1. Introduction

Internet has become a place where people participate in many kinds of interactions, from social experiences (games, dating, communities…) to professional activities (e-commerce, meetings and interviews online, e-learning and education…). Many of these interactions take place in some sort of virtual scenario, usually called virtual world. Virtual worlds are simulated environments with digital resemblance of animated actors and their physical surroundings where they can engage in interactive activities through computer-generated tools (Bainbridge, 2007). The growth of online virtual worlds and the existence of interactions inside them leaded to the need for characters that represented the actual people interacting. In this context the concept of "avatar" emerged. Avatars are defined as "general graphic representations that are personified by means of computer technology" (Holzwarth, Janiszewski, & Neumann, 2006). More specifically, an avatar is defined as a user-created digital representation that symbolizes the user's presence in a metaverse (Davis, Murphy, Owens, Khazanchi, & Zigurs, 2009).

Avatars are used in different contexts and situations, and their morphology may range from a tiny still picture in an online chat to a 3D interactive character in a videogame or virtual world (Lin & Wang, 2014). Avatars may depict human characters, fantastic beings, animals or objects. They may be preconfigured characters or customized by users. In this sense, avatars are also increasingly malleable. Their features may be changed and adapted in many ways to acquire a suitable appearance for a given scenario, just like in the real world but much more flexibly.

Due to the influence that the avatar's appearance may have on the general perception of the avatar's users and their interactions, a careful avatar design is becoming more and more relevant. In this context, the ability of controlling the way an avatar convey messages and emotions is capital. The purpose of this research is to develop a procedure to design avatar faces capable of conveying to the observer the most suitable sensations according to a given context, usually in computer applications. We call "sensation" to the impression that things produce by means of the senses, and it occurs immediately after their excitation. It is an immediate basic experience generated by simple isolated stimuli (Foley & Matlin, 2010). Therefore, the features of an avatar face should produce a series of sensations in the people observing it. The problem is that there is dispersion in the judgment of the observers about these sensations. Although a relation between the attributes of the avatar face and the sensations elicited exists, it varies from an observer to another, due to individual and external factors (motivations, education, personality…) (Allenby & Ginter, 1995; Engel, Blackwell, & Miniard, 1995; Hoch, Kim, Montgomery, Rossi, & Montgomery, 1995). Since the avatars will be used by many people, the objective will be to transmit the desired sensation to most of them.

The proposed system is based on a combination of genetic algorithms (GAs) and artificial neural networks (ANNs) whose training is based on perceptual human responses to a set of faces.

2. Related Work

The types of avatars are very varied and depend on the user, context and situation. The same person may be using different avatars in different scenarios. Some research has been conducted focused on the implications of the social use of avatars as a way to convey self-presentation and communicate online (Chung, DeBuys, & Nam, 2007; Nowak & Rauh, 2005; Schultze, 2010; Taylor, 2002). They may also be used in professional situations like collaborative online design (Koutsabasis, Vosinakis, Malisova, & Paparounas, 2012). In all these cases, the way an avatar looks may have an effect on the opinion and perception of other user in the virtual world. For example, more anthropomorphic and less androgynous avatars are perceived as more credible and trustworthy (Nowak & Rauh, 2008). Large pupil sizes and slow eye blinking frequency make avatars look more sociable and attractive (Weibel, Stricker, Wissmath, & Mast, 2010). Older avatars are perceived by adult users as being more intelligent and safer (Marin, Jo, & Lee, 2013). Attractive and more elaborated avatars have more successful social interactions (Banakou, Chorianopoulos, & Anagnostou, 2009), and even receive more favorable ratings in virtual employment interviews (Behrend, Toaddy, Thompson, & Sharek, 2012). Similar results can be found in Hasler, Tuchman, & Friedman (2013).

People want their avatars to communicate things about themselves, and they spend time choosing or customizing them. Some research has been performed about strategies designing avatars for different purposes (Boberg, Piippo, & Ollila, 2008; Kafai, Fields, & Cook, 2010; Lee, 2014; Vasalou, Joinson, Bänziger, Goldie, & Pitt, 2008). These studies show a balance between the need for aligning the avatar image with the actual self and the strategies of idealization. Vasalou & Joinson (2009), for example, showed that avatars designed for blogging accurately reflected the physical appearance and personality of their creators. When the avatar was intended to be used for gaming or dating, however, people intentionally stressed some characteristics to match the context.

A recent study examining the motivations and strategies for avatar design found these major objectives: virtual exploration (living experiences only possible in the digital world), social navigation (social dynamics online), contextual adaptation and identity representation (Lin & Wang, 2014). The work performed in Ducheneaut, Wen, Yee, & Wadley (2009) describes three main factors in the identity exploration: An idealization of the self, a wish for standing out and following a trend. It concludes that people tends to use the avatar to make physical improvements of their images, while keeping their own personality traits. Moreover, the attitude and behavior of people may be influenced by their avatar's features. Several studies have been conducted in this sense. In Yee & Bailenson (2007) it is shown that the propensity of people to get close to others and their self-confidence when establishing relationships online are affected by the attractiveness and height of their avatars. Similar results were observed in Banakou et al. (2009). The work shown in Merola & Pena (2010) explains this effect by behavioral confirmation, and argues that desired interactions in virtual communities can be shaped and encouraged by using the suitable avatars. Other studies have reported a less remarkable influence in this sense. For example, Chung et al. (2007) show no difference in terms of attitude, presence, empathy and para-social interaction between users that created their own avatar and those who were given a standard one.

Avatars are also increasingly used to convey emotions through non-verbal expressions. It has been proved that the use of emoticons and expressive avatars improves online interactions (M Fabri & Moore, 2005) and may even be useful for education and therapeutic purposes (Marc Fabri & Elzouki, 2007; Orvalho, Miranda, & Sousa, 2009). Also, the transmission of emotions is strongly related to credibility, which in turn affects the ability to persuade (Barbat & Cretulscu, 2003; El-Nasr, Ioerger, Yen, House, & Parke, 1999). This ability is important for applications in areas ranging from e-commerce to e-therapy. The characteristics of an avatar in an e-commerce website may improve the user experience and influence trust building (Keeling, McGoldrick, & Beatty, 2010; Qiu & Benbasat, 2005). In Holzwarth et al. (2006) it is shown that an attractive avatar is more effective selling products of moderate level of involvement, while an expert avatar is better at higher levels. Ethnicity and gender of the avatar also affect its perception as a sociable, competent and enjoyable product recommendation agent (Qiu, 2006). The importance of avatars in online marketing is such that some authors argue that companies should develop specific marketing strategies focused on the avatars' characteristics rather than on their owners' ones (Hemp, 2006). So efforts are being made to achieve more realistic avatars' expressive resources (Di Fiore, Quax, Vanaken, Lamotte, & Van Reeth, 2008; Nasoz & Lisetti, 2006; Wang & Geigel, 2011).

However, the transmission of emotions via avatars' expressions largely depends on the perception and evaluation of these expressions by the viewer. Several studies have reported cultural differences in interpreting avatars' facial expressions (Bartneck, Takahashi, & Katagiri, 2004; T. Koda & Ishida, 2006; Tomoko Koda, Ishida, Rehm, & André, 2009; Tomoko Koda, 2004). Some other works explore ways to make emotional avatars more believable. In El-Nasr et al. (1999) an algorithm is proposed to provide 2D avatars with dynamic emotional expressions. A software called HEFES is proposed in Mazzei, Lazzeri, Hanson, & De-Rossi,

(2012) to create facial expressions in both a robot and its 3D avatar and then used to help children with autism. Other systems to generate expressive gestures in 3D avatars can be found in Gratch & Marsella (2001) or Yu, Garrod, & Schyns (2012).

In this paper, a procedure is proposed to design avatar faces controlling the sensations they have to convey, by means of a combination of GAs and ANNs. Therefore, an overview of these techniques is presented in the next section.

3. Overview of genetic algorithms and artificial neural networks

Although both, GAs and ANNs, are usually included among artificial intelligence techniques, the capabilities and applications of them differ. GAs perform a stochastic guided search based on the evolution of a set of structures as it occurs in natural species evolution (Goldberg, 1989). They are usually employed to solve combinatory optimization problems (Papadimitriou & Steiglitz, 1982) in which an attempt is made to locate the best configuration for a group of variables which, due to certain restrictions, minimize, or if necessary maximize, a target function. The size of the space of the search for solutions will depend on the problem, but in general, the search for acceptable solutions is a complex problem, especially when the number of imposed restrictions is high. Although other techniques exist, such as Integer Programming (Glover, 1986), which are applicable to solving this type of problem, the use of GAs provides good solutions without excessively prolonging the time of calculation.

An ANN is a mathematical model that represents a distributed adaptive system built by means of multiple interconnecting processing elements, just as real neural networks do. ANNs are used in many fields of research (psychology, robotics, biology, production or computer science, to name a few) (Lynch, Dagli, & Vallenki, 1999; Marren, Harston, & Pap, 1990; Principe, Euliano, & Lefebvre, 2000a) due to their ability to adapt, learn[1], generalize, organize or cluster data. Given their ability to learn (in comparison with sequential systems), they are instruments which are suitable for purposes such as those described previously, there being various uses of ANNs in similar environments (Chen & Yan, 2008; Dasgupta, Dispensa, & Ghose, 1994; Hsiao & Huang, 2002; Hsiao & Tsai, 2005; Ishihara, Ishihara, Nagamachi, & Matsubara, 1997; Lai, Lin, & Yeh, 2005; Lai, Lin, Yeh, & Wei, 2006; Tsai, Hsiao, & Hung, 2006; Yang & Shieh, 2010). In this work, the ANNs capacity to establish relationships between the inputs and outputs of a system will be used to guide an AG.

[1] Neural network learning or training is an adaptive procedure in which the weights of the connections between neurons are incrementally modified so as to improve the network performance until reaching a specified criterion.

3.1. Genetic algorithms

As aforementioned, GAs perform a stochastic guided search based on the evolution of a set of structures. The starting point is a set of problem solutions called *individuals*. This first set is randomly generated and called initial population. Each individual is coded by a finite length chain called *chromosome*. Each member of a chromosome is a gene, and codifies one characteristic of the solution that the chromosome represents (e.g. avatar traits). Therefore, there must be as much genes in the chromosome as characteristics must be considered to define a solution. Each gene can take a predefined set of values named *alleles*.

This first set of chromosomes (initial population) is randomly generated assigning a value to each gene of each chromosome. Each chromosome is evaluated using an evaluation function to determine its suitability for the requirements of the problem. Then, the population undergoes several transformations that yield a new population (new generation). These transformations are guided by some genetic operators, being the most common *selection*, *crossover* and *mutation*, which combine or modify the chromosomes representing the individuals. Crossover and mutation operators are applied to create a new generation of individuals that inherit the best characteristics of their predecessors. For this purpose, the individuals that will participate in each of the genetic operators, and those that will survive and pass on to the following generation, are selected previously by means of the selection operator. For this process a *roulette wheel selection* (Goldberg, 1989) could be used, by which the probability of an individual being selected is inversely related to its score in the evaluation, so that those who accomplish the best performance have more chances of being selected.

The crossover process is applied to pairs of individuals chosen from those selected during the previous stage. Two new individuals are generated from the combination of the original individuals. The new individuals replace their parents in the population. The parameter p_c (crossover probability) determines the number of individuals in the next generation that will be created by crossover. The crossover point is usually chosen at random. The offspring is generated by combining the genes that remain on the left and on the right of the crossover point in each of the parents. The mutation operator is applied to genes in a random way. The number of genes that will mutate is determined by the parameter p_m (probability of mutation). The mutation consists in a random change in the allele of the gene. The individuals generated by mutation substitute the originals in the population. Once the new population is conformed each chromosome is evaluated and the process is repeated with the new set of individuals until a certain number of iterations is reached, or until a certain number of iterations without a new best solution have been performed, making the individuals evolve to better solutions to the problem.

Usually, in the evolutive algorithms the entire populations are replaced by individuals chosen by the selection operator, or by others created by crossover or mutation (generational algorithms). There is also another type of strategy which maintains the best individuals discovered by the process from generation to generation. Algorithms of this type are called stable algorithms, since they tend to reach a stable or stationary state (Srinivas & Patnaik, 1994a). In this work, the best individuals in a generation are always selected for the next. This strategy gave satisfactory results in the experiments carried out.

3.2. Artificial Neural Networks

An ANN is a mathematical model built by means of multiple interconnected processing elements, (neurons) distributed in several layers (Figure 1). The intermediate layers are known as the hidden layers, while the first and the last layers are known as the input and output layers, respectively. In general terms, each neuron receives signals processed and transmitted by neurons in the preceding layer and in turn processes and transmits them on to the next layer. The number of layers and the way in which the neurons are connected determine the architecture of the network.

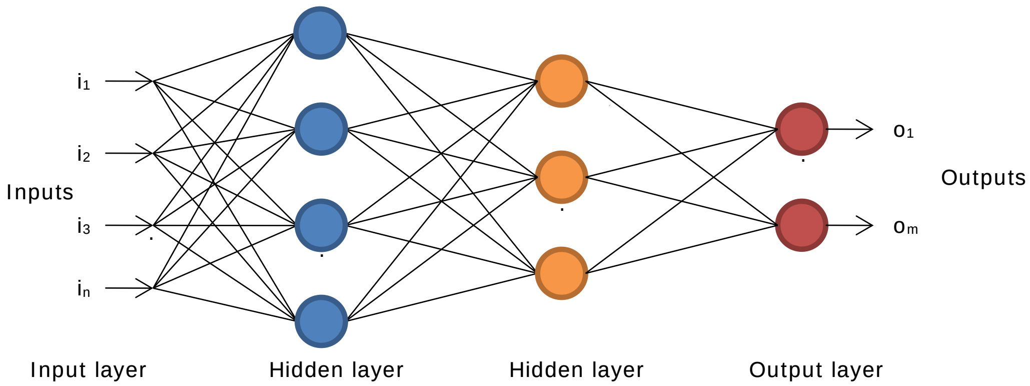


Figure 1: Structure of a neural network with two hidden layers, *n* inputs and *m* outputs.

The input signals (i_1, i_2, …, i_n) are the values of the variables representing an instance of the phenomenon to be modeled (e.g. avatar attributes). They are collected by the input layer, which transmits them through links to the neurons in the first hidden layer. The signals are scaled in each link according to an adjustable parameter associated with each connection between neurons called *weight*. Usually, the initial weight of each link is randomly set. Each neuron in the hidden layer collects the signals from the connections, adds them up and produces an output that is a function of the sum. The most commonly used functions are sigmoids, hyperbolic tangents and linear versions of the latter. The signals traverse the network from the input layer to the output layer, where the network response to the inputs is collected (o_1, o_2, …, o_m). Supervised learning networks are able to learn the relationships between the inputs and outputs through the repeated presentation of input data and the values of the corresponding outputs. Once trained, the network can generalize these relationships to new cases. The training process consists of presenting the network with a sufficient number of input cases and the desired output values (namely, the traits of a number enough of avatars and the corresponding evaluation of these avatars by the users).

The output obtained by the network in each case is compared against the desired output, and the network error is calculated. Then, the weights of neuron connections are modified according to the selected training algorithm in order to minimize this error. This process is repeated until a criterion previously established is reached, for example, when the error value gets to a threshold or stops decreasing. Although there are different training algorithms applicable to different types of networks, the most commonly used to train ANNs is Back-Propagation (BP) (Rumelhart, Hinton, & Williams, 1986).

A significant problem relating to the development of ANNs is overfitting. Overfitting (also known as overtraining) occurs when a model captures the statistical noise in the data rather than the underlying signal (Sarle, 1995), i.e. the model memorizes the correct responses to each pattern rather than learning the relationships between inputs and outputs. To avoid overfitting some regularization procedures can be used such as jitter, weight decay or early stopping. Early stopping is the use of a reduced data set (validation set) to calculate model error periodically during training. These validation data sets are not used to train the network, but rather to determine the moment when the model stops learning and starts memorizing the relationships between training patterns and their resulting outputs. The usual procedure is to divide the available data into three sets: the training set, which is used to train the network; the validation set, which is used to determine the early stopping point; and the test set, which is used to validate the degree of generalization of the trained model.

4. Methodology

The elicitation of a sensation in an observer by the avatar face depends on the traits that conform that given face. In order to make the avatar face conveys some intended sensation, it is needed to find the combination of traits which provoke that precise sensation in the observer. Due to the great diversity of facial traits it is impossible to test each possible combination. The use of GAs to solve this problem is proposed. In these algorithms, the evaluation is usually performed using an analytical formula, but as there is no way of mathematically expressing the elicitation of a sensation by an image, an observer´s evaluation should be used. This is unfeasible, considering that if the GA uses a population size of 50, for instance, in 1000 generations it would produce 50000 faces, which should be evaluated by a real observer. In this work a procedure is proposed to generate an ANN trained to simulate the perceptual response of the observers to the avatars' faces. This ANN will serve as fitness function for a GA which will find the best combination of traits to provoke the desired sensation.

The algorithm starts by generating an initial population of individuals that represent different avatar faces (Figure 2a). Each face is codified through a vector (chromosome). The vectors have as many genes as needed to completely define an avatar face (e.g. as many genes as facial traits). All possible variations of a given facial trait must be coded. For example, if there are 25 available types of mouth, a letter or a number could be assigned to each of these mouths to identify it. The chromosomes of the initial population are produced by randomly selecting these values, obtaining chromosomes that represent random avatar faces.

Each of the faces conforming the initial population is then evaluated by an ANN to determine its capability to elicit the desired sensation (Figure 2b), obtaining a fitness value for each face. The inputs to the ANN are the gene codes of each chromosome, and the output is the evaluation of the resulting face. The way to obtain such ANN will be exposed later.

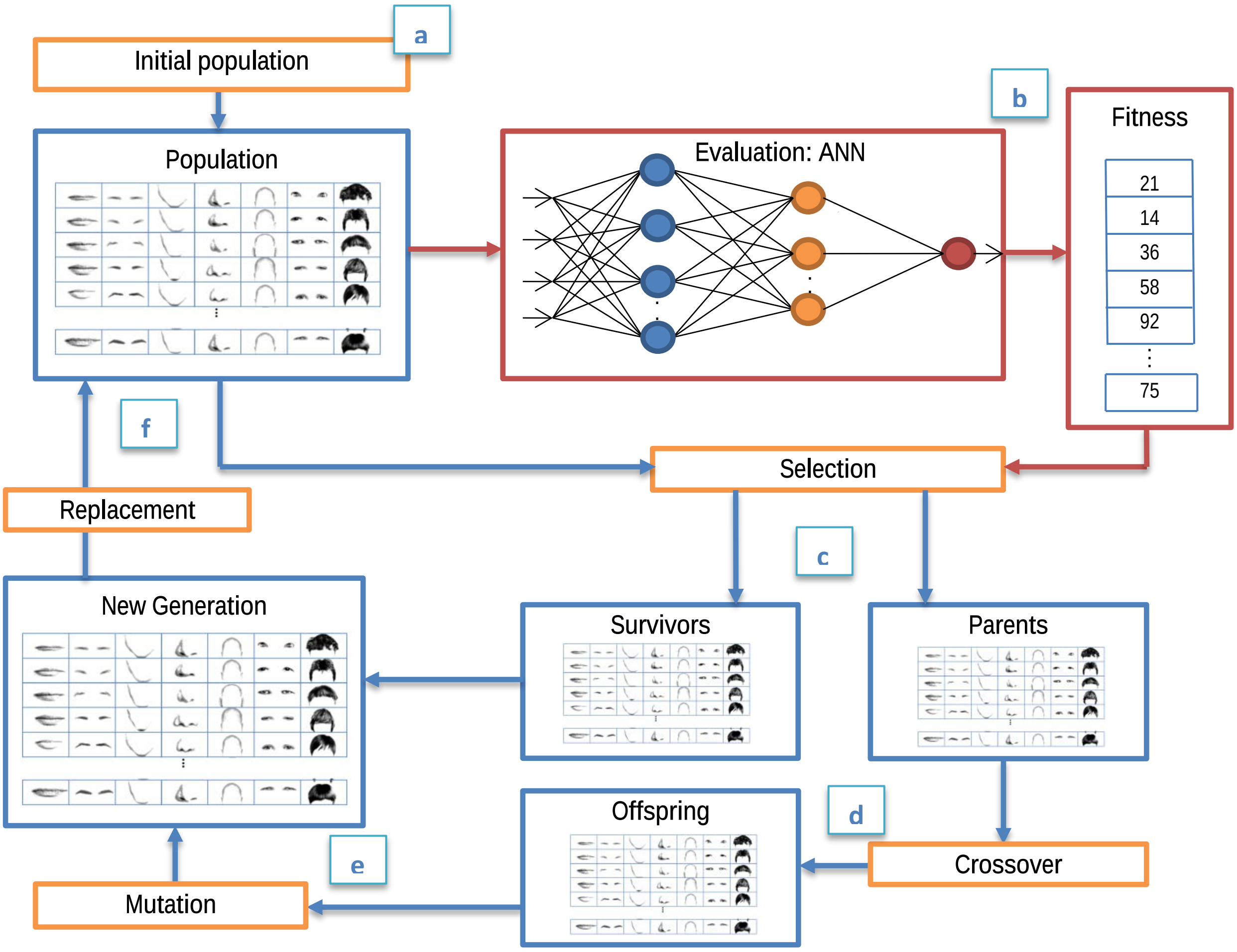


Figure 2. Structure of the GA using an ANN as fitness function.

When all the faces have been evaluated, a selection is made of those that will survive and pass on to the next generation or that will be used as reproducers (Survivors and Parents in Figure 2c). For this process a *roulette wheel selection* (Goldberg, 1989) is used, in which the probability of a face being selected is related to its fitness value, such that those faces who offer the best results have greater chance of being selected. There is an exception for this rule: in our proposal the face with the best fitness is always selected to be part of the Survivors group. This strategy maintains the best face discovered by the process from generation to generation.

Pairs of faces are chosen from the Parents group selected in the previous stage. Reproduction is performed by means of a crossover, in which two new faces (Offspring) are generated from the combination of the solutions represented by each pair of original faces (see Figure 3). In this case, a crossover point is chosen at random (a number between one and the number of traits minus one). The offspring is generated by combining the genes that remain on the left and on the right of the crossover point in each of the parents. Using this procedure, the faces represented by the offspring are a combination of their parent traits. The parameter p_c

(crossover probability) determines the number of faces in the next generation that will be created by crossover and the number of faces that will be survivors. For example, if the population size is 50, and the crossover probability is 0.5, half part of the next generation will be survivors from the previous generation, and the other half one will be descendants from faces from the prior generation. A typical value for this parameter varies between 0.5 and 0.9 (Srinivas & Patnaik, 1994b).

In our proposal, the mutation operator is applied to faces selected at random between the offspring group (Figure 2e). The number of faces that will mutate is determined by the parameter p_m (probability of mutation). This parameter indicates how many faces from the offspring group will suffer a mutation; therefore, if the probability of mutation is 0.5, half part of the descendants will be mutated faces. Mutation is performed by randomly selecting a gene in the chromosome and changing its value for another possible one in the trait.

Avatar faces produced by crossover (and possibly mutation) and survivors from the prior generations integrate the new population replacing the previous one. The process will be repeated with each new set of faces until a predefined number of iterations.

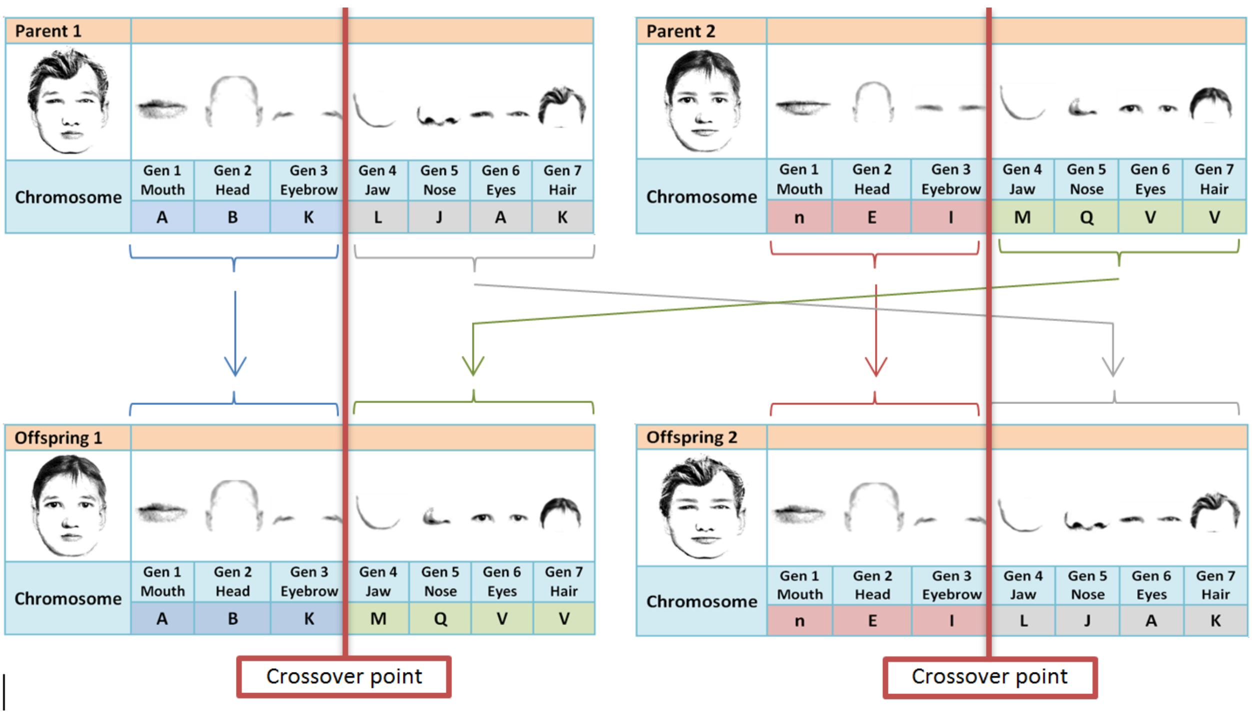


Figure 3. Crossover operator.

4.1. Development of the ANN

In order to build the ANN which will be used as fitness function in the GA, a number of patterns of the process to be modeled is needed. A pattern is in this case a chromosome representing a face plus the rating of this face in the desired sensation given by an observer. These patterns will serve to determine the most suitable type of ANN, the network topology (number of layers, numbers of neurons, transfer functions…) and the training of the chosen network.

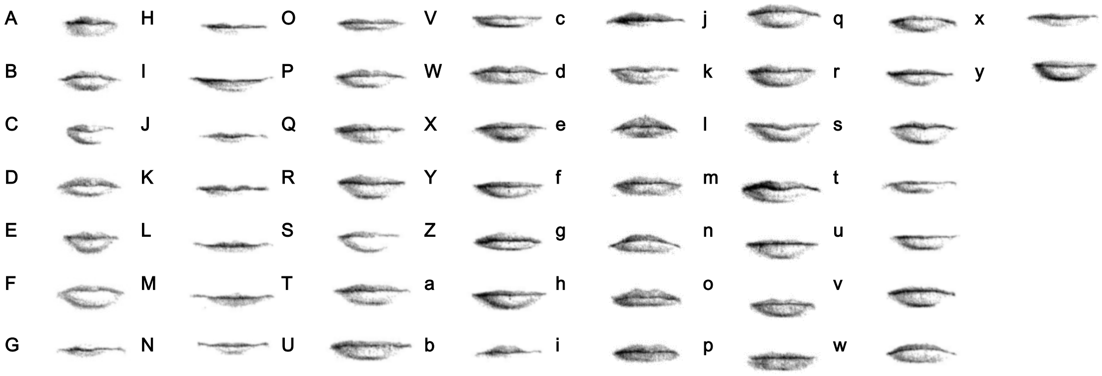
To obtain the patterns to train the ANN a survey must be developed. In this survey randomly generated faces are shown to respondents, asking them for the degree in which the selected emotion is elicited by each face. The size of the survey must be sufficient to obtain enough patterns (chromosome-fitness) to train an ANN. The obtained patterns must be divided in three sets; the training set, which is used to train the network, the validation set used to determine the early stopping point in order to avoid overfitting (see Section 2) and the test set, which is used to validate the degree of generalization of the trained model.

Several network topologies must be trained using these data sets, varying the number of layers the number of neurons by layer and the transfer functions in each neuron. After several tests the best network obtained is selected. In order to measure the performance of a network the Mean Square Error (MSE) produced by the ANN when predicting the users' opinions on the test set faces is used. These faces were not used for training and therefore it is possible to measure whether the relationships between the avatars attributes and the users' opinions according to the models can be generalized to cases which were not used to obtain these relationships.

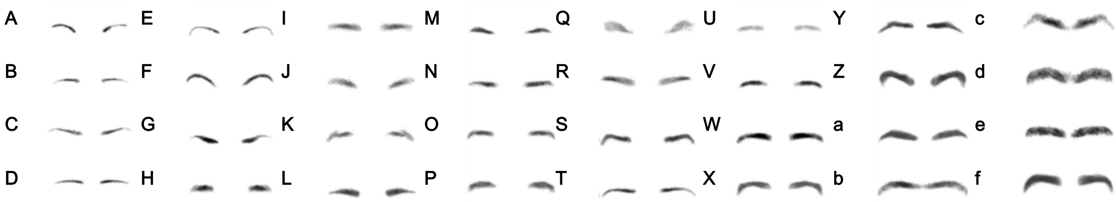
5. Material and methods. Case study.

An experiment using the described procedure is performed as a case study. The experiment consisted in generating three Caucasian avatar faces of people between 20 and 30 years old. These faces should convey an idea of attractiveness, reliability and responsibility, respectively. Seven facial traits were used to build the faces: mouth, eyebrows, jaw, nose, head shape, eyes and hairstyle. A database was used comprising 51 mouths, 32 eyebrows, 29 jaws, 38 noses, 9 head shapes, 22 pairs of eyes and 52 hairstyles (Figure 4). The total number of possible alternative faces is about $1{,}852e^{+10}$.

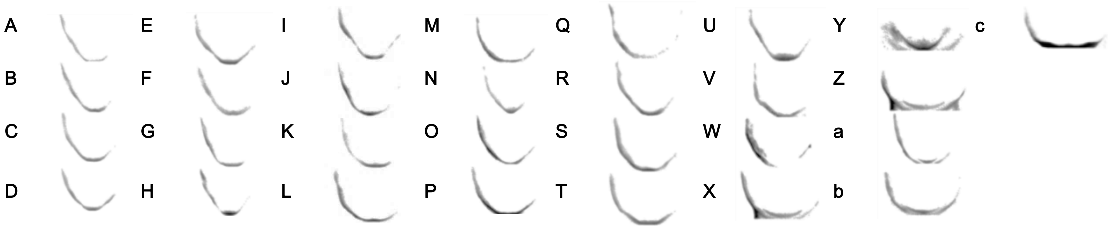
Mouths



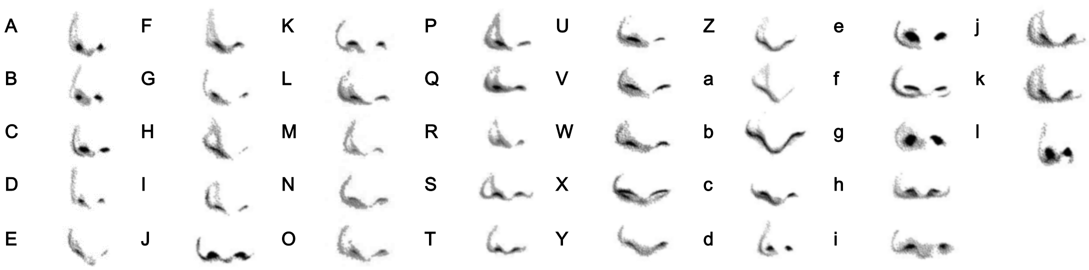
Eyebrows



Jaws

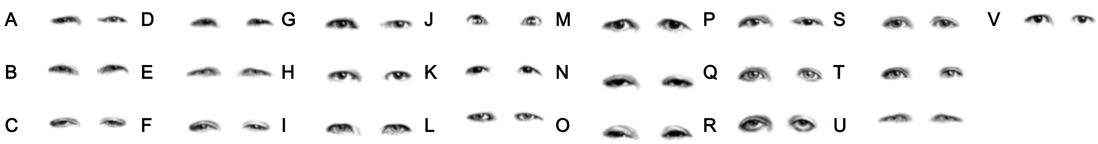


Noses



Heads

Eyes



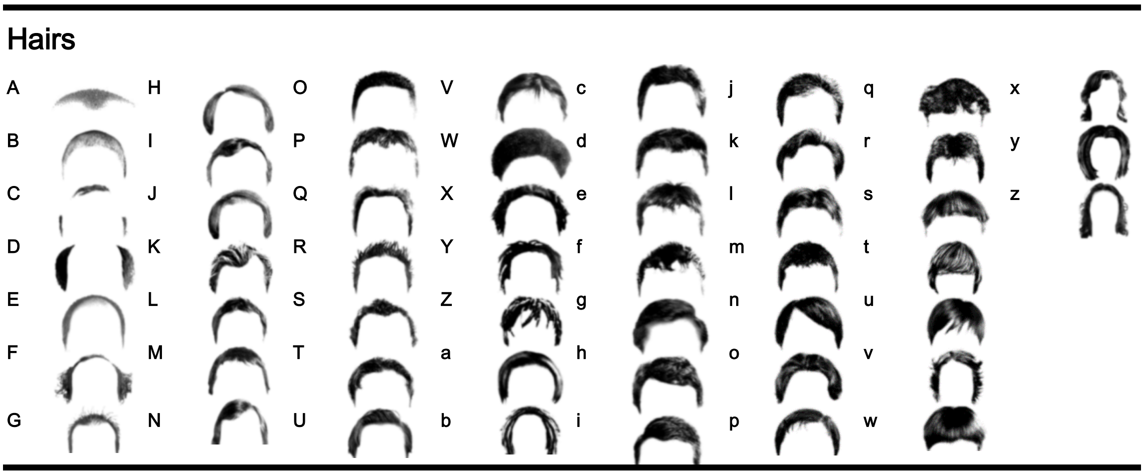


Figure 4. Facial traits database. Each trait is coded by a letter.

A codifying system is established. Each face is codified through a vector. Each vector has a length of seven genes, each one representing a facial trait, namely the mouth, head shape, eyebrows, jaw, nose, eyes and hairstyle. Each trait received a code as shown in Figure 4 (distinguishing between uppercase and lowercase letters). A face is then codified by defining the codes of each trait in the corresponding gene.

Specific software was developed in order to apply the proposed methodology. This software generates faces combining the different traits from the database, gathers the data to train an ANN and uses a GA to produce faces based on the trained ANN (Figure 5). The software was used to gather the data employed to train the ANNs. For the ANNs training and development Neurosolutions 5.0 was employed. Once the ANNs are obtained, Neurosolutions is able to produce “dll” files containing the trained ANNs. The GA implemented in our software is able to send each chromosome to those files and to obtain the response of the network in real time to be used as the fitness of each chromosome.

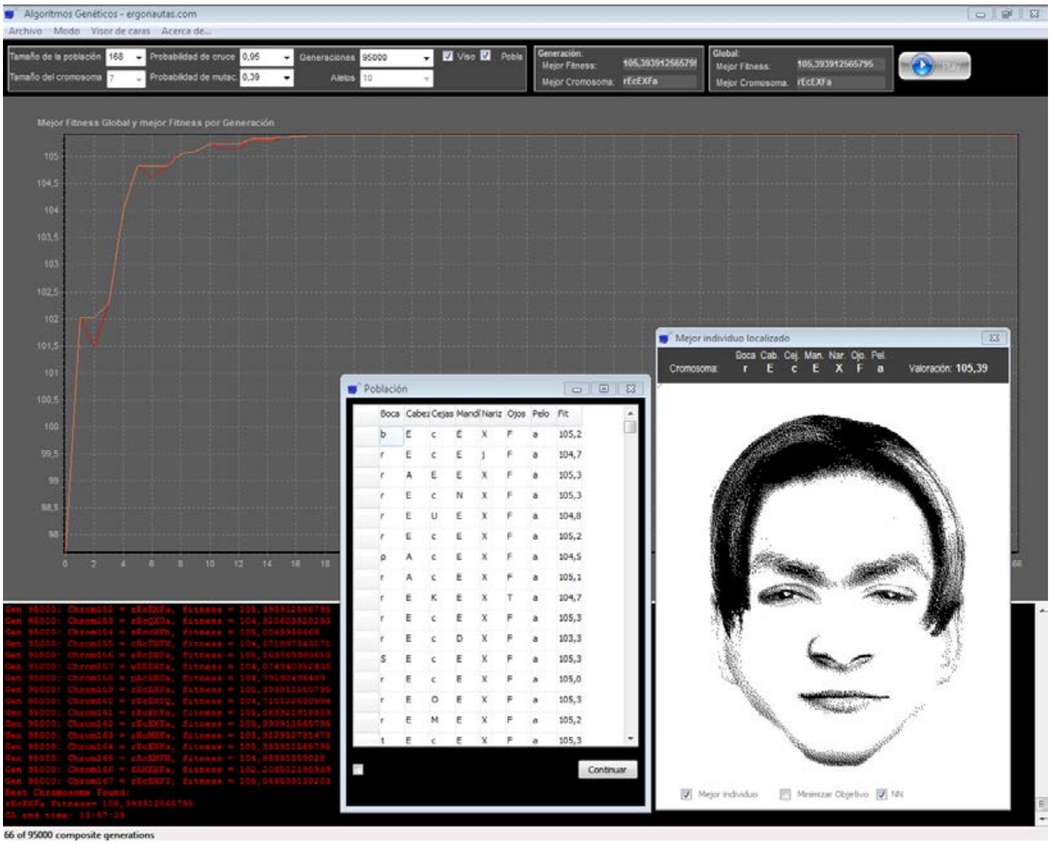


Figure 5. Software to support the face development process.

5.1. Obtaining the ANNs

In order to build the ANN which will be used as fitness function in the GA, a number of patterns of the process to be modeled is needed. A pattern is in this case a chromosome representing a face plus the rating of this face in the desired sensation given by an observer. A total of three networks, one for each sensation, are developed.

5.1.1. Obtaining data to train the ANNs

The sample consisted of 26 people aged between 20-30 years (15 men and 11 women). During the survey, the developed software displayed 40 faces to each participant. These faces were generated randomly using the traits of Figure 4. The generation was forced in order to get all the possible levels of each trait shown at least once.

The participants rated each face in a 0-100 scale, 0 meaning “Completely disagree” and 100 “Completely agree”. The questions were related to each of the three desired sensations to elicit: “Do you consider this person attractive?”, “Do you consider this person reliable?” and “Do you consider this person responsible?” Subjects were permitted to take as long as they needed to answer the survey. The average amount of time it took to complete the rating of the 57 faces forming part of the sample was three minutes and 12 seconds. However, respondents were allowed to take a break if they considered it necessary. 1040 cases were collected for each sensation, comprising both the traits used in the face generation and the respondent rating. These cases were divided in each sensation as follows: 790 cases constituted the training set, which is used to train the network. 226 cases formed the validation set. It is used to determine the early stopping point in order to avoid overfitting (see Section 3.2). The other 24 cases constituted the test set. There is no fixed rule for determining the appropriate number of cases to make up the different sets, although a study in this respect can be found in (Sarle, 1995).

5.1.2. Training the networks

NeuroSolution 5.0 was used to create and train the ANNs. Several tests were performed with different kinds of networks and finally a Multilayer Perceptron network was chosen for this problem. The training algorithm used was BP with momentum (Rumelhart et al., 1986). This algorithm is an improvement on the straight gradient-descent search. The momentum is a term which introduces a certain degree of inertia into the update of connection weights, thus accelerating and stabilizing development and avoiding convergence in a local minimum. The momentum constant was fixed at 0.7. The step size used was 0.01 in the first hidden layer, reduced by a factor of 5 in the second hidden layer, and again in the third (Principe et al. 2000). The initial weights were established randomly with small values (with average 0 and range between -0.1 and 0.1). This procedure is recommended when non-linear activation functions are used to avoid initial saturation of the processing elements (Principe, Euliano, & Lefebvre, 2000b). Hyperbolic tangents were used as activation functions.

To avoid overfitting early stopping was used. The validation data set was used to calculate model error periodically during training. These validation data sets are not used to train the network, but rather to determine the moment when the model stops learning and starts memorizing the relationships between training patterns and their resulting outputs. It must be pointed out that the validation error is not suitable for estimating the degree of generalization obtained by the network; a third data set (test set) is used for this purpose.

After several tests with different topologies, the best networks obtained for each sensation were selected using as the Mean Square Error (MSE) produced by the ANN when predicting the users' opinions on the test set faces. The characteristics of the three best ANNs found are shown in Table 1. The MSE was measured on the standardized outputs of the model ranging from -0.9 to 0.9. ANN1, ANN2 and ANN3 correspond to the networks trained to model the response to the questions “Do you consider this person attractive?”, “Do you consider this person reliable?” and “Do you consider this person responsible?”, respectively. The ANNs obtained differ not only in terms of the weight of the connections among neurons, but also in the number of neurons in each hidden layer. As an example, Figure 6 shows the ratings of the participants for the 24 faces of the set related to the question “Do you consider this person attractive?”, as well as those predicted by the ANN1.

	ANN 1	ANN 2	ANN 3
First Layer Neurons	8	6	9
Second Layer Neurons	4	3	4
Test MSE	0.0367	0.0209	0.0706

Table 1. Characteristics of the best ANNs obtained for each assessment.

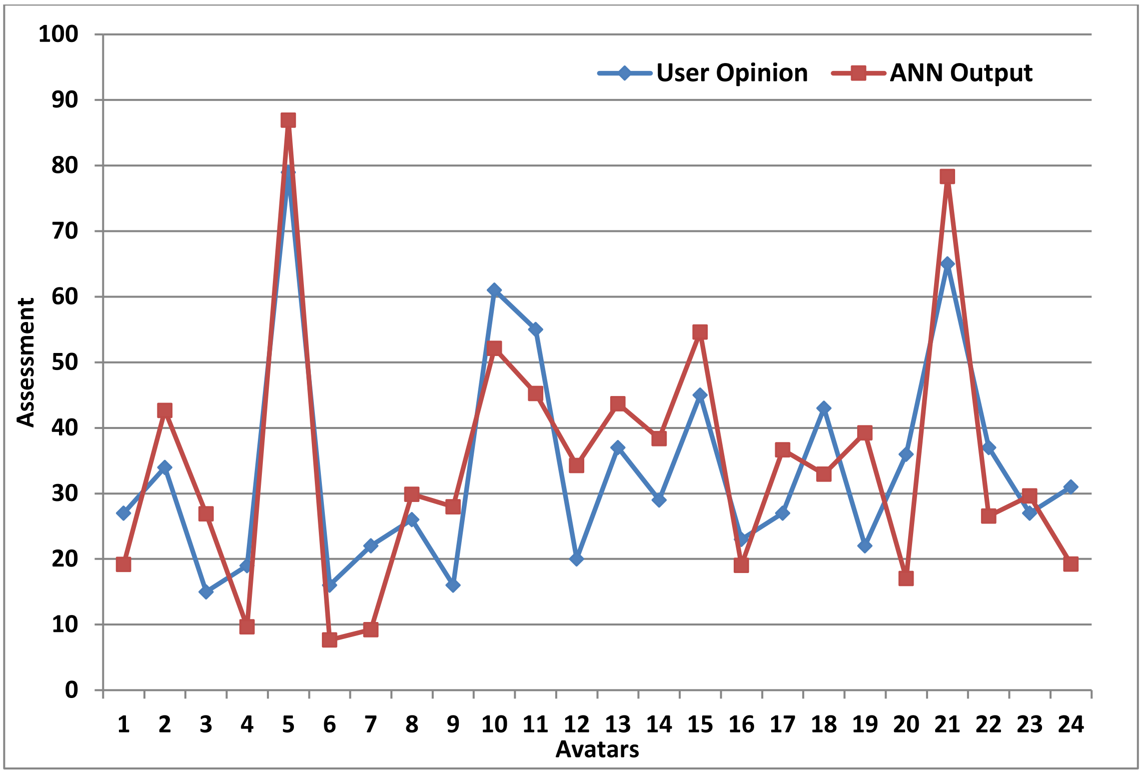


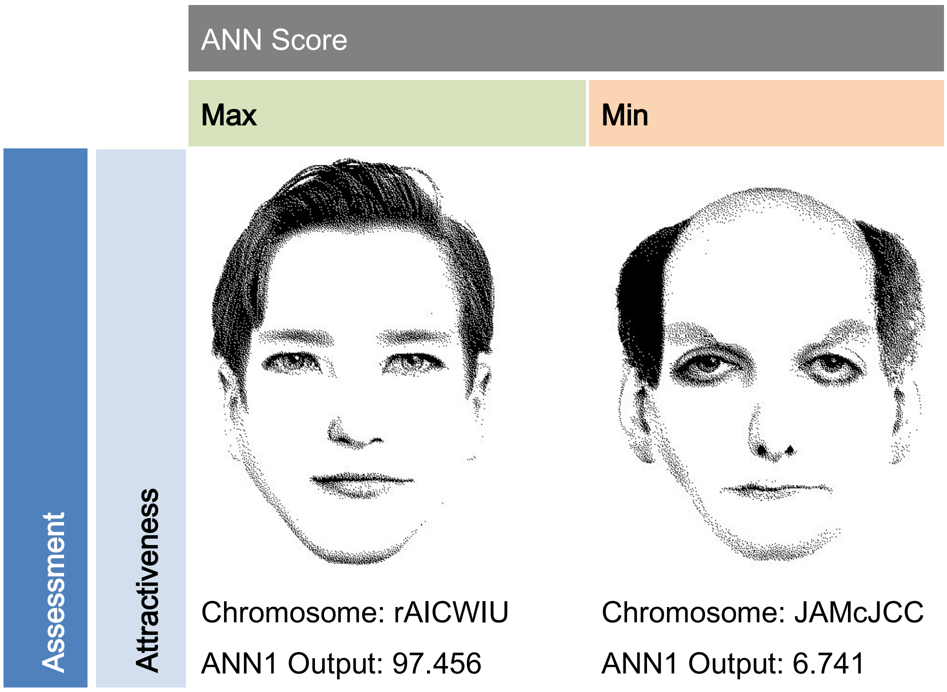
Figure 6. Comparison of the opinions of the users and those predicted for ANN1.

5.2. GA development

The chromosomes of the initial population were produced by randomly selecting the traits values. Size of population was set at 50 individuals. Each of the faces that conforms the initial population was then evaluated by the ANN to determine its capability to elicit the desired sensation. The inputs to the ANN are the genes codes, and the output is the evaluation of the resulting face. Since the inputs are categorical variables, the seven traits were coded in a unary way. Each trait was assigned as many neurons as levels it has. The trait “Head shape”, for example, was assigned 9 neurons as there were 9 different head shapes. So in order to represent the Head A, the first neuron of the trait Head took the value 1, and the other eight got value 0. The crossover probability was set at 0.9. The mutation operator was applied to individuals selected at random from the Parents group and mutation probability was set at 0.4. Although several tests were performed with different parameters (population size, p_c and p_m), the algorithm seems to be robust, and small variations in the parameters do not affect to the results too much. The maximum number of generations was set to1000.

6. Results

The software implementing the GA described in Section 5 was executed three times, once for each ANN, the objective function being to maximize each ANN output. Then the experiment was repeated another three times, but changing the objective to minimize the output. Figure 7 shows the chromosomes that received the highest and lowest score in each case, as well as the corresponding phenotype.



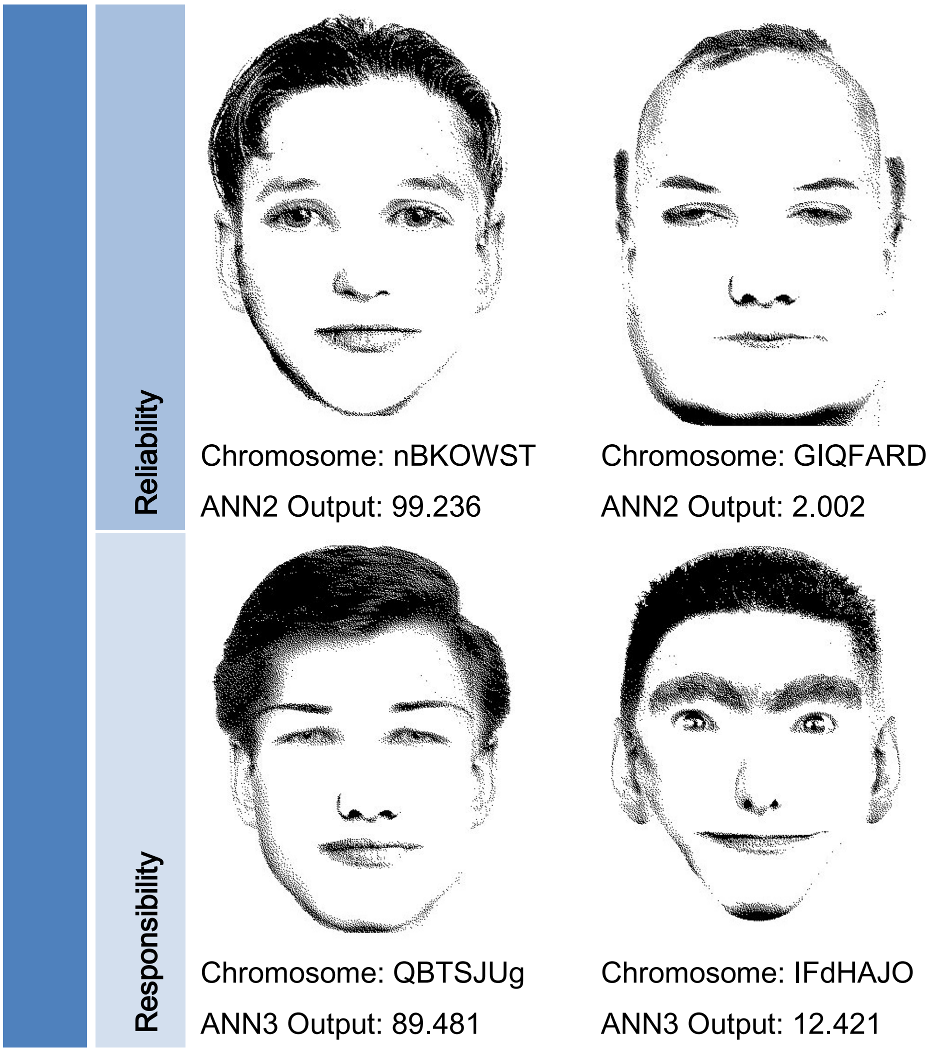


Figure 7. Best and worst rated avatars found by the GA for each sensation.

In order to validate these results with actual users opinions, 9 participants were asked to use the software to manually build the face which best elicited each sensation. They were not given time constraints. Thus, 9 faces with a high user score were obtained for each sensation and added to those found by the GA. These three sets of 10 faces were presented to 10 users. They were told to rank them from the best face conveying the sensation to the worst one, for each sensation. An average ranking was computed for each face.

Table 2 displays the average ranking for the faces found by the GA. In the second column of the Table 2 the average ranking of each face found by the proposed methodology is shown. The rest of the columns show how many times these faces were ranked in first, second and third position respectively. As can be seen, although the faces designed by this procedure did not always get the best score, their average ranking is the highest (1.3, 1.5 and 1.7 respectively). All the manually designed faces obtained worst results. Second best faces got an average ranking of 2.4, 2.8 and 2.9. Moreover, the responses dispersion was lower in the faces found by the GA in the three cases.

	Average ranking	Times in 1st position	Times in 2nd position	Times in 3rd position
Attractiveness	1.3	7	3	0
Reliability	1.5	6	3	1
Responsibility	1.7	4	5	1

Table 2. Average ranking of the faces found by the GA based on actual users opinions.

7. Discussion

The proposed procedure has successfully designed avatar faces conveying the desired sensation, according to the results. This procedure has accomplished this task with three different sensations (attractiveness, reliability and responsibility). The best generated avatars were the best scored in average. It is to be noted that this procedure is based on training the network using responses given by many people, and therefore it tries to maximize the average score.

Figure 6 exposes the differences between users' opinions and values predicted by the ANN1 for faces in the test set. Some cases show significant differences between the real user response and the value given by the network. As the network is trained to maximize the average score, it may in some cases predict values that differ significantly from a particular participant response. To minimize this error, an alternative procedure could be performed in which each user would evaluate all the faces. Then the average value would be used to train the network. But this procedure presents a drawback: Since the face sample size is very large, it would take a long time for the users to respond to the survey. This can lead to fatigue, which reduces the reliability of the data obtained (Brace, 2013; Savage & Waldman, 2008).

The MSE obtained has been very different in each sensation. MSE produced by the ANN in "Responsibility" (0.0706) is more than three times the one produced in "Reliability" (0.0209). There are several explanations to this. Maybe the response to some sensations is more disperse, thus hindering the network fitness. Maybe words utilized to describe a sensation are less accurate in some cases. For example, consensus about the meaning of "Reliability" may be greater than that about the meaning of "Responsible".

The ANNs are trained to produce avatar faces conveying a desired sensation as much as possible. But they can also in turn be used to minimize the transmission of that sensation. It is to be noted that a face generated this way would not necessarily be conveying the opposite sensation, due to the procedure followed to train the network. The ANN programmed to find faces conveying a lot of reliability can generate faces conveying very little reliability, but not necessarily “unreliability”.

Results show that the faces produced by the ANNs get better scores and average rankings than those created by the users, and present a lower dispersion in the users’ judgments. For instance, the face proposed by the GA to transmit the sensation of “Attractiveness” got an average ranking of 1.3, reaching the first place seven times and the second place three times. Second best face got an average ranking of 2.4, reaching the first place only twice, second place four times, and third and fourth place also twice. Thus, the data dispersion was greater in this case.

Even though the results are satisfactory, a more in-depth research is needed. Probably the results can be improved by increasing the number or traits and the levels in each trait. Color of eyes or hair could be considered, as well as the position of different elements in the face. This procedure could also be extended to a full body avatar, thus including a number of other variables related to the body and clothes.

It is also worth noting that the use of 2D static avatars limits the expressive communication. A research using 3D animated characters could enrich the experience and provide a more meaningful interaction. This is especially interesting in cases where the accuracy in the representation of a sensation is important, such as computer based therapeutic applications (as those used for treatment of autism). This issue could also be approached by using different expressive traits meaning to convey not only sensations, but also moods.

Finally, the proposed procedure could be utilized to design an avatar conveying a mix of sensations simultaneously. It could represent a character being attractive, reliable and responsible. In this case, once obtained each ANN, a weighted sum of the outputs of these networks could be used as fitness function for the GA. An example is provided in Figure 8. It shows the six faces found by the three networks working simultaneously, the score for each one in each sensation and the average sensation score

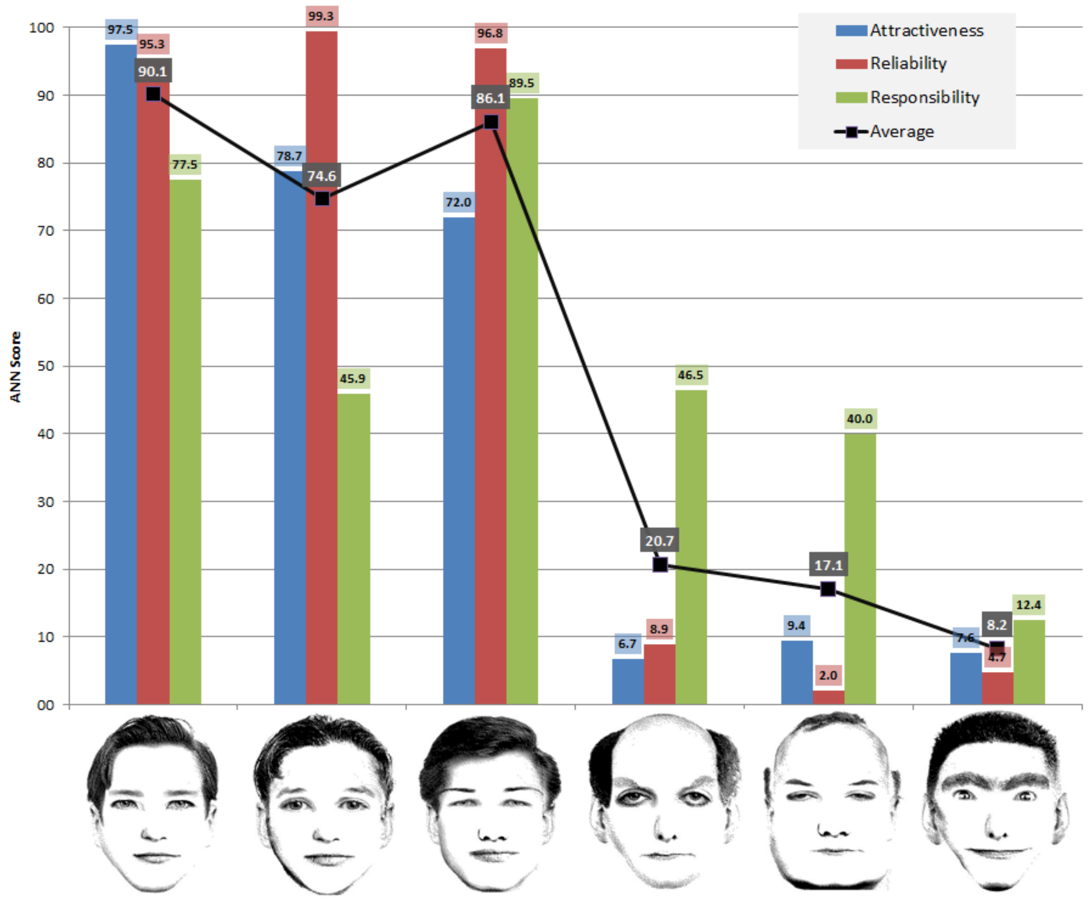


Figure 8. Scores for the six faces found by the three networks working simultaneously.

8. Conclusions

The procedure presented in this paper, based on the combined use of GA and ANN, has successfully produced avatar faces that convey a given sensation. The results have also been satisfactorily contrasted by evaluating system versus user produced faces. According to the study, the system is able to effectively generate and evaluate avatar faces to match the judgments of a set of users and look for the most suitable alternative.

In spite of this, more research is needed in order to generalize these results to other sensations and facial traits. Differences have been found among the accuracy of the results for the three sensations under study, and the use of more realistic avatars could improve the transmission of the desired message. It is to be noted that the present work is based purely on the physical features of the neutral face, with no consideration about gestures and poses. Including them in further research would provide meaningful information, especially about the transmission of mixed sensations.

References

Allenby, G. M., & Ginter, J. L. (1995). Using Extremes to Design Products and Segment Markets. *Journal of Marketing Research*, *32*, 392. doi:10.2307/3152175

Bainbridge, W. S. (2007). The scientific research potential of virtual worlds. *Science (New York, N.Y.)*, *317*(5837), 472–476. doi:10.1126/science.1146930

Banakou, D., Chorianopoulos, K., & Anagnostou, K. (2009). Avatars' Appearance and Social Behavior in Online Virtual Worlds. *2009 13th Panhellenic Conference on Informatics*. doi:10.1109/PCI.2009.9

Barbat, B. E., & Cretulscu, R. (2003). Affordable Affective Avatars. Persuasion, Emotions and Language (s). In *Romania* (pp. 25–35). Publ. Centre Technological Educational Institute of Thessaloniki.

Bartneck, C., Takahashi, T., & Katagiri, Y. (2004). Cross-Cultural Study of Expressive Avatars. In *Proceedings of the Social Intelligence Design 2004* (pp. 21–28). Twente.

Behrend, T., Toaddy, S., Thompson, L. F., & Sharek, D. J. (2012). The effects of avatar appearance on interviewer ratings in virtual employment interviews. *Computers in Human Behavior*, *28*(6), 2128–2133.

Boberg, M., Piippo, P., & Ollila, E. (2008). Designing avatars. In *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts DIMEA 08* (pp. 232–239). ACM. doi:10.1145/1413634.1413679

Brace, I. (2013). *Questionnaire design: how to plan, structure, and write survey material for effective market research*. (K. Page, Ed.) (3rd ed., p. 288). London.

Chen, C., & Yan, W. (2008). An in-process customer utility prediction system for product conceptualisation. *Expert Systems with Applications*, *34*(4), 2555–2567. doi:10.1016/j.eswa.2007.04.019

Chung, D., DeBuys, B., & Nam, C. (2007). Influence of avatar creation on attitude, empathy, presence, and para-social interaction. In *Human-Computer Interaction. Interaction Design and Usability* (pp. 711–720). Springer Berlin / Heidelberg. doi:10.1007/978-3-540-73105-4_78

Dasgupta, C. G., Dispensa, G. S., & Ghose, S. (1994). Comparing the predictive performance of a neural network model with some traditional market response models. *International Journal of Forecasting*. doi:10.1016/0169-2070(94)90004-3

Davis, A., Murphy, J., Owens, D., Khazanchi, D., & Zigurs, I. (2009). Avatars, people, and virtual worlds: Foundations for research in metaverses. *Journal of the Association for Information Systems*, *10*(2), 90–117. doi:1660426061

Di Fiore, F., Quax, P., Vanaken, C., Lamotte, W., & Van Reeth, F. (2008). Conveying Emotions through Facially Animated Avatars in Networked Virtual Environments . In A. Egges, A. Kamphuis, & M. Overmars (Eds.), *Motion in Games* (Vol. 5277, pp. 222–233). Springer Berlin Heidelberg. doi:10.1007/978-3-540-89220-5

Ducheneaut, N., Wen, M., Yee, N., & Wadley, G. (2009). Body and Mind: A Study of Avatar Personalization in Three Virtual Worlds. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1151–1160. doi:10.1145/1518701.1518877

El-Nasr, M. S., Ioerger, T. R., Yen, J., House, D. H., & Parke, F. I. (1999). Emotionally expressive agents. *Proceedings Computer Animation 1999*. doi:10.1109/CA.1999.781198

Engel, J. F., Blackwell, R. D., & Miniard, P. W. (1995). *Consumer behavior* (8th ed.). Fort Worth: Dryden Press.

Fabri, M., & Elzouki, S. (2007). Emotionally expressive avatars for chatting, learning and therapeutic intervention. In J. A. Jacko (Ed.), *Human-Computer Interaction, HCI Intelligent Multimodal Interaction Environments 12th International Conference* (pp. 275–285). Springer Berlin / Heidelberg. doi:10.1007/978-3-540-73110-8

Fabri, M., & Moore, D. (2005). The use of emotionally expressive avatars in Collaborative Virtual Environments. *Sin Nombre*.

Foley, H. J., & Matlin, M. W. (2010). *Sensation and perception* (5ft ed.). Boston: Allyn & Bacon.

Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*. doi:10.1016/0305-0548(86)90048-1

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Boston, MA: Addison-Wesley Longman Publishing Co., Inc.

Gratch, J., & Marsella, S. (2001). Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. *International Conference on Autonomous Agents*, 278–285. doi:10.1145/375735.376309

Hasler, B. S., Tuchman, P., & Friedman, D. (2013). Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds. *Computers in Human Behavior*, *29*, 1608–1616. doi:10.1016/j.chb.2013.01.004

Hemp, P. (2006). Avatar-Based Marketing. *Harvard Business Review*, *84*(6), 48–57. doi:10.1037/e661492007-001

Hoch, S. J., Kim, B., Montgomery, A. L., Rossi, P. E., & Montgomery, L. (1995). Determinants of Store-Level Price. *Journal of Marketing Research*, *32*, 17–29. doi:10.2307/3152107

Holzwarth, M., Janiszewski, C., & Neumann, M. M. (2006). The Influence of Avatars on Online Consumer Shopping Behavior. *Journal of Marketing*. doi:10.1509/jmkg.70.4.19

Hsiao, S.-W., & Huang, H. . (2002). A neural network based approach for product form design. *Design Studies*, *23*(1), 67–84. doi:10.1016/S0142-694X(01)00015-1

Hsiao, S.-W., & Tsai, H.-C. (2005). Applying a hybrid approach based on fuzzy neural network and genetic algorithm to product form design. *International Journal of Industrial Ergonomics*, *35*(5), 411–428. doi:10.1016/j.ergon.2004.10.007

Ishihara, S., Ishihara, K., Nagamachi, M., & Matsubara, Y. (1997). An analysis of Kansei structure on shoes using self-organizing neural networks. *International Journal of Industrial Ergonomics*. doi:10.1016/S0169-8141(96)00006-6

Kafai, Y. B., Fields, D. A., & Cook, M. S. (2010). Your Second Selves: Player-Designed Avatars. *Games and Culture*. doi:10.1177/1555412009351260

Keeling, K., McGoldrick, P., & Beatty, S. (2010). Avatars as salespeople: Communication style, trust, and intentions. *Journal of Business Research*, *63*(8), 793–800.

Koda, T. (2004). Interpretation of Emotionally Expressive Characters in an Intercultural Communication. In M. Negoita, R. Howlett, & L. Jain (Eds.), *Knowledge-Based Intelligent Information and Engineering Systems SE - 113* (Vol. 3214, pp. 862–868). Springer Berlin Heidelberg. doi:10.1007/978-3-540-30133-2_113

Koda, T., & Ishida, T. (2006). Cross-cultural study of avatar expression interpretations. *International Symposium on Applications and the Internet (SAINT'06)*. doi:10.1109/SAINT.2006.19

Koda, T., Ishida, T., Rehm, M., & André, E. (2009). Avatar culture: cross-cultural evaluations of avatar facial expressions. *AI & SOCIETY*. doi:10.1007/s00146-009-0214-5

Koutsabasis, P., Vosinakis, S., Malisova, K., & Paparounas, N. (2012). On the value of Virtual Worlds for collaborative design. *Design Studies*, *33*(4), 357–390. doi:10.1016/j.destud.2011.11.004

Lai, H.-H., Lin, Y.-C., & Yeh, C.-H. (2005). Form design of product image using grey relational analysis and neural network models. *Computers & Operations Research*, *32*(10), 2689–2711. doi:10.1016/j.cor.2004.03.021

Lai, H.-H., Lin, Y.-C., Yeh, C.-H., & Wei, C.-H. (2006). User-oriented design for the optimal combination on product design. *International Journal of Production Economics*, *100*(2), 253–267. doi:10.1016/j.ijpe.2004.11.005

Lee, J.-E. R. (2014). Does virtual diversity matter?: Effects of avatar-based diversity representation on willingness to express offline racial identity and avatar customization. *Computers in Human Behavior*, *36*, 190–197. doi:http://dx.doi.org/10.1016/j.chb.2014.03.040

Lin, H., & Wang, H. (2014). Avatar creation in virtual worlds: Behaviors and motivations. *Computers in Human Behavior*, *34*, 213–218.

Lynch, M. B., Dagli, C. H., & Vallenki, M. (1999). Use of feedforward neural networks for machine vision calibration. *International Journal of Production Economics*, *60*, 479–489. doi:10.1016/S0925-5273(98)00199-6

Marin, A. L., Jo, D., & Lee, S. (2013). Designing robotic avatars Are user's impression affected by avatar's age? In *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on* (pp. 195–196). doi:10.1109/HRI.2013.6483568

Marren, A., Harston, C., & Pap, R. (1990). *Handbook of neural computing applications*. San Diego: Academic Press Inc.

Mazzei, D., Lazzeri, N., Hanson, D., & De-Rossi, D. (2012). HEFES: An Hybrid Engine for Facial Expressions Synthesis to control human-like androids and avatars. *Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS & EMBS International Conference on*, 195–200. doi:10.1109/BioRob.2012.6290687

Merola, N., & Pena, J. (2010). The Effects of Avatar Appearance in Virtual Worlds. *Journal of Virtual Worlds Research*, *2*(5), 3–12.

Nasoz, F., & Lisetti, C. L. (2006). MAUI avatars: Mirroring the user's sensed emotions via expressive multi-ethnic facial avatars. *Journal of Visual Languages and Computing*, *17*(5), 430–444.

Nowak, K. L., & Rauh, C. (2005). The Influence of the Avatar on Online Perceptions of Anthropomorphism, Androgyny, Credibility, Homophily, and Attraction. *Journal of Computer-Mediated Communication*, *11*, 153–178. doi:10.1111/j.1083-6101.2006.tb00308.x

Nowak, K. L., & Rauh, C. (2008). Choose your “buddy icon” carefully: The influence of avatar androgyny, anthropomorphism and credibility in online interactions. *Computers in Human Behavior*, *24*(4), 1473–1493.

Orvalho, V., Miranda, J., & Sousa, A. A. (2009). Facial Synthesys of 3D Avatars for Therapeutic Applications. *Studies in Health Technology and Informatics*, *144*, 96–98.

Papadimitriou, C. H., & Steiglitz, K. (1982). *Combinatorial Optimization: Algorithms and Complexity*. *PrinticeHall New Jersey* (Vol. 91, p. 496). New York: Dover Publications Inc. doi:10.2307/2322374

Principe, J. C., Euliano, N. R., & Lefebvre, W. . (2000a). *Neural and Adaptative Systems: Fundamentals Throngh Simulations*. New York: Wiley and Sons.

Principe, J. C., Euliano, N. R., & Lefebvre, W. . (2000b). *Neural and Adaptative Systems: Fundamentals Throngh Simulations*. New York: Wiley and Sons.

Qiu, L. (2006). *Designing social interactions with animated avatars and speech output for product recommendation agents in electronic commerce*. *ProQuest Dissertations and Theses*. The University of British Columbia (Canada).

Qiu, L., & Benbasat, I. (2005). Online Consumer Trust and Live Help Interfaces: The Effects of Text-to-Speech Voice and Three-Dimensional Avatars. *International Journal of Human-Computer Interaction*. doi:10.1207/s15327590ijhc1901_6

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. . Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1, pp. 318–362). Cambridge: The MIT Press.

Sarle, W. S. (1995). Stopped Training and Other Remedies for Overfitting. *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, 352–360. Retrieved from citeseer.ist.psu.edu/sarle95stopped.html

Savage, S. J., & Waldman, D. M. (2008). Learning and fatigue during choice experiments: a comparison of online and mail survey modes. *Journal of Applied Economics*, *23*, 351–371. doi:10.1002/jae.984

Schultze, U. (2010). Embodiment and presence in virtual worlds: a review. *Journal of Information Technology*. doi:10.1057/jit.2010.25

Srinivas, M., & Patnaik, L. M. (1994a). Genetic algorithms: a survey. *Computer*, *27*. doi:10.1109/2.294849

Srinivas, M., & Patnaik, L. M. (1994b). Genetic algorithms: a survey. *Computer*, *27*. doi:10.1109/2.294849

Taylor, T. L. (2002). Living Digitally : Embodiment in Virtual Worlds. *The Social Life of Avatars Presence and Interaction in Shared Virtual Environments*, (3), 40–62. doi:10.1088/0957-4484/16/9/062

Tsai, H.-C., Hsiao, S.-W., & Hung, F.-K. (2006). An image evaluation approach for parameter-based product form and color design. *Computer-Aided Design*, *38*(2), 157–171. doi:10.1016/j.cad.2005.10.001

Vasalou, A., Joinson, A., Bänziger, T., Goldie, P., & Pitt, J. (2008). Avatars in social media: Balancing accuracy, playfulness and embodied messages. *International Journal of Human Computer Studies*, *66*(11), 801–811.

Vasalou, A., & Joinson, A. N. (2009). Me, myself and I: The role of interactional context on self-presentation through avatars. *Computers in Human Behavior*, *25*(2), 510–520.

Wang, Y., & Geigel, J. (2011). Using Facial Emotional Signals for Communication between Emotionally Expressive Avatars in Virtual Worlds. In *2011 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 297–304). doi:10.1007/978-3-642-24571-8_33

Weibel, D., Stricker, D., Wissmath, B., & Mast, F. W. (2010). How Socially Relevant Visual Characteristics of Avatars Influence Impression Formation. *Journal of Media Psychology: Theories, Methods, and Applications*. doi:10.1027/1864-1105/a000005

Yang, C.-C., & Shieh, M.-D. (2010). A support vector regression based prediction model of affective responses for product form design. *Computers & Industrial Engineering*, *59*(4), 682–689. doi:10.1016/j.cie.2010.07.019

Yee, N., & Bailenson, J. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, *33*(3), 1–38. doi:10.1111/j.1468-2958.2007.00299.x

Yu, H., Garrod, O. G. B., & Schyns, P. G. (2012). Perception-driven facial expression synthesis. *Computers & Graphics*, *36*(3), 152–162. doi:10.1016/j.cag.2011.12.002