Do You Follow?
A Fully Automated System for Adaptive Robot Presenters

Agnes Axelsson
agnaxe@kth.se
KTH Royal Institute of Technology
Division of Speech, Music and Hearing
Stockholm, Sweden

Gabriel Skantze
skantze@kth.se
KTH Royal Institute of Technology
Division of Speech, Music and Hearing
Stockholm, Sweden

Figure 1: Left: A photo of a participant interacting with the robot presenter. Middle: *het Vrolijke Huisgezin* by Jan Steen (1668). Right: *Glorious Entry of the Duke of Anjou into Antwerp on the 19th of February 1582* by M.H.V.H. (undated, between 1584 and 1600)

ABSTRACT

An interesting application for social robots is to act as a presenter, for example as a museum guide. In this paper, we present a fully automated system architecture for building adaptive presentations for embodied agents. The presentation is generated from a knowledge graph, which is also used to track the grounding state of information, based on multimodal feedback from the user. We introduce a novel way to use large-scale language models (GPT-3 in our case) to lexicalise arbitrary knowledge graph triples, greatly simplifying the design of this aspect of the system. We also present an evaluation where 43 participants interacted with the system. The results show that users prefer the adaptive system and consider it more human-like and flexible than a static version of the same system, but only partial results are seen in their learning of the facts presented by the robot.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; User studies; *Collaborative interaction*; • **Computing methodologies** → **Cognitive robotics**.

KEYWORDS

adaptation; multimodal; feedback; learning; behaviour tree; lexicalisation; knowledge graph

HRI '23, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9964-7/23/03.
https://doi.org/10.1145/3568162.3576958

ACM Reference Format:
Agnes Axelsson and Gabriel Skantze. 2023. Do You Follow? A Fully Automated System for Adaptive Robot Presenters. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3568162.3576958

1 INTRODUCTION

In a presentation, a presenter tries to convey some information to one or more recipients. An important task for a presenter is to adapt the presentation to the audience, a process called *audience design* [9]. This adaptation can happen in real time, as the presentation is going on [2, 17], and ahead of time, through rhetoric and planning [46]. When adapting the presentation in real time, the presenter has to continuously take in (positive or negative) feedback from the audience and adapt the presentation accordingly. While current digital tools for presentations (such as pre-recorded lectures) are mostly non-adaptive, an intelligent presentation agent could potentially process user feedback and adapt the presentation in a human-like manner. A social robot, which is physically co-located with the user(s), could further enhance the presentation in that it is easier to perceive if the robot is attending to the user(s) or objects in the environment. When social robots are used as presenters or teachers, adaptation could serve to engage the audience [10], and improve learning [15, 35].

In this paper, we present a fully autonomous system that allows social robots to present information stored in a knowledge graph and which adapts the presentation based on multimodal feedback from the user, as illustrated in Figure 2. The scenario we use is the presentation of paintings in a museum, and our experimental setup can be seen on the left of Figure 1. The museum guide scenario is interesting because it is not entirely clear if its goal should be

Speaker	Behaviour
Robot	"The commander of the French Army is called Francois-Hercule in French and Frans Hercules in Dutch."
Robot	"Did you get that?"
User	(nods) "Yes."
Robot	"The commander of the French Army, born in 1556, died in 1584 from malaria."
User	(shakes head)
Robot	"The Duke of Anjou was the commander of the French Army."

Figure 2: An example of elicitation (first turn), adaptation to positive grounding (second turn) and adaptation to negative grounding (third turn) by our adaptive system (see Section 3).

teaching or entertainment. If the goal of the system is to entertain its audience, then evaluating the system on subjective measures makes sense, and a highly engaging system could be seen as a success. If the goal is learning, then objective measures based on learning outcomes should be part of the evaluation.

This paper has two main contributions. First, we present our system architecture for creating fully automated presenter systems. While several components of the system are based on our prior work [5–7], they were never evaluated in a fully autonomous system interacting with real users. We also introduce a novel approach for turning semi-logical representations of language into natural language (*lexicalisation*) using GPT-3. Our second main contribution is an evaluation that compares the fully automated system with a non-adaptive version of the same system. We are interested in understanding how the two versions compare by the following five criteria:

(1) The perception of how competent and communicative the robot presenter is.
(2) Which presenter the audience prefers.
(3) How much the audience actually learns from the presentation (objective learning).
(4) How much the audience thinks it learns from the presentation (subjective learning).
(5) How much feedback the audience gives to the robot.

2 BACKGROUND AND RELATED WORK

2.1 Feedback and grounding

For a robot presenter to be able to reason about the reactions of its audience, there must be a theoretical model of how to represent those reactions. **Feedback** is communicative information that allows parties in a conversation to know whether information has been perceived, understood, and accepted [2]. This information can flow on a main channel or on a back channel, where information on the back channel (**backchannels**) can pass from the listener to the speaker without taking the turn away from the speaker [50].

Common ground is the information shared by parties engaged in a dialogue [20]. **Grounding** is the process by which a listener signals, through various kinds of feedback, whether something is integrated into the common ground [18]. Clark [17] defines four levels to which feedback may be related: **attention**, **hearing**, **understanding** and **acceptance**. The polarity of feedback is either **positive** or **negative** and can relate to any of these levels. An example of feedback communicating positive understanding can be a simple "Yes" or a nod, while negative hearing could be communicated with a "Sorry?" or frowning, depending on the context [4]. Feedback on these levels is subject to the rules of **upward completion**, where negative feedback on a low level implies negative feedback on all higher levels, and **downward evidence**, where positive feedback on any level implies positive feedback for all lower levels. When humans interact, feedback at higher levels (acceptance and understanding) is not required at every point; the level at which feedback must currently be given is called the **grounding criterion** [17]. This level is continuously agreed upon by the communicating parties and depends on factors like how recently feedback has been given and how important the currently presented information is [17]. The process of updating the grounding criterion can be modelled as a Bayesian process of considering past feedback together with the present feedback [13].

If a speaker does not receive enough feedback from the listener, the speaker can choose to **elicit** feedback, either by direct questions ("Do you follow?"), by gazing at the listener, or through prosodic cues such as rising pitch. This can give rise to a temporary increase in feedback given by the listener [49]. As dialogue partners become used to interacting with each other, the amount of multimodal feedback signals can decrease as the partners optimise their feedback behaviour to provide the signals they expect the partner to be able to sense, a process referred to as **alignment** [17, 21]. Alignment also extends to the modalities and types of signals used: for example, gaze feedback is more common towards an agent that can also produce gaze signals, as compared to smart speakers [36].

Information flow between dialogue partners is more effective when more modalities can be used; for example, it is easier to present information when two dialogue partners can see each other than when they are talking over the phone [11]. In a model-building experiment, it was shown that task performance increases as the opportunities for the partners to give multimodal feedback to each other increase [19].

2.2 Robots as presenters

An early example of a robot presenter that could adapt the timing of its presentation to its audience (depending on for example laughter volume) is the Japanese *manzai* comedy duo system by Hayashi et al. [29]. When compared to a prerecorded video of human comedians, the in-person robot routine was rated higher on measures of presence, but also overall preference and duration of laughter [29]. The robot comedy scenario was also used by Katevas et al. [33], who found that robot gestures timed with joke delivery and aiming gaze at specific individuals in the audience could improve audience happiness in the short term, similarly to methods used by a human stand-up comedian.

Axelsson and Skantze [6] presented a scenario where the Furhat robot, partially controlled by a Wizard of Oz, presented two paintings, and one presentation adapted to the user's feedback while the other did not. Even though they were not told which presentation

was adaptive, users preferred the adaptive mode over the static mode on multiple measures of the Godspeed [8] scale. Masuta et al. [39] set up a scenario where a social robot presented a university lecture. A proportion of the audience was equipped with laptops running a program that let them give feedback such as "interesting" or "boring" to the ongoing presentation. The group of students with access to the feedback program were more likely to respond positively to the question "Do you feel that your response is reflected in [robot]'s lecture?" than those who did not have access to the program.

In a museum setting, Shiomi et al. [43] compared three modes of robot agents deployed in a museum; a mode where the robot would take RFID tags carried by visitors into account to know which objects they were interacting with, and two modes without this interaction. The authors found that the RFID condition was rated more highly in the category of "Experience of science & technology" than the other two modes. Outside of social robotics, Farmer et al. [28] showed that students preferred a learning platform that responded to answers that students got wrong over one that presented static content.

Competence is a dimension of how users perceive conversing agents [24]. The perceived competence of a robotic agent is a component of the commonly-used Godspeed scale [8], and is thus measured in many experiments involving social robots. Factors of agent design and agent behaviour that play into increased perceived competence include anthropomorphism [16] and whether the agent is perceived as having emotions [44]. The context and place in which the agent is found is also a factor in its perceived competence - a teacher robot in a school is perceived differently from a teacher robot in a summer camp setting [41].

2.3 Robots and learning

Subjective learning refers to how much a student or participant in a learning scenario thinks they have learned. Stark et al. [45] found that there was not necessarily a correlation between subjective and objective learning in an economic scenario; students who had studied a scenario independently over-estimated their performance on a task while students who had been guided through their teaching under-estimated their performance on the same task. Koriat et al. [37] found that the subjective learning and objective learning of participants in a word-memorisation task diverged as more tests were administered, with lower subjective learning compared to objective results in later tests. Other factors that can play into subjective learning, or *judgment of learning*, are the time the student spent thinking about their answer [25], as well as previous task performance and whether students know about their previous task performance [3].

Evaluating **objective learning** outcomes from social robots functioning as teachers or tutors appears to be difficult. Robots often increase student engagement without any measurable increase in performance [22, 34]. Additionally, school scenarios require long-term learning results, which require long-term experiments of a type that are challenging to set up. Even if a robotic system with certain social behaviours promotes learning in the short term, it does not necessarily follow that those learning results persist weeks later [15]. In a language café scenario, where the goal is to facilitate learning by getting participants to converse in a language they are learning (i.e. the primary goal is engagement), Engwall et al. [27] found that different factors of participant backgrounds (cultural, gender, proficiency in the language, among others) affected which style of robot conversation partner they preferred, and concluded that adaptation would be required.

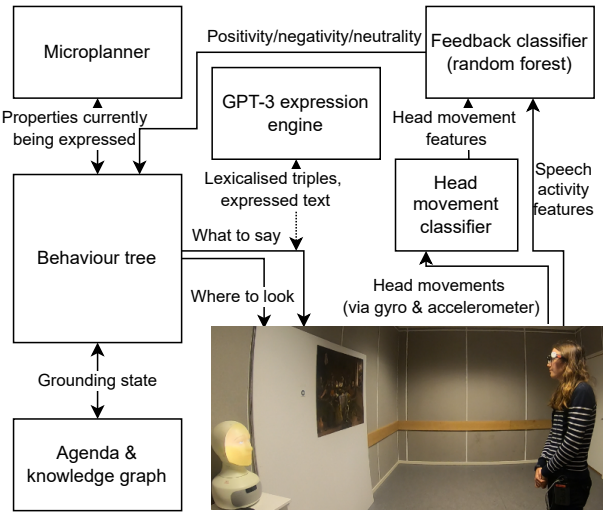

Figure 3: The components of our presenter system and how they connect.

A positive effect on student learning was found by Ishino et al. [32], who presented a comparison between a video recording of a human presenting a university lecture, a robot presenting the same lecture while reproducing body language generated by a model, and a robot presenting the same lecture while replaying the body language of the recorded human. The robot that created body language using a model achieved higher learning results on a post-lecture test than either of the others. No difference in learning results was found between replaying the human presenter's body language and replaying a video of the human.

Konijn and Hoorn [35] showed that social behaviours in a Nao robot could help students being taught multiplication tables. However, increasing the amount of social behaviours in the robot mostly helped students who the authors categorise as "above-average", while "below-average" students seemingly benefit from more neutral, less social behaviour. Konijn and Hoorn note that there is no baseline without a robot, making it hard to conclude that children learned more than they would have by practicing the relevant multiplication tables on their own or with a human teacher.

3 SYSTEM DESCRIPTION

A high-level overview of the system is presented in Figure 3. Central to managing the interaction is a **behaviour tree**, which runs at a frequency of 10 Hz. The tree structure, adapted from [6, 7] coordinates reading data from other modules and executes actions to make the robot gaze at the right target and speak at the right moments. The presentation itself is stored in a **knowledge graph** (see Section 3.1). As execution passes down the behaviour tree, the

system chooses one or more properties from the knowledge graph to present, and puts the logical representation of how it intends to present the properties into a **microplanner**. It then converts the logical representation of its planned utterance into text using **GPT-3** (see Section 3.3). This text is synthesized and spoken. The behavior tree then makes the system wait for the user's feedback, classifying it with the **feedback classifier** (detailed in Section 3.2), elicits feedback if it is not obtained in time, and then updates the **grounding state** based on the feedback which was given. This process is repeated until no more properties exist to present or time has run out.

3.1 Presenting through knowledge graphs

Knowledge graphs, originally proposed by Minsky [40], are a way to represent structured information in a way that can be easily parsed, polled and stored by computers. The edges between nodes in a knowledge graph are called **properties**, and the combination of a property, its source and its target are a **triple** [30].

To track the grounding state, each property in the knowledge graph is labelled with whether it has been positively or negatively attended to, heard, understood and accepted by each user, an approach we previously presented in [7]. This is then used by the system to select which new properties to present. For a property to be be presented, it must be possible to create a referring expression to its source. In Figure 4, the system is able to present three properties about the Duke of Anjou because it can refer to him as the commander of the French army. This referring expression is possible because the property that synthesises it, *commander-of* from *dukeOfAnjou* to *frenchArmy*, is grounded with positive understanding in the dialogue state, a consequence of an earlier exchange where the system presented that this property held and the user reacted positively. As the user reacts with positive or negative grounding, properties in the graph become grounded or ungrounded, giving the system the ability to refer to new facts about previously grounded entities. Negative feedback can also take away the system's ability to use a referring expression; after *commander-of* has been ungrounded in Figure 4, the system needs to use another referring expression if it wishes to talk about the Duke of Anjou. This is illustrated by the third robot line in Figure 2.

If the user has not provided positive or negative feedback matching the grounding criterion by the time the robot takes the turn back after it has spoken, the behaviour tree will make the agent elicit feedback from the user. Attention is elicited by looking at the user, and understanding is elicited by looking at the user and saying *Right?*, *Did you get that?*, *Did that make sense?* or *Did you understand that?* If this also fails to elicit behaviour that the feedback classifier can classify as either positive or negative before the system takes the turn again, the user's feedback is classified as negative by default and the system moves on to another set of properties to present.

3.2 Classifying and eliciting feedback

The feedback classifier is based on previous work [5], where we collected data on robot-human presentations (similar to the setup used here) and trained a classifier to classify feedback as positive, negative or neutral. Various classifiers and combinations of input modalities (head movements, speech, facial expressions, body pose and gaze) were tested, and a random forest classifier focusing on just head, speech and gaze yielded nearly optimal results. However, whereas the classifier from [5] was trained on manually annotated features, we re-implemented it using features that can be automatically extracted in real time. To classify head movements (as single or multiple nods, head shakes, and head tilts), we attached a *MetaMotion S*[1] accelerometer and gyro sensor to the headset worn by our participants, and trained a specific head gesture classifier using a random forest classifier on head movement data we collected in a pre-study. The random forest feedback classifier is then polled for its classification (positive, negative or neutral) through two dedicated leaves in the behaviour tree that run when the user or the robot has the turn, respectively.

A limitation of the feedback classifier we presented in [5] is that it is not capable of separating understanding, hearing or acceptance by the definitions of Clark [17]. Users may expect different types of adaptation depending on whether their feedback constitutes negative hearing or understanding [7]. As a simplifying solution, we treat all feedback identified by our classifier as positive or negative signs of understanding. Upon receiving an indication of positive or negative understanding, our dialogue system adapts in multiple ways, simultaneously addressing hearing and understanding. Speech rate is lowered (by increments of 10%) down to a minimum of 50% upon negative understanding, and raised to a maximum of 100% upon positive understanding[2]. The number of knowledge graph properties the system tries to present at a time is lowered (for negative feedback) or raised (for positive feedback) by one. Additionally, the properties to which the user is reacting are weighted so that they are less (negative feedback) or more (positive feedback) likely to be used in future referring expressions.

3.3 Natural language generation using GPT-3

Lexicalisation is the process of turning structured data into natural language [42]. Here, the structured data is the knowledge graph. Lexicalisation can use neural approaches similar to machine translation [26, 38], or more or less advanced rule-based approaches [7, 31, 42, 48].

All properties in our knowledge graph format are marked with a human-readable label like *has-name* or *father-of*. Optionally, additional labels can be added for use in referring expressions when traversed forwards or backwards (referring either to the source or the target of the property). This gives us a way to uniquely refer to every possible triple in the graph, together with referring expressions to the source and target.

As illustrated in Figure 4, the knowledge graph statement that the Duke of Anjou was born in 1556 is written out as *(commander-of(frenchArmy), birth-year, 1556)*. When the behaviour tree decides to present a new line, the knowledge graph that represents our presentation is queried for ungrounded properties for which such a representation can be created, and for which a reference to the

[1] https://mbientlab.com/store/metamotions/

[2] The speech rate numbers were arrived at by testing for the specific *Amazon Polly* voice we used, and do not necessarily generalise to other speech synthesis or voices.

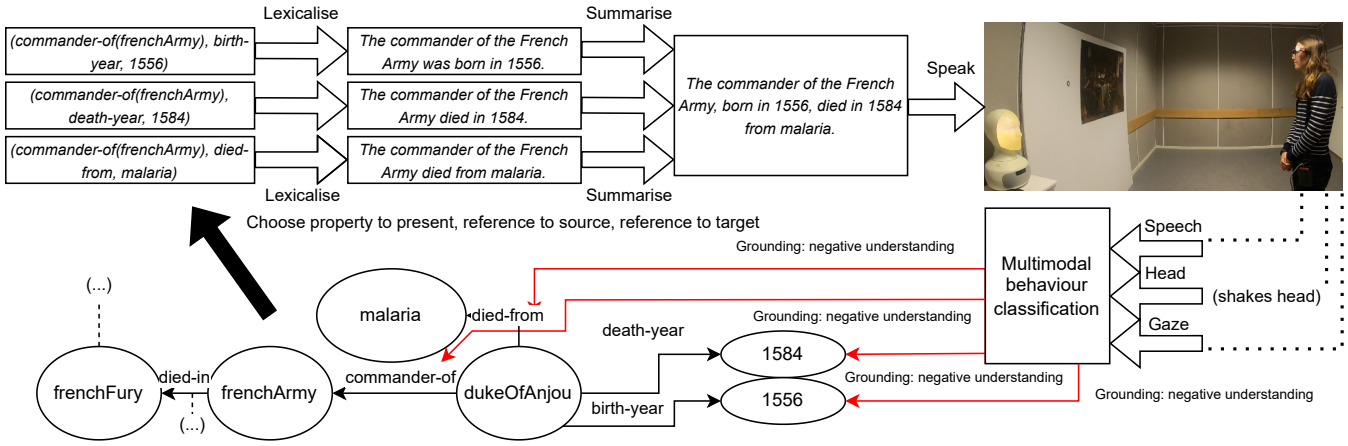

Figure 4: An illustration of how we use GPT-3 to convert knowledge graph triples into a paragraph of text. The example matches the second turn from Figure 2. On the bottom, an excerpt from our knowledge graph on the *Glorious Entry* painting is shown. On the right of the figure, the reaction to the participant's response to the statement is illustrated. The head shake is classified as negative understanding by our feedback classifier, and that negative grounding is mapped back to the knowledge graph properties that were used to synthesise the utterance, in order to update their grounding status. The process for how properties to present are chosen is described in Section 3.1.

source entity can be created based on currently grounded properties. All such properties are sorted by how similar they are to other things that have been successfully presented to the user and positively grounded, and how on-topic they are. The sorting for topic prioritises properties with the same source as things that have been recently presented. One or more of the most suitable properties are taken from this ranking, based on the number of properties the system currently presents at a time, given recent positive or negative feedback. To avoid the system presenting non-sequiturs, presentation of multiple properties at a time is limited to properties that share the same source. For example, all properties that are co-presented in Figure 2 are sourced in the Duke of Anjou.

Each chosen knowledge graph property is passed to GPT-3[12][3] to turn it into text using a prompt. We generate a semi-lexical representation of the chosen statements based on our human-readable labels. Examples of this representation can be seen in the top left of Figure 4. In the prompt passed to GPT-3, we give examples of how this representation can be lexicalised (see Appendix B.1). This turns *commander-of(frenchArmy), birth-year, 1556* into *The commander of the French Army was born in 1556*, as illustrated in Figure 4.

The lexicalisation step generally results in sentences that would make sense to present on their own, and if the system is only presenting one knowledge graph property, the lexicalisation is spoken as-is. When multiple properties are presented at the same time, saying the lexicalised sentences one after the other often leads to redundancies where the same referring expressions are used multiple times. To compress the line the system is about to say, and remove these redundancies, we again use GPT-3 to shorten the lexicalised text. This process, which corresponds to what Tiddi et al. [48] calls *condensing*, is done via a different GPT-3 prompt (see Appendix B.2) that consists of several examples of redundant sentences being written up and summarised into a single sentence. An example of real GPT-3 output resulting from this procedure can be seen in the top-middle of Figure 4.

[3]https://openai.com/api/

4 EVALUATION

To evaluate the presenter system described in Section 3, we designed a within-subject experiment, where the participants were asked to take part in two presentations: one by the adaptive system and one by a non-adaptive version of the system.

4.1 Participants

We recruited participants through self-signup by posters and emailing lists, and we thus did not have a sampling approach to balance for ethnicity or gender. Participants were offered a 250 SEK gift card as compensation for participating. In total, 46 participants were recruited. Three participants had to be ignored - the first participant filled in an incorrect version of our subjective scales, and the sixth and ninth participant did not get the full adaptive presentation because of a system malfunction. We thus ended up with 43 usable participants. Of these, 17 participants self-reported as male, and 26 female. The average age was 25.5 years old (SD = 4.11). All participants signed an information sheet informing them of the right to retract their consent to us storing their personally identifiable data. Participants also optionally agreed to having their pictures included in publications and presentations. All data was collected and stored in accordance with local policy and laws.

4.2 Material and Procedure

Our robotics platform is a Furhat [1] robot. The robot was mounted on a pedestal (see Figure 1) 145 cm to the right of the middle of a 1-metre-wide printout of a painting. The painting was attached to a cardboard sheet to make it possible to switch it out between conditions (see Section 4.3). The user was allowed to stand wherever they felt was natural and comfortable as long as they stayed

within a 45-degree cone in front of the robot, marked with tape on the ground; this was to avoid the robot's camera losing track of the participant. Before each presentation, participants calibrated Tobii gaze-tracking glasses[4]. They were then informed that the system would or would not (depending on the condition) change the presentation in response to their feedback. They were told that they were allowed to react to the static system, but that they could think of it like a human presenter who has to read from a strict script from which they cannot deviate. Before interacting with the adaptive system, we gave them examples of the types of feedback the system would expect, comparing it to the feedback users were giving to the presenter giving them the instructions ("Nodding, saying 'm-hm', like you're doing to me right now"). Recording was then started on both the Tobii gaze-tracking glasses and a wall-mounted GoPro camera, and the researcher left the room to start the system from outside. The system then presented either until it ran out of knowledge graph triples to choose from, or until 4 minutes had passed.

After each painting had been presented, users answered a questionnaire first asking for five answers on multiple-choice fact-based questions about the contents of the presentation. The multiple-choice questions and their answers are listed in Appendix A. The first three questions concerned general information about the painting (the name of the painting and artist), and the final two were about details specific to each painting. The order of the multi-choice answers was shuffled for each user, but the questions themselves were always in the same order.

Following the multiple-choice questions, users were asked to rank the presenter agent on the 18-adjective pair **Partner Modelling Questionnaire (PMQ)** scale [23, 24]. PMQ is a scale where experiment participants rate (from 1 to 7) the agent they interacted with through a number of pairs of adjectives. The pairs of adjectives serve to explain the dimensions of the participant's **partner model** of the system with which they interacted [24]. In the 18-adjective scale [23], 9 adjectives are linked to the dimension of *communicative competence and dependability*, six are linked to *human-likeness* and the remaining three to *communicative flexibility*. The rank in each of these three measures is the average score on all of the adjective pairs. In order for the PMQ scale to estimate the user's partner model, a large amount of adjective pairs were filtered down to 18 [23] to maximise correlation between the adjective pair and one of the three dimensions, and to minimise correlation with the others, while also minimising correlation between different adjective pairs [24]. We chose to use the PMQ scale over the Godspeed [8] scales, more commonly used in HRI settings, as the Godspeed scales evaluate many aspects of the agent that would not be different between our two conditions, such as mechanical smoothness.

After the multi-choice questions and PMQ scales had been filled in for the second presentation, users were finally asked which of the two presenters they preferred (on a 1-7 scale, with one presenter on either end), as well as how confident they felt about their answers to the multiple choice questions on the first and the second presentation, respectively (both 1-7 scales, with "very bad" on one end and "very good" on the other). After each participant finished their session, we gave them their gift card, and debriefed them by asking if they had any comments or questions about the experiment. Some of the comments given by users in the debriefing are discussed in Section 6.

4.3 Experimental conditions

To evaluate the effects of adaptation in the presentation, we compared an **adaptive** mode where the system would use the full architecture described in Section 3 to take the user's feedback into consideration, against a **static** mode where the system would always perceive the user's feedback as positive (even if they did not provide any feedback at all). Additionally, the static mode was set to always present three knowledge graph properties at most. The choice of three knowledge graph properties for the static mode came from pilot testing and does not necessarily extend to other similar systems. The choice depends on how information-dense the knowledge graph is, how many words are needed on average to present a property in the graph, and how verbose the designers want the system to be.

Two paintings, *Glorious Entry of the Duke of Anjou into Antwerp on the 19th of February of 1582* by an unknown artist known only by the signature *M.H.V.H*, and Jan Steen's *Happy Family*, were chosen as objects for the system to present. The paintings can be seen in Figure 1. To balance the order of the paintings and system modes (adaptive or static), four conditions were set up, and we balanced our participants across them. 11 participants were assigned to each condition except *adaptive Glorious Entry first*, which had 10 participants.

Both paintings have WikiData[5] articles, and the knowledge graph data from these pages that was relevant to our scenario was used as part of our presentation graphs. However, since the data for our specific paintings was quite sparse, we manually complemented the graphs with additional information. Because the motifs of the two paintings were inherently different, the presentation about *Glorious Entry* focused more on the historical context of the painting, while the presentation graph about *Happy Family* focused more on the contents of the painting and details about the painter.

5 RESULTS

5.1 PMQ scale and preference

Given that the PMQ scale should be interpreted as ordinal, we fitted a Cumulative Link Mixed Model (CLMM) for each of the three measures of the PMQ scale: *communicative competence and dependence* (based on nine of the adjective pairs), *human-likeness* (based on six separate adjective pairs) and *communicative flexibility* (based on the remaining three) [23]. The identity of the participant was treated as a random factor. The order of the presentation (first or second), the mode of the system (adaptive or static), and the painting being presented (*Happy Family* or *Glorious Entry*) were treated as fixed factors. For communicative competence and dependence, no factors had a statistically significant effect. For human-likeness, only the mode of the system was significant ($\chi^2(1) = 4.75$, $p = .0292$), such that participants rated the adaptive system ($M = 2.94$, SD = 1.34) as more human-like than the static system ($M = 2.47$, SD = 1.27). For

[4]The Tobii gaze data was not used for the analysis presented in this paper.

[5]https://www.wikidata.org

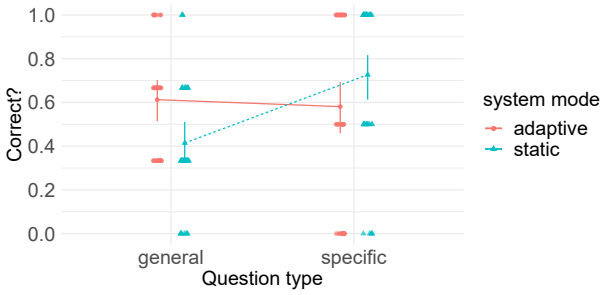

Figure 5: The probability of correct answer (Y axis) depending on whether the question was one of the first three (*general*) or one of the final two (*specific*), presented as the X axis.

communicative flexibility, both the mode ($\chi^2(1) = 7.60, p = .00583$) and order ($\chi^2(1) = 7.174, p = .00740$) were significant; participants considered the adaptive system ($M = 2.86$, SD = 1.22) more flexible than the static system ($M = 2.35$, SD = 1.20), but also considered the second system they interacted with ($M = 2.88$, SD = 1.26) to be more flexible than the first system ($M = 2.33$, SD = 1.14).

60% of the users preferred the adaptive over the static mode, whereas 37% preferred the static mode, and one participant ranked both equally. To further investigate this (taking the degree of preference into account), an equivalent CLMM model to the one above was fitted for which presenter our participants preferred, with the same random and fixed factors as the models fitted for the PMQ scale. The order was significant ($\chi^2(1) = 11.7, p = .000609$) such that participants preferred the second presenter over the first ($M = 4.77$, SD = 2.18, where 1 is the first and 7 is the second). Which painting was being presented was also a statistically significant factor ($\chi^2(1) = 8.89, p = .00286$), such that participants preferred the presenter that presented *Happy Family* ($M = 4.62$, SD = 2.22, where 7 is *Happy Family* and 1 is *Glorious Entry*). The mode of the system was also a statistically significant factor ($\chi^2(1) = 4.29$, $p = .0383$), such that participants preferred the adaptive mode ($M = 4.44$, SD = 2.27, where 7 is the adaptive mode and 1 is the static mode). The interaction between all three factors also held strong significance ($\chi^2(1) = 21.5, p < .0001$).

5.2 Subjective and objective learning outcomes

A CLMM model with the same fixed and random factors as those described in Section 5.1 was fitted for participants' confidence in their answers. The painting being presented was a significant factor ($\chi^2(1) = 17.2, p < .0005$) such that participants had higher confidence for *Happy Family* ($M = 4.37$, SD = 1.45) than for *Glorious Entry* ($M = 3.26$, SD = 1.43). The order was strongly significant ($\chi^2(1) = 14.3, p < .0005$) such that participants were more confident for the second presentation ($M = 4.33$, SD = 1.46) than for the first ($M = 3.30$, SD = 1.46). Whether the mode of the system was adaptive ($M = 4.07$, SD = 1.52) or static ($M = 3.56$, SD = 1.53) was not a statistically significant factor ($\chi^2(1) = 3.53, p = .0603$).

To evaluate the participants' performance on the multi-choice questions, we fitted a binomial Generalized Linear Mixed Model (GLMM) with one data point for each question answered by each participant. The user's identity was used as a random factor, and whether they got the question right was used as the dependent variable. As fixed factors we used (a) whether the presentation was the first or the second, (b) whether the presentation was in the adaptive or static mode, (c) which painting the presentation was about, and (d) whether the question was one of the first three (general or specific information).

The painting ($\chi^2(1) = 12.543, p < .0005$), question type (first three or final two) ($\chi^2(1) = 7.61, p = .00582$), and the interaction between the mode of the system (adaptive/static) and the question type ($\chi^2(1) = 11.267, p < .001$) were significant factors. The significance of the painting is in line with the participants' confidence in their answers and actual performance on the multi-choice questions. To investigate the interaction effect between the system mode and the question type, we plot the outcome for these variables in Figure 5. While the adaptive system has better learning outcomes for the (first three) general questions, there is an opposite effect for the (last two) specific questions.

5.3 Difference in feedback behaviour

To compare the distribution of positive and negative feedback signals between the adaptive and static modes, we extracted the output of the feedback classifier detailed in Section 3.2 at each 10 Hz time step of the presentations. We then compared the proportion of positive and negative signals, respectively, between the adaptive and static modes, for each participant, using a paired t-test. There was no difference ($t(42) = .605, p = .548$) in the proportion of positive feedback for the adaptive mode ($M = .129$, SD = .0648) compared to the static mode ($M = .123$, SD = .0746). There was also no difference ($t(42) = 0.250, p = .804$) in the proportion of negative feedback for adaptive mode ($M = .0598$, SD = .108) compared to the static mode ($M = .0582$, SD = .0953).

6 DISCUSSION

We will now return to the five questions posed in Section 1 and discuss them based on our results.

Did adaptivity affect the participants' perception of competence and communicative ability? Our results show that the adaptive presenter was seen as more human-like and flexible than the static presenter, but not more competent. One explanation for this is that although the static presentation did not take the audience reactions into account, such presentations can still be seen as competent. It is still possible to adapt a static presentation to an imagined audience, a form of audience design Bell [9] refers to as *referee design*. The fact that differences were found for human-likeness and communicative flexibility indicates that users indeed felt that the adaptive system was attending to their feedback more than the static system.

The static mode still performed relatively well on the measures of the PMQ scale. One explanation of this might be that users interpreted the absence of elicitations from the system as establishing a low grounding criterion by the definition of Clark [17]. Similar results were seen previously by Chai et al. [14]. Social factors relating to the peculiarities of the museum guide scenario may also apply here. If a user treats the robot as mostly an entertainer and less of a teacher, there is little incentive to interrupt the robot and explain that the user has not understood something that the system assumed they did understand. For future research, it could

be interesting to implement presenter systems that can move between teaching and entertaining modes, and to identify from the behaviour of users which of the two they expect and prefer.

Did the participants prefer the adaptive version? When asked to indicate which version the participants preferred (and to what degree), there was a significant preference for the adaptive version. This matches earlier results for social robotics [6, 29, 39] and for teaching in general [28]. However, there were also individual differences and 37% of the participants actually preferred the static version. In the post-experiment debriefing, several participants expressed a feeling of being stressed by the adaptive mode, and one participant specifically said that the robot's feedback elicitations made him actively forget the last line presented by the robot. To address participants for whom elicitations and adaptivity detract from the experience of the system and objective learning, a possible future direction is to create systems that sense the user's preferred level of adaptation and adapt to that, mirroring suggestions by Engwall et al. [27]. The adaptive system elicited feedback between 0 and 4 times, depending on the participant ($M = 1.84$, SD = 1.34).

Did adaptivity affect objective learning? Participants were more likely to answer general questions about the painting correctly in the adaptive mode than in the static mode, whereas the opposite effect existed for specific questions, as seen in Figure 5. One possible interpretation of this result is that the adaptive system performed better on the general questions because it would restate basic information about the painting as part of its referring expressions to the sources of more complicated knowledge graph properties. Another possible reason is that the adaptive system presented more properties at the same time, making it hard for users to focus in on one thing the system said of many. We can conclude that the adaptivity of the system does in fact affect users' learning, but that to properly explore this effect, an experiment would have to be designed where the differences between objects being presented is minimised, and ideally where there are more multi-choice questions. Going back to the difference between a museum guide as an entertainer and a museum guide as a teacher, mentioned in Section 1, results by Peters et al. [41] have shown that the environment and setting where a robot is found affects its perceived competence. The fact that our experiment was not set up in a real museum, like earlier work by Thrun et al. [47] and Yousuf et al. [51], could play into how users approached the interaction. In future studies, it would be interesting to evaluate the system in an actual museum, to help participants approach the robot like they would approach a human museum guide, whether that is for entertainment or for learning.

Did adaptivity affect subjective learning? The audience did not significantly believe that their answers were better depending on whether the system was adaptive or static. Instead, as described in Section 5.2, there were two strong effects. Users thought they learned more from the presentation of the *Happy Family* presentation than from the *Glorious Entry* presentation. The second effect was that users thought their answers were better on the second presentation than on the first, which actually did not match the objective learning outcomes. While this effect appears to go against earlier results by Koriat et al. [37], the fact that the questions on the two presentations were slightly related in topic should affect our results – participants would be prepared in the second presentation that there would likely be general questions about the painting's title in Dutch and English, as well as the painter's name.

Did adaptivity affect the amount of positive or negative feedback given by participants? We were expecting participants to provide significantly more positive or negative feedback in response to the adaptive system than towards the static system, since we told them before each presentation started that only one of the two modes would change the presentation in response to the feedback. Despite this, no difference was found between feedback given in the two modes. It does not seem like perceived adaptivity is necessarily a factor in what makes the audience provide multimodal feedback towards a robot presenter. Kontogiorgos et al. [36] have shown that the multimodal feedback signals used by people conversing with a dialogue system partially match the capabilities of the robotic agent with which they interact. Since our agent used the same multimodal behaviours in the static mode as in the adaptive mode (gazing back and forth between the painting and the user, as well as raising the eyebrows and smiling when sensing user speech), this could have served as an implicit elicitation of feedback, even if it was disregarded by the system.

7 CONCLUSIONS

We have presented a system architecture that allows a robot to give adaptive presentations to an audience. The presentation is based on a knowledge graph, which stores the information that is to be presented, but which is also used to track the grounding status of the presented information. The user's multimodal feedback (speech, gaze and head movements) is classified as positive, negative or neutral, and used to update the grounding status in the knowledge graph, and thus affects the way the presentation proceeds. We introduced a novel way to use large-scale language models (GPT-3 in our case) to lexicalise arbitrary knowledge graph triples, greatly simplifying the design of this aspect of the system. We presented an evaluation of the system, comparing it with a static version that does not consider the user's feedback. The results showed that users generally preferred the adaptive system and perceived it as more human-like and flexible. A retention test on some of the facts presented showed that the adaptive version was better for some of the test questions, but not all. We think it is promising to see that these results can be achieved with a fully automated system, but that future tests should be done in more realistic environments and that the system should better detect to what extent users want the system to be adaptive or not.

ACKNOWLEDGMENTS

We would like to thank Fatemeh Sadat Samareh Haheshi for assisting in running our experiment, and the reviewers and area chairs for their valuable comments on the paper. This work was supported by the project *Social robots accelerating the transition to sustainable transport* (50276-1), financed by Furhat Robotics & Swedish Energy Agency.

REFERENCES

[1] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: A Back-Projected Human-Like Robot Head for Multiparty Human-Machine Interaction. In *Cognitive Behavioural Systems*, Anna Esposito, Antonietta M. Esposito, Alessandro Vinciarelli, Rüdiger Hoffmann, and Vincent C. Müller (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 114–130.

[2] Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. 1992. On the Semantics and Pragmatics of Linguistic Feedback. *Journal of Semantics* 9, 1 (01 1992), 1–26. https://doi.org/10.1093/jos/9.1.1 arXiv:https://academic.oup.com/jos/article-pdf/9/1/1/9836977/1.pdf

[3] Robert Ariel and John Dunlosky. 2011. The sensitivity of judgment-of-learning resolution to past test performance, new learning, and forgetting. *Memory & Cognition* 39, 1 (1 Jan 2011), 171–184. https://doi.org/10.3758/s13421-010-0002-y

[4] Agnes Axelsson, Hendrik Buschmeier, and Gabriel Skantze. 2022. Modeling Feedback in Interaction With Conversational Agents—A Review. *Frontiers in Computer Science* 4 (2022), 21 pages. https://doi.org/10.3389/fcomp.2022.744574

[5] Agnes Axelsson and Gabriel Skantze. 2022. Multimodal User Feedback During Adaptive Robot-Human Presentations. *Frontiers in Computer Science* 3 (2022), 19 pages. https://doi.org/10.3389/fcomp.2021.741148

[6] Nils Axelsson and Gabriel Skantze. 2019. Modelling Adaptive Presentations in Human-Robot Interaction using Behaviour Trees. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*. Association for Computational Linguistics, Stockholm, Sweden, 345–352. https://doi.org/10.18653/v1/W19-5940

[7] Nils Axelsson and Gabriel Skantze. 2020. Using Knowledge Graphs and Behaviour Trees for Feedback-Aware Presentation Agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents* (Virtual Event, Scotland, UK) *(IVA '20)*. Association for Computing Machinery, New York, NY, USA, Article 4, 8 pages. https://doi.org/10.1145/3383652.3423884

[8] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (1 Jan 2009), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[9] Allan Bell. 1984. Language style as audience design. *Language in Society* 13, 2 (1984), 145–204. https://doi.org/10.1017/S004740450001037X

[10] Marie-Luce Bourguet, Yanning Jin, Yuyuan Shi, Yin Chen, Liz Rincon-Ardila, and Gentiane Venture. 2020. Social Robots that can Sense and Improve Student Engagement. In *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. IEEE, Takamatsu, Japan, 127–134. https://doi.org/10.1109/TALE48869.2020.9368438

[11] Elizabeth A. Boyle, Anne H. Anderson, and Alison Newlands. 1994. The Effects of Visibility on Dialogue and Performance in a Cooperative Problem Solving Task. *Language and Speech* 37, 1 (1994), 1–20. https://doi.org/10.1177/002383099403700101 arXiv:https://doi.org/10.1177/002383099403700101

[12] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. https://doi.org/10.48550/ARXIV.2005.14165

[13] Hendrik Buschmeier and Stefan Kopp. 2014. A dynamic minimal model of the listener for feedback-based dialogue coordination. In *DialWatt-SemDial 2014: Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue*. Heriot-Watt University, Edinburgh, Scotland, 17–25.

[14] Joyce Y. Chai, Lanbo She, Rui Fang, Spencer Ottarson, Cody Littley, Changsong Liu, and Kenneth Hanson. 2014. Collaborative Effort towards Common Ground in Situated Human-Robot Dialogue. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Association for Computing Machinery, Bielefeld, Germany, 33–40. https://doi.org/10.1145/2559636.2559677

[15] Huili Chen, Hae Won Park, and Cynthia Breazeal. 2020. Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers & Education* 150 (2020), 103836. https://doi.org/10.1016/j.compedu.2020.103836

[16] Lara Christoforakos, Alessio Gallucci, Tinatini Surmava-Große, Daniel Ullrich, and Sarah Diefenbach. 2021. Can Robots Earn Our Trust the Same Way Humans Do? A Systematic Exploration of Competence, Warmth, and Anthropomorphism as Determinants of Trust Development in HRI. *Frontiers in Robotics and AI* 8 (2021), 15 pages. https://doi.org/10.3389/frobt.2021.640444

[17] Herbert H Clark. 1996. *Using language*. Cambridge university press, Cambridge, UK.

[18] Herbert H. Clark and Susan E. Brennan. 1991. *Grounding in communication*. American Psychological Association, Washington, DC, US, 127–149. https://doi.org/10.1037/10096-006

[19] Herbert H. Clark and Meredyth A. Krych. 2004. Speaking while monitoring addressees for understanding. *Journal of Memory and Language* 50, 1 (2004), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

[20] Herbert H Clark, Robert Schreuder, and Samuel Buttrick. 1983. Common ground at the understanding of demonstrative reference. *Journal of verbal learning and verbal behavior* 22, 2 (1983), 245–258.

[21] Moreno I. Coco, Rick Dale, and Frank Keller. 2018. Performance in a Collaborative Search Task: The Role of Feedback and Alignment. *Topics in Cognitive Science* 10, 1 (2018), 55–79. https://doi.org/10.1111/tops.12300 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/tops.12300

[22] Mirjam de Haas, Paul Vogt, and Emiel Krahmer. 2020. The Effects of Feedback on Children's Engagement and Learning Outcomes in Robot-Assisted Second Language Learning. *Frontiers in Robotics and AI* 7 (2020), 17 pages. https://doi.org/10.3389/frobt.2020.00101

[23] Philip Doyle. 2022. *The Dimensions and Adaptation of Partner Models in Human-Machine Dialogue*. Ph. D. Dissertation. University College Dublin.

[24] Philip R Doyle, Leigh Clark, and Benjamin R. Cowan. 2021. What Do We See in Them? Identifying Dimensions of Partner Models for Speech Interfaces Using a Psycholexical Approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 244, 14 pages. https://doi.org/10.1145/3411764.3445206

[25] John Dunlosky and Katherine A. Rawson. 2012. Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction* 22, 4 (2012), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003 Improving Self-Monitoring and Self-Regulation of Learning: From Cognitive Psychology to the Classroom.

[26] Hady Elsahar, Christophe Gravier, and Frederique Laforest. 2018. Zero-Shot Question Generation from Knowledge Graphs for Unseen Predicates and Entity Types. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for Computational Linguistics, New Orleans, Louisiana, 218–228. https://doi.org/10.18653/v1/N18-1020

[27] Olov Engwall, José Lopes, and Anna Åhlund. 2021. Robot Interaction Styles for Conversation Practice in Second Language Learning. *International Journal of Social Robotics* 13, 2 (1 Apr 2021), 251–276. https://doi.org/10.1007/s12369-020-00635-y

[28] E. Christa Farmer, Amy J. Catalano, and Adam J. Halpern. 2020. Exploring Student Preference between Textbook Chapters and Adaptive Learning Lessons in an Introductory Environmental Geology Course. *TechTrends* 64, 1 (1 Jan 2020), 150–157. https://doi.org/10.1007/s11528-019-00435-w

[29] K. Hayashi, T. Kanda, T. Miyashita, H. Ishiguro, and N. Hagita. 2005. Robot Manzai - robots' conversation as a passive social medium. In *5th IEEE-RAS International Conference on Humanoid Robots, 2005.* IEEE, Tsukuba, Japan, 456–462. https://doi.org/10.1109/ICHR.2005.1573609

[30] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. 2021. Knowledge graphs. *ACM Computing Surveys (CSUR)* 54, 4 (2021), 1–37.

[31] Filip Ilievski, Daniel Garijo, Hans Chalupsky, Naren Teja Divvala, Yixiang Yao, Craig Rogers, Rongpeng Li, Jun Liu, Amandeep Singh, Daniel Schwabe, and Pedro Szekely. 2020. KGTK: A Toolkit for Large Knowledge Graph Manipulation and Analysis. In *The Semantic Web – ISWC 2020*, Jeff Z. Pan, Valentina Tamma, Claudia d'Amato, Krzysztof Janowicz, Bo Fu, Axel Polleres, Oshani Seneviratne, and Lalana Kagal (Eds.). Springer International Publishing, Cham, 278–293.

[32] Tatsuya Ishino, Mitsuhiro Goto, and Akihiro Kashihara. 2022. Robot lecture for enhancing presentation in lecture. *Research and Practice in Technology Enhanced Learning* 17, 1 (5 Jan 2022), 1. https://doi.org/10.1186/s41039-021-00176-6

[33] Kleomenis Katevas, Patrick Healey, and Matthew Harris. 2015. Robot Comedy Lab: experimenting with the social dynamics of live performance. *Frontiers in Psychology* 6 (2015), 9 pages. https://doi.org/10.3389/fpsyg.2015.01253

[34] James Kennedy, Paul Baxter, Emmanuel Senft, and Tony Belpaeme. 2016. Social robot tutoring for child second language learning. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Association for Computing Machinery, Christchurch, New Zealand, 231–238. https://doi.org/10.1109/HRI.2016.7451757

[35] Elly A. Konijn and Johan F. Hoorn. 2020. Robot tutor and pupils' educational ability: Teaching the times tables. *Computers & Education* 157 (2020), 103970. https://doi.org/10.1016/j.compedu.2020.103970

[36] Dimosthenis Kontogiorgos, Andre Pereira, and Joakim Gustafson. 2021. Grounding behaviours with conversational interfaces: effects of embodiment and failures. *Journal on Multimodal User Interfaces* 15, 2 (1 Jun 2021), 239–254. https://doi.org/10.1007/s12193-021-00366-y

[37] Asher Koriat, Limor Sheffer, and Hilit Ma'ayan. 2002. Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General* 131 (2002), 147–162. https://doi.org/10.1037/0096-3445.131.2.147

[38] Junyi Li, Tianyi Tang, Wayne Xin Zhao, Zhicheng Wei, Nicholas Jing Yuan, and Ji-Rong Wen. 2021. Few-shot Knowledge Graph-to-Text Generation with Pretrained Language Models. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Association for Computational Linguistics, Bangkok, Thailand, 1558–1568.

[39] Hiroyuki Masuta, Tetsuya Kawamoto, Kei Sawai, Tatsuo Motoyoshi, Takumi Tamamoto, Ken'ichi Koyanagi, and Toru Oshima. 2018. Presentation Robot System with Interaction for Class. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE Press, Bengaluru, India, 1801–1806. https://doi.org/10.1109/SSCI.2018.8628804

[40] Marvin Minsky. 1982. *Semantic information processing*. Ph. D. Dissertation. Cambridge.

[41] Rifca Peters, Joost Broekens, and Mark A. Neerincx. 2017. Robots educate in style: The effect of context and non-verbal behaviour on children's perceptions of warmth and competence. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, Lisbon, Portugal, 449–455. https://doi.org/10.1109/ROMAN.2017.8172341

[42] Jan Pichl, Petr Marek, Jakub Konrád, Petr Lorenc, Van Duy Ta, and Jan Šedivý. 2020. Alquist 3.0: Alexa Prize Bot Using Conversational Knowledge Graph. https://doi.org/10.48550/ARXIV.2011.03261

[43] Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2006. Interactive Humanoid Robots for a Science Museum. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (Salt Lake City, Utah, USA) *(HRI '06)*. Association for Computing Machinery, New York, NY, USA, 305–312. https://doi.org/10.1145/1121241.1121293

[44] Nicolas Spatola and Olga A. Wudarczyk. 2021. Ascribing emotions to robots: Explicit and implicit attribution of emotions and perceived robot anthropomorphism. *Computers in Human Behavior* 124 (2021), 106934. https://doi.org/10.1016/j.chb.2021.106934

[45] Robin Stark, Hans Gruber, Alexander Renkl, and Heinz Mandl. 1998. Instructional effects in complex learning: Do objective and subjective learning outcomes converge? *Learning and Instruction* 8, 2 (1998), 117–129. https://doi.org/10.1016/S0959-4752(97)00005-4

[46] Katja Thieme. 2010. Constitutive Rhetoric as an Aspect of Audience Design: The Public Texts of Canadian Suffragists. *Written Communication* 27, 1 (2010), 36–56. https://doi.org/10.1177/0741088309353505 arXiv:https://doi.org/10.1177/0741088309353505

[47] S. Thrun, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. 1999. MINERVA: a second-generation museum tour-guide robot. In *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C)*, Vol. 3. IEEE Press, Detroit, MI, USA, 1999–2005 vol.3. https://doi.org/10.1109/ROBOT.1999.770401

[48] I Tiddi et al. 2020. Generating Explanations in Natural Language from Knowledge Graphs. *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges* 47 (2020), 213.

[49] Nigel Ward and Wataru Tsukahara. 2000. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics* 32, 8 (2000), 1177–1207. https://doi.org/10.1016/S0378-2166(99)00109-5

[50] Victor H. Yngve. 1970. On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, Mary Ann Campbell et al. (Eds.). Chicago Linguistic Society, Chicago, IL, USA, 567–577.

[51] Mohammad Abu Yousuf, Yoshinori Kobayashi, Yoshinori Kuno, Keiichi Yamazaki, and Akiko Yamazaki. 2019. Social interaction with visitors: mobile guide robots capable of offering a museum tour. *IEEJ Transactions on Electrical and Electronic Engineering* 14, 12 (2019), 1823–1835. https://doi.org/10.1002/tee.23009 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/tee.23009