The Influence of Emotion Detection on Users' Engagement in an Interactive Storytelling Scenario

Stefanos Antaris, Wille Eriksson and Philipp Mondorf

Abstract

Storytelling is fundamental to our human nature and strongly rooted in our culture. It allows us to efficiently communicate knowledge in social contexts. Therefore, it emerges as an essential component to intelligent computer systems, and in particular conversational agents. Typically, storytelling conversational systems interact with another physical person to narrate a story and adjust the story according to the person's actions. The main goal of these conversational systems is to attract the physical person's attention and engage the person during the overall story. Detecting user emotions is essential for the success of such storytelling conversational systems as emotions play an essential role in human communication. In this work, we implemented a conversational storytelling system using the FurhatSDK, to narrate an interactive story in which users can actively change the course of events by communicating their actions to the storyteller. We periodically detect user emotions based on an emotion detection component that employs Convolutional Neural Networks (CNN) on static images. Given the detected emotions, the storytelling conversational system adjusts its responses to attract the user's attention. We conducted experiments on 16 individual persons and validated the influence of the emotion detection on the users' engagement based on a post-experiment questionnaire. Preliminary results from our experiments demonstrate no statistically significant benefit in terms of user engagement when the emotion detection component is activated. However, our results indicate that the participants which exploited the emotion detection component experienced less misunderstandings during the story. Similarly, participants perceived the robotic storyteller as more intelligent when it was capable of detecting emotions than participants who tested the robot without emotion detection. We make our implementation publicly available at https://github.com/stefanosantaris/storyteller.

I. Introduction

When humans communicate with each other, they naturally express emotions either by directly addressing them or indirectly through nonverbal behavior. Even more, interlocutors often mirror emotional expressions as they tend to imitate each others' nonverbal behavior [1], [2]. In this manner, a smile often follows a smile, while a sad facial expression might be mirrored in order to display sympathy. Such an alignment of nonverbal behaviour through time promotes mutual understanding, agreement and acceptance [3]. However, it not only requires the capability of displaying the correctly detected emotion, but also demands a certain sensitivity to timing [4].

In recent years, there have been tremendous advancements in the field of Social Robotics [5], [6]. Robots are no longer merely applied within industrial contexts, but play an increasing role in healthcare, entertainment and education [7]. As such, continuous efforts have been made to improve the multi-modal communication of robots deployed in social contexts. Previous works have shown that social robots and virtual agents which express appropriate emotions are perceived to be more likeable and trustworthy [8], [9], [10]. Moreover, Eyssel et al. [11] showed that robots providing emotional feedback during a Human-Robot interaction are rated as more sympathetic and anthropomorphic compared to robots that are insensible to emotions.

In storytelling, emotions are often an indicator as to how engaged an audience is. When highly transported into a narrative story, emotions are frequently evoked as the audience sympathizes with the protagonists. Nabi and Green [12] further highlighted the importance of emotions in storytelling by showing that changes in the emotional experience across the

course of a narrative promote the engagement into the story. Such emotional shifts seem to increase narrative persuasion, which has been a major focus of consumer research [13], [14], [15]. Previous works have explored the potential of social robots to tell stories to a human audience. One of the driving factors of this field of research is the robots' ability to exhibit facial expressions and other modes of nonverbal behavior to support the narrative flow [16], [17]. Appel et al. [18] designed a storytelling robot that shows emotions which were congruent or incongruent with the events of the narrative. In this manner, the authors could show that displaying congruent emotions with the events of the story increased sympathy and affection towards the social robot as well as the users' engagement into the story.

In this work, we design a robotic storyteller for an interactive storytelling scenario, where users can actively change the course of events by communicating with the robot. While previous studies have mainly focused on the display of emotions to increase the users' narrative engagement, this work is intended to study the influence of detecting and addressing users' emotions to increase this engagement. In particular, we conduct an experiment in which the robotic storyteller guides participants through an interactive story in which they can change the course of events by orally communicating their actions. To investigate the influence of emotion detection on the participants' narrative engagement, we compare two different experimental conditions: i) the users' emotions are detected and incorporated into the story, and ii) the users' emotions are neglected. Finally, we evaluate the experiment by asking participants to fill out a questionnaire.

The remainder of this paper is organized as follows, we review related work in section II, and describe the design and implementation of our robotic storyteller in section III. We further give an overview of the story line designed for this study and explain how we incorporate emotions detected into the narrative. In section IV, we present the results of our study, which are further discussed in section V. Finally, we conclude our study in section VI.

II. Related Work

Our work intersects with research on storytelling, emotion detection, and user engagement. We review recent advancements below.

A. Storytelling

Although storytelling has been developed for years, storytelling in conversational systems is a relatively new subject, thus facing many challenges. Early approaches of storytelling conversational agents have been employed in computer games, where the player interacts with a non-playable character (NPC) [19]. In particular, a storytelling conversational agent provides options to the player and decides the next steps of the game based on the player's choice. Recently, commercial systems such as Charisma.ai allow the generation of storytelling units as part of the digital media projects [20]. By applying storytelling techniques, Charisma.ai enhances their interactive stories, as to encourage the audience to achieve higher engagement. Moreover, conversational storytelling agents have found remarkable success in several other domains, such as education [21], [22], [23], healthcare [24], [25] and recommender systems [26], [27]. Teachers and doctors deploy storytelling conversational agents as assistants to efficiently describe the course and the diagnose to students and patients, respectively. Additionally, storytelling conversational agents are employed in recommender systems to efficiently overcome the cold-start users problem, while providing further explanation of the recommendations [28], [29].

B. Emotion Recognition

The main goal of emotion recognition in conversation (ERC) is to detect the user's emotions during a conversation. Detecting the user's expression in a conversation requires the supervision of multiple sequential behavioral patterns. Therefore, existing approaches exploit

multimodal techniques with recurrent architectures [30], [31], [32]. Zhang et al. [33] exploit text and image as the multimodal input to a quantum multimodal network and a Long Short Term Memory network to analyze sentiment in conversations. Wang et al. [34] employ a multimodal deep regression Bayesian network (MMDRBN) to identify the relationship between audio and visual modalities for emotion recognition. Moreover, Xing et al. [35] detect user emotions by applying an Adapted Dynamic Memory Network on audio, visual and text inputs. Similarly, Hazarika et al. [36] exploit gated recurrent units to analyse audio, visual and text features. Lai et al. [37] supply real-time emotion recognition by employing two Recurrent Neural Networks (RNNs) to compute the utterances before the target and two more RNNs for utterances after the target emotion.

C. User Engagement

Measuring user engagement is essential for the success of interactive storytelling systems. Barber and Kudenko [38] proposed predefined increments and decrements to a vector of user personality traits, such as honesty and selfishness, to measure user personality. At each time step of the story, the authors measure the user's personality traits and exploit these traits to identify the user's engagement level. Similarly, Seif El-Nasr monitors the user's actions to capture tendencies towards character traits, such as heroism, violence, self-interest, and cowardice [39]. In doing so, they are capable of tracking both the player behavior and personality. PaSSAGE [40] models the player's preferences through observations of the player in the virtual world. Based on the modeled user's preferences, the system selects the content of an interactive story dynamically. Ramirez and Bulitko employ a reward function on the model generated by PaSSAGE to select the narrative that maximizes the reward [41].

To assess users' engagement, researchers in psychology have developed several psychometric tests and measurement scales. Among the most widely accepted are those that follow the Big Five proposal, that is (I) Surgency (or Extraversion), (II) Agreeableness, (III) Conscientiousness (or Dependability), (IV) Emotional Stability (vs. Neuroticism), and (V) Culture [42]. Further research in psychology showed that if we ask human subjects a small set of well-designed questions, applied in one minute or less, then we can successfully ascertain their personality traits. In this work, we use the latter approach, by applying a well-defined survey at the end of each experiment to assess the users' engagement into the story.

III. METHOD

With this work, we intend to investigate the influence of emotion detection on a user's narrative engagement in an interactive storytelling scenario. In this manner, we formalize the following research question:

RQ: To what extent does emotion detection influence a user's narrative engagement in an interactive storytelling scenario?

Based on previous works indicating the importance of emotions within conversations and storytelling [1], [4], [12], the following hypothesis was formulated:

H: Emotion detection increases the users' engagement in such a scenario.

The method used to answer the research question will be presented in three sections, corresponding to the approach we took in completing it. First, an interactive storytelling system was built, then an emotion capturing service connected to a camera was implemented and finally, a user study was conducted where users where asked to interact with the storyteller and evaluate their interactions through a questionnaire.

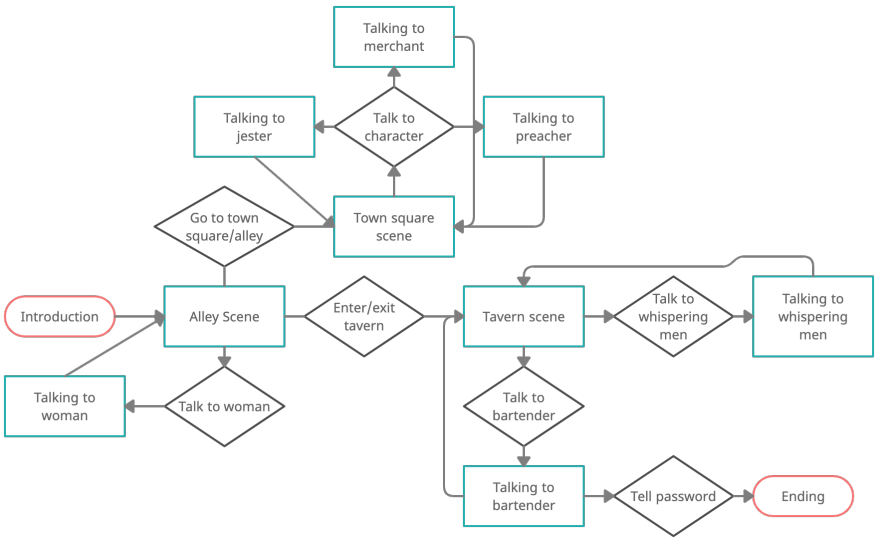

Fig. 1: State chart outlining the flow of the storyline. States are indicated with teal rectangles and intents with black diamonds. Starting and ending states are marked as red rounded rectangles.

A. *Implementation of the storyteller*

For building the storyteller, we used the Furhat robot [43]. The main reason of selecting this robot is that Furhat has gesture and direct gaze capabilities, which benefits the user's attention during a storytelling [44]. Moreover, the state-based modelling approach used in the FurhatSDK further motivated the choice as this fits well with the task of implementing an adventure game.

As illustrated in Figure 1, the story used in the experiment was modelled as a finite-state machine, where each state was defined by an entry action, as well as a number of transition possibilities. A transition from one state to another was triggered by an intent from the user through spoken utterances.

The storyline was inspired by a one page adventure by Tyler Monahan [45], and takes place in a medieval age setting in the fictional town of Millstone. The user plays a detective of the county watchmen who has been sent to the town in order to investigate rumors of an emerging cult. Below, a short description is given for each of the major scenes outlined in the story. A flowchart describing the plot can be found in Figure 1.

1) Introduction: In the introduction, the user is greeted by the narrator of the story, who asks for the user's name and briefly outlines the setting of the experiment. The user is then asked whether they want to proceed by starting the game, and is upon acceptance introduced to the setting of the story. The story begins by the protagonist arriving in the town of Millstone. Upon entering the town gates, the user observes a man with a strange tattoo on his arm. The protagonist decides to follow the man into a dark alley where the man disappears into a tavern.

With this section, the settings of the experiment is added to the common ground, which is accepted by the user agreeing to play the game. Two entities are provided here: a *cult* and a *tattooed man*, both of which may be used in future conversations. The section also has the additional goal of building trust with the robotic narrator, making it to seem more human-like by asking the user for its name and engaging in small talk, actions which have no bearing on the task of completing the adventure.

2) Alley scene: Upon arriving in the alley, the user sees a woman standing next to the door of the tavern. Furthermore the user hears noises from a town square up ahead. In this scene, valid intents are to approach the woman, go to the town square or to enter the tavern.

If the user chooses to speak to the woman, she or he will be introduced to a new entity upon ending the conversation, namely the *preacher* in the town square.

3) Town square scene: If the user chooses to visit the town square, it is presented as a scene where a performing *jester*, a *merchant* selling food in a market stand and a preacher can be seen. Valid intents are to talk to any one of these characters or to return to the alley. The jester and the merchant are mainly used for exploratory purposes, but the preacher can provide the user with a *password*. This entity is needed in order to complete the game. When receiving the password, the user is urged to make its way back to the tavern and present it to the bartender.

4) Tavern scene: If the user chooses to enter the tavern, she or he is entering a scene in which *two men* are sitting by a table, whispering, and a *bartender* is standing behind the counter cleaning dishes. In this scene, valid intents are to speak to the two men, to approach the bartender or to exit to the alley. If the user has received the password by speaking to the preacher in the town square, they can proceed to the next scene by presenting it to the bartender. There is also the possibility of receiving the password from the two men at the table by bribing them. On the other hand, the game may end here if the user decides to start a fight with any of the two entities.

5) Ending: If the user presents the password to the bartender, she or he will be taken to the final sequence of the game, where the bartender leads the user to a corridor in the basement. After walking down the corridor, it splits into three different directions and a riddle is presented in order for the user to know which one to take. If the correct path is chosen, they find the headquarter of the cult and manage to arrest its members, thus successfully completing the game.

In addition to the dialog, features were added to the storyteller in order to make it more human-like. Each character in the story was assigned a unique face and voice. Moreover, features such as facial expressions, gestures and voice tonality were modified to fit the emotional state of the character and the nature of the interaction.

B. Emotion detection

While interacting with the storyteller, the user's face was being recorded with a camera connected to an external device, performing real-time emotion detection. The user's facial expressions were registered and classified once every second, and then used by the FurhatSDK in order to influence the story line.

In order to build our emotion detection service, we used OpenCV [46] for video management and facial recognition, and a prebuilt CNN in Keras [47] for emotion detection. The procedure was heavily influenced by an implementation made by Karan Sethi [48] [49], who also trained the prebuilt model.

Each second, the camera captured one picture of the user. The section with the face of the user was then identified with the Haar Cascade algorithm [50], and resized to 48×48 pixel images. These images were subsequently fed as inputs to the CNN, which classified the emotion displayed on the user's face.

The CNN was trained on a dataset collected for *the ICML 2013 Workshop on Challenges in Representation* [51], a contest where participants were tasked with developing as efficient machine learning algorithms for emotion classification as possible. It consists of 35 887 images, each labeled with one of seven different emotions: *anger*, *disgust*, *fear*, *happiness*, *sadness*, *surprise* and *neutrality* (indifference). It is worth noting that the number of images for each emotion was not equally distributed; for example 8 989 images were labeled as "happiness", while only 547 were labeled as "disgust". For this reason, we chose to focus on

Fig. 2: The experimental setup. The robotic storyteller is placed next to a computer that textually displays the current conversation. An external camera is capturing the facial expressions of the participants to infer their emotional state.

emotions with larger subsets of labeled data in our implementation, as these were easier for the CNN to identify.

We integrated emotion detection into the story by querying the service for the user's current emotion at certain predefined points throughout the story. Depending on the response of the service, different acts of dialog were triggered in accordance with the emotion displayed. Such points were implemented during the introduction, and at least once during each interaction with the different characters.

As an example, we consider the initiation of the interaction with the merchant in the town square. The merchant tells the user about her brother who she has not seen for several days, and that she is beginning to worry. The emotion detection service is then queried for the emotion of the user. If the response indicates either happiness, sadness or neutrality, one of the following pieces of dialog would be triggered:

- Happiness: *"You think this is funny? Either stop smiling or get out."*
- Sadness:*"Lose the gloom, I don't need your pity."*
- Neutrality: *"Not that you seem to care anyway..."*

Note that the emotion detection did not fundamentally alter the path of the story, rather it determined what characters said during conversations which the user would have traversed in any case.

The goal of having these points throughout the story was to raise the characters' awareness of the emotional state of the user, with the hope of increasing user engagement. In the following sections, we outline the methods used for evaluating the results of these efforts.

C. Experimental Design

Our experiment follows a two condition, between-subject design where participants are asked to interact with the robotic storyteller described in III-A. In particular, the storyteller guides participants through an interactive story in which they have the chance to alter the course of events by verbally telling the robot which actions they want to pursue in the story. In this sense, the participants become the main protagonist of the story, following the concept of an interactive story as defined in [52]. In order to address the influence of emotion detection on the users' engagement into the story, we investigate two different experimental conditions - a setup with emotion detection enabled and incorporated into the story line as described in section III-B, and a setup without emotion detection. Finally, we let the participants fill out a questionnaire to evaluate their perception of the story and the robotic storyteller.

D. Procedure

For the experiments, participants were asked to meet the instructor in a dedicated briefing room where they were handed a short information sheet describing the general setup and flow of the experiment. After the instructor had ensured that any upcoming questions had been clarified, an informed consent form was signed by the participant. Thereafter, the instructor guided the participant to the experimental setup depicted in figure 2. The instructor briefly repeated the experimental flow orally to ensure that everything was understood and no open questions remained. Subsequently, the participant started to interact with Furhat. First, Furhat introduced itself and asked for the participant's name. This sequence was intended for the participant to get used to the way she or he can communicate with the robot. After Furhat had asked the participant whether she or he was ready to play the game and the participant affirmed so, the story line was introduced. From this point on, Furhat guided the participant through the story and the participant communicated which action she or he would like to take. Once the participant successfully solved the game or failed to do so, Furhat asked whether or not the participant would like to play again. Depending on the participant's answer, the game was either replayed or not. Finally the instructor asked the participant to fill out a post-experiment questionnaire.

E. Experimental Measure

The experimental measures are designed to capture the users' engagement into the story. Furthermore, they are intended to evaluate how the participants perceive the robotic storyteller.

1) Quantitative Measures: As mentioned briefly in section III-D, participants were asked to fill out a post-experiment questionnaire. The questionnaire consisted of a series of 17 questions that can be grouped into three different categories: the perception of the adventure, the perception of the robot and the users' engagement. In order to fully capture the participants' perception of the robotic storyteller and the story line, we used different scales within the different subcategories of the questionnaire. In particular, we primarily made use of a 5-point Likert scale to evaluate the participants' perception of the story. For evaluating the perception of the robot, we used a 5-point semantic difference scale. An excerpt of the questionnaire can be found in table I.

2) Qualitative Measures: At the end of the experiment, participants were given the chance to freely comment on the interaction with the robotic storyteller. Furthermore, we asked the participants what and how they would improve the system.

IV. Results

In this section, we present the results obtained from the post-experiment questionnaire and interview. As mentioned in section III-E, the questionnaire's items can be grouped into three

TABLE I: An excerpt of the post-experiment questionnaire. The questions are grouped into three different categories: the perception of the adventure, the perception of the robot and the users' engagement. The questions might follow different scales based on the category.

Perception of the Adventure
I felt interested in the adventure (1) Completely disagree - (5) Completely agree
Perception of the Robot
(1) Machinelike - (5) Humanlike? (1) Inarticulate - (5) Eloquent?
Users' Engagement
What is your perception about the duration of the adventure? (1) Short - (5) Long

different categories: the perception of the adventure, the perception of the robot and the users' engagement. For the presentation of the results, we adhere to this subdivision. Finally, we present an evaluation of the qualitative data we collected.

A. *Quantitative Evaluation*

A total of 16 participants took part in the study where 8 participants were assigned to the condition that included emotion detection and 8 participants were testing the experimental setup without emotion detection. Participants' ages ranged from 22 to 28 years (M = 23.78, SD = 1.98). All participants successfully completed the experiment and filled out the post-experiment questionnaire. In the following, we present the evaluation of their answers.

1) Perception of the story: To evaluate the participants' perception of the story, we asked whether they felt interested in the overall adventure. For this, participants had to answer the question *I felt interested in the adventure* on a 5-point Likert scale, where an answer of 1 corresponds to *I completely disagree* and a value of 5 reflects the statemenet *I completely agree*. The overall rating of all participants was comparatively high with a mean value of M = 4.44 (SD = 0.61) and a median of m = 4.50. Figure 3 displays the participants' answers grouped by the different experimental conditions. As we can see, participants subjected to emotion detection (ED) expressed slightly higher interest in the adventure (mean value of M^{ED} = 4.62 (SD = 0.52), median value of m^{ED} = 5) than participants not subjected to emotion detection (mean value of M^{NED} = 4.25 (SD = 0.71), median value of m^{NED} = 4). However, performing a statistical T-test to investigate whether emotion detection increases the overall interest in the adventure shows that the mean interest of participants subjected to ED was not significantly higher than the mean interest of participants not subjected to ED: t(12.8280) = -1.2104, p = 0.2479 $>$ 0.05.

Furthermore, we asked participants whether they experienced any misunderstandings during the adventure. Figure 4 shows that participants interacting with the storyteller that could detect emotions experienced less misunderstandings than participants interacting with the storyteller for whom emotion detection was disabled. While the answers of the former group are reflected in a mean value of M^{ED} = 3.50 (SD = 0.76) and a median value of m^{ED} = 4, the latter group rated the question with a mean value of M^{NED} = 2.38 (SD = 1.41) and a median value of m^{NED} = 2. A subsequent independent samples T-test showed that the mean clarity of the story line might be considered to be statistically correlated with the experimental condition, i.e. whether or not emotion detection was present: t(10.7260) = -1.9912, p = 0.0730 $<$ 0.1.

Finally, we wanted to know whether *the conversations with the characters felt natural for most parts of the story*. Based on the participants' answers, we could observe that both

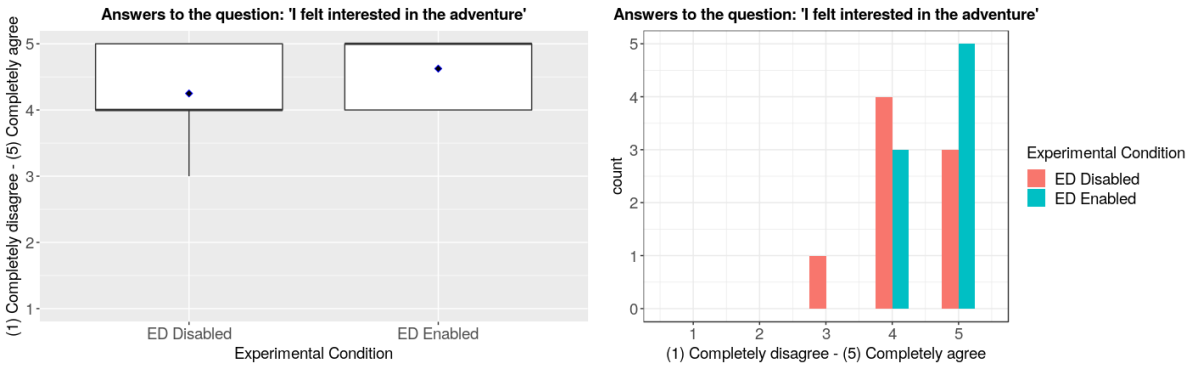

Fig. 3: Participants' answers to the question *I felt interested in the adventure*, based on a 5-point Likert scale. The results are displayed grouped by experimental condition: with emotion detection (ED) and without emotion detection. The blue rhombus depicted in the diagram on the left displays the mean value, while the bold horizontal line indicates the median value.

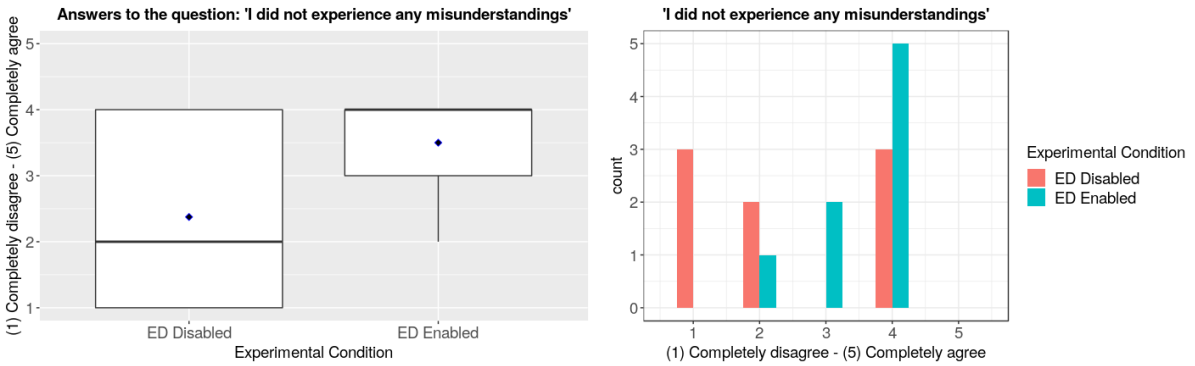

Fig. 4: Participants' answers to the question *I did not experience any misunderstandings during the adventure*, based on a 5-point Likert scale. The results are displayed grouped by experimental condition: with emotion detection (ED) and without emotion detection. The blue rhombus depicted in the diagram on the left displays the mean value, while the bold horizontal line indicates the median value.

experimental groups tended to perceive the conversations as natural. In particular, participants subjected to emotion detection answered this question with a mean of M^{ED} = 4.25 (SD = 0.46) and a median of m^{ED} = 4, while participants not subjected to ED rated the question with a mean of M^{NED} = 3.88 (SD = 0.64) and a median of m^{NED} = 4. However, a respective T-test could not show any statistically significant impact of emotion detection on the participants' perceived clarity of the story.

2) Perception of the robotic storyteller: To evaluate the participants' perception of the robotic storyteller we asked 6 different questions about the participants' conception of the robot. For this, we used a 5-point semantic differential scale in which answering options are grammatically opposite adjectives at each end of the scale. A summary of the participants' answers for each question can be found in table II. We display the mean value with standard deviation and the median for each question, grouped by the experimental condition. When observing the values displayed, we note no significant differences in the mean or median values between the two experimental conditions. This has been confirmed by respective T-tests that registered no statistically significant correlation between experimental condition, i.e. whether or not emotion detection was enabled, and the participants' perception of the robot. Only for the question whether the robot was perceived to be *(1) Unintelligent* or *(5) Intelligent*, we can observe a discrepancy between the two conditions. A respective T-test shows that the mean intelligence of the robot perceived by participants subjected to emotion detection (M^{ED} = 3.62, N = 8) was significantly higher than the mean intelligence perceived by participants not subjected to emotion detection (M^{NED} = 2.38, N = 8): t(12.489) = -3.90, p = 0.001959 $<$ 0.05.

TABLE II: A summary of the participants' answers related to their perception of the robotic storyteller. We display the mean value with standard deviation and the median for each question grouped by experimental condition.

	ED Enabled		**ED Disabled**	
Question	**Mean $\pm$ Stdev**	**Median**	**Mean $\pm$ Stdev**	**Median**
(1) Machinelike - (5) Humanlike?	3.00 ± 0.756	3.0	2.75 ± 1.04	3.0
(1) Artificial - (5) Natural?	3.00 ± 0.756	3.0	2.62 ± 1.06	2.5
(1) Dead - (5) Alive?	3.12 ± 1.13	3.0	3.62 ± 0.916	4.0
(1) Inert - (5) Interactive?	3.5 ± 0.535	3.5	3.5 ± 0.756	4.0
(1) Unintelligent - (5) Intelligent?	3.62 ± 0.518	4.0	2.38 ± 0.744	2.5
(1) Incompetent - (5) Competent?	4.0 ± 0.535	4.0	3.25 ± 1.04	4.0
(1) Inarticulate - (5) Eloquent?	4.0 ± 0.535	4.0	3.88 ± 0.991	3.5

3) Users' Engagement: Finally, we evaluate questions related to the participants' engagement into the story. We asked participants about their *perception of the duration of the adventure*. Figure 5 depicts the participants' answers to this question. As we can see, participants subjected to emotion detection tend to perceive the story to be longer (mean value of M^{ED} = 3.25 (SD = 0.89), median of m^{ED} = 3) than participants not subjected to emotion detection (mean value of M^{NED} = 2.38 (SD = 1.06), median of m^{NED} = 2.50). In addition, a respective T-test indicates a slight correlation between between experimental condition and perceived duration of the story: t(13.572) = -1.79, p = 0.09571 $<$ 0.1.

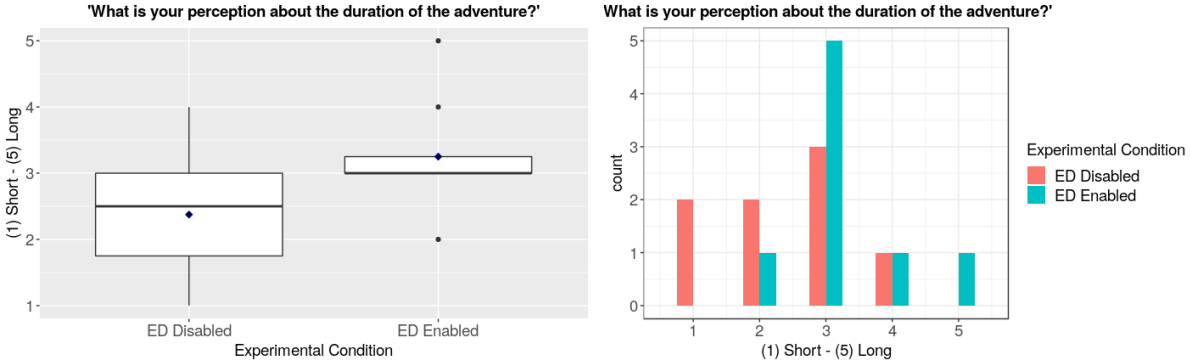

Fig. 5: Participants' answers to the question *What is your perception about the duration of the adventure*, based on a 5-point semantic differential scale. The results are displayed grouped by experimental condition: with emotion detection (ED) and without emotion detection. The blue rhombus depicted in the diagram on the left displays the mean value, while the bold horizontal line indicates the median value.

B. Qualitative Evaluation

In the post-experiment interview, we gave participants the chance to freely comment on the interaction with the robotic storyteller. Furthermore, we asked participants what and how they would improve the system. When having the chance to freely comment on the interaction, participants mainly made positive remarks:

P1: *I really like how the face changed with each character. That helped me getting into the game and imagine the scenes properly.*
P2: *This was fun!*
P3: *You guys did a good job!*

Despite these positive remarks, some participants commented on the restricted options they had during the game:

P4: *I had limited amount of options. Also, the storyteller could possibly wait a little longer for an answer.*
P5: *I asked whether I had a gun in the game, but the storyteller did not understand that.*
P6: *It was clear that the conversational options were limited. However, I liked that I could communicate these options in my own way, and was not constrained to a particular phrase.*

When asking how to improve the system, participants mainly commented on particular parts of the story they had trouble with. One participant mentioned that his facial expressions seemed not always to be recognized correctly:

P4: *I had some issues when I tried to talk to the two shady-looking men in the bar.*
P5: *My facial expressions were not always correctly recognized.*

V. Discussion

We have designed an experiment to study the influence of emotion detection on a user's narrative engagement in an interactive storytelling scenario, and presented the results of this experiment in section IV. In the following, we discuss these results with respect to our research question and try to give an answer to our hypothesis.

A. The influence of emotion detection on users' engagement

In the user study, we tested the experiment on 16 different test subjects in ages ranging from 22-28 years old, 8 of which were in a control group and 8 of which were in the group where emotion detection was used. For quantitative studies, it has been argued that a minimum of 20 test subjects should be used in order to obtain a satisfactory level of statistical significance in the results [35]. Although the number of test subjects in this study was not very far off, with more time more extensive testing could have been made. In regards to the group of test subjects, it should also be noted that they were all within a rather limited age range. Neither was there any control for gender made in the study. It can not be ruled out that such factors would have an impact on the results. In the following, discuss the results presented in IV.

When looking at how participants perceived the story, it can be noted that both the control group and the group using emotion detection rated their interest in the story relatively high, with the ED group scoring slightly higher. These results are in line with our hypothesis, albeit noting a rather weak statistical significance. Of more interest however, is that the strongest correlation in our results seems to suggest that users who interacted with the system using emotion detection experienced a lower level of misunderstandings. The reason for this outcome is not entirely apparent, but it could be hypothesized that an emotively responsive robot acts more in accordance with behaviours expected by the user. Furthermore, we observed that participants reacted surprised when the storyteller commented on their emotions. This promoted the conversational flow as participants gave immediate responses to the robot's comment. However, it should be noted that the standard deviation among both groups was rather high for this question.

Somewhat surprisingly, there was no clear difference between how the two groups perceived the storyteller, an outlier being the question of intelligence where the ED group perceived the robot as being more intelligent. This is also reflected in the question of whether or not the conversations felt natural, where no statistically significant disparity was to be found. A plausible explanation to this is that the emotion detection was employed relatively subtly throughout the story. Furthermore, it was not always triggered, depending on what emotion the user displayed. As a result, the two groups were to a large extent interacting with the same storyteller.

For measuring engagement, the users were asked about their perception of the adventure's duration. For this question, we can see somewhat significant discrepancies between the two groups, with the control group generally perceiving the adventure as being shorter than the ED group. Although the ED group in fact had more dialog, the time these more extensive conversations required should have been negligible in comparison to the entire adventure. How to interpret these answers in terms of engagement could be up for debate. However, how a person might answer this question could be very dependent on their previous expectation about the length of the adventure.

B. Limitations and Future Work

For future work, the improvements that could be made are twofold: the quality of the emotion recognition can be increased and the results of the emotion recognition can be used more extensively in the storytelling. Although facial expressions may decently approximate the emotion of a user, other factors could also be taken into account in order to increase the accuracy of the predictions. Research has shown both body language and voice to have predictive features in terms of emotions [53]. Further, the acoustic properties of the voice can be combined with the information of the language for this purpose [54]. Applying these techniques to our scenario could likely decrease the number of instances where emotions were incorrectly detected, allowing for more appropriate responses from the storyteller.

In addition to this, a greater discrepancy between the results of the two test groups could probably be found if emotion detection were to be used to a greater extent in the storytelling.

For example, actions such as the unlocking of dialog could be used in more places and for a greater range of emotions, but emotions could also potentially be made to have a greater influence on the story and to affect features such as voice tonality and facial expressions of the characters.

VI. Conclusion

In the present work, we have investigated to what extend emotion detection influences the users' narrative engagement into an interactive story. For this, we designed an experiment in which participants interact with a robotic storyteller that guides them through an interactive story where they can change the course of events by verbally communicating their actions within the story. To examine the influence of emotion detection on the engagement of the participants, we followed a two condition, between subject study design, where one group of participants interacted with a robot that incorporated detected emotions into the story line, while the other group interacted with a robot that was insensible to emotions.

From our results, we could observe statistically significant differences between the two study groups. In particular, participants interacting with the emotion detecting storyteller perceived the robot to be significantly more intelligent than participants communicating with the robot insensible to emotions. Furthermore, we observed that participants exposed to emotion detection experienced considerably less misunderstandings during the story than participants not exposed to emotion detection. As mentioned in section V, a possible reason for this might be the immediate and strong reactions we observed when the robotic storyteller commented on the participants' emotions. In particular, we believe that such strong reactions promote the conversation between storyteller and participant and therefore result in less stagnation within the story.

However, despite these findings we could not find further statistically significant differences between the two experimental groups. In particular, we could not observe a clear indication that emotion detection increases the users' narrative engagement. This might have multiple reasons. The most evident factor might be the small batch of participants. In addition, emotions might have been incorrectly detected during the story, resulting in confusion rather than promoting the users' engagement. Furthermore, emotions detected might not have been sufficiently incorporated into the story line.

Nevertheless, the positive feedback we received indicates that there is a demand for such applications. Furthermore, we could see that users perceived the emotion detecting storyteller significantly different compared to the insensible robot, regarding it to be more intelligent. This aligns with previous works, indicating the importance of emotions in conversations [2]. We hope that future work will build on what we have presented in this work, further exploring the impact of emotion detection (and in particular emotion back-channeling) on the users' narrative engagement.

References

[1] G. Rizzolatti and L. Craighero, "The mirror-neuron system," *Annual Review of Neuroscience*, vol. 27, no. 1, pp. 169–192, 2004, pMID: 15217330. [Online]. Available: https://doi.org/10.1146/annurev.neuro.27.070203.144230

[2] F. Bernieri and R. Rosenthal, "Interpersonal coordination: behavior matching and interactional synchrony," *Fundamentals of Nonverbal Behavior. Studies in Emotion and Social Interaction,*, 01 1991.

[3] M. M. Louwerse, R. Dale, E. G. Bard, and P. Jeuniaux, "Behavior matching in multimodal communication is synchronized," *Cognitive science*, vol. 36 8, pp. 1404–26, 2012.

[4] M. Bruijnes, "Social and emotional turn taking for embodied conversational agents," 09 2012, pp. 977–978.

[5] C. Breazeal, K. Dautenhahn, and T. Kanda, *Social Robotics*. Springer International Publishing, 2016, pp. 1935–1972. [Online]. Available: https://doi.org/10.1007/978-3-319-32552-1_72

[6] T. B. Sheridan, "A review of recent research in social robotics," *Current Opinion in Psychology*, vol. 36, pp. 7–12, 2020, cyberpsychology. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352250X2030004X

[7] T. Gnambs and M. Appel, "Are robots becoming unpopular? changes in attitudes towards autonomous robotic systems in europe," *Computers in Human Behavior*, vol. 93, pp. 53–61, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563218305806

[8] M. Ochs and C. Pelachaud, "Model of the perception of smiling virtual character," in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, ser. AAMAS '12. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2012, p. 87–94.

[9] I. Torre, S. Tuncer, D. McDuff, and M. Czerwinski, "Exploring the effects of virtual agents' smiles on human-agent interaction: A mixed-methods study," 09 2021.

[10] D. DeSteno, C. Breazeal, R. H. Frank, D. Pizarro, J. Baumann, L. Dickens, and J. J. Lee, "Detecting the trustworthiness of novel partners in economic exchange," *Psychological Science*, vol. 23, no. 12, pp. 1549–1556, 2012, pMID: 23129062. [Online]. Available: https://doi.org/10.1177/0956797612448793

[11] F. Eyssel, F. Hegel, G. Horstmann, and C. Wagner, "Anthropomorphic inferences from emotional nonverbal cues: A case study," 10 2010, pp. 646 – 651.

[12] R. L. Nabi and M. C. Green, "The role of a narrative's emotional flow in promoting persuasive outcomes," *Media Psychology*, vol. 18, no. 2, pp. 137–162, 2015. [Online]. Available: https://doi.org/10.1080/15213269.2014.912585

[13] J. E. Escalas, "Imagine yourself in the product : Mental simulation, narrative transportation, and persuasion," *Journal of Advertising*, vol. 33, no. 2, pp. 37–48, 2004. [Online]. Available: https://doi.org/10.1080/00913367.2004.10639163

[14] G. Guido, M. Pichierri, and G. Pino, "Place the good after the bad: effects of emotional shifts on consumer memory," *Marketing Letters*, vol. 29, 03 2018.

[15] B. J. Phillips and E. F. McQuarrie, "Narrative and persuasion in fashion advertising," *Journal of Consumer Research*, vol. 37, no. 3, pp. 368–392, 2010. [Online]. Available: http://www.jstor.org/stable/10.1086/653087

[16] A. Augello and G. Pilato, "An annotated corpus of stories and gestures for a robotic storyteller," in *2019 Third IEEE International Conference on Robotic Computing (IRC)*, 2019, pp. 630–635.

[17] J. Ham, R. Bokhorst, R. Cuijpers, D. van der Pol, and J.-J. Cabibihan, "Making robots persuasive: The influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power," in *Social Robotics*, B. Mutlu, C. Bartneck, J. Ham, V. Evers, and T. Kanda, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 71–83.

[18] M. Appel, B. Lugrin, M. Kühle, and C. Heindl, "The emotional robotic storyteller: On the influence of affect congruency on narrative transportation, robot perception, and persuasion," *Computers in Human Behavior*, vol. 120, p. 106749, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563221000716

[19] D. Jackson and A. Latham, "Talk to the ghost: The storybox methodology for faster development of storytelling chatbots," *Expert Systems with Applications*, vol. 190, p. 116223, 2022.

[20] "Charisma.ai : Power conversational characters with charisma," https://charisma.ai/, 2022, [Online; accessed 21-January-2022].

[21] S. Ruan, A. Willis, Q. Xu, G. M. Davis, L. Jiang, E. Brunskill, and J. A. Landay, "Bookbuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot," in *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*, 2019.

[22] D. M. Olson, N. Soliman, A. Wang, M. Price, R. Sahu, and D. F. Harrell, "Breakbeat narratives: A personalized, conversational interactive storytelling system for museum education," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, p. 1–8.

[23] A. Graesser, P. Chipman, B. Haynes, and A. Olney, "Autotutor: an intelligent tutoring system with mixed-initiative dialogue," *IEEE Transactions on Education*, vol. 48, no. 4, pp. 612–618, 2005.

[24] B. Inkster, S. Sarda, and V. Subramanian, "An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study," *JMIR mHealth and uHealth*, vol. 6, 2018.

[25] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial," *JMIR Ment Health*, vol. 4, no. 2, p. e19, Jun 2017.

[26] K. Wegba, A. Lu, Y. Li, and W. Wang, "Interactive storytelling for movie recommendation through latent semantic analysis," in *23rd International Conference on Intelligent User Interfaces*, 2018, p. 521–533.

[27] G. Carenini, J. Smith, and D. Poole, "Towards more conversational and collaborative recommender systems," in *Proceedings of the 8th International Conference on Intelligent User Interfaces*, 2003, p. 12–18.

[28] S. Li, W. Lei, Q. Wu, X. He, P. Jiang, and T.-S. Chua, "Seamlessly unifying attributes and items: Conversational recommendation for cold-start users," *ACM Trans. Inf. Syst.*, vol. 39, no. 4, aug 2021.

[29] K. Christakopoulou, F. Radlinski, and K. Hofmann, "Towards conversational recommender systems," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, p. 815–824.

[30] P. Singh, N. Pisipati, P. R. Krishna, and M. V. N. K. Prasad, "Social signal processing for evaluating conversations using emotion analysis and sentiment detection," in *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, 2019, pp. 1–5.

[31] K.-Y. Huang, C.-H. Wu, Q.-B. Hong, M.-H. Su, and Y.-R. Zeng, "Speech emotion recognition using convolutional neural network with audio word-based embedding," in *2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 2018, pp. 265–269.

[32] S.-L. Yeh, Y.-S. Lin, and C.-C. Lee, "A dialogical emotion decoder for speech emotion recognition in spoken dialog," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 6479–6483.

[33] Y. Zhang, D. Song, X. Li, P. Zhang, P. Wang, L. Rong, G. Yu, and B. Wang, "A quantum-like multimodal network framework for modeling interaction dynamics in multiparty conversational sentiment analysis," *Information Fusion*, vol. 62, pp. 14–31, 2020.

[34] S. Wang, L. Hao, and Q. Ji, "Knowledge-augmented multimodal deep regression bayesian networks for emotion video tagging," *IEEE Transactions on Multimedia*, vol. 22, no. 4, pp. 1084–1097, 2020.

[35] S. Xing, S. Mai, and H. Hu, "Adapted dynamic memory network for emotion recognition in conversation," *IEEE Transactions on Affective Computing*, no. 01, pp. 1–1, 2020.

[36] D. Hazarika, S. Poria, A. Zadeh, E. Cambria, L.-P. Morency, and R. Zimmermann, "Conversational memory network for emotion recognition in dyadic dialogue videos," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Jun. 2018, pp. 2122–2132.

[37] H. Lai, H. Chen, and S. Wu, "Different contextual window sizes based rnns for multimodal emotion detection in interactive conversations," *IEEE Access*, vol. 8, pp. 119 516–119 526, 2020.

[38] H. Barber and D. Kudenko, "Generation of dilemma-based interactive narratives with a changeable story goal," in *Proceedings of the 2nd International Conference on INtelligent TEchnologies for Interactive EnterTAINment*, 2008.

[39] M. El-Nasr, "Interaction, narrative, and drama: Creating an adaptive interactive narrative using performance arts theories," *Interaction Studies*, vol. 8, pp. 209–240, 06 2007.

[40] D. Thue, V. Bulitko, M. L. Spetch, and E. Wasylishen, "Interactive storytelling: A player modelling approach," in *AIIDE*, 2007.

[41] A. Ramirez and V. Bulitko, "Automated planning and player modeling for interactive storytelling," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 4, pp. 375–386, 2015.

[42] L. R. Goldberg, "An alternative" description of personality": the big-five factor structure." *Journal of personality and social psychology*, vol. 59, no. 6, p. 1216, 1990.

[43] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granström, "Furhat: a back-projected human-like robot head for multiparty human-machine interaction," in *Cognitive behavioural systems*. Springer, 2012, pp. 114–130.

[44] C. L. Sidner, C. D. Kidd, C. Lee, and N. Lesh, "Where to look: a study of human-robot engagement," in *Proceedings of the 9th international conference on Intelligent user interfaces*, 2004, pp. 78–84.

[45] T. Monahan, "One page adventures," last accessed 18 January 2022. [Online]. Available: https://drive.google.com/file/d/1jvrAUnW0ijrl3uZHeh3AaxpxV5rlzqdc/view

[46] G. Bradski and A. Kaehler, "Opencv," *Dr. Dobb's journal of software tools*, vol. 3, 2000.

[47] A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing Ltd, 2017.

[48] Emotion detection using opencv and keras. Accessed 2022-01-19. [Online]. Available: https://medium.com/swlh/emotion-detection-using-opencv-and-keras-771260bbd7f7

[49] Emotion detection using deep learning. Accessed 2022-01-19. [Online]. Available: https://github.com/atulapra/Emotion-detection

[50] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. Ieee, 2001, pp. I–I.

[51] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *International conference on neural information processing*. Springer, 2013, pp. 117–124.

[52] A. Stern, "Embracing the combinatorial explosion: A brief prescription for interactive story r&d," in *Interactive Storytelling*, U. Spierling and N. Szilas, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 1–5.

[53] T. Bänziger, D. Grandjean, and K. R. Scherer, "Emotion recognition from expressions in face, voice, and body: the multimodal emotion recognition test (mert)." *Emotion*, vol. 9, no. 5, p. 691, 2009.

[54] C. M. Lee, S. S. Narayanan, and R. Pieraccini, "Combining acoustic and language information for emotion recognition." in *INTERSPEECH*. Citeseer, 2002.