Decision Support in Tourism through Social Robots: Design and Evaluation of a Conversation-Based Recommendation Approach Based on Tourist Segments

Justin Tolle, Alexander Piazza, Carolin Kaiser and René Schallner

Abstract
Tourism recommendation systems can mitigate the potential impact of choice overload on tourists. Social robots are a promising approach to provide recommendations to tourists through an engaging and intuitive user interface on sites like tourist information offices. This study investigates whether tourists perceive tourism recommendations provided via social robots as a satisfying and effective experience and whether tourists respond better to a more human or robotic design of social robot interactions. Therefore, an experiment is conducted at a real-world tourist information office where 60 tourists are exposed to either the more human or robotic version of the social robot recommender system. Their feedback is collected with a survey. The results show that the social robot is perceived positively across all user-centric evaluation dimensions. This indicates that tourists accept social robots in real-world tourist recommendation situations and would also use them in the future.

Keywords
Tourism Recommendation System, Human-Robot Interaction, Social Robot, Furhat

1. Introduction
Tourists are exposed to a wide range of options in their search for a suitable destination and local activities, which can lead to choice overload situations [1]. Recommendation systems can mitigate this choice overload by helping the user make a decision by reducing the number of options for a destination or activity that matches their individual preferences [2]. An emerging and promising approach to offer automated recommendations on sites like tourist information offices is social robots. Social robots can handle complex dialogues and understand and express emotions, e.g., via gestures [3]. These robots interact with the users via speech technology which allows their use in an intuitive and engaging way. Nevertheless, more evidence is needed whether tourists would perceive recommendations provided via social robots as a satisfying and effective experience. According to the Uncanny-Valley-Theory, the degree of human-likeness
determines users’ response to a social robot [4]. Human-like design can be achieved, for instance, by visual features like gestures, auditory features like speech recognition and synthesis, and mental features like content understanding [5]. Nevertheless, the Uncanny-Valley-Theory describes the effect that users tend to respond positively to a human-like robot, but if the robot appears too human, the response becomes negative. The study’s objective is to investigate whether users perceive automated tourism recommendations delivered via social robots as satisfying and effective. A particular focus is whether users prefer a rather robotic or a more human-like social robot interaction. It will also be investigated whether the users accept the recommendation of the social robots and whether they have an influence on their holiday planning. Furthermore, decisive factors in the design of the conversation will be investigated, and the users’ perceptions of a more human-like version and a rather robotic version of the social robot will be compared.

2. Theoretical Framework

2.1. Conversation Design with Social Robots

Unlike industry robots, social robots’ main task is to interact socially with humans. These interactions are generally based on complex dialogues, understanding and expression of emotions, and the social robots’ possession of personality and social competencies to a certain degree [3]. In this study, the social robot Furhat is used to design complex and human-like verbal and non-verbal interactions. This social robot is realized as a head with an animated face using a projector. The interactions are based on facial gestures, the detection of people and their emotions via the camera, and a voice-based multiparty conversational system [6, 7]. Social characteristics should be considered in the interaction design to create satisfying social robot interaction and avoid frustration and dissatisfaction. In the following relevant characteristics for the interaction design of social robot recommender system are introduced, which can be classified into the categories “Conversational Intelligence”, “Social Intelligence” and “Personification” [8].

Conversational Intelligence This describes the social robot’s ability to behave “humanly” in conversations and consists of the following three factors: [8]

- **Proactivity** means that a social robot can act independently in the user’s interest [9] and therefore can grasp the context of the conversation and refer to what is said [8].
- **Conscientiousness** describes the ability to follow the conversation and give appropriate answers [10]. In order to achieve this, the social robot needs to understand the user’s intent [11].
- **Communicability** refers to the ability to communicate effectively and efficiently what it was designed for and how it needs to be interacted with [12]. The robot should be able to present and communicate its capabilities and functions to the user [13]. It can be achieved by stating the purpose of the robot right at the beginning through an opening message [8].

For this study, various dialogues were designed before the programming of the interactions. In both robot versions, care was taken to ensure that the robot can assign the correct intent to
as many different formulations as possible and that it explains its capabilities to the user at the beginning of the dialogue.

**Social Intelligence**  Social Intelligence refers to the individual’s capability to demonstrate an appropriate social behavior for achieving a desired goal [8, 14]. The following three main concepts were considered for the implementation of the recommender system:

- **Damage control** refers to the robot’s ability to deal with conflicts [15]. When talking to a robot, users are more conflict-prone [16], invest less time [17] and use vulgar language more often [18]. To counter abuse, the robot can show an emotional reaction [19], can react authoritatively [20] or escalate the conversation to a human employee [21].

- **Manners** describes the ability to be polite. [10]. To meet social norms of conversation, the robot should be at least as polite as his counterpart [15]. The ability to make small talk supports the impression of a “polite” robot [22] and is often expected [23].

- **Personalization** means the ability to tailor functionality, interface, content and behavior to the user [24].

The humanoid version of the social robot in the study was equipped with extensive damage control functionality to be able to react appropriately to insults. Here, an emotional reaction supported by facial expressions was combined with an authoritarian reaction. In addition, emphasis was placed on the humanoid social robot being always polite and being able to give an answer to a question about the robot’s well-being. The personalization of the social robot is mainly achieved through the individual recommendation that is given at the end of the dialogue, which was the case for both versions.

**Personification**  The identity and personality of a social robot are influenced by its appearance and how it speaks [25] as well as the choice of gender, age, and name [8]. The presence of identity leads to more engagement and perceived human-likeness [26].

The more humanoid version of the social robot has a female face and voice, which sounds neutral in German and has an American accent in English (as the majority of native English-speaking tourists in Rothenburg are from the USA). In addition, the robot is designed to be friendly and helpful. The social robot was given the name “Victoria” because this fits its feminine appearance and voice and is also the name of the text-to-speech voice used. The robotic version of the robot was given the same name to exclude influences here but was equipped with a different face that visually resembles a robot. In addition, the neural voice was replaced by a monotone-sounding standard voice.

### 2.2. Types of Tourists

The social robot aims to give users a recommendation for a suitable activity in the city of Rothenburg ob der Tauber at the end of the conversation. Since the creation of a personality profile individually tailored to each person is too complex, the test persons are assigned to a specific tourist type, based on which an activity is then recommended. This goes hand in hand with the theory of segmentation in marketing, as products cannot be specifically adapted to
Assuming that tourism to Rothenburg is generally characterized by the fact that visitors are looking for a cultural experience, a classification of cultural tourists is selected. A widely used theory for classifying cultural tourists is the classification according to McKercher [28]. McKercher divides cultural tourists into five different types, each differing in the experience sought and the importance of cultural tourism in deciding on a destination (see Figure 1). For the purposeful cultural tourist, learning about the culture and having in-depth experiences is central. For the sightseeing cultural tourist, the former is also important. However, for them, entertainment is in the foreground. For the casual cultural tourist, culture plays a secondary role, and he or she has a superficial experience at the destination. The fourth type only happens to participate in cultural activities during his holiday and has only a superficial experience. In the fifth case, culture does not play a role in selecting the holiday destination, but this type of tourist has a profound experience at the destination by serendipity.

3. Design and Implementation of the Prototype

To enable a comparison between a humanoid and a robotic version of the recommender robot, two skills were programmed for the Furhat. Both versions offer the possibility to interact with the social robot in English and German language. A classic menu structure was used to design the dialogues. After a greeting, which also serves to set the language, the user is led to a "main menu" from which the functions can be accessed. The range of functions was determined in accordance with the research goals and the requirements of the Rothenburg Tourism Information Office. In the first version, the social robot was able to make a recommendation for activities based on the type of tourist and provide simple information about guided tours. The humanoid and robotic versions of the robot differ primarily in the extent to which the aspects mentioned in the literature have been implemented to achieve the most human interaction possible. For example, the robotic version cannot respond to insults, does not engage in small talk, and gives less varied answers to the user’s questions. In addition, a robotic appearance and a rather monotonous voice was chosen. Also, the knowledge about the dialogue process is limited to its basic functions so that the user cannot, for example, ask for a repetition of the last piece of information.
3.1. The Recommender System

The recommender system is based on McKercher’s findings on cultural tourists and adapts the matrix for classifying cultural tourist types. The user is asked two questions to assign him or her to a tourist type based on the dimensions of the matrix (see Fig. 1) and to make a recommendation. The dimension "experience sought" is determined by asking the user whether he or she seeks an in-depth cultural experience in Rothenburg (yes/no). The second dimension is determined by the question, how important the aspect of culture was when deciding to travel to Rothenburg (high, medium or low). Based on this, the user is presented with a recommendation which, depending on the type, offers a profound cultural experience or is more for entertainment purposes. The similar cultural tourism type model supported the selection of specific recommendations according to Pröbstle [29], which, in contrast to McKercher, explains the individual types in more detail.

Pre-test and Optimizations  A pre-test was conducted before the main study to evaluate the experiment setup. Various changes were made to the programming of the social robot to ensure a smooth dialogue and to implement the feedback from the users and the tourism service. In contrast to the recommendation in the literature that the robots should explain their functions at the beginning and thus manage expectations, this was found to be annoying and too lengthy in the pre-test. For this reason, the greeting by the social robot was simplified to only explain its functions when asked. In addition, the dialogue path to get the information about the city tours was shortened to increase efficiency. After the pre-test, the recommender’s functional scope was expanded to provide information on the nearest ATM and toilets as well as with minor small talk functionalities. Further, the recommendation system now provides two additional tourist recommendations per user type if the first recommendation is not considered relevant by the user. In addition, numerous gestures and facial expressions were implemented in the dialogue of the human version of the robot to promote a more natural appearance and to integrate functions such as the confirmation of a statement intuitively by nodding. Finally, in line with the theoretical research on conversational intelligence, a function was implemented that enables the robot to remember and repeat previous user statements during a conversation.

4. Design and Results of the Experimental User Study

A questionnaire is developed to measure the perceived satisfaction and effectiveness of the developed recommender system within the experiment based on a selection of predefined constructs from the Godspeed questionnaire [3] and the ResQue (Recommender systems’ Quality of user experience) model [30]. During the questionnaire design, it was assumed that tourists would only participate in the study if the questionnaire is short. Therefore, only a subset of constructs and items was selected to shorten the time effort per participant. The selected constructs are illustrated in Table 1.

The experiment was conducted for two days in the tourist information office in Rothenburg ob der Tauber. The experiment setup is illustrated in Figure 2. Random tourists were invited to participate as test users and guided through the conversation process. As part of the experiment, 60 people were interviewed who had interacted with the social robot in the tourist information
Table 1
Used constructs in the user study to measure the users’ perception and intentions

<table>
<thead>
<tr>
<th>Category</th>
<th>Constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>User perceived Human-Robot-Interaction</td>
<td>Intelligence [3]</td>
</tr>
<tr>
<td>User perceived Human-Robot-Interaction</td>
<td>Likeability [3]</td>
</tr>
<tr>
<td>User perceived Qualities</td>
<td>Accuracy [30]</td>
</tr>
<tr>
<td>User perceived Qualities</td>
<td>Novelty [30]</td>
</tr>
<tr>
<td>User Beliefs</td>
<td>Ease of Decision Making [30]</td>
</tr>
<tr>
<td>User Beliefs</td>
<td>Perceived Usefulness [30]</td>
</tr>
<tr>
<td>User Attitudes</td>
<td>Overall Satisfaction [30]</td>
</tr>
<tr>
<td>Behavioral Intentions</td>
<td>Intention to use the system [30]</td>
</tr>
<tr>
<td>Behavioral Intentions</td>
<td>Recommendations to Friends [30]</td>
</tr>
<tr>
<td>Behavioral Intentions</td>
<td>Purchase Intentions [30]</td>
</tr>
</tbody>
</table>

Figure 2: The setup of the experimental user study at the Rothenburg Tourist Information Office

office beforehand. Of these, 31 participants had contact with the robotic version of the social robot and 29 participants with the human version. The mean age of the respondents was 44.27, with the youngest being 14 years old and the oldest being 79 years old. From the language perspective, 38 respondents used the German version of the robot, and 22 used the English one. The resulting answers are illustrated in Table 2. The intelligence aspect was rated with an average of 3.93 out of 5.00 and is thus in the upper range. The different versions of the robot differ only minimally with a mean difference value of 0.08. The robot was largely perceived as likeable, as the mean value for the question category "likeability" is 4.24. For example, only one person stated that the robot was "unfriendly". The most remarkable difference between the robot versions can be observed in likeability, where the mean for the human version is 4.66 and for the robotic version 3.79. The distribution of the likeability values is illustrated in Figure 3.

Regarding the accuracy of activity recommendation, the social robot was rated 3.84 on average. On the question of whether the robot suggested novel and interesting activities for the user, the average rating is 3.25. While the accuracy of the human version of the robot was rated 0.47 worse on average, the novelty is valued 0.35 better on average. It should be noted that the recommendations are identical in both versions and no adjustments were made to the dialogues.
Table 2
Overview of all measurements and the measurements of the human (H) and robotic (R) version

<table>
<thead>
<tr>
<th>Construct</th>
<th>Mean all</th>
<th>Std all</th>
<th>Mean (H)</th>
<th>Mean (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intelligence</td>
<td>3.93</td>
<td>0.77</td>
<td>3.89</td>
<td>3.97</td>
</tr>
<tr>
<td>Likeability</td>
<td>4.24</td>
<td>0.80</td>
<td>4.66</td>
<td>3.79</td>
</tr>
<tr>
<td>Accuracy</td>
<td>3.84</td>
<td>1.23</td>
<td>3.61</td>
<td>4.09</td>
</tr>
<tr>
<td>Novelty</td>
<td>3.25</td>
<td>1.19</td>
<td>3.42</td>
<td>3.07</td>
</tr>
<tr>
<td>Ease of Decision Making</td>
<td>3.96</td>
<td>0.93</td>
<td>3.97</td>
<td>4.19</td>
</tr>
<tr>
<td>Perceived Usefulness</td>
<td>3.31</td>
<td>0.98</td>
<td>3.23</td>
<td>3.39</td>
</tr>
<tr>
<td>Overall Satisfaction</td>
<td>3.87</td>
<td>0.95</td>
<td>3.77</td>
<td>3.97</td>
</tr>
<tr>
<td>Intention to use the system</td>
<td>3.46</td>
<td>1.18</td>
<td>3.32</td>
<td>3.62</td>
</tr>
<tr>
<td>Recommendation to Friends</td>
<td>4.12</td>
<td>1.13</td>
<td>3.90</td>
<td>4.34</td>
</tr>
<tr>
<td>Purchase Intentions</td>
<td>3.60</td>
<td>1.21</td>
<td>3.58</td>
<td>3.62</td>
</tr>
</tbody>
</table>

Figure 3: Boxplot of the likeability construct per social robot version

According to the testers, finding an activity they want to do with the recommendation system is mostly easy. The value for this criterion is 3.96, with a standard deviation of 0.93. Overall, 38 out of 60 participants were (very) satisfied with the advice provided by the social robot, which corresponds to a mean value of 3.87 on the rating scale. When it comes to wanting to use a similar recommendation system again in the future, the average score is 3.46 out of 5.00 and 35 out of 60 people rated this question as 4 or 5. The perceived usefulness of the recommendation was rated at 3.31 and is thus in the medium positive range. Within this construct, however, there is wide variation within the used 3 questions. For example, 40 out of 60 people feel supported by the social robot in finding what they like, but only 5 participants said that the recommendations influenced their previous holiday planning. Therefore, the average score for this question is only 2.60 out of 5.00. Looking at the correlations between the individual constructs, few correlations can be identified. The greatest correlation is found in the combination of the accuracy of the recommendation and the satisfaction with the counseling. Here, the correlation value is 0.84. For the aspects intelligence and perceived usefulness, there is also a certain positive correlation in connection with satisfaction with the recommendation (0.72, 0.77). Furthermore, the different mean value for the different application versions regarding sympathy is confirmed here by a negative correlation (-0.55) between sympathy and version. It is also striking that likeability does not show a meaningful correlation with any of the other aspects mentioned. The questionnaire also offered the possibility to leave comments on the experience with the social robot in a free text at the end. The evaluation of these comments showed that 10 study participants stated that the robot did not understand them well. Eight participants said the robot had been “very friendly”, which fits with the scores from the rest of the questionnaire regarding sympathy.
percent (6) of respondents said that the conversation overall was not interactive enough and that they did not have time to explore the robot’s recommendations because of pre-existing plans. Five participants lacked human contact or personal interaction.

5. Discussion

Overall, the user study results indicate that the participants perceive tourist recommendations via social robots well. Nevertheless, two important limitations are not directly evident from the results of the questionnaire data:

- Without an active approach by the study organizers, only a fraction of the people would have sought a conversation with the robot on their own initiative.
- Without active support during the conversation, the dialogue would not have gone through properly in most cases. This had technical reasons, as the speech recognition could not always provide reliable results at the beginning of the conversation.

When comparing the results of the two versions of the recommender system, the sympathy value for the human version is higher than for the robotic version. However, it also showed that sympathy does not influence the users’ behavioral intentions. The two versions have no remarkable differences in the other measured constructs. It should be noted, however, that some strategies proposed and implemented in the literature are only useful when the user interacts freely with the robot, which was not the case in this study due to the limitations. Nevertheless, it became clear in the personal conversations with the users that these extended functionalities can positively affect the perception of the social robot if they are sufficiently well elaborated. An important result of the survey is that although the advice was very positively received, the tourists’ holiday planning was not influenced by the recommendations, as in their case, the planning was already done before the trip. In future research, validating and deepening these findings with a larger sample based on a longer time frame is recommended. Furthermore, the design of the experiment setup should enable a completely independent interaction of the test users through, e.g., improved interaction design and robust speech recognition.

6. Conclusion

The study aimed to evaluate the users’ perception of a tourism recommender system using a social robot and to examine the influence on the users’ behavioral intentions. In addition, the change in users’ perception of a human-like and a robotic version of the social robot was compared. The questionnaire evaluation indicates that the tourists perceived the social robot recommender as satisfying and effective in both versions. The majority of users also stated that they would use such a recommendation service again in the future. However, the recommendation only had a minor influence on the users’ holiday planning. The sympathy value for the human version is higher than for the robotic version, but the perceived sympathy does not influence the users’ behavioral intentions. The differences between the two versions of the robot in the individual criteria are only marginal and suggest that a lower human-like design of the social robot only has a minor influence on the quality of the decision support.
References


