

Research Article

A Text-Based Chat System Embodied with an Expressive Agent

Lamia Alam and Mohammed Moshiul Hoque

Department of Computer Science & Engineering, Chittagong University of Engineering & Technology, Chittagong 4349, Bangladesh

Correspondence should be addressed to Mohammed Moshiul Hoque; moshiulh@yahoo.com

Received 31 May 2017; Accepted 14 November 2017; Published 26 December 2017

Academic Editor: Carole Adam

Copyright © 2017 Lamia Alam and Mohammed Moshiul Hoque. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Life-like characters are playing vital role in social computing by making human-computer interaction more easy and spontaneous. Nowadays, use of these characters to interact in online virtual environment has gained immense popularity. In this paper, we proposed a framework for a text-based chat system embodied with a life-like virtual agent that aims at natural communication between the users. To achieve this kind of system, we developed an agent that performs some nonverbal communications such as generating facial expression and motions by analyzing the text messages of the users. More specifically, this agent is capable of generating facial expressions for six basic emotions such as happy, sad, fear, angry, surprise, and disgust along with two additional emotions, irony and determined. Then to make the interaction between the users more realistic and lively, we added motions such as eye blink and head movements. We measured our proposed system from different aspects and found the results satisfactory, which make us believe that this kind of system can play a significant role in making an interaction episode more natural, effective, and interesting. Experimental evaluation reveals that the proposed agent can display emotive expressions correctly 93% of the time by analyzing the users' text input.

1. Introduction

The ability to express emotions or displaying expression is essential characteristic for both the human-human and human-agent interactions. To simulate emotional expressions in an interactive environment, an intelligent agent needs both a model for generating persuasive responses and a visualization model for mapping emotions into facial expressions. Expressive behavior in intelligent agents provides several important purposes. First, it permits an entity to interact effectively with other expressive beings in social contexts. Second, expressive behavior may provide a visualization tool for monitoring the complex internal states of a computer system. Moreover, displaying emotions play a positive role in several applications that allow users to have a more fruitful and enjoyable experience with the system [1].

Face is the part of human body that is most closely observed during the interaction. When the people interact with one another, they tend to adapt their head movements and facial expressions in response to each other [2]. However, this is not the same when people interact with each other

through computer nonverbally. A major challenge here is to develop an expressive intelligent agent that can generate facial expressions and motions, while people are interacting through it. In order to do so, the agent should understand the emotional content of the user's input that is expressed through his/her text message and should respond accordingly to that text. It requires both the model to recognize the emotions and the model to generate the facial expressions to produce the appropriate emotional response.

People's thoughts, feelings, or intents are not expressed completely while communicating with each other by exchanging text messages (especially in chatting scenario) through computer. Moreover, communicating via text messages alone often seems boring to interacting partners. In order to make the chatting interaction amusing, few systems use virtual characters to represent chatting partners. These are mostly 2D or 3D models of human, cartoon, or animal like characters. But most people find these kinds of virtual characters unrealistic and incompatible with their personality because of their static and unsuitable representations.

In this paper, we developed a system embodied with an intelligent agent that can exhibit various emotions by generating appropriate facial expressions and motions. According to El-Nasr et al. [3], this type of agent requires a visual interface, typically a human face, which is familiar and nonintimidating, making users feel comfortable when interacting with a computer. We have decided to use six basic expressions (i.e., happy, sad, fear, surprise, anger, and disgust) due to their universality [4] and two more expressions, determined and irony. In order to visualize these emotions through facial expression, we developed a life-like character that understands the emotional intent of user that expressed through textual input (as a form of words or sentence) [5]. In addition to these emotions, we adopted eye blinks and head movements to make the interaction more engaging and to maintain more natural communication.

2. Related Work

A few research activities have been conducted on life-like characters. There are some 2D or 3D virtual agents with very limited ability to demonstrate nonverbal behavior such as displaying predefined facial expression controlling through emotion wheel [6], movement of face components (lips, eye brows, etc.) to conduct input speech [7], and low level communication like gaze direction [8]. Other interactive animated creatures called woggles are autonomous and self-controlled and have the ability to jump, slide, move eyes, and change body and eye shape while interacting among themselves and with the outside world [9]. Most of these studies are focused on generating static facial expressions, while the motions of some of the face components have been neglected, in particular the eyes and the rigid motion of the head. Another aspect that is not addressed in these previous works is adaption of expressions with intermediate states.

Not only these virtual agents are used for entertainment, but also these agents are widely used for customer service functionality. Two such types of agents are Anna (virtual assistant) [10] and Rea (the real estate agent) [11]. These agents interact with the users in a human-like way with a very limited emotion.

Another facial animation system LUCIA is developed by Riccardo and Peiro [12] works on standard Facial Animation Parameters and speaks with the Italian version of FESTIVAL TTS and can copy a real human by reproducing the movements of passive markers positioned on his face and recorded by the ELITE device or can be driven by an emotional XML tagged input text.

Kramer et al. [13] explored design issues of the conversational virtual human as a socially acceptable autonomous assistive system for elderly and cognitively impaired users.

Ameixa et al. [14] introduced Filipe, a chatbot that answers users' request by taking advantage of a corpus of turns obtained from movies subtitles (the Subtle corpus). Filipe is based on Say Something Smart, a tool responsible for indexing a corpus of turns and selecting the most appropriate answer.

Youssef et al. [15] presented a *socially adaptive virtual agent* that can adapt its behavior according to social constructs (e.g., attitude and relationship) that are updated depending on the behavior of its interlocutor. They considered the context of job interviews with the virtual agent playing the role of the recruiter.

In their work, Straßmann et al. [16] tested four different categories of nonverbal behavior: dominant, submissive, cooperative, and noncooperative nonverbal behavior. Most of the behaviors were created using motion capturing with a postprocessing of bones, gaze, and hand shape and were mapped onto the embodied conversational agent Billy (Social Cognitive System Group, Citec Bielefeld Germany), while some behaviors were created with a key-frame editor. The virtual agent Billie is humanoid, male, more childish-looking and has a medium degree of realism (between cartoon and photo-realistic)

Kopp et al. [17] describe an application of the conversational agent Max in a real-world setting. The agent is employed as guide in a public computer museum, where he engages with visitors in natural face-to-face communication, provides them with information about the museum or the exhibition, and conducts natural small talk conversations

In their work, Vosinakis and Panayiotopoulos [18] presented SimHuman, a platform for the generation of real-time 3D environments with virtual agents. SimHuman is highly dynamic and configurable, as it is not based on fixed scenes and models and has an embedded physically based modeling engine. Its agents can use features such as path finding, inverse kinematics, and planning to achieve their goals.

Formolo and Bosse [19] developed a new system that captures emotions from human voice and, combined with the context of a particular situation, uses this to influence the internal state of the agent and change its behavior.

Gratch et al. [20] describe a system, based on psycholinguistic theory, designed to create a sense of rapport between a human speaker and virtual human listener.

In order to deal with the problems of managing chat dialogues in the standard 2D text-based chat, Kim et al. [21] proposed a more realistic communication model for chat agents in 3D virtual space. In their work, they measure the capacity of communication between chat agents by considering the spatial information and applied a novel visualization method to depict the hierarchical structure of chat dialogues. They also proposed a new communication network model to reveal the microscopic aspect of a social network. But their work is limited to only seven users.

Weise et al. [22] presented a system for performance-based character animation that enables any user to control the facial expressions of a digital avatar in real-time. The user is recorded in a natural environment using a nonintrusive, commercially available 3D sensor Kinect.

mocitoTalk! [23] is a solution to stream videos in real-time for instant messaging and/or video-chat applications developed by a company named Charamel. It provides real-time 3D avatar rendering with direct video-stream output and automatic lip synchronization live via headset. It is widely used for messaging services, video production, server-side

video generating, e-mail marketing, and interactive digital assistants (IDA).

Alam and Hoque [24] focused on developing a human-like virtual agent that produces facial expressions with motions (such as head movement and eye blinks) during human-computer interaction scenarios through analyzing the input text messages of users. This agent is able to display six facial expressions, namely, happy, sad, angry, disgust, fear, and surprise, based on the chatting partner's input text that makes the interaction enjoyable. Table 1 summarizes the works.

Sentiment analysis or opinion mining is one of the most popular research topics today. Sentiment analysis has its roots in natural language processing and linguistics and has become popular due to widespread Internet usage and the texts freely available online on social media. Sentiment analysis or opinion mining deals with using automatic analysis to find sentiments, emotions, opinions, and attitudes from a written text towards a subject. This subject may be a product, an organization, a person, a service, or their attributes. On the other hand, for a successful and effective human-human communication, emotion plays a very significant role. In fact, for an engaging interaction, sometimes emotion is more important than IQ. According to the study of Cambria et al. [25], affective computing and sentiment analysis are key for the advancement of AI and all the research fields that stem from it. In their work, they pointed out various application areas where affective computing and sentiment analysis has great potentials. They also described the existing approaches to affective computing and sentiment analysis.

Chandra and Cambria [26] developed a system to enhance the chat experience by using an intelligent adaptive user interface (UI) that exploits semantics and sentics, which is the cognitive and affective information, associated with the ongoing communication. In particular, their approach leverages sentiment analysis techniques to process communication content and context and, hence, enable the interface to be adaptive in order to offer users a richer and more immersive chat experience. For example, if the detected mood of the conversation is "happy," the UI will reflect a clear sunny day. Similarly, a gloomy weather reflects a melancholy tone in the conversation. Anusha and Sandhya [27] proposed an approach which adds natural language processing techniques to improve the performance of learning based emotion classifier by considering the syntactic and semantic features of text. They added an extra module to traditional learning system architecture. This extra module (NLP module) focuses on analyzing syntactic and semantic information using NLP techniques. Mohammad [28] summarized the diverse landscape of problems and applications associated with automatic sentiment analysis. He also explained several manual and automatic approaches to creating valence- and emotion-association lexicons and described work on sentence-level sentiment analysis.

Dai et al. [29] proposed a computational method for emotion recognition and affective computing on vocal social media to estimate complex emotion as well as its dynamic changes in a three-dimensional PAD (Position-Arousal-Dominance) space; furthermore, this paper analyzes the propagation characteristics of emotions on the vocal social media site WeChat. Poria et al. [30] propose a novel

methodology for multimodal sentiment analysis to harvest sentiments from Web videos by demonstrating a model that uses audio, visual, and textual modalities as sources of information. They used both feature- and decision-level fusion methods to merge affective information extracted from multiple modalities. For their textual sentiment analysis module they have used sentic-computing-based features. In his work, Younis [31] presented an open source approach, throughout which twitter Microblogs data has been collected, preprocessed, analyzed, and visualized using open source tools to perform text mining and sentiment analysis for analyzing user contributed online reviews about two giant retail stores in the UK, namely, Tesco and Asda stores, over Christmas period 2014.

In this work, sentiment analysis refers to perhaps the most basic form, which is detecting emotions such as joy, fear, and anger, from a text and to determine if a text or sentence is positive or negative. We mainly focused on developing an expressive agent that displays different facial expressions by analyzing textual contents of the user.

3. Proposed System Framework

A schematic representation of the proposed system is shown in Figure 1. The proposed system consists of some main modules: text processing, database, emotion recognition, intensity determination, facial expression visualization, and movement generation. A brief overview of these modules follows.

3.1. Text Processing. In order to produce facial expression, the module recognizes the emotional word of sender's text. The module tokenizes the sender's input text into tokens by splitting the input sentence on specified delimiter characters (such as " ", ",", ".", ":", ";", "?" etc.).

For example, consider a text message: "Hi, How are you doing today?"

Our system will at first take this message as text input and will tokenize the input based on delimiters " " (space), ",", and "?". Output from this module will be tokens: "Hi," "How," "are," and "you" which will be sent as input to the next module.

3.2. Recognizing Emotion. From a set of tokens, it searches for keywords and related modifiers by matching each token with the keywords and modifiers stored in the database. After recognition of keywords and modifiers, the module will represent the emotional state against each word. But in case no keyword or modifier is recognized, it assumes that it is a normal text and maintains a neutral emotional state.

In the current implementation, we have used eight emotional states: happy, sad, anger, surprise, fear, disgust, irony, and determined. The recognized words are mainly adjectives (like happy, sad, furious, serious, horrid, etc.) that give a clearer idea of emotion. For example, consider two people who are chatting to each other and trying to express their feelings:

User A: My friend met a *terrible* accident today.

User B: Sorry to hear about the accident.

TABLE 1: Summarized related works in the literature.

Article	Agent's appearance	Inputs/sensors	Actions	Agent's communication type	Application
Kurlander et al. [6]	2D comic characters	Text; emotional wheel	Display gesture; generate text balloons	Both verbal and nonverbal	Entertainments
Nagao and Takeuchi [7]	3D face model	Human voice	Displays facial expressions and gestures; generates voice	Both verbal and nonverbal	Entertainments
Vilhjalmsson and Cassell [8]	3D upper body of a cartoon like character	Cursor key; mouse; text	Generates gestures: glance, smile, wave, head nod, raising brows	Nonverbal	Chatting
Loyall and Bates [9]	An ellipsoid with eyes	Observes the virtual environment	Jump, turn, squash, puff up, slide, change color, move eyes, and so on	Nonverbal	Entertainment
Mount [10]	Photo unreal representation of a female	Text input (queries by customer or visitors)	Answers queries; generates gestures: blink eyes, raising eye brows, smiles, tilts, and so on	Both verbal and nonverbal	Virtual customer care agent (Anna) in IKEA.com
Cassell et al. [11]	Computer generated human	Camera; audio; input text	Generates voice; displays facial expression and gestures	Both verbal and nonverbal	Virtual real estate agent (REA)
Riccardo and Peiro [12]	3D female facial model	Input text; camera; human voice	Speaks Italian language; copies human movement	Both verbal and nonverbal	Research
Kramer et al. [13]	Virtual male agent	Audio	Speak; displays appointments	Both verbal and nonverbal	Assistive system for elderly and cognitively impaired users
Ameixa et al. [14]	Animated agent	Input text (user requests)	Replies using knowledge based system	Verbal	Research
Youssef et al. [15]	Female animated agent	Camera; audio	Leads discussions; proposes conversations; reacts in response to interviewee	Both verbal and nonverbal	Virtual recruiter
Straßmann et al. [16]	Humanoid male agent	-	Displays nonverbal behavior: dominant, submissive, cooperative, and noncooperative	Nonverbal	Research
Kopp et al. [17]	Animated agent	Keyboard; camera; audio	Generates voice; displays nonverbal behavior	Both verbal and nonverbal	Virtual guide in a public computer museum
Vosinakis and Panayiotopoulos [18]	3D model of an agent/avatar or any object	-	Can use features such as path finding, inverse kinematics, and planning to achieve their goals	Nonverbal	Research
Formolo and Bosse [19]	Animated agent	Audio; input text	Changes behavior	Nonverbal	Research
Gratch et al. [20]	Animated agent	Camera; audio	Speaks German language; copies the gesture of one user and shows it to the other user	Both verbal and nonverbal	Entertainment

TABLE 1: Continued.

Article	Agent's appearance	Inputs/sensors	Actions	Agent's communication type	Application
Kim et al. [21]	3D humanoid agent	Input text	Generates movements and dialogues; generates gaze	Both verbal and nonverbal	Social interaction
Weise et al. [22]	Digital character	Kinect; audio	Copies the user gesture and voice	Both verbal and nonverbal	Digital game play; social interaction
mocitotalk [23]	Animated agent	Camera; audio	Generates an animate agent, copies action, and so on	Both verbal and nonverbal	Entertainment
Alam and Hoque [24]	Male and female agent	Input text	Generates facial expressions with motions (such as head movement and eye blinks)	Nonverbal	Chatting

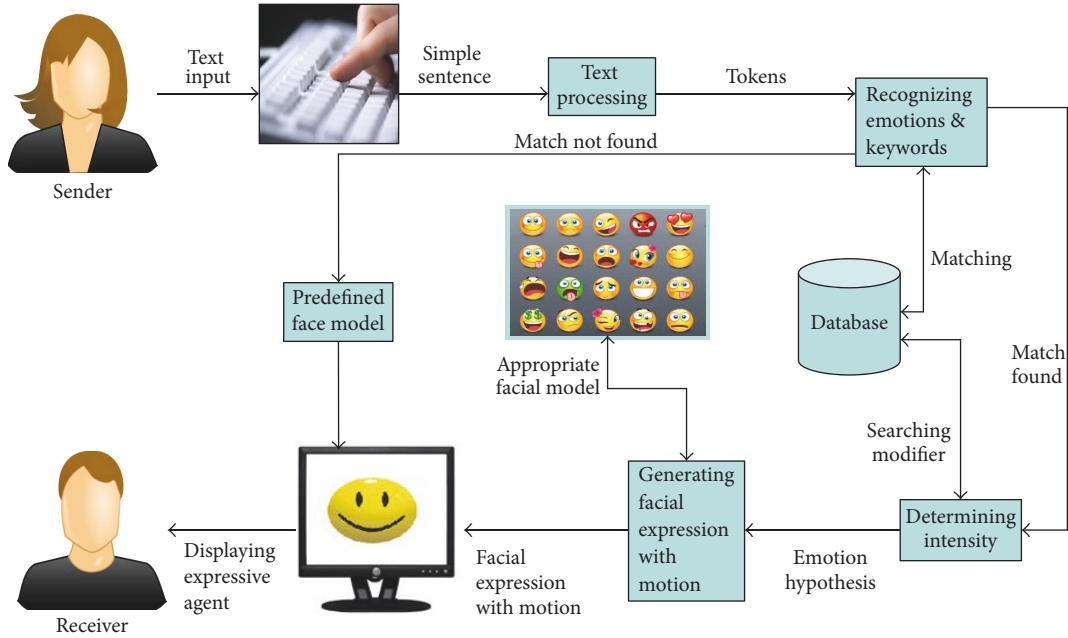


FIGURE 1: Proposed framework for the system.

Here the sentence “My friend met a terrible accident today” is first divided into tokens: “My,” “friend,” “met,” “a,” “terrible,” “accident,” and “today.” Then the keyword “terrible” is extracted from these tokens to characterize User A’s emotional state. Similarly, User B’s emotional state is characterized with the word “Sorry.”

3.3. Determining Intensity. This module will assign the level of intensity with the corresponding emotional state. Consider another example: User B’s statement with different emotional intensity:

User B: Very sorry to hear about the accident.

In this case, the modifier “very” is used to determine the intensity level for this emotional state. The intensity level for the emotional state “very sorry” and “sorry” is different.

Therefore, their corresponding facial expression and intensity level will change with high to low values. Once the emotional strength of a particular category passes a certain threshold, then user’s agents’ representation can be changed to show the appropriate expression. This system also generates facial motions such as eye blink and head movements based on certain text input such as yes, ya, no, and nope.

3.4. Database. Database is an important module of our system. It contains keywords which are used to characterize different emotional states and to express acceptance and rejection. It also contains modifiers to define the intensity level of emotion. Each token found from text processing module is matched with database in order to identify the emotional state and its density. Table 2 contains some sample keywords used to identify the emotional state from text message of user.

TABLE 2: Example of keywords used to identify emotional state.

Emotional state	Happy	Sad	Anger	Surprise	Disgust	Fear	Irony	Determined
Keywords	Glad, delighted, amazing	Unhappy, poor, sorry	Crazy, annoying, furious	Amazed, unexpected, wonder	Outrageous, disgusting, dislike	Fearful, horrified, direful	Humorous, contempt, twisted	Serious, determined, intended

TABLE 3: Example of keywords to express acceptance and rejection.

Express	Acceptance	Rejection
Keywords	Yes, yeah, ya	No, nope, never

Some keywords which are used to express acceptance and rejection in form of head nodding and shaking are shown in Table 3.

Some examples of modifiers used to determine the intensity levels of an emotion are *so*, *very*, *extremely*, *highly*, and so on.

3.5. Generating Facial Expression with Motion. We have designed this module by dividing it into two steps as follows.

3.5.1. Facial Expression Visualization. In our work, we mainly focused on generating facial expression for eight emotions: happy, sad, fear, surprise, anger, disgust, irony, and determined. According to El-Nasr et al. [3], facial expression is controlled through the action of sixteen parameters. Five of these control the eyebrows. Four control the eyes. Five control the mouth and two control face orientation. We manipulated these parameters using the cues for facial expression of six basic emotions as suggested by Ekman and Friesen [32].

To visualize expression, we created two 3D life-like human characters both a male and a female using software MakeHuman [33]. We designed the characteristic of the agents in a way so that an Asian can easily relate him/herself to the characters.

Blender [34] is used to make these agents more realistic and to generate the facial expressions for each emotion. In order to make the agents more natural, we manipulated the texture of the agent and performed cloth simulation.

We generated eight facial expressions and motions (i.e., eye blinks and head movements) by manipulating the facial parameters based on the involvement of facial muscles and other nonverbal behaviors. A brief description of these emotions is given bellow along with the cues of facial parameters used to generate facial expression:

- (i) *Happy* is an emotion of feeling or showing pleasure or contentment. When we are happy corners of our lips are pulled up, mouth may or may not be parted with teeth exposed or not, corners of cheeks are raised, and lower eyelid shows wrinkles below it and maybe raised but not tense.
- (ii) *Sad* is an emotion of being affected with or expressive of grief or unhappiness. For sad inner corners of eyebrows are drawn up, skin below the eyebrow is

triangulated, corners of the lips are drawn, or lip is trembling.

- (iii) *Angry* is an emotion that highly contrasts and disagrees with the emotion happy. When angry, brows are lowered and drawn together, vertical lines appear between brows, eyes have a hard stare and may have a bulging appearance, and lips are either pressed firmly together with corners straight or down or open.
- (iv) *Fear* is an unpleasant emotion caused by the threat of danger, pain, or harm. We can define the expression fear by brows raised and drawn together and forehead wrinkles drawn to the center, mouth is open, and lips are slightly tense or stretched and drawn back.
- (v) *Surprise* is an emotion to strike or occur with a sudden feeling of wonder or astonishment, as through unexpectedness. Brows are raised, eyelids are opened and more of the white of the eye is visible, and jaw drops open without tension or stretching of the mouth in case of surprise.
- (vi) *Irony* is the expression of one's meaning by using language that normally signifies the opposite, typically for humorous or emphatic effect. It involves neutral eyebrows and upper lips stretched.
- (vii) *Determined* is the emotion of having made a firm decision and being resolved not to change it. We can define the expression determined by inner brows raised and lips rolled.

Figure 2 shows the facial expression of eight emotions along with neutral expression for both male and female character. From this figure, we can easily understand the difference between the expressions of each emotion.

3.5.2. Facial Motion and Head Movements. Humans prefer to interact with virtual agents as naturally as they do with other. To facilitate this kind of interaction, agent behavior should reflect life-like qualities. For naturalness, we included three kinds of movements in facial components: eye blinks, head nodding, and head shaking.

- (i) *Eye blink* is defined as a temporary closure of both eyes, involving movements of the upper and lower eyelids. Figure 3 shows the frames used to generate eye blink. We choose to blink every 3 seconds and one blink lasts about 1/3 seconds.
- (ii) *Head movements* in human mainly involve motion in head. There are two major kinds of head movements:

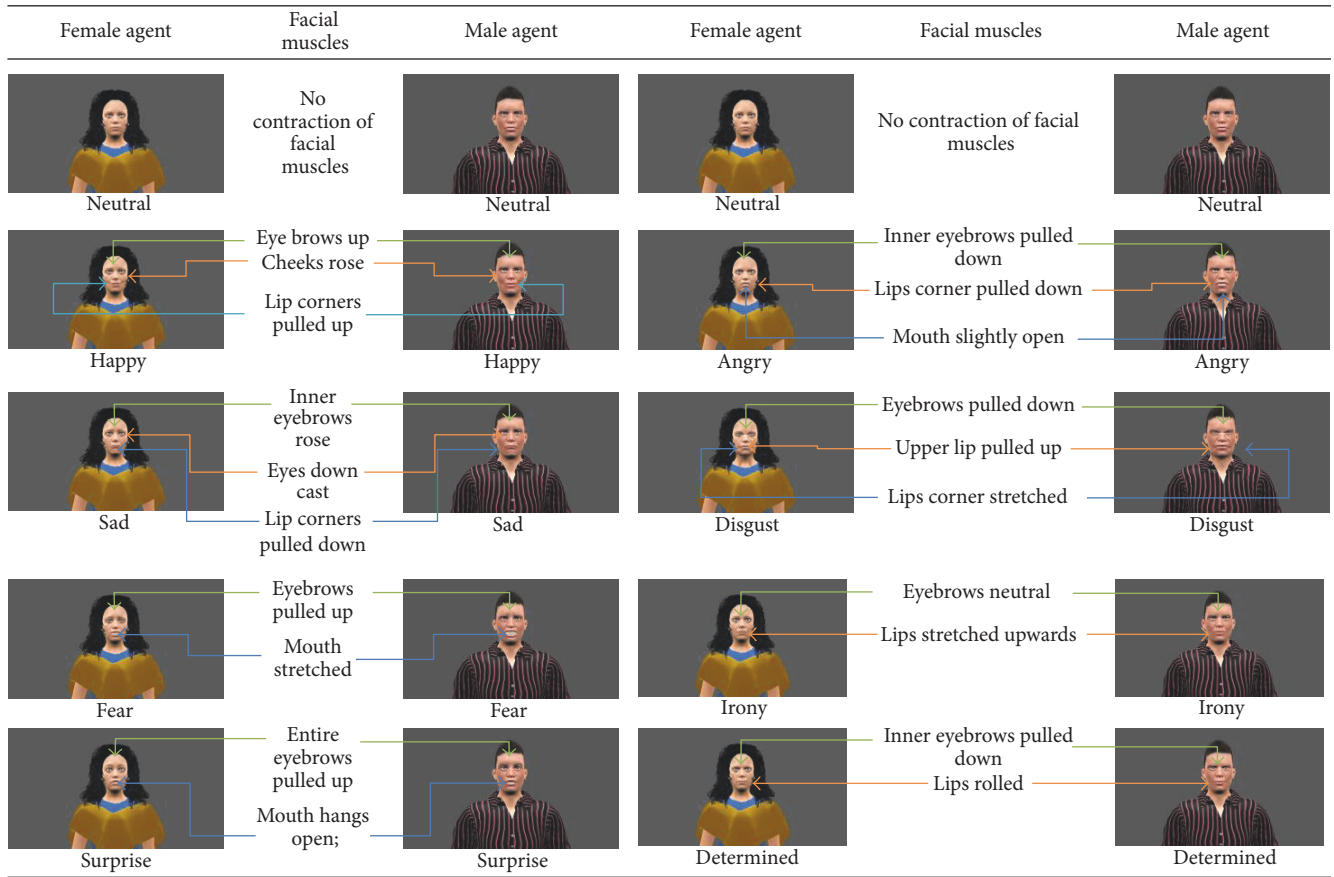


FIGURE 2: A fragment of generated facial expressions.

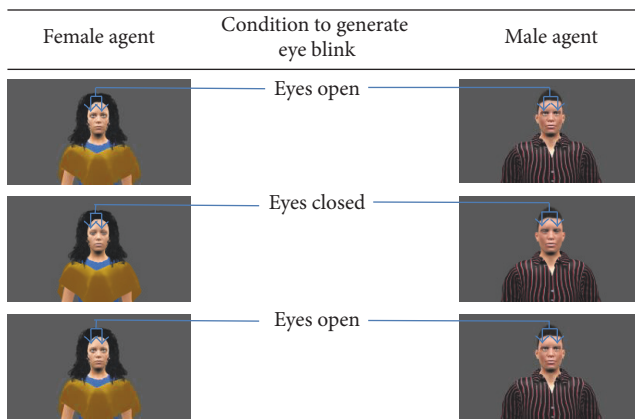


FIGURE 3: Example of typical frames used to generate eye blinks.

- (a) *Nodding*: head is tilted in alternating up and down arcs along the sagittal plane. Nodding of head is used to indicate acceptance.
- (b) *Shaking*: head is turned left and right along the transverse plane repeatedly in a quick succession. Shaking of head is used to indicate disagreement, denial, or rejection.

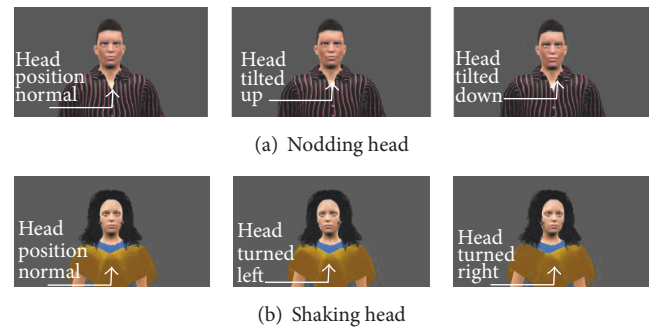


FIGURE 4: Frames used to generate head movements.

In order to generate the head movements, we manipulated the bone group of neck. To achieve these head movements, frame-by-frame animation technique is used. Figure 4 shows the frames used to generate head movements.

4. Graphical User Interface

In our project, designing the GUI is the most important part of the system as through this the users will interact with each other by text message and will be able to see the facial

expressions and facial motion generated by the agents. Three graphical user interfaces for this system are as follows:

- (i) *Login window*: this is the main interface that user's see when the application is launched. This window requires users to input a user name and server IP address and to specify user's gender in order to select suitable avatar representation for the user. This interface checks if the user name already exists in the network and also generates error message in case of blank input, unspecified gender, and wrong server address. For correct input information, this interface establishes a connection with the server and directs users to user homepage. From this window, user first gets access to the database.
- (ii) *User homepage*: a successful connection to the server leads the user to this window. This interface allows users to change his/her availability (Available/Away/Busy/Offline) and to update his/her status. This interface also shows the other users who are currently online and available for chat. The interface runs updates periodically for all kind of changes to take place and to make it visible to the user, for example, change in availability by other user or user itself and any user leaving or joining the list. Changes are also made to the database from this interface. User can select any user from the list to initiate a personal chat and this will open a chat window. This interface also offers the feature to logout which in turn requests the server to remove the user from active user list and to close the connection with the corresponding user.
- (iii) *Chat window*: this window is opened when a user wishes to chat with another user in the network whose name is visible in the online friend list of the user who initiated the chat.

Figure 5 shows the graphical user interfaces of our system.

5. Experimental Results and Analysis

To evaluate the system, we conducted two experiments. The purpose of first experiment is to evaluate the appropriate expression for each basic emotion. In the second experiment, we evaluated the overall performance of our proposed system by comparing the system with other emoticon-based systems.

5.1. Experiment 1: To Evaluate Appropriate Expressions. To evaluate the system, we conducted an experiment to achieve the appropriate expression for each basic emotion. A total of ten participants participated in this experiment. The average age of participants is 32.4 years (SD = 4.45).

5.1.1. Experimental Design. To select the appropriate facial expression, we designed different types of expressions of each. For happy, sad, angry, fear, and disgust, we designed five different types of expression and for surprise, irony, and determined we used three types. Before the experiment, we explained that the purpose of this experiment is to

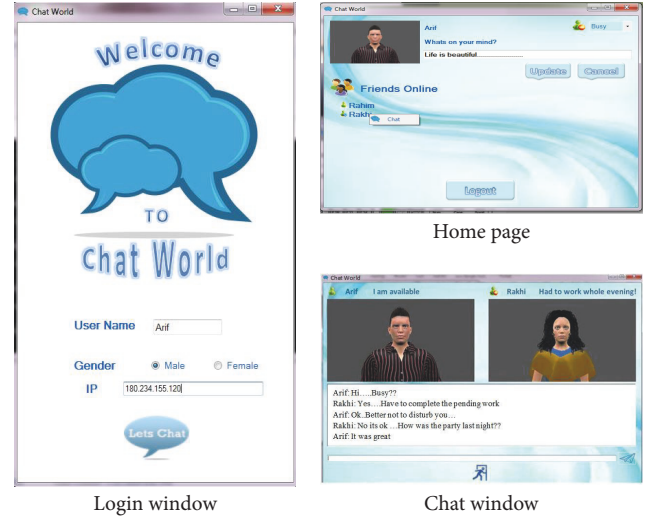


FIGURE 5: GUI of our proposed system.

evaluate the suitable expression for each emotion. Each trial was started with showing the participants different types of expressions for each emotion twice. Each participant interacted with both male and female agents and each session took approximately 45 minutes. After experiencing all expressions, participants were asked to rate their feelings for each type in terms of 1-to-7-point Likert scale.

5.1.2. Evaluation Measures. We measured the following two items in this experiment:

- (i) *Appropriateness*: we asked a question (“Which type of expression do you like most preferable to represent an emotion?”) to all participants.
- (ii) *Accuracy*: to evaluate the system performance, we counted total number of emotive keywords in the texts (N), total number of emotive words that are correctly recognized (C), and total number of incorrect emotive words (M). We used the following equations to measure the accuracy (A) in recognizing emotion:

$$A = \frac{N - M}{N} \times 100\%. \quad (1)$$

5.1.3. Results. The results of the experiments conducted to measure the appropriateness of each facial expression and the overall accuracy of the system are explained as follows.

(1) Appropriate Expressions. We collected a total of 800 (10 [participants] \times 8 [expressions] \times 5 [types] \times 2 [agents]) interaction responses for both agents. We conducted a repeated measure of analysis of variance (ANOVA) on the participants' scores for both male and female agents. Figure 6 shows the results of analysis.

For happy expression of female agent, the results show that the differences among conditions were statistically significant [$F(4, 49) = 31.5, p < 0.0001, \eta^2 = 0.75$]. Results also indicate that type 3 expression gained the higher scores than

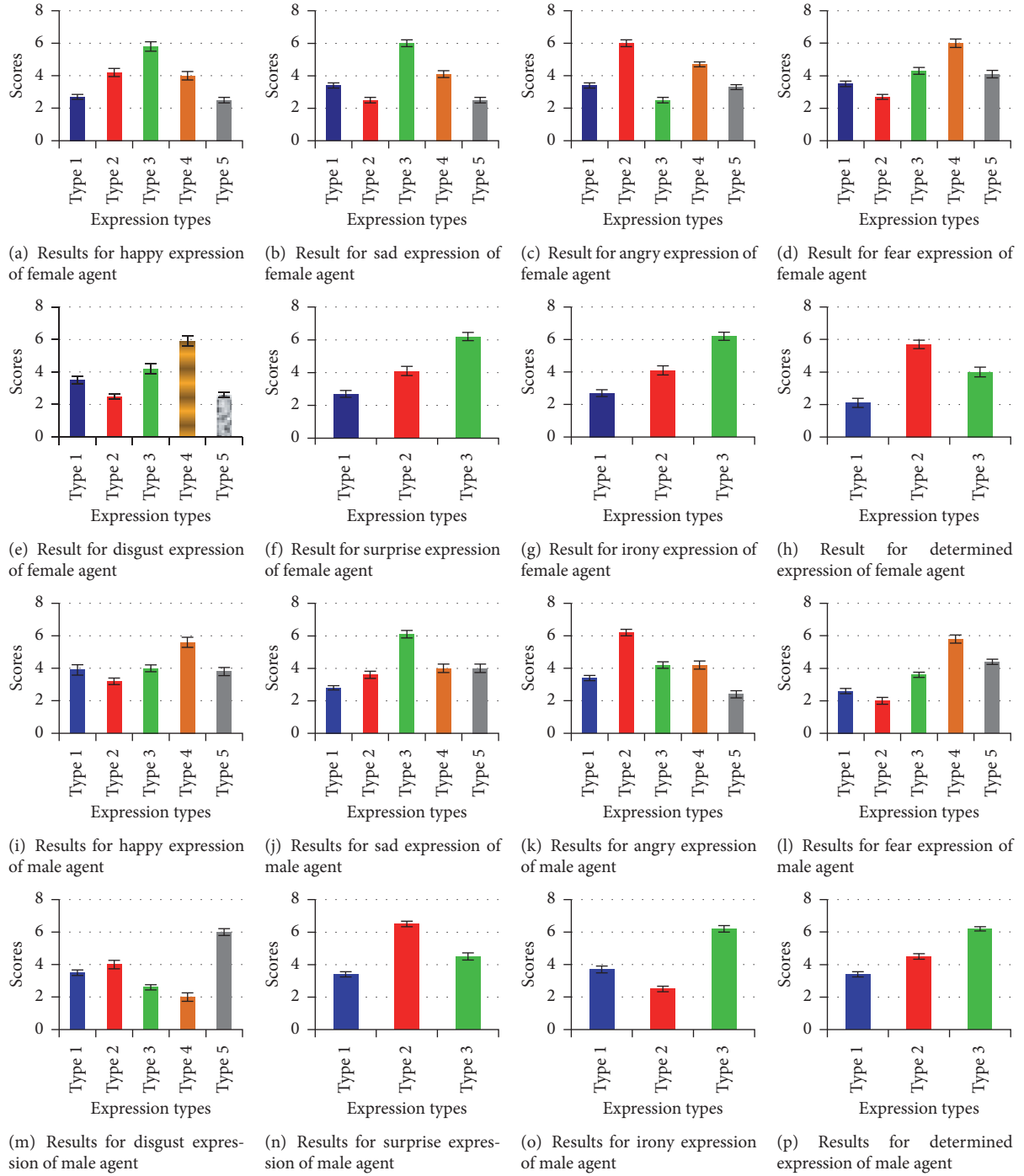


FIGURE 6: Evaluation results for both female and male agents.

other types (Figure 6(a)). Thus, we chose type 3 expression for happy expression of female agent. On the other hand, for same expression of male agent, we also found significant differences among conditions [$F(4, 49) = 11.8, p < 0.0001, \eta^2 = 0.51$]. The result also reveals that type 4 expression gained the higher scores than other types (Figure 6(i)). Thus, we have to decide to use expression type 4 for producing the happy expression of male agent.

In case of sad expression of female agent [$F(4, 49) = 58.1, p < 0.0001, \eta^2 = 0.83$] and male agent [$F(4, 49) = 29.3, p < 0.0001, \eta^2 = 0.72$], the results show that the differences among conditions were statistically significant. Results (Figures 6(b) and 6(j)) also indicate that for both female and male agent the expression type 3 gained the higher scores than other types. Thus, we chose type 3 expression for sad expression of both female and male agents.

TABLE 4: Summarization of analysis results for male and female agents.

Expression types	Type 1		Type 2		Type 3		Type 4		Type 5	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Happy	2.7	3.9	4.2	3.2	5.8	4.0	4.0	5.6	2.5	3.8
Sad	3.4	2.8	2.5	3.6	6.0	6.1	4.1	4.0	2.5	4.0
Angry	3.4	3.4	6.0	6.2	2.5	4.2	4.7	4.2	3.3	2.4
Fear	3.5	2.6	2.7	2.0	4.3	3.6	6.0	5.8	4.1	4.4
Disgust	3.5	3.5	2.5	4.0	4.2	2.6	5.9	2.0	2.6	6.0
Surprise	2.7	3.4	4.1	6.5	6.2	4.5	-	-	-	-
Irony	3.4	3.7	3.9	2.5	5.6	6.2	-	-	-	-
Determined	2.1	3.4	5.7	4.5	4.0	6.2	-	-	-	-

Similarly for angry expression of female agent [$F(4, 49) = 65.2$, $p < 0.0001$, $\eta^2 = 0.85$] and male agent [$F(4, 49) = 44.8$, $p < 0.0001$, $\eta^2 = 0.80$], the results show that the differences among conditions were statistically significant. Results (Figures 6(c) and 6(k)) revealed that type 2 expression gained the higher scores than other types for both female and male agents. Thus, we have to decide to use expression type 2 for producing the angry expression.

In case of fear expression of female agent [$F(4, 49) = 34.3$, $p < 0.0001$, $\eta^2 = 0.75$] and male [$F(4, 49) = 60.3$, $p < 0.0001$, $\eta^2 = 0.80$] agent, the results (Figures 6(d) and 6(l)) show that the differences among conditions were statistically significant and indicate that type 4 expression gained the higher scores than other types for both agents. Thus, we chose type 4 expression for fear expression.

For disgust expression, the results show statistically significant differences among female [$F(4, 49) = 31.3$, $p < 0.0001$, $\eta^2 = 0.74$] and male agents [$F(4, 49) = 51.1$, $p < 0.0001$, $\eta^2 = 0.82$]. The results also indicate that type 4 expression gained the higher scores than other types for female agent (Figure 6(e)). On the other hand, for the same expression of male agent, the result indicates that type 5 expression gained the higher scores than the other types (Figure 6(m)).

For surprise expression, the results show statistically significant differences among conditions for female [$F(2, 29) = 50.4$, $p < 0.0001$, $\eta^2 = 0.78$] and male [$F(2, 29) = 32.8$, $p < 0.0001$, $\eta^2 = 0.84$] agents. The results also indicate that type 3 expression gained the higher scores than the other types for female agent (Figure 6(f)) and type 2 expressions gained the higher scores than the other types for male agent (Figure 6(n)). Thus, we have to decide to use expression types 3 and 2 for producing the surprise expression of female and male agent, respectively.

For irony expression, the results show statistically significant differences among conditions for female [$F(2, 29) = 30.69$, $p < 0.0001$, $\eta^2 = 0.69$] and male [$F(2, 29) = 34.32$, $p < 0.0001$, $\eta^2 = 0.87$] agents. The results also indicate that type 3 expression gained the higher scores than the other types for both female and male agents (Figures 6(g) and 6(o)).

For determined expression, the results show statistically significant differences among conditions for female [$F(2, 29) = 41.7$, $p < 0.0001$, $\eta^2 = 0.75$] and male

TABLE 5: Accuracy of recognizing emotive words.

Number of emotive keywords (N)	Correctly recognizes emotive words	Incorrectly recognized words (M)
100	93	7

[$F(2, 29) = 32.4$, $p < 0.0001$, $\eta^2 = 0.86$] agents. The results also indicate that type 2 expression gained higher scores than other types for female agent (Figure 6(h)) and type 3 expressions gained higher scores than the other types for male agent (Figure 6(p)). Thus, we have to decide to use expression types 2 and 3 for producing the surprise expression of female and male agents, respectively.

Table 4 summarizes the results of analysis. It shows the mean score for different expression types for both male and female agent. For angry, sad, fear, and irony, the same expression type for both male and female agents was chosen, that is, type 2, type 3, type 4, and type 3, respectively. For happy, disgust, surprise, and irony, different expression type was selected. It is due to the fact that in some cases intensity of facial expression shown by male and female is different.

(2) *Accuracy*. Table 5 summarizes the results of data analysis. We calculated the accuracy of the system in recognizing emotions and generating corresponding expressions using (1). This result revealed that the system is about 93% accurate in recognizing emotion from the input texts. From Table 5 we can see that some sentences produce wrong output expressions, because expressions are generated using the adjective detected first. For example, for the sentence “His sudden death made everyone sad,” the keyword “sudden” is detected first and the system produced the expression *surprise* rather than the expression *sad*.

5.2. *Experiment 2: To Evaluate Overall System*. In order to measure the acceptability and usability of the system, we design experiment and compared the proposed system with other emoticon-based systems (such as Yahoo Messenger).

5.2.1. *Experimental Design*. A total of 10 participants interacted with different professions that have experienced using emoticon-based chatting system [mean = 38, SD = 9.56]. We

TABLE 6: Comparative analysis based on the subjective evaluation between the proposed and emoticon based approaches.

Questionnaire items	Proposed method	Emoticon based method	Significant level (p)
	Mean (Std.)	Mean (Std.)	
Ease of operation	5.4 (0.87)	3.9 (0.84)	$p = 0.0003$
Reliability	5 (0.48)	1.7 (0.47)	$p < 0.0001$
Expressiveness	5.7 (0.82)	2.3 (0.48)	$p < 0.0001$
Suitability	4.4 (0.42)	1.2 (1.07)	$p < 0.0001$
Effectiveness	5.6 (0.82)	2.7 (0.51)	$p < 0.0001$
Appropriateness	5.6 (0.94)	2.7 (0.69)	$p < 0.0001$
Interestingness	6.3 (1.2)	3.5 (0.48)	$p < 0.0001$



FIGURE 7: A scene where the participant is interacting with the proposed system.

explained to the participants that the purpose of experiment was to evaluate the performance agent's behaviors to make them feel that the agent can effectively display its expression. The experiment had a within-subject design, and the order of all experimental trials was counterbalanced. We asked the participants in pairs (i.e., 10 participants are paired into 5 groups) to use the proposed system and also an existing emoticon-based system to chat with their respective partners. There was no remuneration for participants. Figure 7 illustrates a scene, where a participant is interacting with the proposed agent.

During their communication through the system, they interacted with each other using the agents and emoticons. Figure 8 shows the conversation window.

5.2.2. Measurements. After interacting, we asked participants to fill out a questionnaire for each condition. The measurement was a simple rating on a Likert scale of 1-to-7 where 1 stands for the lowest and 7 for the highest. The questionnaire had the following items:

- (i) *Ease of operation*: how easy was it to interact with the agent?
- (ii) *Reliability*: how reliable was the agent in generating the expression?
- (iii) *Expressiveness*: do you think that the agent is able to display emotions according to your emotive word(s)?
- (iv) *Suitability*: was the agent suitable for your interaction?
- (v) *Effectiveness*: was the agent effective in replicating your textual emotion into its facial expression?

(vi) *Appropriateness*: are the expressions generated by the agent appropriate against your input emotive text?

(vii) *Interestingness*: was the agent interesting or boring?

5.2.3. Results. Table 6 shows the results of the questionnaire assessment. We compared the 10 resultant pairs for each questionnaire item using t -test. The result shows significant differences for all items between proposed and traditional emoticon-based methods.

Figure 9 also illustrates these results. For ease of operation, the result shows the significant differences between two methods [$t(9) = 5.58, p = 0.0003$]. For reliability, the result indicates the significant differences between two methods [$t(9) = 21.6, p < 0.0001$]. In case of expressiveness, we also found the significant differences between methods [$t(9) = 12.7, p < 0.0001$]. Concerning suitability, analysis reveals that there is significant differences between two methods [$t(9) = 7.6, p < 0.0001$]. Significant differences between two methods are also found for effectiveness [$t(9) = 10.4, p < 0.0001$] as well as for interestingness [$t(9) = 8.3, p < 0.0001$]. For appropriateness, the result shows the significant differences between two methods [$t(9) = 7.2, p < 0.0001$].

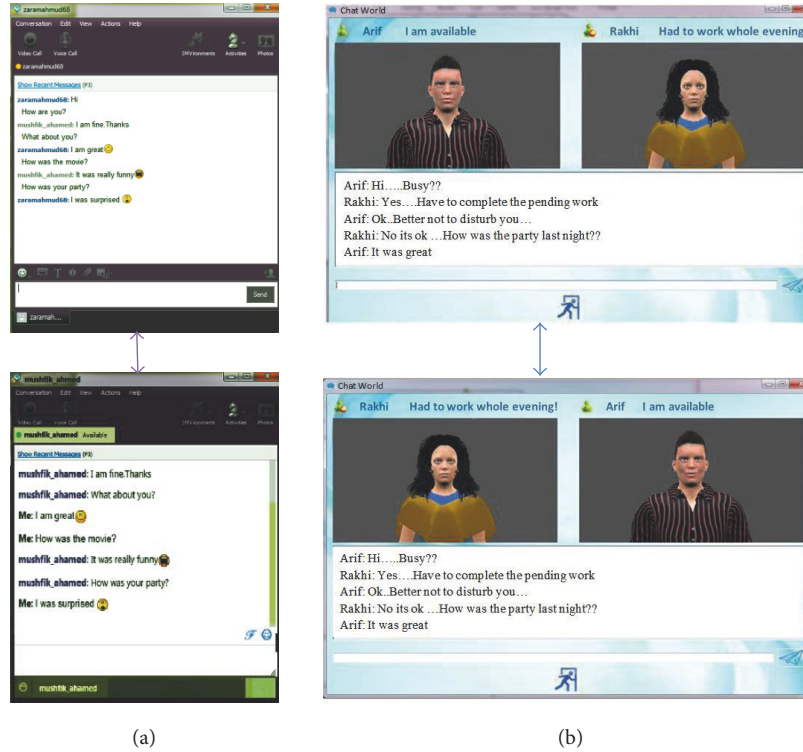
Although the current system has limited capabilities, the above analysis revealed that the proposed system outperforms the traditional emoticon-based system.

6. Discussion

The primary focus of our work is to develop a life-like character that can generate facial expressions with some facial movements as a means of communication. For this purpose, we developed an agent that can display eight facial expressions and three motions (eye blinks, head shaking, and head nodding) depending on the textual input of the users. This agent can be used in chatting scenarios in place of emoticons that will make the text-based chatting more interesting and enjoyable.

Results from experiment 1 were used to map appropriate facial expressions for corresponding emotions. Here, we carried out an evaluation of expression as psychological satisfaction is relative matter and varies from man to man.

Experiment 2 was conducted in order to measure the acceptability and usability of the system in terms of ease of operation, reliability, expressiveness, suitability, effectiveness, appropriateness, interestingness, and overall evaluation.



zaramahmud68: Hi
How are you?
mushfik_ahmed: I am fine. Thanks
What about you?
zaramahmud68: I am great.
How was the movie?
mushfik_ahmed: It was really funny
How was your party?
zaramahmud68: I was surprised

(c) Emoticon-based system

Arif: Hi.....Busy??
Rakhi: Yes.....Have to complete the pending work
Arif: Ok.....Better not to disturb you....
Rakhi: No its ok.....How was the party last night??
Arif: It was great

(d) Proposed agent based system

FIGURE 8: An example scenario of user's chatting through (a) existing emoticon-based system and (b) our proposed system and a sample of conversation between participants interacting through (c) emoticon-based system and (d) our proposed system.

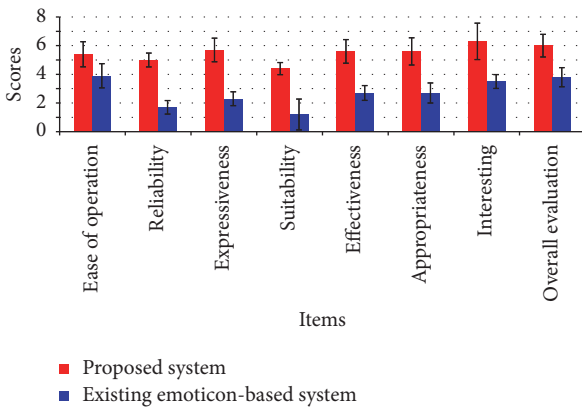


FIGURE 9: Results of questionnaire assessment based on the subjective evaluation.

Results revealed that the system is quite satisfactory to serve its purpose.

Although the current version of the agent has some limitations, it can be a better way of recreation for people and may represent themselves via this agent online. Full body embodiment with various gestures will enhance the interaction quality of the agent. Moreover, the relationship between cognition and expression is not yet well understood in the current work. These are left as future issues.

7. Conclusion

Virtual agent plays an important role in human-machine interaction, allowing users to interact with a system. In this paper, we focused on developing a virtual agent with expressive capability. To increase life-likeness of the agent, we tried to combine a model of emotions with a facial model, by providing a mapping from emotional states onto facial expressions. It is extremely complex and difficult to model. Moreover, emotions are quite important in human interpersonal relations and individual development. Our proposed

framework may not express a vast range of expression but the expression with some facial motion made by the agent is quite satisfactory. The overall experimental results show that the project is functioning quite well. Finally, we can say that the people using the system get some mental satisfaction for a moment. So the main motivation of providing an expressive intelligent is quite fulfilled. Adding chat sounds and animated emoticon may improve the quality and enjoyment of interaction in chatting scenario. These are left as future issues of this work.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] K. Höök, A. Bullock, A. Paiva, M. Vala, R. Chaves, and R. Prada, "FantasyA and SenToy," in *Proceedings of the Conference on Human Factors in Computing Systems, CHI EA 2003*, pp. 804–805, ACM Press, Lauderdale, Fla, USA, April 2003.
- [2] S. M. Boker, J. F. Cohn, B.-J. Theobald, I. Matthews, T. R. Brick, and J. R. Spies, "Effects of damping head movement and facial expression in dyadic conversation using real-time facial expression tracking and synthesized avatars," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3485–3495, 2009.
- [3] M. El-Nasr, T. Ioerger, J. Yen, D. House, and F. Parke, "Emotionally expressive agents," in *Proceedings of the Computer Animation 1999*, pp. 48–57, Geneva, Switzerland.
- [4] P. Ekman, E. R. Sorenson, and W. V. Friesen, "Pan-cultural elements in facial displays of emotion," *Science*, vol. 164, no. 3875, pp. 86–88, 1969.
- [5] S. Helokunnas, *Neural responses to observed eye blinks in normal and slow motion: an MEG study [M.S. Thesis]*, Cognitive Science, Institute of Behavioural Sciences, University of Helsinki, Finland, 2012.
- [6] D. Kurlander, T. Skelly, and D. Salesin, "Comic chat," in *Proceedings of the 1996 Computer Graphics Conference, SIGGRAPH*, pp. 225–236, August 1996.
- [7] K. Nagao and A. Takeuchi, "Speech dialogue with facial displays: multimodal human-computer conversation," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 102–109, USA, June 1994.
- [8] H. H. Vilhjalmsen and J. Cassell, "BodyChat: Autonomous communicative behaviors in avatars," in *Proceedings of the 1998 2nd International Conference on Autonomous Agents*, pp. 269–276, Minneapolis, USA, May 1998.
- [9] A. B. Loyall and J. Bates, "Real-time control of animated broad agents," in *Proceedings of Conference of the Cognitive Science Society*, USA, 1993.
- [10] I. Mount, "Cranky consumer: testing online service reps," *The Wall Street Journal*, 2005, <https://www.wsj.com/articles/SB110721706388041791>.
- [11] J. Cassell, T. Bickmore, M. Billinghurst et al., "Embodiment in conversational interfaces: Rea," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1999*, pp. 520–527, USA, May 1999.
- [12] L. Riccardo and C. Peiro, "A Facial Animation Framework with Emotive/Expressive Capabilities," in *Proceeding of IADIS International Conference Interfaces and Human Computer Interaction*, pp. 49–53, Italy, 2011.
- [13] K. Kramer, R. Yaghoubzadeh, S. Kopp, and K. Pitsch, "A conversational virtual human as autonomous assistant for elderly and cognitively impaired users? Social acceptability and design consideration," in *Lecture Notes in Informatics (LNI)*, vol. 220 of *Series of the Gesellschaft für Informatik (GI)*, pp. 1105–1119, 2013.
- [14] D. Ameixa, L. Coheur, P. Fialho, and P. Quaresma, "Luke, I am your father: Dealing with out-of-domain requests by using movies subtitles," in *Proceedings of the 14th International Conference*, vol. 8637 of *on Intelligent Virtual Agents (IVA)*, pp. 13–21, Boston, MA, USA, August 27–29, 2014.
- [15] A. B. Youssef, M. Chollet, H. Jones, N. Sabouret, C. Pelachaud, and M. Ochs, "Towards a socially adaptive virtual agent," in *Proceedings of the 15th International Conference*, vol. 9238 of *on Intelligent Virtual Agents (IVA)*, p. 3, Delft, Netherlands, August 26–28, 2015.
- [16] C. Straßmann, A. R. Von Der Pütten, R. Yaghoubzadeh, R. Kaminski, and N. Krämer, "The effect of an intelligent virtual agent's nonverbal behavior with regard to dominance and cooperativity," in *Proceedings of the 16th International Conference on Intelligent Virtual Agents (IVA)*, vol. 10011, 28 pages, Los Angeles, CA, USA, September 20–23, 2016.
- [17] S. Kopp, L. Gesellensetter, N. C. Krämer, and I. Wachsmuth, "A conversational agent as museum guide: design and evaluation of a real-world application," in *Lecture Notes in Computer Science*, T. Rist, T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, and P. Olivier, Eds., pp. 329–343, Springer-Verlag, London, UK, 2005.
- [18] S. Vosinakis and T. Panayiotopoulos, "SimHuman: A Platform for Real-Time Virtual Agents with Planning Capabilities," in *Proceedings of the of International Workshop on Intelligent Virtual Agents (IVA)*, vol. 2190 of *Lecture Notes in Computer Science*, pp. 210–223, Madrid, Spain, September 10–11, 2001.
- [19] D. Formolo and T. Bosse, "A Conversational Agent that Reacts to Vocal Signals," in *Proceedings of the 8th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2016)*, vol. 178, Springer International Publishing, Utrecht, Netherlands, June 28–30, 2016.
- [20] J. Gratch, A. Okhmatovskaia, F. Lamothe et al., "Virtual Rapport," in *Proceedings of the 6th International Conference of Intelligent Virtual Agents (IVA)*, vol. 4133, pp. 14–27, Marina Del Rey, CA, USA, August 21–23, 2006.
- [21] J.-W. Kim, S.-H. Ji, S.-Y. Kim, and H.-G. Cho, "A new communication network model for chat agents in virtual space," *KSII Transactions on Internet and Information Systems*, vol. 5, no. 2, pp. 287–312, 2011.
- [22] T. Weise, S. Bouaziz, H. Li, and M. Pauly, "Realtime performance-based facial animation," *ACM Transactions on Graphics*, vol. 30, no. 4, p. 1, 2011.
- [23] https://www.charamel.com/en/solutions/avatar_live_chat_moc-itotalk.html.
- [24] L. Alam and M. M. Hoque, "The design of expressive intelligent agent for human-computer interaction," in *Proceedings of the 2nd International Conference on Electrical Engineering and Information and Communication Technology, iCEEiCT 2015*, Bangladesh, May 2015.
- [25] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, "Affective computing and sentiment analysis," in *A Practical Guide to Sentiment Analysis*, vol. 5 of *Socio-Affective Computing*, pp. 1–10, Springer, Cham, Switzerland, 2017.
- [26] P. Chandra and E. Cambria, "Enriching social communication through semantics and sentics," in *Proceedings of the Workshop*

- on *Sentiment Analysis where AI meets Psychology (SAAIP)*, pp. 68–72, Chiang Mai, Thailand, November 13, 2011.
- [27] V. Anusha and B. Sandhya, “A learning based emotion classifier with semantic text processing,” *Advances in Intelligent Systems and Computing*, vol. 320, pp. 371–382, 2015.
 - [28] S. M. Mohammad, “Sentiment analysis: detecting valence, emotions, and other affectual states from text,” *Emotion Measurement*, pp. 201–237, 2016.
 - [29] W. Dai, D. Han, Y. Dai, and D. Xu, “Emotion recognition and affective computing on vocal social media,” *Information Management*, vol. 52, pp. 777–788, 2015.
 - [30] S. Poria, E. Cambria, N. Howard, G.-B. Huang, and A. Hussain, “Fusing audio, visual and textual clues for sentiment analysis from multimodal content,” *Neurocomputing*, vol. 174, pp. 50–59, 2016.
 - [31] E. M. G. Younis, “Sentiment analysis and text mining for social media microblogs using open source tools: an empirical study,” *International Journal of Computer Applications*, vol. 112, no. 5, 2015.
 - [32] P. Ekman and W. Friesen, *Unmasking the face: A Guide to Recognizing Emotions from Facial Clues*, Prentice-Hall, 1975.
 - [33] MakeHuman. <http://www.makehuman.org/>.
 - [34] Blender. <http://www.blender.org/>.

