Emotional virtual agents: how do young people decode synthetic facial expressions?

Anna Esposito [a], Terry Amorese [a], Marialucia Cuciniello [a] and Gennaro Cordasco [a]

[a] *Università degli studi della Campania "L. Vanvitelli", Caserta, Italy and International Institute for Advanced Scientific Studies (IIASS), Vietri sul Mare, Italy*

Abstract

Given the need of remote learning and the growing presence of virtual agents within online learning environments, the present research aims at investigating young people' ability to decode emotional expressions conveyed by virtual agents. The study, involves 50 healthy participants aged between 22 and 35 years (mean age=27.86; SD= ±2.75; 30 females) which were required to label pictures and video clips depicting female and male virtual agents of different ages (young, middle-aged and old) displaying static and dynamic expressions of disgust, anger, sadness, fear, happiness, surprise and neutrality. Depending on the emotional category, significant effects were observed for the agents' age, gender, and type of administered (static vs dynamic) stimuli on the young people' decoding accuracy of the virtual agents' emotional faces. Anger was significantly more accurately decoded in male rather than female faces while the opposite result was observed for happy, fearful, surprised, and disgusted faces. Middle aged faces were generally more accurately decoded than young and old emotional faces except for sadness and disgust. Significantly greater accuracy was observed for dynamic vs static faces of disgust, sadness, and fear, in contrast to static vs dynamic neutral and surprised faces.

Keywords

Virtual agents, gender differences, age differences, online learning, emotional decoding of synthetic facial expressions

1. Introduction

The Covid-19 pandemic disease, and subsequently the need to respect social distancing, has changed every aspect of our lives, including school and teaching. From a traditional educational environment, based on face-to-face lectures in a classroom, we have witnessed a sudden shift toward an online teaching modality [1]. Adaption to this new situation posed huge challenges to teachers and students [2], and suggested investigations devoted to identifying effective and efficient strategies to enhance the effectiveness of online teaching methodologies. A possibility is represented by intelligent user interface as virtual agents. Online learning environments in which learners are allowed to interact with a virtual agent might have a positive effect on students' performances and satisfaction [3], ensuring that the learning environment is perceived by students as pleasant and enjoyable [4]. When dealing with conversational technologies as virtual agents it is necessary to take into account users' acceptance of these systems. Virtual agents' acceptance depends on several factors, such as agents' gender, voices, and appearance [5]. A key aspect related to the appearance and the design of a virtual agent concerns their ability to manifest emotional expressions. It has been shown [6] that people prefer to communicate with virtual agents which can show emotional facial expressions rather than with "emotion-less" agents, although the expressions are subtle and not particularly marked.

EMAIL: {anna.esposito, terry.amorese, marialucia.cuciniello,gennaro.cordasco}@unicampania.it

Emotions play an essential role in learning [7], consequently even in online learning contexts, the ability of a virtual agent to express emotions is fundamental [8]. Agents conveying emotional expressions could affect students' emotional state [9] and help students surmount learning constraints [10], increasing their enthusiasm towards the covered topics and therefore also their interest, self- efficacy, and motivation [11]. Given the relevance of virtual agents' capability to express emotions, even within blended education contexts, it is necessary to understand and investigate how people recognize synthetic emotions, namely emotional expressions conveyed by virtual agents. So far, studies carried out with the aim of comparing peoples' ability to decode emotions represented by virtual agents and real humans reached discordant results, bringing out two main trends: on the one hand is highlighted the absence of significant differences in the recognition of emotions expressed by virtual agents and humans [12], on the other hand several studies outlined that people better recognize human emotional expressions rather than synthetic ones [13,14]. However, results can vary depending on the emotional category considered. For instance, as regards disgust, it's better recognized when expressed by real human faces [15], while sadness and fear seem to be better recognized if represented by synthetic faces [16]; instead in the case of anger, positions are conflicting since some studies showed that this emotion is better recognized when expressed by synthetic faces [15]while others have found a greater accuracy in the recognition of anger when expressed by human faces [17].

With the intent of further investigate this issue, and also to provide useful guidelines for the implementation of virtual agents which can be exploited in educational contexts, a study is proposed in which young participants have been required to recognize emotional expressions conveyed by virtual agents. Unique humanoid female and male agents of different ages were developed for the experiment, conveying both unimodal and multimodal expressions of disgust, anger, sadness, fear, happiness, surprise, and neutrality.

2. Materials and Methods

This study aims at investigating young participants' decoding ability of virtual agents' emotional expressions and the effect of stimuli's age (young, middle-aged, and old virtual agents), gender (female and male virtual agents), and type (static pictures and dynamic video clips) on participants' decoding accuracy.

2.1. Participants

The experiment involved a total of 50 young participants aged between 22 and 35 years (mean age=27.86; SD= ±2.75; 30 females and 20 males), all in a good health status, which were administered a decoding task of virtual agents' emotional expressions. Participants, all Italians living in the Campania region, due to the Covid-19 pandemic disease and the need to respect social distancing, were recruited exploiting telematic tools as social networks and e-mails; they voluntarily joined the study and read and agreed to an informed consent formulated according to the Italian and European laws about privacy and data protection. The research was authorized by the ethical committee of the Department of Psychology at the Università degli Studi della Campania "Luigi Vanvitelli" with the protocol number 25/2017.

2.2. Stimuli

The experiment consisted in an emotion decoding task in which participants were showed pictures and video clips depicting female and male differently aged (young, middle-aged and old) virtual agents expressing disgust, anger, sadness, fear, happiness, surprise and neutrality. Virtual agents were developed in collaboration with Paphus Solutions Inc., a Canadian corporation specialized in the development of bots, artificial intelligence, and deep learning products and services. The proposed virtual agents were developed using Daz3D software which provides emotion pose presets for most emotions. For each emotion, the expression morph from Daz3D has been used at the 100% magnitude setting. A video clip for each agent was developed, then to get every agent talk, audio files were extracted from video clips of the AFEW database [18] and embed within each agent video clip. Each agent was assessed in a qualitative manner by experts in the field of emotional interactional exchanges and human-machine interaction. For the current experiment, 90 virtual agents were exploited. Six of such agents were used within the trial session (3 static pictures and 3 dynamic video clips), and the remaining 84 (42 static pictures and 42 dynamic video clips) in the experimental session of the present study. Fig.1 illustrates some examples of virtual agents used within the experiment.

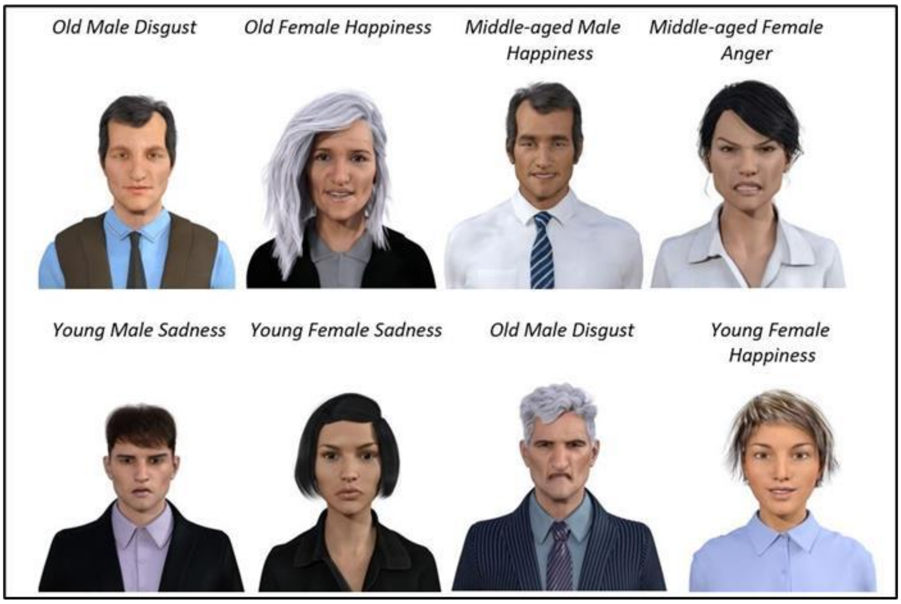

Figure 1: A sample of stimuli exploited for the current investigation.

2.3. Tools and Procedures

An emotion decoding task was developed using the online study builder Lab.js and then exported on JATOS (Just Another Tool for Online Studies), which allows to generate links which can be sent to participants. Volunteers were recruited exploiting social networks and e-mail addresses. Each participant who agreed to join the study was provided with a link and asked to open the link from a laptop. Once opened the link, participants were asked to give their consent to a personal data processing form, subsequently participants' demographic data and information about their degree of experience with technology were collected. After this, on the screen appeared the instructions required to carry out the experiment, followed by a trial session (composed by 6 stimuli, 3 static pictures and 3 dynamic video clips) and by an experimental session (composed by 84 stimuli, 42 static pictures and 42 dynamic video clips). Both the trial and the experimental session consisted in the presentation of randomized stimuli. For each stimulus, participants had to attach an emotional label choosing among these options: disgust, anger, sadness, fear, happiness, surprise, neutrality, or other emotion.

1. Results

The following sections describe the assessment of the collected data through statistical analyses.

1.1. Results- Age, gender, and type of stimuli effects for each emotional category

Data acquired from participants were analyzed in order to explore how stimuli's age, gender and type affected their ability to decode the proposed virtual agents' emotional faces. Separate ANOVA repeated measures analyses were carried out on the decoding accuracy of each emotional category (disgust, anger, sadness, fear, happiness, surprise, and neutrality) considering participants' gender as between subjects factor. Age (young, middle-aged, and old agents), gender (female and male agents) and type (static pictures and dynamic video clip) of stimuli were considered as within subjects factors. The significance level was set at α <.05 and differences among means were assessed through Bonferroni's post hoc tests. Results for each emotional category are reported below.

Disgust

No significant effects of participants' gender (F (1,48) =1.870, p= .178) emerged. No significant effects of the age of stimuli (F (2,96) =2.647, p=.076) were observed. Significant effects of the gender of stimuli were observed (F (1,48) =29.214, p<<.01). Bonferroni's post hoc tests revealed that participants were significantly more accurate in decoding synthetic emotional faces of disgust when expressed by female stimuli (mean=.503, p<<.01) rather than when expressed by male stimuli (mean=.311). Significant effects of the type of stimuli were observed (F (1,48) =45.004, p<<.01). Bonferroni's post hoc tests revealed that participants were significantly more accurate in decoding synthetic emotional faces of disgust when conveyed by dynamic stimuli (mean=.535, p<<.01) rather than when conveyed by static stimuli (mean=.279), these results are showed in Fig.2. A significant interaction between age and gender of stimuli (F (2,96) =9.555, p<<. 01) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and gender of stimuli). These tests revealed that:

a) Concerning the age of the faces: old male virtual agents (mean=.463) were more accurately decoded rather than young male (mean=.271, p=.007) and middle-aged male virtual agents' faces of disgust (mean=.200, p<<.01).
b) Concerning the gender of stimuli: participants were more accurate in decoding young female (mean=.479) rather than young male virtual agents' disgusted faces (mean=.271, p=.002), and moreover they were more accurate in decoding middle-aged female (mean=.575) rather than middle-aged male virtual agents' disgusted faces (mean=.200, p<<.01).

A significant interaction between age and type of stimuli (F (2,96) =6.733, p=.002) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were more accurate in decoding static middle-aged (mean=.338) rather than static young disgusted synthetic faces (mean=.175, p=.011), and moreover they were more accurate in decoding dynamic old (mean=.592) rather than dynamic middle-aged synthetic disgusted faces (mean=.438, p= .038).

b) Concerning the type of stimuli: participants were more accurate in decoding dynamic young (mean=.575) rather than static young synthetic disgusted faces (mean=.175, p<< .01), and in addition, they were more accurate in decoding dynamic old (mean=.592) rather than static old synthetic disgusted faces (mean=.325, p<<.01).

A significant interaction among age, gender, and type of stimuli (F (2,96) =3.751, p=.027) was found. Since this was a triple interaction, Bonferroni's post hoc tests were performed for each single factor (age, gender, and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were more accurate in decoding middle-aged female static (mean=.533) rather than young female static (mean=.292, p=.018), and old female static virtual agents' disgusted faces (mean=.217, p=.001). Moreover, they were more accurate in decoding old male static (mean=.433) rather than young male static (mean=.058, p<<.01), and middle-aged male static virtual agents' disgusted faces (mean=.142, p=.001). In addition, old male dynamic faces (mean=.492) were better recognized than middle-aged male virtual agents' dynamic stimuli (mean=.258, p= .050).

b) Concerning the gender of stimuli: participants were more accurate in decoding young static female (mean=.292) rather than young static male disgusted virtual agents' faces (mean=.058, p=.001). In addition, they were more accurate in decoding middle-aged static female (mean=.533) rather than middle-aged static male virtual agents' disgusted faces (mean=.142, p<<.01) and they better recognized middle-aged dynamic female (mean=.617) rather than middle-aged dynamic male virtual agents' disgusted faces (mean=.258, p<<.01). Conclusively, they were more accurate in decoding old static male (mean=.433) rather than old static female disgusted faces (mean=.217, p=.009) and they better recognized old dynamic female (mean=.692) rather than old dynamic male virtual agents' disgusted faces (mean=.492, p=.042).

c) Concerning the type of stimuli: participants were more accurate in decoding dynamic female young (mean=.667) rather than static female young synthetic disgusted faces (mean=.292, p<< .01), and in addition, they were more accurate in decoding dynamic male young (mean=.483) rather than static male young synthetic disgusted faces (mean=.058, p<<.01). Conclusively, participants better recognized dynamic female old (mean=.692) rather than static female old virtual agents' disgusted faces (mean=.217, p<< .01).

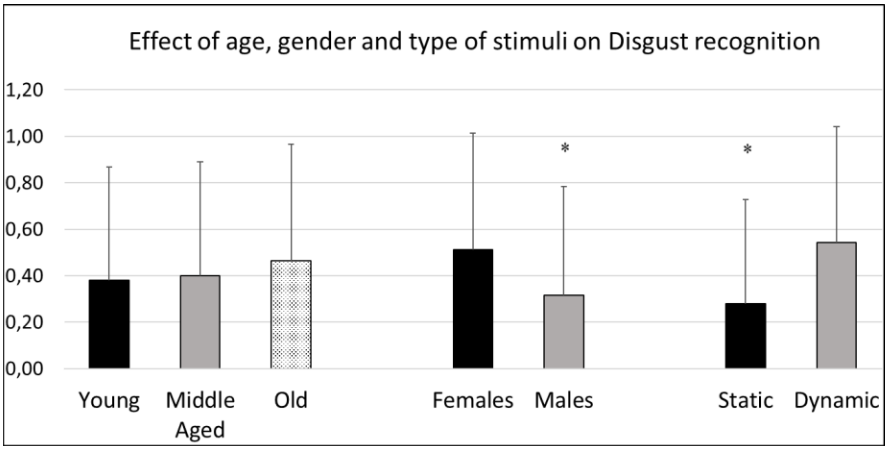

Figure 2: Agents' age, gender, and type (static vs dynamic) of stimuli effects on young participants' decoding accuracy of disgusted faces.

Anger

No significant effects of participants' gender (F (1,48) =1.375, p= .247) emerged. Significant effects of the age of stimuli (F (2,96) =7.417, p=.001) were observed. Bonferroni's post hoc tests revealed that participants were less accurate in decoding old (mean=.744) rather than young (mean=.840, p=.016) and middle-aged virtual agents' angry faces (mean=.869, p= .005). Significant effects of the gender of stimuli were observed (F (1,48) =40.806, p<<.01). Bonferroni's post hoc tests revealed that participants were significantly more accurate in decoding synthetic faces of anger when expressed by male stimuli (mean=.907, p<<.01) rather than when depicted by female stimuli (mean=.728), these results are summarized in Fig.3. No significant effects of the type of stimuli were observed (F (1,48) =1.458, p=.233). A significant interaction between participants' gender and gender of stimuli (F (1,48) =4.966, p=.031) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (participants' gender and gender of stimuli). These tests revealed that:

a) Concerning participants' gender: male participants better recognized male (mean=.908) rather than female synthetic faces expressing anger (mean=.667, p<<.01). This happened also for female participants which better recognized male (mean=.906) rather than female synthetic faces conveying anger (mean=.789, p= .002). A significant interaction between age and gender of stimuli (F (2,96) =3.401, p=.037) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and gender of stimuli). These tests revealed that:

a) Concerning the age of the faces: old female synthetic faces expressing anger (mean=.608) were less accurately decoded by participants compared to young (mean=.767, p=.011) and middle- aged synthetic female faces (mean =.808, p=.003).

b) Concerning the gender of stimuli: participants were more accurate in decoding young male (mean=.913) rather than young female virtual agents' angry faces (mean=.767, p=.001), moreover they were more accurate in decoding middle-aged males (mean=.929) rather than middle-aged females virtual agents' angry faces (mean=.808, p=.005). The same happened for old males (mean=.879) that were better decoded by participants rather than old female synthetic angry faces (mean=.608, p<<.01).

A significant interaction between age and type of stimuli (F (2,96) =8.179, p=.001) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were more accurate in decoding static middle-aged (mean=.921) rather than static young (mean=.763, p=.002) and static old virtual agents' angry faces (mean=.708, p<<.01), moreover they were more accurate in decoding dynamic young (mean=.917) rather than dynamic old virtual agents' angry faces (mean=.779, p=.008).

b) Concerning the type of stimuli: participants were more accurate in decoding dynamic young (mean=.917) rather than static young virtual agents' angry faces (mean=.763, p=.003), and in addition, they were more accurate in decoding static middle-aged (mean=.921) rather than dynamic middle-aged virtual agents' angry faces (mean=.817, p=.022).

A significant interaction among age, gender, and type of stimuli (F (2,96) =5.770, p=.004) was found. Since this was a triple interaction, Bonferroni's post hoc tests were performed for each single factor (age, gender, and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were more accurate in decoding middle-aged female static (mean=.883) rather than young female static (mean=.617, p=.001) and old female static virtual agents' angry faces (mean=.533, p<<.01). Moreover, they were more accurate in decoding young female dynamic (mean=.917) rather than middle-aged female dynamic (mean=.733, p=.024), and old female dynamic virtual agents' angry faces (mean=.683, p=.001).

b) Concerning the gender of stimuli: participants were more accurate in decoding young static male (mean=.908) rather than young static female virtual agents' angry faces (mean=.617, p<<.01). In addition, they were more accurate in decoding middle-aged dynamic male (mean= .900) rather than middle-aged dynamic female virtual agents' angry faces (mean=.733, p=.011) and they better recognized old static male (mean=.883) rather than old static female virtual agents' angry faces (mean=.533, p<<.01). Conclusively, they were more accurate in decoding old dynamic male (mean=.875) rather than old dynamic female virtual agents' angry faces (mean=.683, p=.022).

c) Concerning the type of stimuli: participants were more accurate in decoding young female dynamic (mean=.917) rather than young female static virtual agents' angry faces (mean=.617, p=.001) while they better recognized middle-aged female static (mean=.883) rather than middle-aged female dynamic virtual agents' angry faces (mean=.733, p=.027).

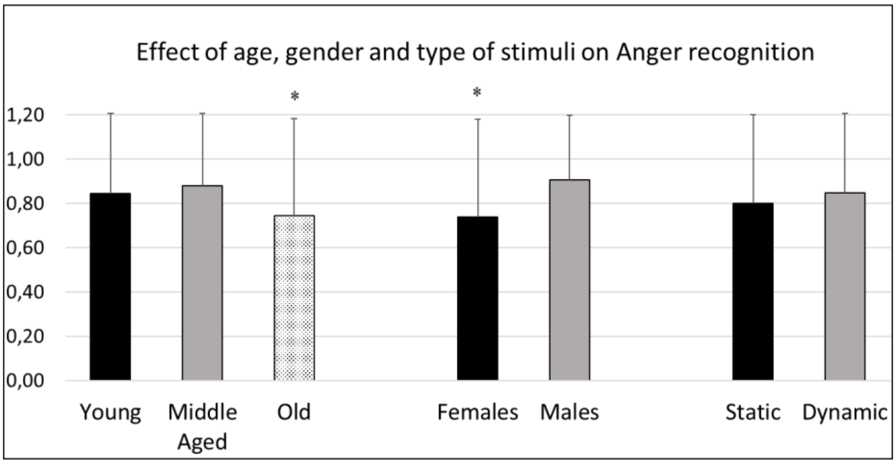

Figure 3: Agents' age, gender, and type (static vs dynamic) of stimuli effects on young participants' decoding accuracy of angry faces.

Sadness

No significant effects of participants' gender (F (1,48) =.518, p= .475) emerged. Significant effects of the age of stimuli (F (2,96) =19.179, p<<.01) were observed. Bonferroni's post hoc tests revealed that participants were less accurate in decoding young (mean=.406) rather than old (mean=.596, p<<.01) and middle-aged virtual agents' sad faces (mean=.656, p<<.01). No significant effects of gender of stimuli were observed (F (1,48) =.516, p=.476). Significant effects of the type of stimuli were found (F (1,48) =42.926, p<<.01). Bonferroni's post hoc tests revealed that participants were significantly more accurate in decoding synthetic emotional faces expressing sadness when conveyed by dynamic stimuli (mean=.669, p<<.01) rather than when depicted by static stimuli (mean=.436), these results are showed in Fig.4. A significant interaction between age and type of stimuli (F (2,96) =20.935, p<<.01) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were less accurate in decoding static young (mean=.125) rather than static middle-aged (mean=.558, p<<.01) and static old virtual agents' sad faces (mean=.625, p<<.01), and moreover they were more accurate in decoding dynamic middle-aged (mean=.754) rather than dynamic old virtual agents' sad faces (mean=.567, p=.032).

b) Concerning the type of stimuli: participants were more accurate in decoding dynamic young (mean=.688) rather than static young virtual agents' sad faces (mean=.125, p<<.01), and in addition, they were more accurate in decoding dynamic middle-aged (mean=.754) rather than static middle- aged virtual agents' sad faces (mean=.558, p=.003).

A significant interaction among participants' gender and age and type of stimuli (F (2,96) =3.295, p=.041) was found. Since this was a triple interaction, Bonferroni's post hoc tests were performed for each single factor (participants' gender, age, and type of stimuli). These tests revealed that:

a) Concerning participants' gender: female participants (mean=.750) were more accurate in decoding old static virtual agents' sad faces rather than male participants (mean=.500, p=.009).

b) Concerning the age of stimuli: male participants were less accurate in decoding young static (mean=.100) rather than middle-aged static (mean=.550, p<<.01) and old static virtual agents' sad faces (mean=.500, p<<.01). The same occurred for female participants which were less accurate in decoding young static (mean=.150) rather than middle-aged static (mean=.567, p<<.01) and old static virtual agents' sad faces (mean=.750, p<<.01). Moreover, female participants were more accurate in decoding middle-aged dynamic (mean=.783) rather than old dynamic virtual agents' sad faces (mean=.483, p=.005).

c) Concerning the type of stimuli: male participants were more accurate in decoding young dynamic (mean=.675) rather than young static virtual agents' sad faces (mean=.100, p<<.01). On the other hand, female participants were more accurate in decoding young dynamic (mean=.700) rather than young static virtual agents' sad faces (mean=.150, p<<.01). In addition, female participants were more accurate in decoding also middle-aged dynamic (mean=.783) rather than middle-aged static virtual agents' sad faces (mean=.567, p =.009). Conclusively, female participants better recognized old static (mean=.750) than old dynamic virtual agents' sad faces (mean=.483, p=.006).

A significant interaction among age, gender, and type of stimuli (F (2,96) =15.183, p<<.01) was found. Since this was a triple interaction, Bonferroni's post hoc tests were performed for each single factor (age, gender, and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were less accurate in decoding young female static (mean=.125) rather than middle-aged female static (mean=.383, p=.007) and old female static virtual agents' sad faces (mean=.675, p<<.01). Moreover, they were more accurate in decoding old female static (mean=.675) rather than middle-aged female static (mean=.383, p=.013) facial expressions and more accurate in decoding middle-aged female dynamic (mean=.858) rather than old female dynamic (mean=.475, p<<.01) virtual agents' faces. Moreover, participants were less accurate in decoding young male static (mean=.125) rather than middle-aged male static (mean=.733, p<<.01) and old male static virtual agents' sad faces (mean=.575, p<<.01).

b) Concerning the gender of stimuli: participants were more accurate in decoding middle-aged static male (mean=.733) rather than middle-aged static female virtual agents' sad faces (mean=.383, p<<.01). In addition, they were more accurate in decoding middle-aged dynamic female (mean= .858) rather than middle-aged dynamic male virtual agents' sad faces (mean=.650, p=.006) and they better recognized old dynamic male (mean=.658) rather than old dynamic female virtual agents' sad faces (mean=.475, p=.029).

c) Concerning the type of stimuli: participants were more accurate in decoding young female dynamic (mean=.708) rather than young female static virtual agents' sad faces (mean=.125, p<<.01) and they better recognized young male dynamic (mean=.667) rather than young male static virtual agents' sad faces (mean=.125, p<<.01). Moreover, participants were more accurate in decoding middle-aged female dynamic (mean=.858) rather than middle-aged female static virtual agents' sad faces (mean=.383, p<<.01) while they better recognized old female static (mean=.675) rather than old female dynamic virtual agents' sad faces (mean=.475, p=.033).

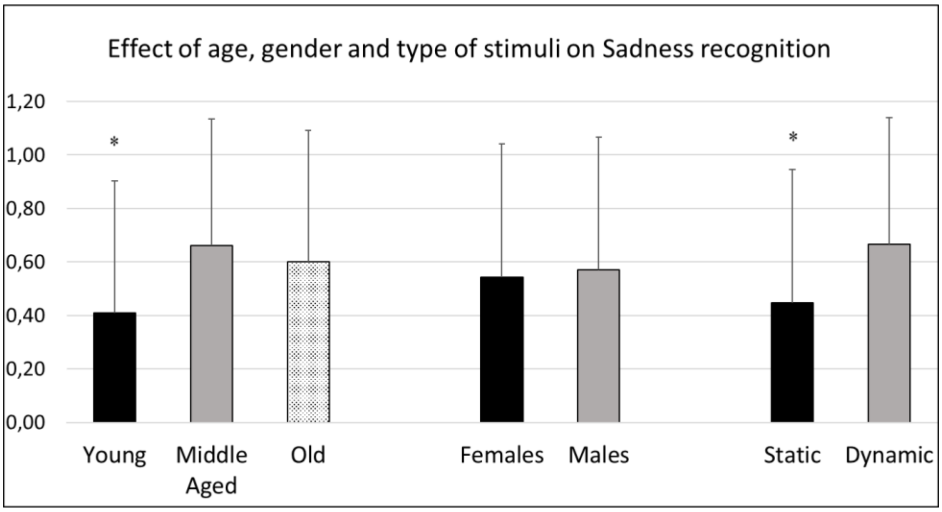

Figure 4: Agents' age, gender, and type (static vs dynamic) of stimuli effects on young participants' decoding accuracy of sad faces.

Fear

Significant effects of participants' gender (F (1,48) =7.973, p= .007) emerged. Bonferroni's post hoc tests revealed that female participants (mean=.400) were more accurate than male participants (mean=.283, p=.007) in decoding synthetic faces depicting fear. Significant effects of the age of stimuli (F (2,96) =12.497, p<<.01) were observed. Bonferroni's post hoc tests revealed that participants were more accurate in decoding middle-aged (mean=.446) than young (mean=.315, p=.001) and old virtual agents' faces of fear (mean=.265, p<<.01). Significant effects of the gender of stimuli were observed (F (1,48) =47.474, p<<.01). Bonferroni's post hoc tests revealed that participants were significantly more accurate in decoding virtual agents' faces of fear when expressed by female stimuli (mean=.447, p<<.01) rather than depicted by male stimuli (mean=.236). Significant effects of the type of stimuli were found (F (1,48) =105.817, p<<.01). Bonferroni's post hoc tests revealed that participants were significantly more accurate in decoding synthetic faces expressing fear when conveyed by dynamic stimuli (mean=.531, p<<.01) rather than when depicted by static stimuli (mean=.153), these results are showed in Fig. 5. A significant interaction between age and type of stimuli (F (2,96) =6.218, p=.003) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were more accurate in decoding static middle-aged (mean=.308) rather than static young (mean=.046, p<<.01) and static old virtual agents' faces of fear (mean=.104, p<<.01).

b) Concerning the type of stimuli: participants were more accurate in decoding dynamic young (mean=.583) rather than static young virtual agents' faces of fear (mean=.046, p<<.01). In addition, they were more accurate in decoding dynamic middle-aged (mean=.583) rather than static middle- aged virtual agents' faces of fear (mean=.308, p<<.01) and they were also more accurate in decoding dynamic old (mean=.425) rather than static old synthetic faces expressing fear (mean=.104, p<<.01).

A significant interaction among age, gender, and type of stimuli (F (2,96) =34.508, p<<.01) was found. Since this was a triple interaction, Bonferroni's post hoc tests were performed for each single factor (age, gender, and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were more accurate in decoding middle-aged female static (mean=.617) rather than young female static (mean=.058, p<<.01) and old female static virtual agents' faces of fear (mean=.175, p<<.01). Moreover, they were more accurate in decoding young female dynamic (mean=.825) rather than middle-aged female dynamic (mean=.500, p=.002) and old female dynamic virtual agents' faces of fear (mean=.508, p=.001). In addition, participants were more accurate in decoding middle-aged male dynamic (mean=.667) rather than young male dynamic (mean=.342, p=.004) and old male dynamic virtual agents' faces of fear (mean=.342, p=.001).

b) Concerning the gender of stimuli: participants were more accurate in decoding female dynamic young (mean=.825) rather than male dynamic young virtual agents' faces of fear (mean=.342, p<<.01). In addition, they were more accurate in decoding female middle-aged static (mean= .617) rather than male middle-aged static virtual agents' faces of fear (mean=.000, p<<.01) and they better recognized female old static (mean=.175) rather than male old static virtual agents' faces of fear (mean=.033, p=.022).

c) Concerning the type of stimuli: participants were more accurate in decoding young female dynamic (mean=.825) rather than young female static synthetic faces expressing fear (mean=.058, p<<.01) and they better recognized young male dynamic (mean=.342) rather than young male static synthetic faces expressing fear (mean=.033, p<<.01). Moreover, participants were more accurate in decoding middle-aged male dynamic (mean=.667) rather than middle-aged male static virtual agents' faces of fear (mean=.000, p<<.01). Conclusively, participants were more accurate in decoding old female dynamic (mean=.508) rather than old female static synthetic faces expressing fear (mean=.175, p=.001) and they better recognized old male dynamic (mean=.342) rather than old male static synthetic faces expressing fear (mean=.033, p<<.01).

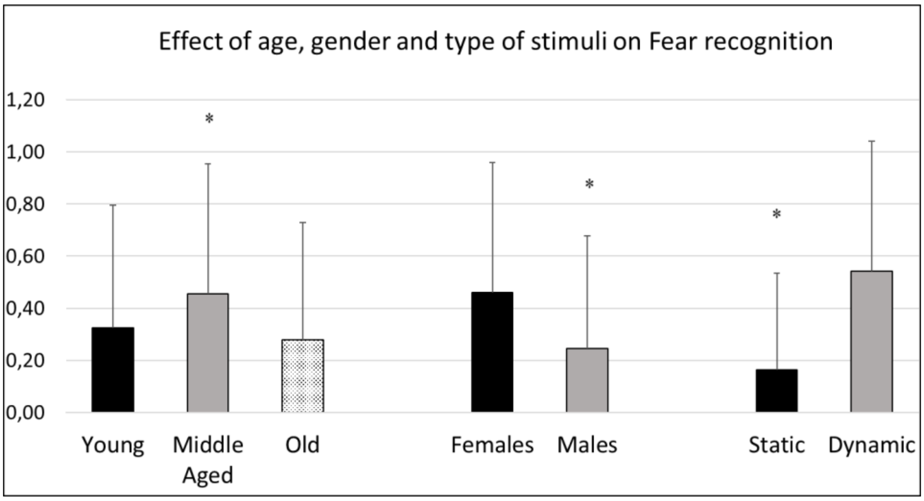

Figure 5: Agents' age, gender, and type (static vs dynamic) of stimuli effects on young participants' decoding accuracy of fearful faces.

Happiness

No significant effects of participants' gender (F (1,48) =.576, p= .452) emerged. Significant effects of the age of stimuli (F (2,96) =19.017, p<<.01) were observed. Bonferroni's post hoc tests revealed that participants were more accurate in decoding middle-aged (mean=.854) rather than young (mean=.725, p=.005) and old happy synthetic faces (mean=.625, p<<.01). Significant effects of the gender of stimuli were observed (F (1,48) =27.459, p<<.01). Bonferroni's post hoc tests revealed that participants were significantly more accurate in decoding virtual agents' expressing happiness conveyed by female (mean=.824, p<<.01) rather than by male stimuli (mean=.646). No significant effects of the type of stimuli were observed (F (1,48) =.416, p=.522), these results are summarized in Fig.6. A significant interaction between age and gender of stimuli (F (2,96) =10.840, p<<.01) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and gender of stimuli). These tests revealed that:

a) Concerning the age of the faces: middle-aged male (mean=.871) were more accurately decoded by participants rather than young male (mean=.592, p<<.01) and old male synthetic faces of happiness (mean =.475, p<<.01).

b) Concerning the gender of stimuli: participants were more accurate in decoding young female (mean=.858) rather than young male synthetic happy faces (mean=.592, p<<.01). Moreover, they were more accurate in decoding old female (mean=.775) than old male synthetic happy faces (mean=.475, p<<.01).

A significant interaction between age and type of stimuli (F (2,96) =20.868, p<<.01) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were more accurate in decoding static middle-aged (mean=.850) rather than static young synthetic happy faces (mean=.646, p=.001). In addition, they showed lower accuracy in decoding dynamic old (mean=.504) rather than dynamic young (mean=.804, p<<.01) and dynamic middle-aged synthetic happy faces (mean=.858, p<<.01).

b) Concerning the type of stimuli: participants were more accurate in decoding dynamic young (mean=.804) rather than static young synthetic happy faces (mean=.646, p=.008). In addition, they were more accurate in decoding dynamic old (mean=.746) rather than static old virtual agents' expressing happiness (mean=.504, p<<.01).

A significant interaction between gender of stimuli and type of stimuli (F (1,48) =5.266, p=.026) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and gender of stimuli). These tests revealed that:

a) Concerning the gender of stimuli: participants were more accurate in decoding static young female (mean=.803) rather than static young male synthetic happy faces (mean=.692, p=.024), moreover they were more accurate in decoding dynamic young female (mean=.844) rather than dynamic young male synthetic happy faces (mean=.600, p<<.01).

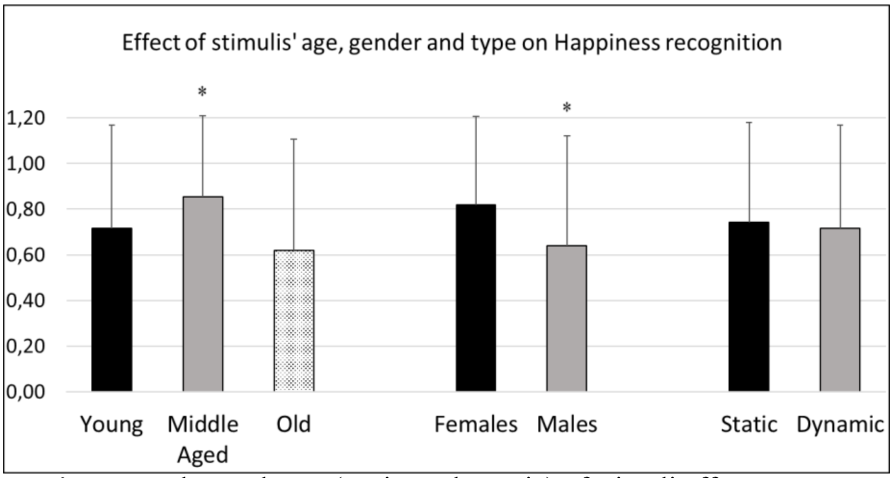

Figure 6: Agents' age, gender, and type (static vs dynamic) of stimuli effects on young participants' decoding accuracy of happy faces.

Surprise

No significant effects of participants' gender (F (1,48) =.079, p= .780) emerged. Significant effects of the age of stimuli (F (2,96) =41.532, p<<.01) were observed. Bonferroni's post hoc tests revealed that participants were less accurate in decoding old (mean=.285) rather than young (mean=.567, p<<.01) and middle-aged virtual agents' surprised faces (mean=.613, p<<.01). Significant effects of the gender of stimuli were observed (F (1,48) =21.726, p<<.01). Bonferroni's post hoc tests revealed that participants were significantly more accurate in decoding synthetic faces of surprise when expressed by female (mean=.563, p<<.01) rather than by male stimuli (mean=.414). Significant effects of the type of stimuli were found (F (1,48) =12.488, p=.001). Bonferroni's post hoc tests revealed that participants were significantly more accurate in decoding synthetic faces expressing surprise when conveyed by static (mean=.561, p=.001) rather than dynamic stimuli (mean=.417), these results are depicted in Fig.7. A significant interaction between age and gender of stimuli (F (2,96) =3.415, p=.037) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and gender of stimuli). These tests revealed that:

a) Concerning the age of the faces: old females (mean=.304) were less accurately decoded compared to young (mean=.696, p<<.01) and middle-aged female synthetic faces of surprise (mean =.688, p<<.01). Moreover, participants were also less accurate in decoding old male (mean=.267) rather than young male (mean=.438, p=.002) and middle-aged male synthetic expressions of surprise (mean =.538, p<<.01).

b) Concerning the gender of stimuli: participants were more accurate in decoding young female (mean=.696) rather than young male virtual agents' surprised faces (mean=.438, p<<.01). Moreover, they were more accurate in decoding middle-aged female (mean=.688) than middle-aged male virtual agents' surprised faces (mean=.538, p=.028).

A significant interaction between age and type of stimuli (F (2,96) =19.233, p<<.01) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were more accurate in decoding static young (mean=.692) rather than static middle-aged (mean=.533, p=.021) and static old virtual agents' surprised faces (mean=.458, p=.002). In addition, they showed greater accuracy in decoding dynamic middle-aged (mean=.692) rather than dynamic young (mean=.442, p<<.01) and dynamic old virtual agents' surprised faces (mean=.113, p<<.01). Moreover, participants were more accurate in decoding dynamic young (mean=.442) rather than dynamic old virtual agents' surprised faces (mean=.113, p<<.01).

b) Concerning the type of stimuli: participants were more accurate in decoding static young (mean=.692) rather than dynamic young virtual agents' surprised faces (mean=.442, p<<.01). In addition, they were more accurate in decoding dynamic middle-aged (mean=.692) rather than static middle-aged synthetic faces expressing surprise (mean=.533, p=.022). Moreover, participants were more accurate in decoding static old (mean=.458) rather than dynamic old virtual agents' surprised faces (mean=.113, p<<.01).

A significant interaction between gender and type of stimuli (F (1,48) =4.193, p=.046) was found. Since this was an interaction, Bonferroni's post hoc tests were performed for each single factor (age and gender of stimuli). These tests revealed that:

a) Concerning the gender of stimuli: participants were more accurate in decoding static female (mean=.600) rather than static male virtual agents' surprised faces (mean=.522, p=.046), moreover they were more accurate in decoding dynamic female (mean=.525) rather than dynamic male virtual agents' surprised faces (mean=.306, p=.048).

b) Concerning the type of stimuli: participants were more accurate in decoding static male faces (mean=.522) rather than dynamic male virtual agents' surprised faces (mean=.306, p<<.01).

A significant interaction among age, gender, and type of stimuli (F (2,96) =14.261, p<<.01) was found. Since this was a triple interaction, Bonferroni's post hoc tests were performed for each single factor (age, gender, and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were less accurate in decoding old female dynamic (mean=.142) rather than young female dynamic (mean=.733, p<<.01) and middle-aged female dynamic virtual agents' surprised faces (mean=.700, p<<.01). Moreover, they were more accurate in decoding young male static (mean=.725) rather than middle-aged male static (mean=.392, p=.002) and old male static virtual agents' surprised faces (mean=.450, p=.003). In addition, participants were more accurate in decoding middle-aged male dynamic (mean=.683) rather than young male dynamic (mean=.150, p<<.01) and old male dynamic virtual agents' surprised faces (mean=.083, p<<.01).

b) Concerning the gender of stimuli: participants were more accurate in decoding female dynamic young (mean=.733) rather than male dynamic young virtual agents' surprised faces (mean=.150, p<<.01). In addition, they were more accurate in decoding female middle-aged static (mean= .675) rather than male middle-aged static virtual agents' surprised faces (mean=.392, p=.006).

c) Concerning the type of stimuli: participants were more accurate in decoding young male static (mean=.725) rather than young male dynamic (mean=.150, p<<.01) synthetic faces. Moreover, participants were more accurate in decoding middle-aged male dynamic (mean=.683) rather than middle-aged male static virtual agents' surprised faces (mean=.392, p=.005). Conclusively, participants were more accurate in decoding old female static (mean=.467) rather than old female dynamic faces (mean=.142, p=.001) and they better recognized old male static (mean=.450) rather than old male dynamic synthetic faces expressing surprise (mean=.083, p<<.01).

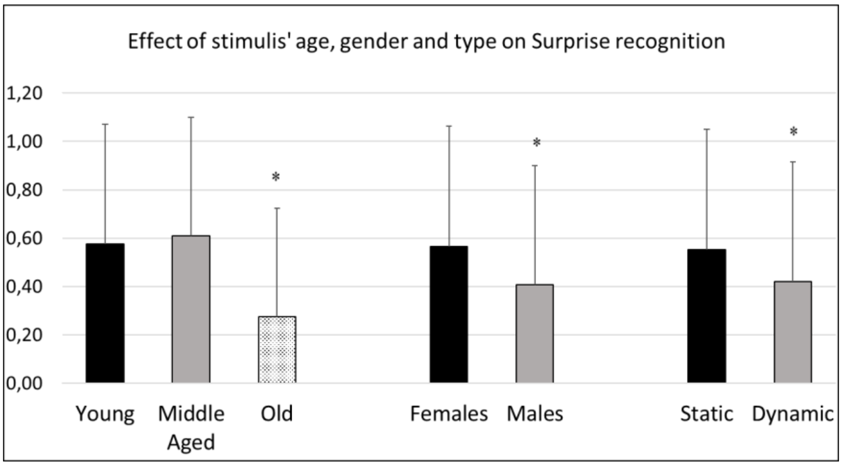

Figure 7: Agents' age, gender, and type (static vs dynamic) of stimuli effects on young participants' decoding accuracy of surprised faces.

Neutrality

Significant effects of participants' gender (F (1,48) =7.000, p= .011) emerged. Bonferroni's post hoc tests revealed that male participants (mean=.750) were more accurate than female participants (mean=.622, p=.011) in decoding virtual agents' faces expressing neutrality. Significant effects of the age of stimuli (F (2,96) =19.370, p<<.01) were observed. Bonferroni's post hoc tests revealed that participants were less accurate in decoding old (mean=.538) than young (mean=.727, p<<.01) and middle-aged synthetic faces of neutrality (mean=.794, p<<.01). No significant effects of gender of stimuli were observed (F (1,48) =1.744, p=.193). Significant effects of the type of stimuli were found (F (1,48) =6.144, p=.017). Bonferroni's post hoc tests revealed that participants were significantly more accurate in decoding synthetic emotional faces of neutrality when conveyed by static stimuli (mean=.742, p=.017) rather than when depicted by dynamic stimuli (mean=.631), these results are showed in Fig.8. A significant interaction among age, gender, and type of stimuli (F (2,96) =18.452, p<<.01) was found. Since this was a triple interaction, Bonferroni's post hoc tests were performed for each single factor (age, gender, and type of stimuli). These tests revealed that:

a) Concerning the age of the faces: participants were less accurate in decoding old female dynamic (mean=.300) rather than young female dynamic (mean=.808, p<<.01) and middle-aged female dynamic synthetic neutral faces (mean=.742, p<<.01). Moreover, they were less accurate in decoding old male static (mean=.533) rather than young male static (mean=.925, p<<.01) and middle- aged male static neutral virtual agents' faces (mean=.867, p=.003). In addition, participants were more accurate in decoding middle-aged male dynamic (mean=.817) rather than young male dynamic neutral synthetic faces (mean=.517, p=.001).

b) Concerning the gender of stimuli: participants were more accurate in decoding male static young (mean=.925) rather than female static young neutral synthetic faces (mean=.658, p=.001). In addition, they were more accurate in decoding female dynamic young (mean= .808) rather than male dynamic young neutral faces (mean=.517, p=.001) and they better recognized male dynamic old (mean=.600) rather than female dynamic old virtual agents' neutral faces (mean=.300, p=.003).

c) Concerning the type of stimuli: participants were more accurate in decoding static young male (mean=.925) rather than dynamic young male faces expressing neutrality (mean=.517, p<<.01) and they better recognized old female static (mean=.717) rather than old female dynamic virtual agents' neutral faces (mean=.300, p<<.01).

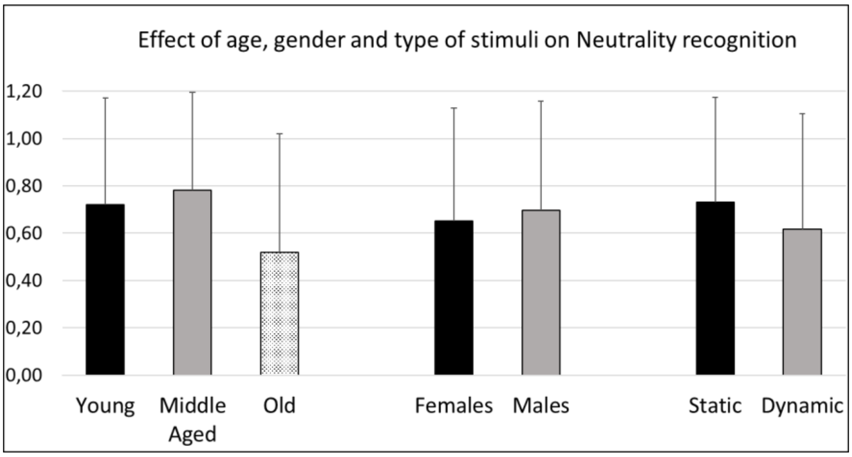

Figure 8: Agents' age, gender, and type (static vs dynamic) of stimuli effects on young participants' decoding accuracy of neutral faces.

1.2. Results- Young participants' decoding accuracy of synthetic emotional faces

The decoding scores obtained from young participants were then analyzed to assess their ability to correctly decode the proposed emotional categories conveyed by virtual agents, independently from the age, gender, and type of stimuli. In order to carry out these analyses, recognition scores obtained from all stimuli's age, gender and type were added up together for each emotional category. Repeated measures ANOVA were performed on these data, considering participants' gender as between subject factor and emotional categories as within subject factors. The significance level was set at $\alpha <.05$ and differences among means were assessed through Bonferroni's post hoc tests.

Results showed no significant effects of participants' gender ($F (1,48) =.324$, $p=.572$). Significant differences emerged in the recognition of the proposed emotional category ($F (6,288) =50.097$, $p<<.01$). Bonferroni's post hoc tests revealed that participants were less accurate when decoding synthetic facial expressions of disgust (mean=4.883) compared to: anger (mean=9.808, p<<.01), sadness (mean=6.633, p=.004), happiness (mean=8.817, p<<.01) and neutrality (mean=8.233, p<<.01). Anger was better decoded (mean=9.808), compared to: disgust, sadness (mean=6.633, p<<.01), fear (mean=4.100, p<<.01), surprise (mean=5.858, p<<.01) and neutrality (mean=8.233, p=.002). Synthetic facial expressions of sadness (mean=6.633) were better decoded compared to disgust and fear (mean=4.100, p<<.01), while were worse decoded compared to anger, happiness (mean=8.817, p<<.01) and neutrality (mean=8.233, p=.021). Fear (mean=4.100) was worse decoded compared to: anger, sadness, happiness (mean=8.817, p<<.01), surprise (mean=5.858, p=.001) and neutrality (mean=8.233, p<<.01). Happiness (mean= 8.817) was better decoded when compared to disgust, sadness, fear, and surprise (mean=5.858, p<<.01). Surprise (mean=5.858) was better decoded compared to fear and worse decoded compared to anger, happiness, and neutrality (mean=8.233, p<<.01). These results are showed in Fig.9.

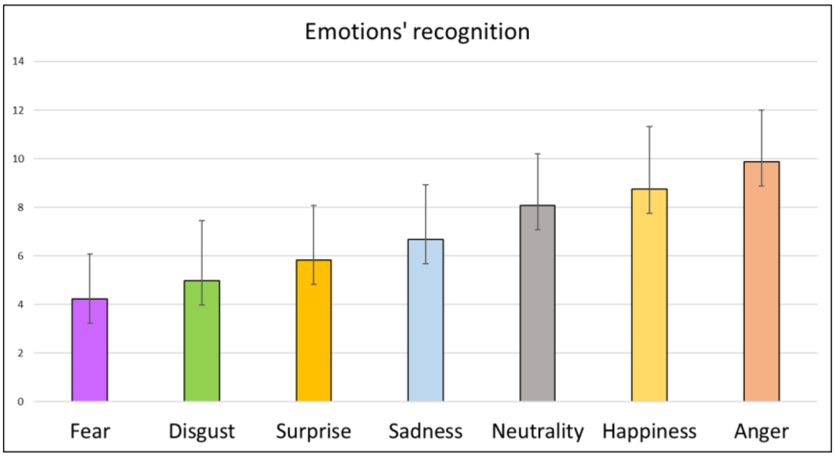

Figure 9: Young participants' ability to decode virtual agents' emotional expressions.

A significant interaction between emotional categories and participants' gender (F (6,288) =2.834, p=.011) emerged. Bonferroni's post hoc tests were performed for each single factor (participants' gender and emotional categories). These tests revealed that:

a) Concerning participants' gender: female participants (mean= 4.800) were more accurate in decoding synthetic expressions of fear compared to male participants (mean= 3.400, p=.007). On the other hand, male participants (mean= 9.000) were more accurate in decoding synthetic expressions of neutrality compared to female participants (mean= 7.467, p=.011).

b) Concerning the emotional categories: male participants were significantly less accurate in decoding synthetic expression of disgust (mean=4.400) compared to anger (mean=9.450, p<<.01), happiness (mean=9.100, p<<.01) and neutrality (mean=9.000, p<<.01). On the other hand, male participants were more accurate in decoding anger (mean=9.450) rather than disgust, sadness (mean=6.400, p=.001), fear (mean=3.400, p<<.01) and surprise (mean=5.950, p<<.01). Sadness (mean=6.400), was worse decoded compared to anger, happiness (mean=9.100, p=.001) and neutrality (mean=9.000, p=.013) while was better decoded compared to fear (mean=3.400, p<<.01). Concerning the synthetic expression of fear, male participants showed lower ability to decode it compared to anger and sadness. Moreover, they showed less accuracy in decoding fear (mean=3.400) compared to happiness (mean=9.100, p<<.01), surprise (mean=5.950, p=.003) and neutrality (mean=9.000, p<<.01) and greater accuracy in decoding happiness compared to disgust, sadness and fear. Moreover, they better recognized happiness (mean=9.100) when compared to surprise (mean=5.950, p=.002). Conclusively, as mentioned above, male participants significantly differed in decoding surprise compared to anger, fear, happiness and in addition, they were less accurate in decoding surprise (mean=5.950) compared to neutrality (mean=9.000, p<<.01). As regards as female participants, they were significantly less accurate in decoding synthetic expression of disgust (mean=5.367) compared to anger (mean=10.167, p<<.01), happiness (mean=8.533, p<<.01) and neutrality (mean=7.467, p=.017), while showed more accuracy in decoding anger (mean=10.167) rather than disgust, sadness (mean=6.867, p<<.01), fear (mean=4.800, p<<.01), happiness (mean=8.533, p=.036), surprise (mean=5.767, p<<.01) and neutrality (mean=7.467, p<<.01). Sadness (mean=6.867) was worse decoded compared to anger and happiness (mean=8.533, p=.019) and better decoded when compared to fear (mean=4.800, p=.002). Female participants showed a lower ability to decode fear (mean=4.800) compared to anger and sadness, happiness (mean=8.533, p<<.01) and neutrality (mean=7.467, p<<.01). Female participants showed less accuracy in decoding happiness (mean=8.533) compared to anger, while happiness was more accurately decoded compared to disgust, sadness and fear, surprise (mean=5.767, p=.001).

2. Conclusions

The present study is aimed at exploring young people's decoding ability of emotional expressions conveyed by virtual agents and specifically, the effects of age (young, middle-aged, and old virtual agents), gender (female and male virtual agents), and type of stimuli (static pictures and dynamic video clips). Different effects were observed depending on the emotional category examined. Regarding participants' gender, results showed that females and males differed in particular on the ability to recognize fear and neutrality. More in detail, female participants were more accurate than males in decoding synthetic faces depicting fear, while males better decoded synthetic faces of neutrality compared to female participants. Differences in the recognition of diversely aged virtual agents were observed, young participants were more accurate in decoding middle-aged than young and old virtual agents' faces of fear and happiness. Furthermore, results revealed that participants better recognized anger, surprise and neutrality when depicted by young and middle-aged virtual agents rather than old agents. Additionally, the emotional category of sadness was better decoded by participants when conveyed by old and middle-aged compared to young virtual agents; the same occurred concerning disgust recognition, even if not significant from a statistical point of view. Participants' accuracy in decoding happiness, fear, anger, neutrality and surprise brings out a difficulty in recognizing these emotional expressions when conveyed by faces belonging to an age group far from that of the participant, namely old faces, although facial expressions were conveyed by synthetic faces. Surprisingly, this is not true as regard as disgust and sadness, since in this case participants better decoded them when conveyed by old synthetic faces. A possible psychological interpretation of this result concern the fact that young people could be reluctant in attributing this emotions, which are socially considered as "negative", to their age group as to try distancing from negative moods. The study also highlighted that the gender of a face, even if a virtual face, affects the way people recognize facial emotional expressions; more specifically disgust, fear, happiness, and surprise were more accurately decoded when represented by female rather than male virtual agents. Conversely, participants better recognized the emotional category of anger when depicted by male compared to female agents; this could reflect a cultural heritage that tend to accept more easily a man expressing anger, rather than a women because of the stereotype attributed to their gender. As regards as the effect that the type of stimulus administered could exert on people's decoding accuracy, the proposed investigation underlined that participants were significantly more accurate in decoding synthetic emotional faces of disgust, sadness and fear, when conveyed by dynamic rather than static stimuli, while they showed a greater accuracy in decoding the synthetic emotional expressions of surprise and neutrality, when conveyed by static rather than dynamic stimuli.

Lastly, it was observed that young participants when required to recognize synthetic emotional expressions, independently from the age and the gender of the face and from the typology of stimulus administered, better decoded anger, happiness and neutrality while showed greater difficulty in the recognition of fear, disgust, and surprise. In conclusion, future research perspective should be focused on comparing these data with that obtained from experiments investigating young people's decoding accuracy of naturalistic emotional expressions, in other words emotions conveyed by real human beings, including also differently aged group of participants. Possible practical implications of the present study consist in providing useful information to assistive technology, such as virtual assistants, developers, since a crucial factor in the interaction between the user and a virtual agent is the way in which the user interprets and recognizes agents' facial expressions. Based on these studies and results, the developer could adapt the agent features to the user's expectations and needs, trying to render it more user-friendly. Moreover, given the possibility to exploit emotional virtual agents in on-line learning contexts, result of this investigation could be exploited with the aim of increasing students' involvement toward blended educational environments.

3. Acknowledgements

The research leading to these results has received funding from the European Union Horizon 2020 research and innovation programme under grant agreement N. 769872 (EMPATHIC) and N. 823907 (MENHIR), from the project SIROBOTICS that received funding from Ministero dell'Istruzione, dell'Università, e della Ricerca (MIUR), PNR 2015-2020, Decreto Direttoriale 1735 July 13 2017, and from the project ANDROIDS that received funding from Università della Campania "Luigi Vanvitelli" inside the programme V:ALERE 2019, funded with D.R. 906 del 4/10/2019, prot. n. 157264, October 17, 2019.

4. References

[1] S. Dhawan, Online learning: A panacea in the time of COVID-19 crisis. Journal of Educational Technology Systems, 49(1), 5-22, 2020.

[2] N. Dietrich, K. Kentheswaran, A. Ahmadi, J. Teychené, Y. Bessière, S. Alfenore, M. Sperandio, Attempts, Successes, and Failures of Distance Learning in the Time of COVID-19. Journal of Chemical Education, 97(9), 2448-2457, 2020.

[3] D. Song, M. Rice, E.Y. Oh, Participation in online courses and interaction with a virtual agent. International Review of Research in Open and Distributed Learning, 20(1), 2019.

[4] A. Serenko, Are interface agents scapegoats? Attributions of responsibility in human–agent interaction. Interacting with Computers, 19(2), 293–303, 2007.

[5] A. Esposito, T. Amorese, M. Cuciniello, M. T. Riviello, A. M. Esposito, A. Troncone, G. Cordasco, Elder user's attitude toward assistive virtual agents: the role of voice and gender. Journal of Ambient Intelligence and Humanized Computing, 1-8, 2019.

[6] S. Gobron, J. Ahn, D. Thalmann, M. Skowron, A. Kappas, Impact Study of Nonverbal Facial Cues on Spontaneous Chatting with Virtual Humans. Journal of Virtual Reality and Broadcasting, 10(6), 2013.

[7] Y. Hernández, J. Noguez, E. Sucar, G. Arroyo-Figueroa, Incorporating an affective model to an intelligent tutor for mobile robotics. In Proceedings. Frontiers in Education. 36th Annual Conference (pp. 22-27). IEEE, 2006.

[8] A. Ortiz, D. Oyarzun, M. del Puy Carretero, Elein: E-learning with 3d interactive emotional agents. In International Conference on Technologies for E-Learning and Digital Entertainment (pp. 294-305). Springer, Berlin, Heidelberg, 2009.

[9] C.N Moridis, A.A. Economides, Affective learning: Empathetic agents with emotional facial and tone of voice expressions. IEEE Transactions on Affective Computing, 3(3), 260-272, 2012.

[10] K. Chatzara, C. Karagiannidis, D. Stamatis, Students attitude and learning effectiveness of emotional agents. 2010 10th IEEE International Conference on Advanced Learning Technologies. IEEE, 2010.

[11] Y. Kim, A.L. Baylor, E. Shen, Pedagogical agents as learning companions: the impact of agent emotion and gender. Journal of Computer Assisted Learning, 23(3), 220-234, 2007.

[12] C. C. Joyal, L. Jacob, M.H. Cigna, J. P. Guay, P. Renaud, Virtual faces expressing emotions: an initial concomitant and construct validity study. Frontiers in Human Neuroscience, 8(787), 2014.

[13] A. Mollahosseini, H. Abdollahi, T.D. Sweeny, R. Cole, M.H. Mahoor, Role of embodiment and presence in human perception of robots' facial cues. International Journal of Human Computer Studies, 116, 25–39, 2018.

[14] M. Paetzel, G. Varni, I. Hupont, M. Chetouani, C. Peters, G. Castellano, The Attribution of Emotional State - How Embodiment Features and Social Traits Affect the Perception of an Artificial Agent. 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 495–502, IEEE, 2018.

[15] J. Gutiérrez-Maldonado, M. Rus-Calafell, J. González-Conde, Creation of a new set of dynamic virtual reality faces for the assessment and training of facial emotion recognition ability. Virtual Reality, 18(1), 61–71, 2014.

[16] M. Dyck, M.Winbeck, S. Leiberg,Y. Chen, R.C. Gur, K. Mathiak, K., Recognition profile of emotions in natural and virtual faces. PloS One, 3(11), 2008.

[17] E. Costantini, F. Pianesi, M. Prete, Recognising Emotions in Human and Synthetic Faces: The Role of the Upper and Lower Parts of the Face. In Proceedings of the 10th international conference on Intelligent user interfaces, 20-27, 2005

[18] R. Dhall, S.L. Goecke, T. Gedeon. Acted Facial Expressions in the Wild Database. In Technical Report, 2011.