



Socially Assistive Robot for Providing Recommendations: Comparing a Humanoid Robot with a Mobile Application

Silvia Rossi¹ · Mariacarla Staffa² · Anna Tamburro¹

Accepted: 21 January 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract

Socially assistive robotics has gained a valuable role in assisting, influencing and motivating the human behavior in many Human–Machine Interaction contexts. This finding suggested the use of social robots as recommendation interfaces. Despite many other types of recommending technologies exist (e.g., virtual agents, apps on cell phones, etc.), experimental studies convey that human beings result to be more engaged and influenced by the interaction with robots with respect to these other technologies. In particular, this has been shown by comparing social robots with virtual agents; a poor literature copes with the comparison between social robots and applications on mobile phones. To this extent, in this work, we address the comparison between these latest two technologies in the context of movie recommendation, where the two considered interfaces are programmed to provide the same contents, but through different communication channels. We provide the results of two experimental studies with the aim of evaluating the quality of the interaction from both the point of view of the application (by considering the users' acceptance rate of the recommendations) and of the users (by analyzing the users' evaluations) while interacting with the two interfaces. The main result arising from this study is that the social robot is preferred by users although, apparently, it does not change the acceptance rate of the proposed movies.

Keywords Socially assistive robotics · Interfaces for recommendations · Non-verbal cues · Mobile applications

1 Introduction

Social assistive robotics (SAR) [9] is receiving a great attention from the scientific, technological and industrial communities for its potential value in improving the quality of life of a large segment of the population [6]. SAR applications focus on assisting humans through social interaction rather than offering a physical support [9]. In this context, the robot has potentially a role in assisting people for increasing the well-being, suggesting something to do, as for example, to perform a particular activity [28] or to do exercises [34].

Social robots are effective in influencing and motivating human behavior [17]. Several projects have shown motivational capabilities of robots also in cases real people have difficulty in motivating users (e.g., interaction with autistic

people [8]). Robots present strong differences with respect to other types of assistive technologies, namely applications on cell phones or tablets [17,33]. When a person interacts with an embodied physical agent, he/she is typically more engaged and influenced by the interaction with respect to other technologies [17], and it has been shown that subjects are more likely to value their experience as more satisfying when confronting with human-like interfaces [33]. People perceive robots as intelligent agents and for this reason, they relate to them as they would with other people, but with less embarrassment [4]. This is extremely important in the case of health-care problems.

These robots' peculiarities become even more valuable in the extent that these can be used to affect the choices of users interacting with them. Humanoid robots can, thus, be used to provide recommendations which can make the interaction more meaningful with respect to simple uni-modal interfaces. However, it must be highlighted that it is not the complexity of the robot, but the richness of the interaction that influences a person. Previous studies on the evaluation of recommending interfaces focused on the comparison between social robots and virtual agents. Only a few works, instead, have been

✉ Silvia Rossi
silvia.rossi@unina.it

¹ Department of Physics, University of Naples Federico II, Naples, Italy

² Department of Electrical Engineering and Information Technology, University of Naples Federico II, Naples, Italy

presented that treat the applications on mobile phones (from now on apps) as a comparison term for social robots which provide recommendations. In our opinion, since nowadays tablets and smart-phones represent the most commonly used interfaces by humans, it makes a lot of sense to provide a comparative study with respect to these interfaces.

In this work, we present the results of two experiments conducted between 2015 and 2016 to observe if the users' acceptance rate of recommendations is increased, as well as if the users' engagement is elicited, when using a social robot rather than a mobile app. To analyze the role on providing recommendations, we tailored the provided suggestions on the users' needs and preferences. We considered, hence, the domain of movie recommendation that is a very popular domain allowing to rely on public available data sets to develop a real recommendation system. The two analyzed recommending interfaces have been designed to provide the same information to the users, but using different communication channels. Namely, the robot will use a multimodal behavior to interact with the participants by establishing a social binding expressed via head orientation, gaze direction, speech and gestures, while the app will provide the presentation of movies by means of a textual and graphical exposure. Robot non-verbal cues are manipulated with respect to a mapping between the suggested movie genres, the main associated emotion, the speed, and the extent parameters [16]. The main purpose of these experiments is to witness whether the presence of non-verbal cues, such as eye contact or genre-driven motion primitives, can have alone (thus without any graphical displays) a potential impact on the users' recommendation acceptance rate, by consequently attesting that a social robot constitutes a more valuable interface for providing recommendations. The results stated that although the social robot is preferred by users, it does not affect the acceptance rate of proposed movies so much to make it preferable with respect to the app. However, while several factors could influence the interaction, some stability in the acceptance rate (with respect to the English proficiency and the robotics skills) is found for the case of the social robot.

2 Related Works

Recent research started to investigate the use of social robots as recommendation interfaces, by focusing on the comparison between physical robots and their virtual counterpart [30]. The main findings of these comparative studies highlighted that the embodiment condition provided by the robot affects human perception of social engagement with respect to 2D/3D virtual agents [3,23,29]. The capabilities of a robot to persuade human users, with respect to a computer agent, in tasks such as following indications, has been addressed in [3], by showing an increased confidence and trust for the physical

robot, concluding that real robots affect subjects' decision-making more effectively than computer agents in real-world environments. User's behavior in accepting advises was investigated also in [23] showing that the robot was more effective in providing recommendations, so leading to a preference for the robot with respect to the computer agent. The effect of persuasion of a robot with respect to a 2D/3D computer agent was addressed in a laboratory environment in [29]. The authors showed that an important factor influencing the interaction was the possibility for the robot of showing a geometric coherence with the environment. This coherence was not achieved by both the 2D and 3D screen agents. However, even if they revealed the effects of the robot's presence on recommendation uses, it remains unknown whether such differences affect advertisements. Finally, the role of the robot appearance with respect to the acceptance of recommendation was investigated in [30]. Robots of different sizes were used in a shopping mall to provide advertisements. While a bigger robot could attract more the customer attention, results showed that smaller robots are more effective since the users found easier to interact with them.

With respect to the current literature, here, we investigate the effects of a robotic recommendation interface with respect to an app. Furthermore, we state that an efficient comparison between a robot and other types of human-machine interfaces should not only consider the differences in terms of embodiment condition, but should also consider the degree of interactive capabilities showed by the different interfaces, aiming to analyze in what extent the presence of communication abilities (also including the non-verbal ones) can influence the humans' choices and feelings. In fact, the ability of a human-machine interface to build a meaningful relationship depends on its capacity to help humans understand it, and this is made, in part, through non-verbal behaviors. It has been shown that emotional related non-verbal signals, as for example the modulation of the voice pitch, as well as the adaptation of gaze and head behavior [27] influence the human trust towards the robot [5] and the provided recommendation [19].

It has been well-documented that humanoid robots, due to their physical characteristics to exhibit social intelligent responses, are perceived as more engaging than a 2D virtual agent, and sometimes as engaging as a human. Hence, they could influence the humans' perception of the interaction and, as consequent, of the provided recommendation through the use of different verbal and non-verbal communication abilities. In Kidd and Breazeal's work [14], a particular communication channel, e.g., the agent's eyes, accompanied by verbal instructions, was used to interact with the users. The considered agents were a cartoon robot, a robot, and a human. The goal of the experimental study was to evaluate the users' engagement, reliability, usefulness, and trust during the interaction by using a questionnaire. Results showed

that the robot was considered more engaging, credible and informative, as well as being more pleasant as an interaction partner. As in [14], here, the considered interfaces provided the same information contents, but on the contrary, by using different communication channels. Related to the persuasive effect of a robot, in [10,13], the authors considered a robot that uses gaze and gestures and a robot using only the gaze or the gestures. Results showed that the gaze played a fundamental role in persuasiveness, while gesture may also provide a contribution, but only in addition to gaze.

3 A Movie Recommendation Case Study

The main functionality of a Recommender System (RS) is to guess and propose items which are likely to be of interest for the users, basing on their preferences or profiles. Our starting hypothesis is that the way a recommendation is provided is as much important as the quality of the recommending algorithm. It has been shown, in fact, that users are more inclined to accept recommendations when feeling a positive experience [15,20] accompanied with a sense of trust [21]. Furthermore, social responses are more prevalent if the system is personified since the presence of a humanoid virtual agent induces trust [36,38]. According to these latest findings, we developed a client/server application, where the server provides the recommendation service (i.e., the recommending algorithm is the same) and the possible clients can be either a humanoid robot or a mobile application (i.e., the provision of the recommendations occurs through different communication channels). This diversity should be reflected in a different perception of the recommendations by the users and it will presumably affect their experience. In particular, the possible clients are of three types: two humanoid robots with a different degree of social interactive capabilities (i.e., movie vocal description, with or without accompanying gestures, eyes color changes, and pitch manipulations) and a mobile application (see Fig. 3) showing textual and graphical description of the recommended movies (such as names of actors, director, movie genre, plot, and movie poster).

In Fig. 1, the developed architecture is shown. Summarizing, the core of the server layer is characterized by the Recommendation Engine module that generates rating predictions through an item-based collaborative filtering approach [25]. It needs an initial set of at least 20 movie ratings from the users provided by using the mobile app. Additional movie specific information, such as director, actors, and genres are retrieved through the OMDb¹ web service. For a detailed description please refer to [7].

¹ <http://www.omdbapi.com>—The Open Movie Database is a free web service to obtain movie information.

3.1 A Social Robot for Movie Recommendations

The *Robot* used for our experimentation is the NAO T14 robot model (see Fig. 1). It consists of a humanoid torso with 14 degrees of freedom (2 for the head and 12 for the arms), endowed with different sensors and actuators: such as colored eyes, tactile sensors, camera, microphones, sonars, speakers, which we controlled via the Robotic Operating System (ROS).

Robot Sensory Modules Two modules have been developed to allow NAO interaction with the user: (1) a *Face Detection* module based on a face recognition solution provided by OKI and included in the Python SDK for NAO, which is used to process camera frames to detect the users' presence and track them by accordingly directing Robot gaze [32] and (2) a *Speech Recognition* module which relies on sophisticated speech recognition technologies provided by NUANCE for NAO Version 4 and used to process sounds obtained from the microphone for user authentication.

Robot Actuator Modules The *Behavior Selector* module is the main module of the robotic control system. It is in charge to select the most suitable behavior including gestures, gaze, and the voice feedback to users. The vocal feedback relies on the Speech Synthesis module which provides movie description with different speech intonations, depending on the movie plot. It can be accompanied by arms gestures and facial expressions (e.g., different eyes colors) generated through Motion Controller module. The Behavior Selector gets recommendations from the Web Service and maps the movie genre into a predefined set of animations (gathered from the Animations Repository) and eyes colors (see Sect. 3.2). As example, the robot describes a comedy with expansive gestures accompanied with yellow colored eyes. The pitch of the voice is accordingly manipulated by the Speech Synthesis module.

3.2 Mapping Movie Genres into Robot Non-verbal Cues

The recommendation system domain is designed to use the embodiment condition and voice-based communication to enhance the users' trusting beliefs, perceptions of enjoyment, and ultimately, their intentions to use the agent as a decision aid [24]. Nao acts like a friend in these tests, so gestures have been specially chosen to look "friendly"; there are no sudden shots in the movements [11], Nao looks in the direction of the user, and the voice tends to make the idea of what it recommends. It does not try to scare the user, even while it is recommending a horror movie. Robot can express emotions for a better human interaction, but it does not have the abilities to express emotions through modalities such as facial

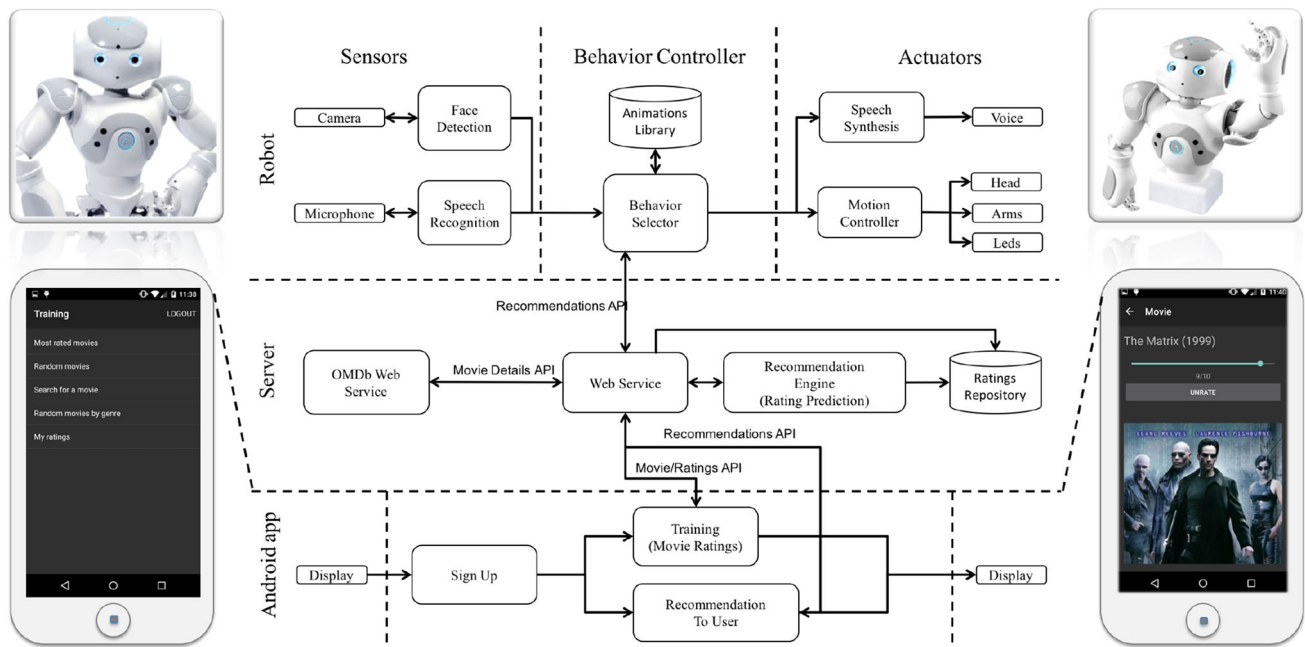


Fig. 1 The general architecture of the movies recommendation system

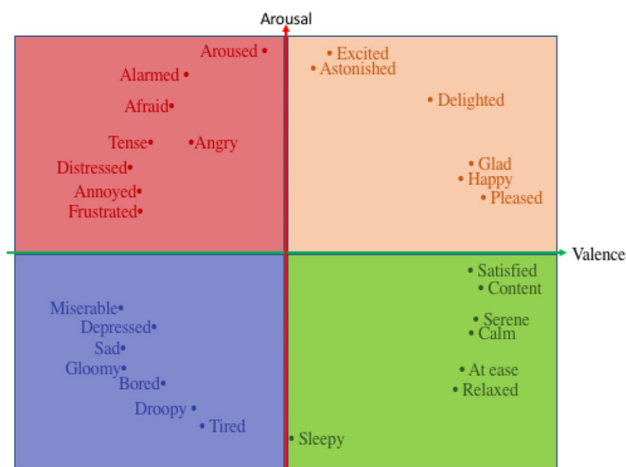


Fig. 2 Circumplex model of affect

expression. Furthermore, it is very complex to identify an emotion through a single expressive channel, so the use of multiple cues is better understandable than using a single modality alone. Hence, in this work, we use three modalities for expressing emotions: color, sound, and motion [31]. The circumplex model of affect is used to map emotions onto a valence-arousal space [22] (see Fig. 2). We decide on a set of basic rules that represent the mappings between each single modality and the emotions. Mixed-modality expressions were built upon these basic rules.

Emotion can be modeled through four dynamic parameters: *Speed*, *Intensity*, *Regularity* and *Extend* [16]. This parameters set is called SIRE. We manipulate *Speed* and

Extend for *Sound* and *Motion* channels, and *Intensity* and *Regularity* for *Color* channel.

Regarding color, it is possible to combine the three primary colors (red, green and blue) to obtain different colors and to control the LEDs intensity. There are several mappings between emotional states and colors. We selected the mapping that is most commonly accepted, so, for example, white means peaceful, blue means depressed, and red means angry.

The robot voice plays a very important role to express emotion. In this work, the robot voice is used to express emotion by manipulating two basic modalities: speed and pitch. For example, with an increasing intonation, people perceive the robot attitude as showing anger, while, with a decreasing intonation, people perceive the robot attitude as showing sadness [31]. Anger related emotions can be interpreted as an emotion consisting of negative affection and a high level of arousal, while sadness consists of a negative emotion with a low level of arousal (see Fig. 2).

Expressing emotions through gestures provides an important feature in communicating your state to others. Through spatial expansiveness and contraction index, it is possible to identify a particular emotion in a robot gesture. For example, an outstretched arm increases the hand traveling distance and the arm rigidity indicating a positive mood; an unextended arm shows unconcern or reluctance indicating a negative mood. Regarding motion-speed, fast motion speed expresses positive moods (e.g., happiness and excitement), while slow motion speed expresses negative moods (e.g., sadness) [37].

Table 1 Interactive channels values of the considered genre clusters

Cluster Genre classification	Color		Sound		Gesture	
	RGB (intensity)	Fade (regularity)	Speed	Pitch (extend)	Speed	Spatial expansiveness
Group 1	Yellow	Low	High	High	High	High
Group 2	Orange	Fast	Fast	Mid	Fast	Mid
Group 3	Blue	Fast	Low	Low	Low	Low
Group 4	Cyan	Mid	Mid	Mid	Mid	Mid
Group 5	Green	Mid	Mid	Mid	Mid	Mid
Group 6	Red	Low	Low	Low	Low	Low

To cluster movie genres into the different emotional states through the considered three channels (*Color*, *Sound* and *Motion*), we needed to group movie genres into macro categories. The relationship between genres is calculated from the eight emotions of the Plutchik's values (Joy, Trust, Fear, Surprise, Sadness, Disgust, Anger, and Anticipation) for each genre and is very close to the common perception about movie genres [2].

The genres clusters are classified as follow:

- *Group 1* = Animation, Family, Comedy, Romance
- *Group 2* = Fantasy, Sci, Adventure, Action, Drama, Crime, Thriller, Mystery
- *Group 3* = Biography, Documentary, History, War
- *Group 4* = Music
- *Group 5* = Sport
- *Group 6* = Horror

The chosen mapping among movie genre groups, emotions, and the three modalities (gestures, sound, and color) is shown in Table 1.

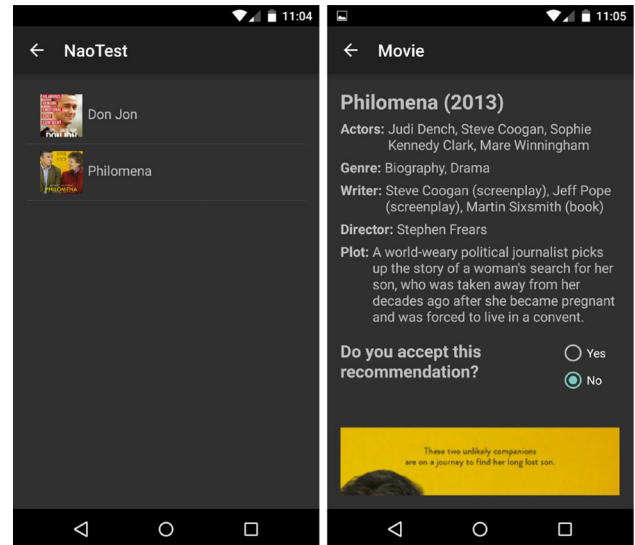
4 Experimental Setup

In the following sections, we discuss the results of two different tests conducted between 2015 and 2016. These case studies aimed at evaluating the importance of interactive non-verbal capabilities (or behaviors) in affecting recommendations acceptance and the subjective evaluation of the interaction, by comparing the use of three interfaces: a mobile app, and two social robots endowed with different non-verbal capabilities.

Here, we present the details of the experimental setup that are common to both case studies.

4.1 Recommendation Interfaces

We considered three possible interactive conditions:

**Fig. 3** Snapshots from the APP recommendation interface

- **APP**: in this setting, the user interacts with the mobile application. The application provides two different movies suggestions to the user. For each movie, the application shows on the screen the title, the movie poster, and additional information regarding the actors, the genre, the director, and the movie plot. For each movie, the user has to express his/her intention on accepting or not the proposed recommendation (see Fig. 3).
- **NAO**: in this setting, the user interacts with the NAO robot. The robot is located on a table in front of the user (see Fig. 4 left). The interaction starts as soon as NAO recognizes a face in his field of view. NAO greets the person, introduces itself, and presents two movie recommendations. The recommendations are presented only by using the robot voice. Hence, the same information provided by the APP (title, plot, genre, actors, and so on) will be presented only through speech. Finally, the robot asks the user if she/he intends to see the recommended movie.
- **ENAO**: in this setting, the user interacts with the ENAO robot. Differently from the NAO setting, in this case the robot is endowed with the motion controller module that

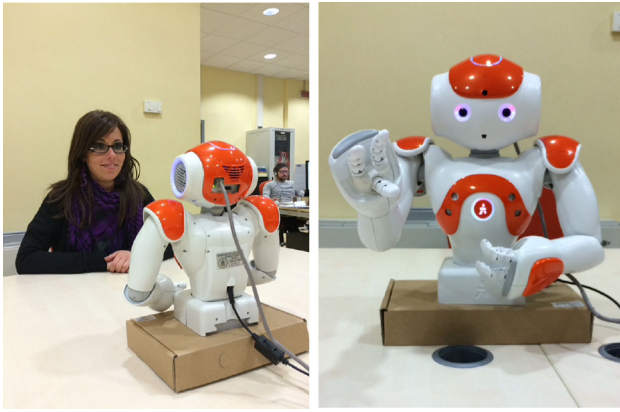


Fig. 4 Snapshots from the NAO (left) and ENAO (right) recommendation interfaces

correlates the recommended movie genres with respect to the leds' color, the speed, the pitch of the voice, and the gestures (see Fig. 4 right); for this reason, we called it ENAO, that states for Emotional NAO.

4.2 Experimental Hypotheses

Our starting hypotheses are here summarized:

- **H1:** a humanoid robot interacting through natural modalities will be considered by the user more engaging and will be better liked with respect to a commonly used application on a mobile phone. This qualitative value has been evaluated by providing a questionnaire to the user;
- **H2:** As a consequence of H1, the animated robot obtains a persuasive effect on the user choices. Hence, the recommendations provided by the robot should be more likely to be accepted. This analysis considered the recommendations acceptance rate.

With respect to these starting hypothesis, we considered an additional one related to the embodiment condition:

- **H3:** the embodiment condition only (NAO) will not be preferred to the APP condition.

Finally, we considered another additional one related to the subject personality:

- **H4:** The personality of the subjects will have an effect on the acceptance rate and on the evaluation of the interaction.

4.3 Testing Procedure

Before the beginning of the interaction, the user is required to provide 10 movie ratings, by using the mobile application,

in order to train the recommender system with respect to the user preferences. Moreover, the users has to provide personal information that depends on the considered case study. In the first one: gender, age, instruction level, familiarity with robotic skills, and a self-evaluation of English language proficiency; while in the second: gender, age, instruction level, familiarity with Android applications, familiarity with the movie domain, and a personality questionnaire. After the training phase, the recommendation engine provide six movies to be recommended according to the user profile. Such recommendation will be randomly assigned to one of the three interaction modalities APP, NAO, and ENAO (two for each). For each experimental setting, the user has to complete a satisfaction and usability questionnaire. Finally, at the end of the test, each participant is requested to express a single preference for one of the presented interfaces.

Qualitative Questionnaire To collect the participants' explicit impressions on the interaction about the three experimental conditions, we proposed a qualitative questionnaire composed of the following six questions:

- **Q1. (Ease)** How easy was to perform the task?
- **Q2. (Expectation)** Did the system react accordingly to your expectations?
- **Q3. (Naturalness)** How natural is this kind of interaction?
- **Q4. (Satisfaction)** How satisfying do you find the interactive system?
- **Q5. (Naturalness of Motion)** Did agents motions are natural (5) or unnatural (1)?
- **Q6. (Consistency)** Are you sure (5) or unsure (1) about your answers?

We adopted a classical likert scale from 1 to 5.

For both the case studies, we evaluated the number of recommendations that were accepted by the users and explicitly asked which was the interface the users' preferred to interact with. Finally, we made some considerations starting from the statistical analysis of between groups data grouped per features. Furthermore, according to other authors [34], we also analyzed the users' personality traits, in order to see if these human characteristics can affect the previous considered values and opinions.

5 Experimental Results

5.1 The First Case Study

In the first experiment, the three conditions (APP, NAO, and ENAO) are compared with respect to the recommendations

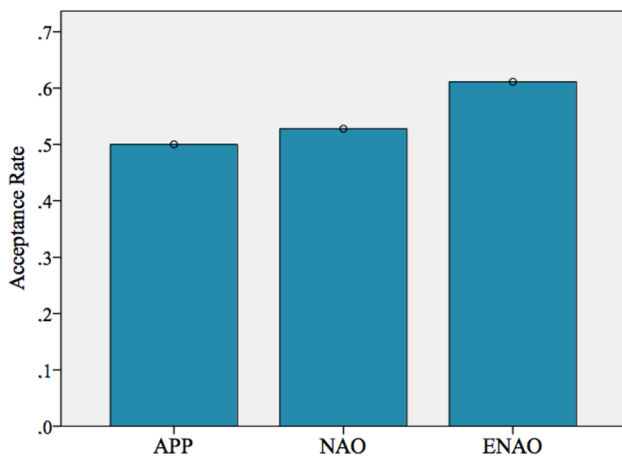


Fig. 5 Acceptance rates of movies recommended by APP, NAO, and ENAO

acceptance rate and the subjective evaluation of the interaction. We designed this study as a within-subjects repeated measures experiment. The considered independent variable is the interaction condition, respectively, a humanoid robot (NAO), a humanoid robot with increased non-verbal communication capabilities (ENAO), and a mobile application (APP). Every single participant is subjected to every interaction condition and the order with which the participants interact with the APP, NAO and ENAO is random and balanced between the three interfaces.

Participants In the first experiment 18 Italian native speaker were recruited (67% Males and 33% Females) with an average age of 31.8. The self-assessed English proficiency was medium/high for 39% and medium/low level for the remaining 61%. The 56% of the participants declared an high confidence with robotic applications. The language adopted for the experiment was the English language both for the movies descriptions and for the robot's voice synthesizer.

5.1.1 Results Analysis

Acceptance Rate Due to the limited number of participants and recommendations provided to each participant the differences in acceptance rates, evaluated using ANOVA with repeated measures, are not statistically significant ($F(2, 70) = 0.513$, $p = 0.601$). Results of the acceptance rate are shown in Fig. 5. With respect to the acceptance rate of the APP condition, the NAO condition acceptance rate shows similar results. A slightly greater acceptance rate is shown in the case of the ENAO condition. These results are in accordance with our hypothesis H2 that people are inclined to accept more recommendations provided through a more natural interaction and with the hypothesis H3, so that the sole

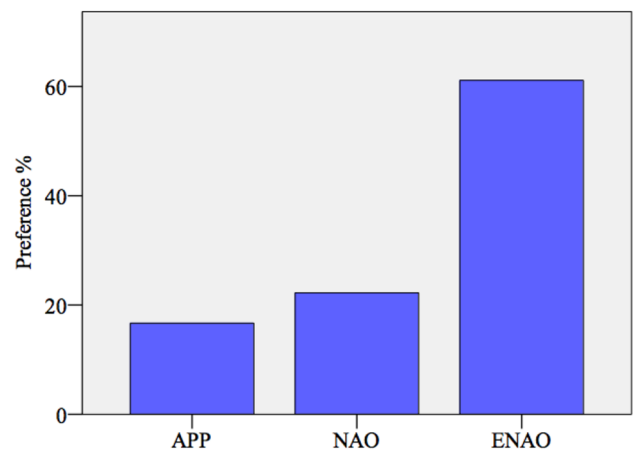


Fig. 6 Preference rates of APP, NAO, and ENAO

embodiment condition does not imply significant changes in the acceptance rate with respect to the APP condition.

Preferences On the contrary, subjective evaluations of the interaction (see Fig. 6) show a clear and statistically significant preference for the ENAO condition with respect to the other two (from 17% to 60%, with a one sample t-test significance $p < 0.05$). These results are in support of our hypothesis H1 that people are inclined to prefer an interface with natural interaction modalities.

English Proficiency Results of one-way ANOVA statistical test show that English proficiency level does have a significant main effect on the acceptance with $F(1, 33) = 7.748$ and $p = 0.009$. More precisely, as shown in Fig. 7, where the rate of accepted recommendations are plotted in the case of low/high English proficiency for APP, NAO, and ENAO, there is a significantly higher movie acceptance rate in general in the case of a low-English level (62% is the average acceptance rate computed among the three conditions), while the high-English proficiency level differently impacts on the acceptance rate associated with the three conditions. Namely, the level of English proficiency affects the acceptance rate with respect to the considered interface with a statistical significance contrast (repeated measures ANOVA, within-subject effect $F(2, 68) = 2.598$ and $p = 0.082$, and within-subject linear contrast $F(1, 34) = 5.926$ and $p = 0.02$, that corresponds to considering only the APP and ENAO interfaces).

There is a moderate negative correlation, for the APP condition, between the English proficiency and the acceptance rate (Pearson correlation with $\rho = -0.46$ and $p = 0.005$). This is to say that users with a high level of English proficiency accept fewer recommendations (21%) with respect to users with a low level (68%). This is the same for the NAO condition, but with a smaller difference. In the case

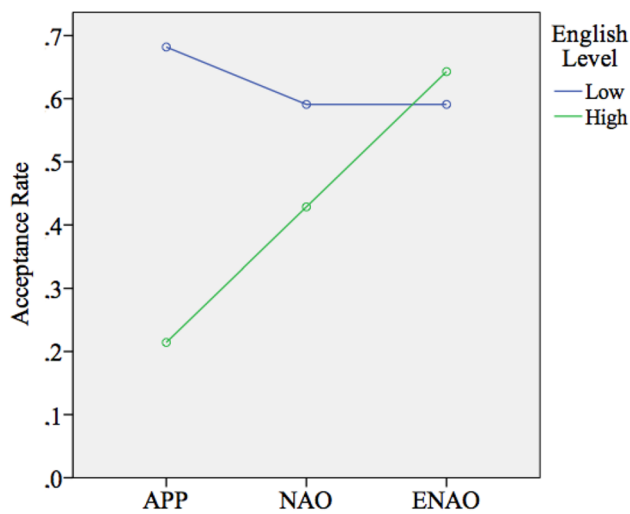


Fig. 7 Acceptance rate evaluated by grouping participants per low/high English proficiency

of ENAO, such difference is even smaller (59% for low and 64% for high) and it is not statistically significant with $F = 0.09$ and $p = 0.76$. Since ENAO provides a similar and good acceptance rate in both cases, the use of such interface could be valuable whenever the audience, in terms of different level of English proficiency, is wide; in the sense that it can address a wider range of people characterized by a different level of English proficiency. Furthermore, users with a good understanding of English are more inclined to accept recommendations provided by ENAO than by APP. Finally, since the high-English level cases obviously corresponds to the subjects that better understood the recommended content, the trend of the high-level acceptance rate clearly shows an improvement in the acceptance rate for the case of ENAO, with respect to NAO and APP, and so in accordance with our experimental hypothesis H2.

Robotics Skills Robotics skills do not to have a significant impact on the acceptance rate (one-way ANOVA with $F(4, 31) = 1.398$ and $p = 0.258$). This result endorses the previous one, stating the robustness and stability of this interface under certain potential influencing factors (such as English proficiency and robotics skills levels).

Gender and Age Concerning the gender aspect, the statistical analysis shows that the users' gender affects the acceptance rate only for the APP interface (one-way ANOVA with $F = 4.86$ and $p = 0.034$). Finally, no statistically significant differences were found grouping by age.

Questionnaire Evaluation We also analyzed the interaction from the users' point of view (see Table 2, where bold values represent the higher average evaluations with statistically sig-

nificant differences). In detail, concerning the ease of use of the interface (Q1), users perceived easier to use the APP with a statistically significant difference ($p = 0.048$). Robotics skills showed a moderate correlation with respect to the evaluation of the easiness of the interaction (Pearson $\rho = 0.512$ with $p = 0.03$ in the case of ENAO). This means that the users evaluated as more easy the use of the APP interface (that is commonly used by all the participant), while with respect to the robot such unfamiliarity is reflected in the evaluation (the ones that are more familiar with the robot evaluated with higher score the ease). Giving socially intelligent responses, the humanoid robot did not disappoint the expectations of participants, who judged better the interaction with ENAO with respect to the expectation (Q2), but with no significant differences (ANOVA with repeated measures $F(2, 34) = 1.109$ and $p = 0.341$). The users found more natural to interact (Q3) with the APP, since they are accustomed using it, with no significant difference (ANOVA with repeated measures $F(2, 34) = 2.902$ and $p = 0.069$). We only found a moderate Pearson correlation between robotics skills and Q3 evaluations in the case of NAO ($\rho = 0.551$ with $p = 0.018$), meaning that people who are more familiar with robotic applications found the interaction with NAO more natural. Finally, the interaction with each interface has been evaluated as satisfactory (Q4), with a slightly increased preference in the case of APP (ANOVA with repeated measures $F(2, 34) = 2.944$ and $p = 0.066$). The evaluation of the satisfaction has a moderate correlation with the acceptance rate in the case of NAO (Pearson $\rho = 0.517$ with $p < 0.05$). Q5 was evaluated only with respect to the ENAO configuration showing good appreciation of the naturalness of the proposed gestures and motions, while Q6 was used as a consistency check. Thanks to the questionnaire, we can state that the users were, on the whole, satisfied about the interaction.

5.2 The Second Case Study

In order to validate the hypothesis that the robotic application is a valuable solution for recommendation systems with respect to a mobile application, we must avoid experimental constraints that could possibly affect the interaction performance. Hence, in the second case study, we wanted to avoid that a low English proficiency could have an impact on the movie acceptance rate. Using the mother tongue ensures a complete comprehension, while, by introducing the diversity of the language factor, we are able to better understand the effects of non-verbal behavior on users' understanding, acceptability, and engaging. For this purpose, we collected the same data in a second experiment using the Italian language.

Moreover, in the second experiment, we also evaluated the personality traits of the participants in order to understand

Table 2 HRI qualitative questionnaire ratings for the first case study

	Q1	Q2	Q3	Q4	Q5	Q6
	Ease	Expectation	Naturalness	Satisfaction	Naturalness of Motion	Consistency
APP	4.28 ± 0.75	3.39 ± 1.24	4.11 ± 0.83	3.89 ± 0.96	–	4.67 ± 0.48
NAO	3.67 ± 0.69	3.33 ± 0.91	3.56 ± 0.78	3.28 ± 0.89	–	4.67 ± 0.59
ENAO	3.83 ± 0.71	3.72 ± 1.18	3.67 ± 0.84	3.67 ± 0.91	3.56 ± 0.86	4.72 ± 0.46

whether personality could have an impact on the evaluation of the interaction. Research has shown that personality is a primary factor which influences human behaviors. Research in psychology has come out with different models to describe the human beings' personality. Among them, the most popular approach among psychologists for studying personality traits is the Five-Factor Model (FFM), which describes human personality using five factors (OCEAN), also known as Big-Five [18]. *Openness* represents the inclination to openness to new experiences, having an active imagination and a preference about the will to find new ideas. Closed people are less flexible and rarely understand others' point of views. *Conscientiousness* describes how much an individual is responsible, disciplined and dutiful. *Extraversion* is an indicator of assertiveness and trust. Extroverted people easily create interpersonal relationships and love working and being together with others. *Agreeableness* describes the level of sympathy, availability, and cooperativeness. People with low level of this factor are competitive, skeptics and antagonistic. It measures how much a person is nice and altruistic. Finally, *Neuroticism* represents an emotional instability characterized by negative emotions like fear, anger, sadness and low self-esteem. People with an high Neuroticism trait rarely are able to control their impulses and cope with stress. For the personality evaluation, among the many models of personality, the Five Factor Model appears suitable for usage both in recommender systems and human–robot interaction as it can be quantitatively measured (i.e. numerical values for each of the considered factors, namely, openness, conscientiousness, extraversion, agreeableness, and neuroticism) [35].

Since the second experiment required an additional time for the completion of a personality questionnaire, we limited the evaluation with respect only to the APP and ENAO conditions. As the first case study, this one is designed as a within-subjects repeated measures experiment, where the independent variable, in each experiment, is the interface used for providing the recommendation, respectively, the robot with non-verbal communication abilities (ENAO) and the mobile application (APP).

Personality Questionnaire For our experimental study, we used the Italian BFI questionnaire reported in [12], consisting of 44 sentences, where users were asked to define a certain

number of characteristics that may or may not be applied to themselves (see Fig. 1), by associating a rate from 1 (Disagree Strongly) to 5 (Agree Strongly) to each question. The scoring analysis of the results will provide the score calculated through specific formulas that relate some of these answers, and it will generate the membership with a certain percentage to one of the 5 possible traits (Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness).

Participants 30 participants (63% males and 37% females) were involved in the test, with an average age of 28.5. The 40% of the participants had high robotics skills, while the 60% had very low proficiency with robotic devices (with an average value of 2.8). The average familiarity value for the movie domain is 4.3, while the average familiarity with the android applications domain is 3.8. We followed the same procedure as the first case study, but adopting the Italian language both for text description and for the robot's voice synthesizer.

5.2.1 Results Analysis

Acceptance Rate As in the previous test, results show that there is a minimum difference in the acceptance rate between the recommendations provided by APP and ENAO not yet statistically significant (ANOVA with repeated measure $F(1, 59) = 1.616$, $p = 0.209$). This result could come from the used recommendation algorithm that is able to propose movies, which, presumably, fit very well the users' preferences, independently from the used interface (we had more of 60% of acceptances). This is also supported by the fact that there is a moderate significant Pearson correlation between the acceptance of the movie suggested by the APP and the ones suggested by ENAO ($\rho = 0.642$ with $p < 0.001$). Results are shown in Fig. 8.

Personality and Acceptance Rate Concerning the possible main effect of the personality on the acceptance rate, in Fig. 9 the variation of such rate for APP and ENAO cases are shown with respect to High or Low values of each personality trait. Also, in this case, such differences are not statistically significant. However, in the case of extraversion subjects with high extraversion values accepted more recommenda-

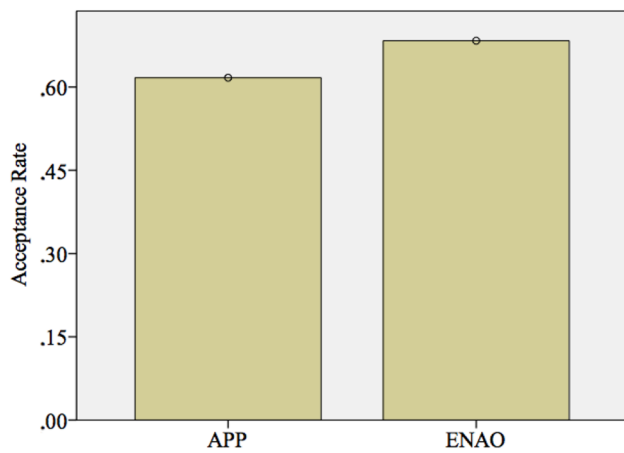


Fig. 8 Acceptance rates of movies recommended by APP and ENAO

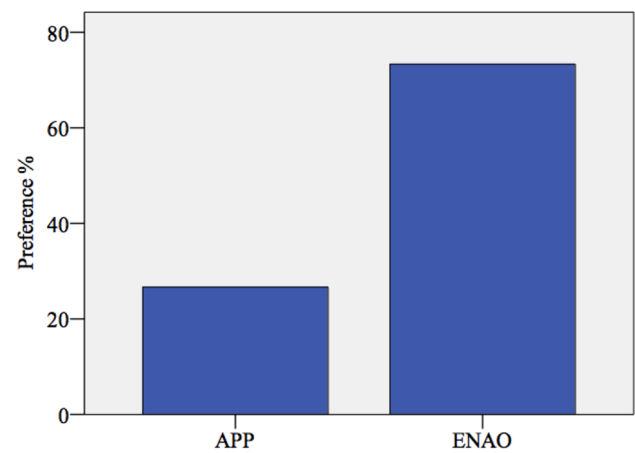


Fig. 10 Preference rates for APP and ENAO

tions with respect to subjects with lower values in the case of ENAO condition, but there is not a significant interaction effect on the acceptance rate with respect to the two conditions ($F(1, 42) = 0.057$ with $p = 0.812$). No significant result are found also for the case of Agreeableness ($F(1, 42) = 0.286$ with $p = 0.596$) and Conscientiousness ($F(1, 42) = 0.093$ with $p = 0.762$). In the case of neuroticism, as we expected, low neuroticism values correspond to a higher acceptance rate ($F(1, 42) = 0.462$ with $p = 0.500$).

Finally, the openness does not seem to have an impact on the acceptance rate in the case of low values, by while it produces a higher acceptance in the case of ENAO ($F(1, 42) = 1.964$ with $p = 0.168$) for high values.

Preferences In Fig. 10, the percentage of preferences expressed by the users at the end of the interaction with the two interfaces (APP and ENAO) is presented. The histogram clearly denotes an enhanced experience of users when inter-

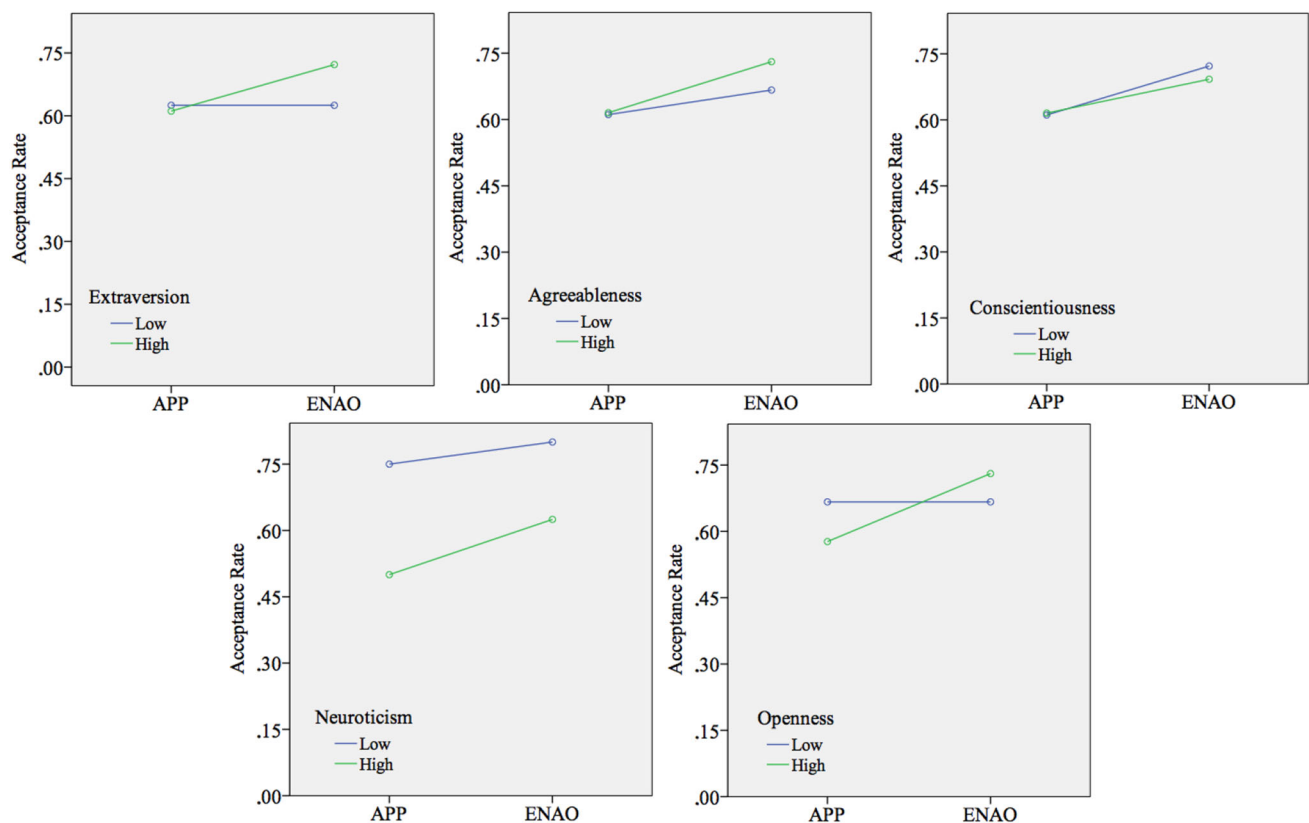


Fig. 9 Personality traits grouped for high/low values and acceptance rate

acting with the humanoid robot (i.e., 73% of participants preferred to interact with ENAO). This result is in accordance with our hypothesis H1 that people are inclined to prefer an interface interacting through more natural modalities. This consideration is also supported by statistical analysis computed by through a one sample t-test that stated a significance of $p < 0.008$.

Personality and Preferences With respect to H4, there is a moderate Pearson correlation of the neurotic personality trait and the expressed preference on the interface ($\rho = 0.545$ with $p = 0.009$). Namely, subjects with a higher neurotic trait prefer more ENAO with respect to subjects with a low value that preferred the APP.

Movie Familiarity Regarding the familiarity with the movie domain, between groups indicates that the variable familiarity is significant ($F(1, 26) = 53.33$ with $p < 0.001$). Movie familiarity has a weak Pearson correlation with the acceptance rate ($\rho = 0.38$ with $p = 0.038$). However, the within-subjects test indicates that there is not a significant interaction effect on the acceptance rate with respect to the two conditions (ANOVA with repeated measure $F(3, 26) = 0.199$ with $p = 0.896$). This means that the movie familiarity has an impact on the acceptance rate in general, but there are no statistically significant differences in the acceptance rates with respect to the two considered conditions.

Robotics Skills As in the previous test, Robotic skills do not have a significant main effect on the acceptance rate ($F(4, 25) = 1.183$ with $p = 0.343$). The same holds for Android skills ($F(4, 25) = 1.784$ with $p = 0.164$). Concerning the gender aspect, the statistical analysis shows that the users' gender does not affect the acceptance rate ($F(1, 28) = 0.977$ and $p = 0.331$).

Questionnaire Evaluation We also analyzed the interaction from the users' point of view (see Table 3). Concerning the ease of use of the interface (Q1), users perceived easier to use the ENAO interface, but without statistically significant difference with respect to the evaluation of the other conditions (ANOVA). Moreover, there is a strong Pearson correlation between the evaluations of Q1 provided for ENAO and APP ($\rho = 0.84$ with $p < 0.001$) meaning that the users evaluated the ease of use of the two interfaces in the same way. Such correlation becomes moderate in the case of Q2 ($\rho = 0.52$ with $p = 0.003$), where the fulfillment of the expectations with respect the two interfaces is greater for ENAO than the APP with no statistically significance ($p = 0.246$). As, in the previous case study, the users found more natural to interact (Q3) with the APP ($F(1, 29) = 3.777$ with $p = 0.062$).

Table 3 HRI qualitative questionnaire ratings for the second case study

	Q1	Q2	Q3	Q4
	Ease	Expect.	Natural.	Satisf.
APP	4.60 \pm 0.89	3.87 \pm 1.04	4.20 \pm 0.80	3.83 \pm 0.83
ENAO	4.70 \pm 0.79	4.07 \pm 0.78	3.83 \pm 1.12	4.10 \pm 0.96

The interaction with both the interfaces has been evaluated as satisfactory ($F(1, 29) = 2.220$ with $p = 0.147$).

Finally, with respect to the considered personality traits, while agreeableness is the typical factor describing the ease of a person to trust others and to be easy to be satisfied, we found that the conscientiousness personality trait, that is typically related the ability to complete tasks successfully and follow the rules, but also socially prescribed norms, has a moderate Pearson correlation with the satisfaction evaluation in using ENAO ($\rho = 0.409$ with $p = 0.025$). So while an agreeable person has the same behavior when interacting with an APP of ENAO, in the case of contentiousness, the interaction of the human-robot is more meaningful with respect to the one with the APP.

6 Discussion

From the performed statistical analysis came out that even if the rate of acceptance obtained with ENAO condition is always higher than the one obtained by using the APP condition (and the NAO condition), there are not statistically significant differences. Hence, the H2 hypothesis cannot be sustained. Hopefully, these results could be validated more in the case of a larger considered population. Moreover, in our opinion, the considered movie domain implies a high variability in the length of the interaction that may have an impact on the acceptance rate and the evaluations of the interaction. In particular, during the experimental evaluation, we noticed that there were different reactions with respect to the length of the plots. Longer plots require longer interactions with the robotic device, while, sometimes, reading from a screen can be a faster solution.

The physicality of NAO robot alone does not significantly affect the participants' acceptability (H3) with respect to the APP modality, however, this observation evaluated using ANOVA with repeated measures is not supported by statistically significant value. Moreover, while there are several factors affecting the interaction with APP, we noticed, in the first case study, that the acceptance rate remains stable when interacting with ENAO. In particular, according to other studies [1] providing the evidence that using robots for teaching English as a foreign language can be a valuable solution, we observe that results obtained while interacting with ENAO are more stable as the participants' robotic skills and English proficiency level change.

Regarding the expressed preference with respect to the considered conditions, results clearly denote an enhanced experience of users when interacting with ENAO, this time with a statistically significant difference. As hypothesized in H1, the participants prefer to interact with an interface providing a more natural interaction mode. This trend is observed in both the case studies.

Concerning the ease of use of the interface, in the first test, users perceived easier to use the APP, but we found that there was a moderate correlation of robotics skills with respect to the evaluation of the easiness of the interaction. Hence, this result could be affected by the considered population.

Because of the familiarity with applications on mobile phones, in both the presented studies, the participants perceived the interaction with the APP more natural, even if no statistically significant values emerged from this observation. This can also be due to a sort of habituation to the use of the app arising from the first phase of the experiments where the users are forced, before interacting with the one or the other interface, to interact with the android application for the ratings. However, the social behavior expressed by the robot through speech, movements, and gaze, has been easily interpreted by humans and resulted to enhance the users' satisfaction with respect to their expectations.

Finally, with respect to the personality traits (H4), we found that subjects with a higher neurotic trait preferred more to interact with ENAO with respect to subjects with a low value that preferred the interaction with the APP. Moreover, we found a moderate Pearson correlation between the Satisfaction evaluation in using ENAO and the conscientiousness personality trait.

7 Conclusions

Social robots are starting to be used for advertisements in public spaces such as shops and museums, primarily for their greater ability to grab the users' attention with respect to displays. Lately, Socially assistive robotics applications are starting to be developed for home environment to proactively assist the users in their daily activities by providing reminders and recommendations.

This work aims to provide a further validation of the role of socially assistive robots in suggesting effective recommendations and in motivating human users. In particular, we want to demonstrate that such a kind of social interaction is preferred by humans than other types of commonly used interactive interfaces and to evaluate in what extent they can affect the humans' choices. Our assumption is that the potential of a humanoid robot to portray a rich repertoire of non-verbal behaviors could make the interaction more credible and engaging since these behaviors express a social meaning that is very familiar to human users.

In previous works, the study on the users' engagement in interactive tasks were conducted mainly by comparing the advantages of the embodiment condition as provided by a physical robot with respect to its virtual counterpart. Only lately, non-verbal cues are getting more attention within the HRI community for their role in providing a more natural interaction [26]. According to this research trend, we designed two pilot studies, where we considered the effects on human users engagement of a social robot capable to exploit different non-verbal communication channels. However, rather than versus virtual agents, we focused on the comparison of this robot with respect to a well-known interface such as that of a mobile application. In order to confront these interactive interfaces, we considered the recommendation providing task. In particular, we took into consideration the movie recommendation domain, where the use of mobile apps is of common use. We evaluated the impact of three information providing interfaces: a humanoid robot with only voice interaction, a humanoid robot that adds to the voice interaction non-verbal cues, and a mobile-phone application, with respect to different factors, including the personality, which could have an impact on the results. Results stated that although the social robot is preferred by users, it does not affect the acceptance rate of proposed movies, but other factors are involved that could make it preferable with respect to the app.

Compliance with Ethical Standards

Conflicts of interest The authors declare that they have no conflict of interest.

Funding This study has been partially supported by MIUR (Italian Ministry of Education, Universities, and Research) within the PRIN 2015 research project "UPA4SAR - User-centered Profiling and Adaptation for Socially Assistive Robotics" (Grant no. 2015KB-L78T).

References

1. Alemi M, Meghdari A, Ghazisaedy M (2015) The impact of social robotics on 12 learners' anxiety and attitude in english vocabulary acquisition. *Int J Soc Robot* 7(4):523–535
2. Bader N, Mokryn O, Lanir J (2017) Exploring emotions in online movie reviews for online browsing. In: *Proceedings of the 22nd international conference on intelligent user interfaces companion*, pp 35–38. ACM
3. Bainbridge W, Hart J, Kim E, Scassellati B (2008) The effect of presence on human–robot interaction. In: *The 17th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pp. 701–706
4. Bartneck C, Bleeker T, Bun J, Fens P, Riet L (2010) The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn J Behav Robot* 1(2):109–115
5. Boone RT, Buck R (2003) Emotional expressivity and trustworthiness: the role of nonverbal behavior in the evolution of cooperation. *J Nonverbal Behav*

6. Cavallo F, Aquilano M, Carrozza M, Dario P (2012) Robot-era project: the vision of 3d service robotics. *Gerontechnology* 11(2)
7. Cervone F, Sica V, Staffa M, Tamburro A, Rossi S (2015) Comparing a social robot and a mobile application for movie recommendation: a pilot study. In: *Proceedings of the 16th workshop from objects to agents, CEUR workshop proceedings*, pp 32–38
8. Dautenhahn K, Werry I (2004) Towards interactive robots in autism therapy: background, motivation and challenges. *Pragmat Cognit* 12(1):1–35
9. Feil-Seifer D, Mataric MJ (2005) Defining socially assistive robotics. In: *9th International conference on rehabilitation robotics*, pp 465–468
10. Ham J, Bokhorst R, Cuijpers R, van der Pol D, Cabibihan JJ (2011) Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In: *Proceedings of the third international conference on social robotics, ICSR'11*. Springer, Berlin, pp 71–83
11. Iengo S, Rossi S, Staffa M, Finzi A (2014) Continuous gesture recognition for flexible human–robot interaction. In: *ICRA. IEEE*, pp 4863–4868
12. John OP, Naumann LP, Soto CJ (2008) Paradigm shift to the integrative big five trait taxonomy: history, measurement, and conceptual issues. In: John OP, Robbins RW, Pervin LA (eds) *Handbook of personality: theory and research*. Guilford, New York, pp 114–156
13. Johnson DO, Cuijpers RH, Juola JF, Torta E, Simonov M, Frisiello A, Bazzani M, Yan W, Weber C, Wermter S, Meins N, Oberzaucher J, Panek P, Edelmayer G, Mayer P, Beck C (2014) Socially assistive robots: a comprehensive approach to extending independent living. *Int J Social Robot* 6(2):195–211
14. Kidd C, Breazeal C (2004) Effect of a robot on user perceptions. In: *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS)*, vol 4, pp 3559–3564
15. Konstan JA, Riedl J (2014) Recommender systems: from algorithms to user experience. *User Model User-Adapt Interact* 22(1):101–123
16. Lim A, Ogata T, Okuno HG (2012) The desire model: Cross-modal emotion analysis and expression for robots. 74(5):4
17. Mataric MJ (2017) Socially assistive robotics: human augmentation versus automation. *Sci Robot* 2(4)
18. McCrae RR, Costa PT (1987) Validation of the five-factor model of personality across instruments and observers. *J Pers Soc Psychol* 52(1):81
19. de Melo CM, Zheng L, Gratch J (2009) Expression of moral emotions in cooperating agents. In: *Proceedings of the 9th international conference on intelligent virtual agents (IVA)*
20. Merritt SM, Ilgen DR (2008) Not all trust is created equal: dispositional and history-based trust in human–automation interactions. *Hum Factors* 50(2):194–210
21. Murphy-Hill E, Murphy G (2014) Recommendation delivery. In: Robillard MP, Maalej W, Walker RJ, Zimmermann T (eds) *Recommendation systems in software engineering*. Springer, Berlin, pp 223–242
22. Posner J, Russell JA, Peterson BS (2005) The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev Psychopathol* 17(03):715–734
23. Powers A, Kiesler S, Fussell S, Torrey C (2007) Comparing a computer agent with a humanoid robot. In: *2nd ACM/IEEE international conference on human–robot interaction (HRI)*, pp 145–152
24. Qiu L, Benbasat I (2009) Evaluating anthropomorphic product recommendation agents: a social relationship perspective to designing information systems. *J Manag Inf Syst* 25(4):145–182
25. Rossi S, Cervone F (2016) Social utilities and personality traits for group recommendation: a pilot user study. In: *Proceedings of the 8th international conference on agents and artificial intelligence*, pp 38–46
26. Rossi S, Ferland F, Tapus A (2017) User profiling and behavioral adaptation for HRI: a survey. *Pattern Recognit Lett* 99(Supplement C):3–12
27. Rossi S, Staffa M, Giordano M, De Gregorio M, Rossi A, Tamburro A, Vellucci C (2015) Robot head movements and human effort in the evaluation of tracking performance. In: *24th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pp 791–796
28. Salichs MA, Encinar IP, Salichs E, Castro-González Á, Malfaz M (2016) Study of scenarios and technical requirements of a social assistive robot for alzheimers disease patients and their caregivers. *Int J Soc Robot* 8(1):85–102
29. Shinozawa K, Naya F, Yamato J, Kogure K (2005) Differences in effect of robot and screen agent recommendations on human decision-making. *Int J Hum Comput Stud* 62(2):267–279
30. Shiomi M, Shinozawa K, Nakagawa Y, Miyashita T, Sakamoto T, Terakubo T, Ishiguro H, Hagita N (2013) Recommendation effects of a social robot for advertisement-use context in a shopping mall. *Int J Soc Robot* 5(2):251–262
31. Song S, Yamada S (2017) Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In: *Proceedings of the 2017 ACM/IEEE international conference on human–robot interaction*, pp 2–11. ACM
32. Staffa M, Gregorio MD, Giordano M, Rossi S (2014) Can you follow that guy? In: *22th European symposium on artificial neural networks ESANN*, Bruges, Belgium, April 23–25, pp 511–516
33. Staffa M, Rossi S (2016) Recommender interfaces: the more human-like, the more humans like. In: *Social robotics: 8th international conference. ICSR 2016, Kansas City, MO, USA, November 1–3, proceedings*. Springer, Cham, pp 200–210
34. Tapus A, Tăpus C, Mataric MJ (2008) User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intell Serv Robot* 1(2):169–183
35. Tkalcic M, Chen L (2015) *Personality and recommender systems*. Springer, Berlin, pp 715–739
36. Wang YD, Emurian HH (2005) An overview of online trust: concepts, elements, and implications. *Comput Hum Behav* 21(1):105–125
37. Xu J, Broekens J, Hindriks K, Neerincx MA (2013) Mood expression through parameterized functional behavior of robots. In: *2013 IEEE RO-MAN*, pp 533–540
38. Yoo KH, Gretzel U (2011) Creating more credible and persuasive recommender systems: The influence of source characteristics on recommender system evaluations. In: *Recommender systems handbook*. Springer, US, pp 455–477

Silvia Rossi is currently assistant professor at the University of Naples Federico II (Department of Electrical Engineering and Information Technologies). She is currently co-chief manager of the PRISCA (Intelligent Robotics and Advanced Cognitive System Projects) lab. She received the M.Sc. degree in Physics from University of Naples Federico II, Italy, in 2001, and the Ph.D. in Information and Communication Technologies from the University of Trento, Italy, in 2006. During her career, she was research assistant at the Division on Cognitive and Communication Technologies—ITC-irst (Italy), at the institute of Cybernetics E. Caianiello—CNR (Italy), and visiting researcher at the Center for Human–Computer Communication—Oregon Health and Science University, Oregon (USA). She is in the program/organization committee of several international conferences on Human–robot interaction, Artificial Intelligence, and Multi-Agent Systems, and Associate Editor of *IEEE Robotics and Automation Letters* and *Intelligent Service Robotics journal*.

Mariacarla Staffa is currently Research Fellow at the Department of Physics of the University of Naples Federico II, Italy. She received the M.Sc. degree in Computer Science from the University of Naples Federico II with honors, in 2008. She got a Ph.D. in Computer Science and Automation Engineering from the University of Napoli in 2011. She is member of the PRISMA (Projects of industrial and service robotics, mechatronics and automation) and of the PRISCA (Projects of intelligent robotics and advanced cognitive systems) Laboratories, making research in the fields of Cognitive Robotics, Artificial Intelligence and Human–Robot Interaction. She also collaborates with the FITNESS (Fault and Intrusion Tolerant Networked Systems) Research Group

of the University Parthenope, where her main interest is to explore the use of computational neuroscience and artificial intelligence algorithms to generate innovative solutions for addressing cyber-security problems.

Anna Tamburro is currently a student of the M.sc. in Computer Science of the University of Naples Federico II. Her main interests are humanoid robots, Human–robot interaction and Computer Vision. She is currently collaborating with the PRISCA laboratory (Projects of intelligent robotics and advanced cognitive systems) for the development of her M.sc. thesis.