotion.salk.edu/Emotion/History/Hgeneral.html

[2] Buck, R., 2001. *Emotional Expression, Communication, and Competence: Applications of a Developmental-Interactionist Theory of Biological, Social, Cognitive and Moral Emotions*. University of Connecticut. http://wattlab.coms.uconn.edu/ftp/users/rbuck/Germany01/

[3] Dautenhahn, K. & Werry, I., 2001. The AURORA Project: Using mobile Robots in Autism Therapy. *Learning Technology*, IEEE Computer Society Learning Technology Task Force, Vol. 3(1), January 2001, ISSN 1438-0625

[4] Zhe, X. & Boucouvalas, A.C., 2002. Text-to-Emotion Engine for Real Time Internet Communication. *International Symposium on Communication Systems, Networks and DSPs*, 15-17 July 2002, Staffordshire University, UK, pp 164-168

[5] Zhe, X, John,D & Boucouvalas, A.C., 2002. Text-to-Emotion Engine: tests of user preferences.ISCE 2002, 22-25 September 2002, Erfurt, Germany, pp B25-B30

[6]. Kuehn, S.A., 1993. Communication innovation on BBS, Interpersonal Computing & Technology, 1(2), ISSN: 1064-4326

[7] Leech, G. 1997. A Brief Users' Guide to the Grammatical Tagging of the British National Corpus. http://info.ox.ac.uk/bnc/what/gramtag.html

[8] Francis, W.N. & Kucera, H., 1979. *Brown Corpus Manual*, Department of Linguistics, Brown

[9] Russell, S. & Norvig, P., 1995. *Artificial Intelligence - a modern approach*, Prentice Hall, New Jersey

[10] Parke, F.I., & Waters, K., 1996. *Computer Facial Animation*, A.K. Peters, Wellesley, Massachusetts

[11] Paul Ekman and W.V. Friesen "facial action coding system"

[12] Cambridge digital research laboratory : faceworks. http://interface.digital.com/

[13] Frederic Pighin, JLOE Auslander, Dani Lischinski and David Salesin. Realistic Facial Animation Using Image-Based 3D Morphing. Technical report UW-CSE-97-01-03.

[14] "digital Image warping", George Wolberg ISBN: 0-8186-8944-7

[15]. A. Goshtasby, "Piecewise Linear Mapping Functions for Image Registration", Pattern Recognition, Vol. 19, No. 6, pp. 459-466, 1986.

[16] Jocelyn Scheirer, R. Fernandez, J. Klein, and R. W. Picard (2002), "Frustrating the User on Purpose: A Step Toward Building an Affective Computer", Interacting with Computers, 14, 2 (2002), 93-118.

[17] Maes, P. Velfisquez, J.D. "Cathexis: A Computational Model of Emotio ns.