We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,300

Open access books available

172,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

The Role of Facial Expressions and Body Gestures in Avatars for e-Commerce Interfaces

Dimitrios Rigas and Nikolaos Gazepidis
University of Bradford
United Kingdom

1. Introduction

As e-commerce user interfaces continue to expand, the need for interaction with multimodal content becomes noticeable (Böszörményi et al., 2001; Jalali-Sohi & Baskaya, 2001). User interfaces for e-Commerce (EC) applications occasionally use speaking avatars with facial expressions and body gestures to deliver information. The challenge is to provide a series of guidelines for e-commerce interfaces that attract the users' interest (Nemetz, 2000) and usable. Facial expressions and body gestures, as part of an expressive avatar may play an important role in interfaces. This paper describes three experiments that investigated the role of expressive avatars with facial expressions and body gestures in e-Commerce interfaces in terms of effectiveness, efficiency and user satisfaction.

2. Multimedia metaphors

2.1 Avatars

Avatars are interactive characters in real time virtual environments (VEs) (Dix et al., 2003; Sengers et al., 1999; Benyon et al., 2005; Bartneck et al., 2004). They are often depicted as three-dimensional (3D) animated human-like models (Theonas et al., 2008a). Users from any physical location can interact, communicate and cooperate with each other (Burford & Blake, 1999; Prasolova-Førland & Monica Divitini, 2002; Krenn et al., 2003) in Collaborative Virtual Environments or CVEs (Fabri et al., 1999). Other researchers (Hobbs et al., 1999; Thalmann, 1997) remark that VEs cover a wide range of applications such as educational, edutainment, e-commerce and simulation.

Fabri and Moore argue that when CVEs incorporate facial expressive avatars, they can be beneficial for people with special needs (e.g. autism) in terms of achievement and performance (Fabri & Moore, 2005). Emotional expressive avatars can also help with the interaction process of communicating information (Fabri et al., 1999). Fabri et al. experimented in a two-person messaging application that aimed to measure "richness of experience" (Fabri & Moore, 2005; Fabri et al., 2007). Their results indicated that users who interacted with the expressive avatars were more active, provided positive feedback, and enjoyed the experience. On the other hand, users who did not come into contact with

expressive avatars were less active and they did not appear to be more intensive in completing their tasks.

The reason for experimenting with these human-like characters in EC websites is mainly commercial; to see how much they attract individuals and encourage them to visit the site again in the future (Krenn et al., 2003).

In commercial web sites, there is a lack of a face to face communication between the customer and the seller. In addition, referring to a website that sells for example, electronic products are not tangible. Hence, it is necessary to experiment and evaluate some multimedia aspects that can be embedded in the designing of an EC site. It will measure the effect of audio-visual stimuli and multimodal interactions by introducing other modes of communication (see multimedia metaphors) between the user and the EC application. These metaphors will be used either separately or simultaneously, or with some combinations so as to have precise results (Ostermann & Millen, 2000).

It has been speculated, that human-like interface agents with a combination of facial expressions like moving eyebrows, head movement, smiling, quiver of the eyelids, opening and closing eyes, or lip movement synchronised with a text-to-speech generation system (Krenn et al., 2003; Cassell et al., 2001; Nijholt et al., 2000) will have a positive effect on consumers' decisions (Paradiso & L'Abbate, 2005) as they simulate a real-life character interacting in a realistic way with them. Several projects have been developed such as the BEAT (Cassell et al., 2001), the COGITO (Paradiso & L'Abbate, 2005), and the SoNG (Guerin et al., 2000).

Facial expressions give a more realistic interaction in human computer interfaces. The face is a means of expressing emotions, feelings, and linguistic information and due to the improvement of computer hardware (high performance graphics and processing speed) instances of cartoon-like and human-like synthesized faces have been investigated and developed for use in computer applications (Beskow, 1996a).

Animated or realistic characters are used in spoken dialogue interfaces, conveying information for verbal and non-verbal communication by several means of facial modalities e.g. lip-synchronization, eyes gazing and blinking, turn taking and further advanced modelling capabilities such as the use of gestures and motion (Beskow, 1996b). Beskow (Beskow, 1996a) developed a three-dimensional human face model complimented by a rule-based audio-visual text-to-speech synthesis system. The benefit of such an auditory-visual (bimodal) system as far as speech perception is concerned is that it could have a better success in an environment with a reduced level of acoustics and could be significantly used in applications for hearing impaired people (Cohen & Massaro, 1993).

Theonas et al. (Theonas et al., 2008a; Theonas et al., 2008b), conducted an experiment in a real academic lecture environment to study the relation of the facial expressions of three lecturers with the way students react based on lecturers' expressions. An observer attended all the lectures to study the role of lecturers' expressions and whether their expressions motivate students and increase their interest during the lecture. He also experimented in a virtual classroom aiming to study virtual lecturers' expressions to evaluate students'

performance. His experimental study indicated that being more expressive (smiling) can increase students' motivation and interest and can also make them more enthusiastic towards the lecturer. As a result, it positively affects the students' learning and performance.

Li Gong (Gong, 2006) from Ohio State University conducted a research on digital characters to measure the effect of whether a happy expression is better than a sad one when being used from talking-head interface agents. He states that "...emotion is an essential factor in human psychology for it conveys feelings and attitudes, regulates motivation colours cognition and affects performance" (Gong, 2006). According to Massaro as well (Massaro, 1998) when synthetic faces are combined with speech to generate an animated-talking face in order to convey information or emotion, the results are more effective.

Human communications often involve the use of verbal communication (speech, writing) (Wiki, 2005; Hartley, 2003) that means direct or indirect contact with a person or a number of people, and in some other circumstances it involves non-verbal signals (eye contact, facial expressions, hand, arm, and leg gestures, body postures) where several times could be more expressive and meaningful in a conversation than the use of words or talks. The study of the body language and its non-verbal parameters is widely used in our everyday communication or business life and whoever knows how to use it, realises what a fascinating skill it is, enhancing his communication in every bit of life (Kyle, 2001).

Beskow (Beskow, 1996b) talked for a set of parameters in communication, the verbal and the non-verbal. The verbal refers to the control of speech articulation and the morphing and shaping of lips and mouth during a speech process whereas the non-verbal communication refers to facial expressions and gestures.

In order to give emphasis to our speech or to point at an object the movement of the body, of hands and head, play a major role. Gestures are used widely in our everyday life and they are also known as 'body language' communication (Cohen & Massaro, 1993).

McBreen (McBreen, 2001) with the ANOVA taking agent gender test experiments found that human-like gestures complement the agent's human-like appearance and she states that "gestures promote friendliness, politeness and lifelikeness" and that subjects significantly preferred female gestures than male ones. Cassell (Cassell, 2001) states that hands rather than eyes and speech are an excellent representation tool of events or objects and more specifically on describing ambiguous contexts. Pease (Pease, 1981) states that "many researchers rely on the idea that verbal communication is primarily used for conveying information where as the non-verbal aspect is for negotiating interpersonal attitude and sometimes they can substitute some verbal meanings". Regardless the culture, geographical parts, traditions and folkways there are some common gestures that people use. For instance people are showing that they are happy with a smile on their face, mad by frown their eyebrows, or a negation by shaking the head from side to side. On the other hand, there are some non-verbal gestures that differ from country to country or it is possible the same gesture to have a different meaning. The O.K. gesture for instance, in the USA it has the meaning that everything is fine, in France it has the meaning of zero notation or nothing, in

Japan means money and in Brazil has the meaning of insulting someone. Pease also states that "When in Rome do as the Romans do" (Pease, 1981).

2.2 Multimodal metaphors

Speech metaphor is often used in multimodal user interfaces so as to provide users with feedback along with the graphical environment about system's current state (Preece et al., 1994) and it is a very useful tool especially for visually impaired users (Lines & Home, 2002). We distinguish two types of speech; natural and synthesised. Natural speech output is a digitally recorded message of a male or female spoken word. It is often useful for applications that require short sentences to be spoken but a dynamic use of recorded speech by incorporating short recorded messages (as building blocks) is a complex process given the need for grammatical structure, context, tone changes and phonemes. Large storage capacity is also required due to the vast vocabulary. Synthetic speech output is produced using a speech synthesizer. It can be generated mainly by two methods: Concatenation or Synthesis by Rule (also referred to as formant synthesis) (Preece et al., 1994; Lines & Home, 2002). Using the concatenation method, digital recordings conducted by real human speech are stored and later on controlled as single words or sentences in a computer (Preece et al., 1994). An example based on concatenation is when someone uses a phone card and a recorded voice informs the user how many minutes left in the card. The audio message is digitally recordings of each digit separately, controlled by the computer system generating the spoken message. Synthesis by rule, involves the combination of synthesised words generated by rules of phoneme. It is useful for large vocabularies and as a result the quality of speech produced is poorer compared to the concatenation method (Lines & Home, 2002). Janse (Janse, 2002) had studied the perception of natural speech compared to a synthetic speech and derived that although synthetic speech is becoming more and more intelligent, natural speech still more comprehensible for listeners. A numerous of studies on synthesised speech have shown that natural speech still more intelligible and comprehensible than the synthetic. Voltrax, Echo, DECTalk, Voder, are some good examples of speech synthesizers developed in the past but with poor quality (Reynolds et al., 2002; Lemmetty & Karjalainen, 1999).

Computer-based systems offer a speech technology, known as Text-to-Speech (TTS) synthesis technology (Dutoit, 1999). TTS systems have the ability to read any arbitrary text, they analyze it and after converting it, they output it as a synthesised spoken message comprehensible by the user (Schroeter et al., 2000; Wouters et al., 1999). TTS technology is widely used in software applications, and many corporations are taking into account the benefits of involving this technology in EC websites (Kouroupetroglou & Mitsopoulos, 2000). TTS brings out new issues for the development of EC systems (Xydas & Kouroupetroglou, 2001) and provides the scope for new innovative applications.

2.2.1 Facial expressions and body gestures in avatars

This experiment aimed to investigate usability (effectiveness and user satisfaction) of avatars with facial expressions only or avatars with facial expressions and body gestures, and compare the use of expressive avatars with a typical textual and graphical metaphors in an e-commerce interface. The experiment was performed with 42 users in a specially

developed experimental platform. This platform presented three products using three different conditions. The first two conditions involved avatars with facial expressions and the same facial expressions but with the addition of body gestures. The third condition provided typical textual and graphical representations. During the experiment, three products were presented randomly to the users using, again randomly, the three different conditions of presentation.

Figure 1 shows users' preferences for each presentation method used in the experiment and their preferences for the multi-modal metaphors. Almost 50% of the users described the text and graphic presentations of the products as poor. Another 33% and 12% described that particular presentation as good and very good respectively. The avatar with the facial expressions was chosen and ranked positively almost by every user. The total of good and very good ranking for the avatar with the facial expressions almost reached the ultimate acceptance level of 90% satisfaction and only a 10% described it as poor but none of the users judged it as very poor. Similarly to the avatar with the facial expressions, the avatar with the facial expressions and body gestures had 89% of positive views.

As far as the multimedia metaphors' ability to play an important role when shopping online, users decided that Speech output could be considered very important getting the majority of the users' preference with an 87%. Moreover, graphics are the most essential part in a website and play an important role for all the 42 users that participated in the experiment. Furthermore, avatars with facial expressions is viewed as positive with an 80% and it is believed that it could help users when shopping online whilst avatar with facial expressions and body gestures comes with a 67% of positive views and a 33% of negative ones.

Figure 1 also shows the percentages of the way that users selected a presentation method for each of the three products presented in random order. It also presents the figures for the way users selected or rejected a presentation method, regardless of the product. Initially, users selected product 1 mostly with the avatar with facial expressions (78.57%) and as a second preference the avatar with facial expressions and body gestures (50%); in the third place comes the text and graphics way with only 28%. Here it is necessary to mention that the avatar with facial expressions and body gestures was not well manipulated it could be compared with the avatar with facial expressions and body gestures of the other two products. Product 2 and product 3 were chosen by almost every user with avatar with facial expressions and body gestures getting over 85%-90% of their preference, while text and graphics was chosen by a 40% or less. Figure 3 also shows that regardless of the product, about 64% of the users in total did not choose the text and graphics presentation, 88% chose the avatar with facial expressions and only a 12% rejected it. Lastly, the avatar with facial expressions and body gestures presentation was selected by more than 75%.

Fig. 1. Users’ preferences and views regarding the presentation methods and multimodal metaphors as well as the results in percentages of the way users selected the presentation according to product and regardless of it.

These results were significant for both facial expressions ($\chi^2 = 24.38$) and body gestures ($\chi^2 = 11.52$), but not significant for the textual with graphics presentation ($\chi^2 = 3.42$) at p-value 0.05 and critical value at 3.84.

Table 1 shows the percentages and the details related to three products. When P1 was selected with the text and graphics, it achieved 100% of correct answers by the users. However, there was only a small number of users that selected that way of presentation minimising error-making (distinctly 4 out of 14). For the successful non-selected answers for the text and graphics way the average dropped to 80%.When the avatar with facial expressions was chosen, the corresponding success for the selected way was over 80% almost 10% more than the avatar with facial expressions and body gestures. When users did not select these two ways of presentation the success rate of answering the product’s question was about 65% for the avatar with facial expressions and about 94% on average for the avatar with facial expressions and body gestures. On the other hand, for the successful non-selected answers the percentages are about 65% for the avatar with facial expressions and quite higher for the avatar with facial expressions and body gestures reaching almost 95%.

As far as product 2 (P2) is concerned, there are some fluctuations regarding the successful answers within the 3 methods. Based on average correct answers it can be seen that the text

and graphics reached high as 80%, the avatar with facial expressions a bit above 81% and the avatar with facial expressions and body gestures presentation is at 76.92%, almost 77% for the selected questions. When the corresponding presentation methods were not selected the correct answers for text and graphics were 81.25%, for the avatar with facial expressions 50% due to the very small number of users that did not select that method, and for the avatar with facial expressions and body gestures 75% for the same reason as the avatar with facial expressions.

Product 3 (P3) illustrates absolute success (100%) for the text and graphics when selected, again due to the small number of users who chose it, and at the same time almost total success for the avatar with facial expressions with 96.15%. The average slightly dropped for those users who chose the avatar with facial expressions and body gestures that is 79.16%. For the non-selected ways of presentations the mistaken answers for all of the presentation methods are null for avatar with facial expressions and body gestures and nominal for the text and graphics method.

			Product 1	Product 2	Product 3	All Products
Text and Graphics	Question 1	Selected	100%	100%	100%	100%
		Not Selected	70%	87.5%	88.8%	81.4%
	Question 2	Selected	100%	60%	100%	86.3%
		Not Selected	90%	75%	100%	88.8%
	Mean	Selected	100%	80%	100%	93.3%
		Not Selected	80%	81.2%	94.4%	85.1%
Avatar with facial expressions	Question 1	Selected	81.8%	69.2%	92.3%	81%
		Not Selected	66.6%	100%	100%	80%
	Question 2	Selected	90.9%	84.6%	100%	91.8%
		Not Selected	66.6%	0%	100%	60%
	Mean	Selected	86.3%	76.9%	96.1%	86.4%
		Not Selected	66.6%	50%	100%	70%
Avatar with facial expressions and body gestures	Question 1	Selected	71.4%	46.1%	66.6%	59.3%
		Not Selected	100%	50%	100%	92.3%
	Question 2	Selected	71.4%	92.3%	91.6%	87.5%
		Not Selected	85.7%	100%	100%	80%
	Mean	Selected	71.4%	69.2%	100%	73.4%
		Not Selected	94.11%	75%	100%	86.9%

Table 1. Results in percentages for successful answers for selected and non-selected ways of presentations for each product & all products.

Table 1 also shows the total results of all products for both selected and non-selected presentations. It can be observed that in total the text and graphics reached 93% of successful answers from the users, 86.48% for the avatar with facial expressions and lastly for the avatar with facial expressions and body gestures users reached a score of 73.48% when selecting that presentation. On the other hand, the non-selected answers range as follows: about 85% for the text and graphics, 70% for the avatar with facial expressions and 86% for the avatar with facial expressions and body gestures.

During the experiment two main observations were made. First, when the avatar with facial expressions and body gestures was demonstrated to the users with speech metaphor, they focused more on the speech rather than the text description of the products. Second, many users were pleasantly surprised when they were shown the avatar with facial expressions and body gestures concentrating more on the animation (facial expressions, lips synchronisation, body and hand gestures, and body postures) rather than the specifications of the products.

2.2.2 User preference of specific facial expressions and body gestures

(a) Experiment with Interactive Context

This experiment measured effectiveness user satisfaction in realistic interface circumstances of an avatar with 13 facial expressions and 9 body gestures that demonstrated two products. The first product (Windows Vista) was presented with facial expressions and the second product (Nintendo Wii) with full body gestures. Users ($n=42$) were asked to indicate their perception (positive or negative) after each random presentation that used a specific facial expression and body gesture.

Figure 2 shows that all positive facial expressions obtained high results in positive views by users (i.e. over 85%). On the other hand, the negative facial expressions had only 5% of positive views by users. Lastly, according to the users' ratings for the neutral expressions, the neutral expression had 83% (35 out of 42 users) of a positive view and two thirds (67% or 28 out of 42) of the users rated the thinking expression positively.

Fig. 2. Users' positive views on facial expressions and body gestures.

Figure 2 also shows that the positive body gestures were viewed positively by the users with some fluctuations in the percentages. The open palms obtained a positive score of 98% (41 out of 42) of users' answers, followed by an 88% (37 out of 42 users) for the hands steeping. An approximately 80% of the users viewed the hands clenching (34 out of 42 users) positively and the head up (35 out of 42 users) and another 74% of users found the chin stroking positive (31 out of 42 users).

According to the results for the negative body gestures, the majority of the users (i.e. over 85%) expressed a negative impression during their demonstration, especially for the face scratching (93% or 39 out of 42 users) and hiding face (98% or 41 out of 42 users).

2.2.3 Comparing facial expressions, body gestures and textual presentations

Lastly, the second section compared three interfaces (text and graphics, avatar with facial expressions and avatar with facial expressions and body gestures), where users had to select 2 out of 3 high-tech products according to the presentation method and regardless of the product information and then answer a number of questions upon each presentation of a product. The brand names of the products were not presented to avoid any influence on users' opinion. The description of each product was taken from CNET website. Each user was presented these products in a different order (GPS Navigator, Multimedia Player and PDA Phone) by using these methods of presentation, which followed 42 different combinations. Each product was presented twice. Product descriptions were categorised into short, medium and long in terms of text in order to investigate the error rating on users' answers. The effectiveness of communication metaphor, recall of information by users and learnability by users of each product according to interface were measured. This section discusses the users' views on the presentation method and the multimodal metaphors, the selection of users according to the presentation of products and the successful answers for the three product descriptions used in the experiment.

Figure 3 shows users' preferences for each presentation method used in the experiment and users' general viewpoint of multimodal metaphors. More than 60% (26 out of 42) of the users described the text and graphics method of presentation as very poor or poor expressions and about 1/3 of them as good. The presentation method with the avatar with facial expressions reached a level of acceptance of 96% (40 out of 42 users). The avatar with facial expressions and body gestures achieved a 46% of good, 40% as a very good and 14% poor.

Concerning the role of the multimodal metaphors could play an important role, the Speech output was chosen by a 90% (38 out of 42) of the users. At the same level was the preference for the Graphics with a 98% (41 out of 42 users), making it the most important metaphor when browsing a website. Moreover, avatar with facial expressions is being viewed by 83% (35 out of 42 users) of the sample positively, whilst avatar with facial expressions and body gestures is ranked at similar levels with an 81% (34 out of 42) of positive views from the users.

Figure 3 also depicts the percentages of the way users selected a presentation method among the three products presented in a random order as well as the percentages of the

presentation method regardless of the products. The text presented was 112 words for product 3, 207 for product 1, and 289 for product 2.

Fig. 3. Preferences and views of users regarding the presentation methods and multimodal metaphors, as well as the results in percentages of the way users selected the presentation method according to the product and regardless of it.

Results show that P1, P2, and P3 were mostly selected when an avatar with facial expressions was used, with percentages over 85%. The percentage was also large for all products when the avatar with facial expressions and body gestures was chosen by users, with an average score of 78%. As far as the text and graphics presentation is concerned, the highest score was achieved for P3 (i.e. the shortest length) with a 46%. P1 has a percentage of 35.71% and P2 was chosen by the smallest number of users compared to any other product with the low score of 15%. Figure 3 also shows the percentages of users that chose a presentation method regardless of the product. The avatar with the facial expressions was chosen by the majority of the users, nearly 90% followed by the 78% of the avatar with facial expressions and body gestures and only a 33% of the users selected the textual method of presentation. These results were significant for both facial expressions ($\chi^2 = 24.38$) and body gestures ($\chi^2 = 13.71$), but not significant for the textual with graphics presentation ($\chi^2 = 1.52$) at p-value 0.05 and critical value at 3.84.

Table 2 shows the percentages and the details derived for P1 upon the selection of a presentation method. When P1 was selected with the text and graphics, the mean value of correct answers from the users was approximately 65%. Compared to the other presentation methods, only 5 users out of 14 selected this method. However, for the non-selected presentation method results, the score was approximately at 63%. When the avatar with the facial expressions was selected, the successful rate was relatively high at about 83%. A

similar percentage was achieved when the avatar with the facial expressions was not selected which occurred with a small number of users (2 out of 14). When the avatar with facial expressions and body gestures was selected, the corresponding success was slightly lower than the avatar with facial expressions at about 79%. Likewise the percentage is high at 78% but the number of users (3 out of 14) and the result obtained is not representative.

Product 2 shows some fluctuations regarding the successful answers of the presentation methods chosen that were selected by users due to the long length of text. The method that achieved the highest score was the avatar with facial expressions and body gestures with 63.88%, followed by the avatar with facial expressions and body gestures with a 60.60%, and lastly the text and graphics with the lowest percentage of 33%. Approximately 50% of the users failed to answer correctly when the text and graphics was not selected (11 out of 13). Equal percentage applies for the avatar with facial expressions when it was not selected. Only 2 users though did not select it, therefore the error rates are not representative. For the not-selected answers given with the avatar with facial expressions and body gestures the mean rate of success was 66%.

			Text and Graphics		Avatar with facial expressions		Avatar with facial expressions and body gestures	
			Correct answers	mean	Correct answers	mean	Correct answers	mean
Product 1 (medium)	Selected	Q1:	4/5=80%	66.6%	9/12=75%	83.3%	7/11=63.6%	78.7%
		Q2:	2/5=40%		11/12=91.6%		9/11=81.8%	
		Q3:	4/5=80%		10/12=83.3%		10/11=90.9%	
	Not Selected	Q1:	4/9=44.4%	62.9%	1/2=50%	83.3%	2/3=66.6%	77.7%
		Q2:	6/9=66.6%		2/2=100%		2/3=66.6%	
		Q3:	7/9=77.7%		2/2=100%		3/3=100%	
Product 2 (long)	Selected	Q1:	0/2=0%	33.3%	7/12=58.3%	63.8%	5/11=45.4%	60.6%
		Q2:	2/2=100%		10/12=83.3%		8/11=72.7%	
		Q3:	0/2=0%		6/12=50%		7/11=63.6%	
	Not Selected	Q1:	4/11=36.3%	48.4%	2/2=100%	50%	3/4=75%	66.6%
		Q2:	7/11=63.6%		0/2=100%		3/4=75%	
		Q3:	5/11=45.4%		1/2=50%		2/4=50%	
Product 3 (short)	Selected	Q1:	4/7=57.1%	80.9%	11/13=84.6%	92.3%	10/11=90.9%	87.8%
		Q2:	6/7=85.7%		12/13=92.3%		8/11=72.7%	
		Q3:	7/7=100%		13/13=100%		11/11=100%	
	Not Selected	Q1:	5/8=62.5%	70.8%	1/1=100%	100%	1/2=50%	50%
		Q2:	4/8=50%		1/1=100%		1/2=50%	
		Q3:	8/8=100%		1/1=100%		1/2=50%	

Table 2. Results in percentages for correct answers for selected (A) and non-selected (B) ways of presentation for all products.

When P3 was selected due to its short length of text, there were no major statistical differences across the three presentation methods. Correct answers were higher than any other product. A remarkable high rate of successful answers was observed in the text and

graphics when selected, approaching the 81%. Taking into account the other two methods, avatar with facial expressions and body gestures was around 88% and avatar with facial expressions was as high as 92%. For the text and graphics results were quite successful for the non-selected answers with a 70%. The avatar with facial expressions and body gestures, due to the small number of users that had not selected them cannot be fully judged.

2.2.4 Combining facial expressions and body gestures

This experiment aimed to verify the positive, negative, and combined effect and to measure effectiveness (i.e. correct answers), efficiency (i.e. time taken to answer with 60 sec being the maximum), and user satisfaction of expressive avatars with the best and least suitable facial expressions and body gestures to e-commerce interfaces. Table 3 shows the 13 facial expressions and 9 gestures used in the four experimental conditions. These were the best rated facial expressions (BRFE), the best rated body gestures (BRBG), the least rated facial expressions (LRFE) and the least rated body gestures (LRBG).

Conditions Expressive Avatars	Facial Expressions												Body Gestures									
	Positive						Negative						Positive				Negative					
	Amazed	Happy	Interested	Pos. Surprised	Neutral	Thinking	Angry/Mad	Disapproving	Sad	Upset	Neg. Surprised	Disgusted	Tired/Bored	Head Up	Open Palms	Chin Stroking	Hands steeping	Hands Clenching	Arms Folded	Face Scratching	Hiding Face	Legs Crossed
BRFE	✓	✓	✓	✓	✓	✓																
BRBG	✓	✓	✓	✓	✓	✓								✓	✓	✓	✓	✓				
LRFE							✓	✓	✓	✓	✓	✓	✓									
LRBG							✓	✓	✓	✓	✓	✓	✓						✓	✓	✓	✓

Table 3. List of the facial expressions and body gestures consist each interface.

For the statistical analysis, the normal distribution of the continuous variables (e.g. time) was assessed by the nonparametric Kolmogorov-Smirnov test. The non-parametric Friedman test was applied as the time taken by users to answer each question in all conditions was not normally distributed. Furthermore, paired comparisons between experimental conditions (BRFE vs. BRBG, BRFE vs. LRFE, BRFE vs. LRBG, BRBG vs. LRFE, BRBG vs. LRBG and LRFE vs. LRBG) were performed by the non-parametric Wilcoxon test. Dichotomous variables (e.g. percentage of correct or incorrect answers for each question) were compared between interfaces by the McNemar test. The significance level was set at 0.05 with p-values < 0.05 indicating statistically significant differences.

Figure 4 shows the percentages of the users that correctly answered questions within the criterion time (60 sec) in each task for each experimental condition. The BRFE was the best performing condition with correct answers that reached, 96% of the users for task 4, approximately 91% for tasks 1 and 3, and 86% for task 2. The second best performing

condition was the BRBG with percentages of correct answers that varied between 85% and 90%. In the LRFE and LRBG conditions, percentages dropped and ranged between 72% to 63% with the exception of 75% in task 2 in the LRBG. Similar performance of four conditions can also be observed using the mean values which are 91.3%, 87.82%, 70%, and 68.26% for BRFE, BRBG, LRFE, and LRBG respectively. Results indicated that 85% to 96% of the sample taken up to 60 seconds to answer all questions in the conditions in the BRFE and BRBG conditions respectively. These figures drop to approximately 65% to 75% in the LRFE and LRBG conditions.

Table 4 shows a comparison of the correct answers of the recall and recognition questions between the experimental conditions. The McNemar test was used to calculate the p-values obtained and determine the significance. The BRFE and BRBG conditions were significant when compared with the LRFE and LRBG. There was no significance difference between BRFE and BRBG as well as between the LRFE and LRBG conditions.

In the BRFE condition, 45 out of 46 (97.8%) users answered correctly questions 1 (recall) and 3 (recognition) and 40 out of 46 (87%) users questions 2 (recall), 4 (recognition) and 5 (yes/no). On overall, users performed well in all tasks in this condition. The results for the BRBG are similar to the BRFE. Users achieved the highest score in question 3 (recognition), as well as question 1 (recall) with correct answers being 97.8% and 91.3% respectively. Question 2 (recall) was answered by 40 out of 46 users (87%) and the rest of the results for BRBG condition ranged between 80% to 82%

Fig. 4. Correct answers and mean values for each experimental interface and for each case used (n=46).

Conditions	All		Recall				Recognition				Yes / No	
	Questions		Question 1		Question 2		Question 3		Question 4		Question 5	
	%	p	%	p	%	p	%	p	%	p	%	p
BRFE vs. BRBG	91.3 vs. 87.8	0.280	97.8 vs. 91.3	0.375	87 vs. 87	1.000	97.8 vs. 97.8	1.000	87 vs. 82.6	0.774	87 vs. 80.4	0.581
BRFE vs. LRFE	91.3 vs. 70	<0.001	97.8 vs. 78.3	0.012	87 vs. 71.7	0.092	97.8 vs. 95.7	1.000	87 vs. 54.3	0.003	87 vs. 50	0.001
BRFE vs. LRBG	91.3 vs. 68.3	<0.001	97.8 vs. 71.7	0.001	87 vs. 69.6	0.077	97.8 vs. 95.7	1.000	87 vs. 58.7	0.004	87 vs. 45.7	0.001
BRBG vs. LRFE	87.8 vs. 70	<0.001	91.3 vs. 78.3	0.146	87 vs. 71.7	0.143	97.8 vs. 95.7	1.000	82.6 vs. 54.3	0.007	80.4 vs. 50	0.003
BRBG vs. LRBG	87.8 vs. 68.3	<0.001	91.3 vs. 71.7	0.035	87 vs. 69.6	0.115	97.8 vs. 95.7	1.000	82.6 vs. 58.7	0.043	80.4 vs. 45.7	0.002
LRFE vs. LRBG	70 vs. 68.3	0.752	78.3 vs. 71.7	0.648	71.7 vs. 69.6	1.000	95.7 vs. 95.7	1.000	54.3 vs. 58.7	0.845	50 vs. 45.7	0.839

Table 4. P-values among interfaces for all questions. (for p-values with many decimals close to 0 “<” used).

The results of the LRFE condition are considerably lower. It can be seen that users achieved a 78% for the first question and approximately 72% for the second question of the recall tasks. For the recognition tasks, users performed well in Question 3 with 44 out of 46 users (95.7%) answered correctly. However, the correct answers in the remaining questions were equally distributed and percentages range up to 50%. The LRBG condition was the least performing. Approximately 70% of the users answered the recall questions correctly. The third question (recognition) was answered correctly, as in the other three conditions by 44 out of 46 users (95.7%). However, the number of users who answered the fourth question (recognition) correctly dropped to 27 users out of 46 (58.7%). The fifth question (yes/no choice) was answered correctly by a small number of users (i.e. 21 out of 46 or 45.7%).

Table 5 shows that over 89% of the users answered the recall and recognition questions under the BRFE and BRBG and a range between 70% to 77% for the other two conditions. As for the last yes/no question, the BRFE experimental interface comes first on users’ correct answers with an 87%, followed by the BRBG, with a slightly lower percentage of around 80%. Lastly, the percentages of the correct answers for the yes/no question have noticeably dropped for the other two conditions as half the number of the users answered correctly when the presentations were made using the LRFE, whereas the percentage was below 50% for LRBG, at 45.65%.

	Recall	Recognition	Yes/No
BRFE	92.39%	92.39%	87%
BRBG	89.13%	90.22%	80.43%
LRFE	75%	75%	50%
LRBG	70.65%	77.17%	45.65%

Table 5. Users’ correct answers (n=46) shown in percentages for recall, recognition and yes/no question for each experimental interface (i.e. BRFE, BRBG, LRFE and LRBG)..

Figure 5 shows the results of user satisfaction for each experimental interface as well as the importance of the multimedia metaphors in e-Commerce applications. The most popular condition was the BRFE with the majority of users (93.48%) being satisfied with the facial expressions used. The second most preferred condition by users was the BRBG with approximately 85% of user preference. The level of user satisfaction in LRFE and LRBG was approximately 30.43% and 18%. Figure 5 also shows user satisfaction of other multimodal metaphors used in the interface. The use of speech output (recorded) was considered to be a good feature by 93.5% and graphics by 96% of the users. The use of avatars that incorporated facial expressions was considered by approximately 80% of the users as an essential metaphor for an e-Commerce interface. Finally, about 70% of the users preferred the full animated body that incorporated gestures with facial expressions. All these results are in agreement with the empirical data.

Fig. 5. User satisfaction for each experimental interface and specific multimodal metaphors.

3. Discussion

The first experiment showed that users preferred the avatar with facial expressions (88% or 37 out of 42 users) followed by the avatar with facial expressions and body gestures (76.1% or 32 out of 42 users). The text and graphics presentation was the least preferred method (35.7% or 15 out of 42 users). These findings were demonstrated consistently for all three

products or presentation methods used in the experiments. One can therefore assume that this type of simulated face-to-face communication would positively contribute to an e-Business system.

Users were observed to focus more on the products themselves when these products were presented by using the avatar with facial expressions and body gestures. Post-experimental interviews suggested that users thought that the avatar with facial expressions and body gestures was amusing and entertaining and their attention was directed more towards the technical aspects of the presentation than towards the specifications of the products presented. This result is therefore confirmed by empirical observations as well as by user views. Users demonstrated the ability to remember the information communicated by using the avatar with facial expressions and body gestures could better than when the information was communicated with text and graphics.

The experiment with an interactive context indicated the users' views on avatar with facial expressions and body gestures with the presentation of two products. Results showed that positive expressions generally got over 85% of positive views; distinguishably the positively surprised expression (95%) was followed by the amazed and the happy expressions (90%). The negative expressions got high percentages of negative impressions from the users. Namely, the angry/mad, disgusted, sad and upset got 98% of negative users' views, whereas the remaining negative expressions got 95%. Neutral and thinking got 83% and 67% of positive views respectively.

As for the positive gestures the open palms once again got a high percentage with 98% of positive views followed by the hands steeping with 88%. The rest of the positive gestures reached approximately 88%. The negative gestures confirmed the initial hypothesis and results showed negative views by users for all of them. However, the most negatively rated body gestures were the hiding face with 98%, the face scratching and the legs crossed with 93% and lastly the arms folded with 86%.

The third experiment investigated the "best rated" and "least rated" by users facial expressions and body gestures in a human-like expressive avatar in a simulated e-commerce interface. The results indicated that on overall the percentage of correct answers in BRFE and BRBG conditions ranged from 85% to 96%, but dropped to 65% and 75% for the LRFE and LRBG. The mean values of the users' correct answers ranged between 87% to 91% for the BRFE and the BRBG and 68% to 70% for the LRFE and the LRBG conditions. The correct answers per interface per question showed an exceptional performance of the users for BRFE and BRBG. The BRFE results ranged from 87% to almost 98%, and the BRBG results ranged from 80% to approximately 98% for correct answers. On the other hand, correct answers are lower for the other two interfaces with results ranging from approximately 45% to 71%. However, 95% of the users managed to answer correctly the third question (recognition) using the LRFE and LRBG experimental interfaces. The level of user acceptance and satisfaction was also high for the BRFE and BRBG conditions (93.48% and 84.78% respectively) but low for the LRFE and the LRBG (less than 31%) conditions.

The results also showed that users prioritise metaphors in the order of speech output (93%) and graphics (95%), avatars with facial expressions (78%), and avatars with facial expressions and body gestures (69%). Users' correct answers of the recall, recognition and yes/no answers for the BRFE and the BRBG differed significantly from the LRFE and the LRBG. Avatars with a combination of positive facial expressions or facial expressions with body gestures contribute positively to the interaction process between the user and the interface and can successfully communicate product information in an e-commerce interface. On the other hand, a combination of negative facial expressions or facial expressions with body gestures could distract the user from perceiving the information.

4. Conclusion

The results obtained from this experimental study were interpreted and concluded with some key points on usability aspects of B2C interfaces. These empirical findings can be interpreted into a set of usability guidelines that would effectively enhance the usability of B2C interfaces. The guidelines derived are structured into length of text, speech metaphor, use of facial expressions, use of body gestures and combination of facial expressions and body gestures.

4.1 Length of text

When short text descriptions needed to be communicated, expressive avatars did not effectively enhance the usability of an interface. When short text descriptions needed to be communicated, expressive avatars did not effectively enhance the usability of an interface. User error ratings did not significantly differ among presentation methods (i.e. text with graphics, avatar with facial expressions, avatar with facial expressions and body gestures). In cases where the text was longer, results obtained indicated that facial expressions and body gestures effectively and efficiently contributed to the interface becoming more usable. However, a designer of a B2C platform should always have an option of textual method of presentation no matter what the length of the text is, as there are some users who find animated avatars annoying and do not want to familiarise themselves with the multimedia methods of presentation to communicate information.

4.2 Speech metaphor

Recorded speech enhanced the usability of the interface and kept the attention of the user to the presentation when being used efficiently. A clear, "crisp" speech articulation pattern maintained the interest of the user throughout the presentation. When speech tone was used with positive facial expressions, users showed more interest and they paid more attention to the text description. Results obtained indicate a significant increase in the efficiency and effectiveness of the users' interaction with the interface. This, however, did not happen with the negative facial expressions. In addition to the text, speech also enhanced the usability of the interface.

4.3 Use of facial expressions

Expressive avatars give a sense of presence in an interface and 'engage' users in the interface as users were often observed to mimic the avatars' expressions. The 13 facial expressions

investigated in the experiments, in both the absence and presence of an interactive context, demonstrated that only the positive facial expressions contributed positively to the interaction process between the user and the interface. Expressions that are advised to be used with an avatar are the happy (not too excessively), interested, positive surprised or amazed. Users also thought that these expressions were suitable metaphors to communicate information. The results obtained indicate a significant increase in the efficiency and effectiveness for users in their interface tasks. Users were also more motivated and focused on product presentations when positive facial expressions were used. The same was not observed with negatively rated by users facial expressions. Negative facial expressions should not be used by an avatar. Users showed an antipathy for all the negative expressions and they were distracted from the communication of information. As a result, the use of negative expressions would only negatively influence users' decision when an avatar with negative facial expressions is used. These negative facial expressions are the angry or mad, disapproving, sad, upset, negatively surprised, disgusted and tired or bored. The results derived from the neutral facial expressions showed that users like the neutral expression, but not the thinking expression and in all cases (experiment with interactive context. When the thinking expression was used, it did not effectively persuade a user (prospective buyer) when shopping online, as a feeling of uncertainty was portrayed. Therefore, the thinking facial expression should not be used in B2C interfaces.

4.4 Use of body gestures

The experiments investigated 9 body gestures in an interface in both the absence and presence of an interactive context. It can be suggested that positive body gestures such as open palms, head up, or hands clenching could only improve and enhance the usability of an interface and can be used extensively by an avatar. Also experimental results indicate that hands steepling efficiently contributed to a presentation, although the exact use of this gesture has not been fully specified in literature. These gestures express openness and honesty, influencing positively users' decisions. Lastly, chin stroking should be carefully used as it often communicates a feeling of uncertainty by an avatar. Experimental results obtained about the use of negative body gestures indicate that negative gestures should not be used by an avatar. Body gestures such as the arms folded, face scratching, hiding face and legs crossed only convey a defensive and dishonest message when used by avatars. Body gestures however, may distract the attention of the user if they are used extensively, failing as a result, to achieve their main goal to successfully communicate information and persuade users to purchase specific products in an e-Commerce interface.

5. References

- Böszörményi, L.; Hellwagner, H. & Kosch, H. (2001). Multimedia Technologies for E-Business Systems and Processes, Institut für Informationstechnologie, Universität Klagenfurt.
- Jalali-Sohi, M. & Baskaya, F. (2001). A Multimodal Shopping Assistant for Home E-Commerce, Fraunhofer Institute for Computer Graphics.
- Nemetz, F. (2000). The Missing Link: Multimedia and E-Commerce, Department of Mathematical Sciences, University of Bath.

- Dix, A.; Finlay, J.; Abowd, G. D. & Beale, R. (2003). *Human-Computer Interaction* 3rd Edition, Prentice Hall, (379-381).
- Sengers, P.; Penny, S. & Smith, J. (1999). *Traces: Semi-Autonomous Avatars*.
- Benyon, D.; Turner, P. & Turner, S. (2005). *Designing Interactive Systems - People, Activities, Contexts, Technologies*, Addison-Wesley.
- Bartneck, C.; Takahashi, T. & Katagiri, Y. (2004). *Cross-Cultural Study of Expressive Avatars*.
- Theonas, G.; Hobbs D. & Rigas, D. (2008a). Employing Virtual Lecturers' Facial Expressions in Virtual Educational Environments, *The International Journal of Virtual Reality*, (31-44).
- Burford, D. & Blake, E. (1999). *Real-Time Facial Animation for Avatars in Collaborative Virtual Environments*, Collaborative Visual Computing, Department of Computer Science, University of Cape Town.
- Prasolova-Førland, E. & Divitini, M. (2002). Supporting learning communities with collaborative virtual environments: Different spatial metaphors, IDI, Norwegian University of Science and Technology.
- Krenn, B.; Neumayr, B.; Gstrein, E. & Grice, M. (2003). *Life-like Agents for the Internet: A Cross-Cultural Case Study*.
- Fabri, M.; Gerhard, M.; Moore, D.J. & Hobbs, D.J. (1999). *Perceiving the Others: Cognitive Processes in Collaborative Virtual Environments*, Virtual Learning Environment Research Group, School of Computing, Leeds Metropolitan University.
- Hobbs, D.; Perera S. & Moore, D. (1999). The use of Metaphors to enhance Educational Hypermedia, *Proceedings of the 7th International Conference on Computers in Education*, Tokyo, Japan.
- Thalmann, D. (1997). *Virtual Humans in Virtual Environments: A New View of Multimedia Applications*, Computer Graphics Lab (LIG), Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.
- Fabri, M. & Moore, D. (2005a). The use of emotionally expressive avatars in Collaborative Virtual Environments, ISLE Research Group, School of Computing, Leeds Metropolitan University.
- Fabri, M.; Moore, D.J. & Hobbs, D.J. (1999). The Emotional Avatar: Nonverbal Communication between Inhabitants of Collaborative Virtual Environments, in Braffort et al (eds.) *Gesture-Based Communication in Human-Computer Interaction*, Springer Lecture Notes in Artificial Intelligence 1739, pp. 245-248.
- Fabri, M. & Moore, D.J. (2005b). Is Empathy the Key? Effective Communication via Instant Messaging, in *Proceedings of 11th EATA International Conference on Networking Entities*, St. Pölten, Austria.
- Fabri, M.; Elzouki, S. Y. & Moore, D. (2007) Emotionally Expressive Avatars for Chatting, Learning and Therapeutic Intervention, *Human-Computer Interaction, Part III, HCII 2007*, (275-285), Springer-Verlag Berlin Heidelberg.
- Ostermann, J. & Millen, D. (2000). *Talking Heads and Synthetic Speech: An Architecture for Supporting Electronic Commerce*, AT&T Labs - Research, New Jersey, USA.
- Cassell, J.; O'Vilhjálmsson, H. H. & Bickmore, T. (2001). *BEAT: the Behavior Expression Animation Toolkit*, MIT Media Laboratory.

- Nijholt, A.; Heylen, D. & Vertegaal, R. (2000). Inhabited interfaces: attentative conversational agents that help, Deptment of Computer Science, University of Twente, The Netherlands.
- Paradiso, A. & L'Abbate, M. (2005). A Model for the Generation and Combination of Emotional Expressions, GMD – German National Research Center for Information Technology Institute (IPSI), Darmstadt, Germany.
- Guerin, F.; Kamyab, K.; Arafa, Y. & Mamdani, E. (2000). Conversational Sales Assistants, Intelligent and Interactive Systems, Department of EEE, Imperial College of Science.
- Beskow, J. (1996a). Talking heads – communication, articulation and animation, Department of Speech, Music and Hearing, KTH.
- Beskow, J. (1996b). Animation of Talking Agents”, Centre for Speech Technology, KTH, Stockholm, Sweden.
- Cohen, M. M. & Massaro, D. W. (1993). Modeling Coarticulation in Synthetic Visual Speech.
- Theonas, G.; Hobbs D. & Rigas, D. (2008b). The Effect of Facial Expressions on Students in Virtual Educational Environments, International Journal of Social Sciences, Volume 2, (42-49).
- Gong, L. (2006). Is Happy better than sad even if they are both non-adaptive? Effects of emotional expressions of talking-head interface agents, School of Communication, Ohio State University, USA.
- Massaro, D. W. (1998). Perceiving Talking Faces: From Speech Perception to a Behavioral Principle, MIT Press, Cambridge, MA.
- Wiki, (2005). http://en.wikipedia.org/wiki/Talk:Nonverbal_communication.
- Hartley, M. (2003). Body Language At Work, Sheldon Press.
- Kyle, D. R. (2001). Reading Body Language for the Sales Professional.
- McBreen, H. M. & Jack, M. A. (2001). Evaluating Humanoid Synthetic Agents in E-Retail Applications.
- Cassell, J. (2001). Embodied Conversational Agents – Representations and Intelligence in User Interfaces.
- Pease, A. (1981). Body Language – How to read others’ thoughts by their gestures”, Sheldon Press.
- Preece, J.; Rogers, Y.; Sharp, H.; Benyon, D.; Holland, S. & Carrey, T. (1994). Human-Computer Interaction. Addison-Wesley, (252-253).
- Lines, L. & Home, K. (2002). Older Adults’ Evaluations of Speech Output. Department of Information Systems & Computing, Brunel University, Uxbridge, Middlesex, U.K.
- Janse, E. (2002). Time-Compressing Natural and Synthetic Speech, University of Utrecht, The Netherlands.
- Reynolds, M. E.; Isaacs-Duvall, C. & Haddox, M. L. (2002). A Comparison of Learning Curves in Natural and Synthesized Speech Comprehension, Marshall University, Huntington, West Virginia.
- Lemmetty, S. & Karjalainen, M. (1999). Review of Speech Synthesis Technology, Department of Electrical and Communications Engineering, Helsinki University of Technology.
- Dutoit, T. (1999). A Short Introduction to Text-to-Speech Synthesis, TTS research team, TCTS Lab.

- Schroeter, J.; Ostermann, J.; Graf, H.P.; Beutnagel, M.; Cosatto, E.; Syrdal, A.; Conkie, A. & Stylianou, A. (2000). Multimodal Speech Synthesis, AT&T Labs - Research, New Jersey, USA.
- Wouters, J.; Rundle, B. & Macon, M W. (1999). Authoring Tools for Speech Synthesis Using the Sable Markup Standard.
- Kouroupetroglou, G. & Mitsopoulos, E. (2000). Speech-enabled e-Commerce for Disabled and Elderly Persons, University of Athens, Greece.
- Xydas, G. & Kouroupetroglou, G. (2001). Text-to-Speech Scripting Interface for Appropriate Vocalisation of e-Texts, University of Athens, Greece.

IntechOpen

IntechOpen

Human-Computer Interaction

Edited by Inaki Maurtua

ISBN 978-953-307-022-3

Hard cover, 560 pages

Publisher InTech

Published online 01, December, 2009

Published in print edition December, 2009

In this book the reader will find a collection of 31 papers presenting different facets of Human Computer Interaction, the result of research projects and experiments as well as new approaches to design user interfaces. The book is organized according to the following main topics in a sequential order: new interaction paradigms, multimodality, usability studies on several interaction mechanisms, human factors, universal design and development methodologies and tools.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Dimitrios Rigas and Nikolaos Gazepidis (2009). The Role of Facial Expressions and Body Gestures in Avatars for e-Commerce Interfaces, Human-Computer Interaction, Inaki Maurtua (Ed.), ISBN: 978-953-307-022-3, InTech, Available from: <http://www.intechopen.com/books/human-computer-interaction/the-role-of-facial-expressions-and-body-gestures-in-avatars-for-e-commerce-interfaces>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen