User Modeling and User-Adapted Interaction
https://doi.org/10.1007/s11257-020-09267-3

Effects of adapting to user pitch on rapport perception, behavior, and state with a social robotic learning companion

Nichola Lubold[1,2] · **Erin Walker**[3] · **Heather Pon-Barry**[4]

Received: 25 February 2019 / Accepted in revised form: 28 May 2020

Abstract
Social robots such as learning companions, therapeutic assistants, and tour guides are dependent on the challenging task of establishing a rapport with their users. People rarely communicate with just words alone; facial expressions, gaze, gesture, and prosodic cues like tone of voice and speaking rate combine to help individuals express their words and convey emotion. One way that individuals communicate a sense of connection with one another is entrainment, where interaction partners adapt their way of speaking, facial expressions, or gestures to each other; entrainment has been linked to trust, liking, and task success and is thought to be a vital phenomenon in how people build rapport. In this work, we introduce a social robot that combines multiple channels of rapport-building behavior, including forms of social dialog and prosodic entrainment. We explore how social dialog and entrainment contribute to both self-reported and behavioral rapport responses. We find prosodic adaptation enhances perceptions of social dialog, and that social dialog and entrainment combined build rapport. Individual differences indicated by gender mediate these social responses; an individual's underlying rapport state, as indicated by their verbal rapport behavior, is exhibited and triggered differently depending on gender. These results have important repercussions for assessing and modeling a user's social responses and designing adaptive social agents.

Keywords Rapport · Social robot · Pitch · Gender · Adaptation · Social dialog

✉ Nichola Lubold
nichola.lubold@asu.edu

[1] School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ, USA

[2] Human Centered Systems Group, Honeywell Labs, Deer Valley, AZ, USA

[3] School of Computing and Information, University of Pittsburg, Pittsburg, PA, USA

[4] Department of Computer Science, Mount Holyoke College, South Hadley, MA, USA

Published online: 26 June 2020

1 Introduction

Rapport is a feeling of connection, harmony, and potential friendship, and when it occurs between people it can lead to better communication, collaboration, and task success (Stewart et al. 1999; Drolet and Morris 2000; Ogan et al. 2012). For *social robots*, robots that rely on some form of social interaction, establishing rapport is an important factor to success. However, modeling effective rapport-building behavior can be challenging. People build rapport with one another in many ways, including via facial expressions, gestures, gaze, verbal expressions, and prosodic cues such as tone of voice and speaking rate. This combination of rapport-building behaviors has been found to be critical to establishing successful social connections (Mehrabian 1972). In human–robot interactions, most research exploring how robots can combine rapport-building behaviors across multiple channels has focused on gesture and facial expressions (Lakens and Stel 2011; Saint-Aimé et al. 2007; Breazeal 2003; Brown and Levinson 1987). Prosodic cues such as tone of voice and speaking rate have been less explored, even though prosodic cues are known to contain important meta-information and have been linked to rapport (Lubold and Pon-Barry 2014). Incorporating prosodic cues into a multi-channel model of rapport-building behavior has potential to enhance social responses and increase the success of interactions.

We explore a model of multi-channel rapport-building behavior by combining prosodic cues with social dialog in a social robot. Social dialog has been shown to have important rapport-building effects in human–robot and human–agent interactions, including improving trust, motivation, and engagement (Bickmore and Cassell 2001; Bickmore 2003; Kanda et al. 2004; Gulz et al. 2011). Prior work has also shown that the effects of social dialog can vary and be enhanced with non-verbal behaviors such as gesture and facial expressions (Bickmore and Picard 2005; Csapo et al. 2012). We focus on the effects of social dialog when combined with prosodic cues.

We model prosodic behavior utilizing the phenomenon of acoustic-prosodic entrainment. Acoustic-prosodic entrainment, also known as accommodation, occurs when individuals adapt their prosodic features of speech, such as pitch or tone of voice, loudness, or speaking rate, to one another over the course of a conversation. According to the Communication Accommodation Theory (CAT), entrainment may facilitate social responses because individuals accommodate to their partner to achieve social approval (Giles and Smith 1979). This theory suggests an individual on the receiving end of a high level of prosodic adaptation is likely to feel a greater sense of self-esteem, satisfaction, and rapport for their partner than if they were a receiver of low adaptation. A robot which models the prosodic fluctuations of a human conversational partner and adapts prosodically might build more rapport than a robot which builds rapport through social dialog alone. We pose the following research question to explore this:

RQ 1 How does a social robot which prosodically adapts to a user build rapport?

Assessing how a social, entraining robot builds rapport can be difficult because people exhibit rapport in many ways. While generally we can infer an individual's rapport from self-reported questionnaires (i.e., "I felt a connection with the robot"), we can also assess rapport by observing their behavior (Pantic et al. 2007; De Carolis et al. 2015; Foster et al. 2013; Bechade et al. 2015). Behavioral measures can be an important indicator of how an individual is feeling in the moment while self-reported responses provide more distinct and accurate insight into an individual's feelings at the end of an interaction. Behavioral responses can potentially be assessed real-time and used as input to a model of social, adaptive rapport-building behavior. For this work, we utilize both self-reported and behavioral measures of rapport to assess the effects of an entraining, social robot. We explore the following research question to gain insight into the nuances of different measures of rapport responses:

RQ 2 How do different measures of rapport reflect responses to a social, prosodically adaptive robot?

Additionally, factors such as personality, age, and gender might lead to varying rapport responses for any particular individual. We use gender as a proxy for certain individual differences that may mediate users' self-reported and behavioral rapport responses. Prior work suggests gender plays a particularly important role in interactions with social robots. Males appear to be more likely to see robots as human-like and socially present (Schermerhorn et al. 2008), but females are more sensitive to verbal behavior (Strait et al. 2015). Depending on the role of gender, it is possible a social robot should engage in social behaviors differently based on the individual. With this work, we hope to provide insight into how social dialog and prosodic adaptation influence individual responses and provide insight into an eventual model for monitoring rapport responses in real-time and responding accordingly. We pose a third research question:

RQ 3 How are the effects of a social, entraining robot mediated by the gender of the user?

To investigate the effects of an entraining social robot, we introduce a teachable robot named Quinn comprised of a LEGO® Mindstorms® NXT robot with an iPod Touch that displays facial expressions and outputs speech. Teachable robots are a form of social robot for learning and are based on the principle of "learning by teaching" where students learn from teaching others because they attend more to the problem, reflect on their own misconceptions when correcting their peers' errors and then elaborate on their knowledge as they construct explanations. Teachable robots, where students teach the robot, have proven that learning by teaching can improve domain learning, student self-efficacy, and motivation (Leelawong and Biswas 2008; Walker et al. 2016; Jacq et al. 2016; Tanaka and Matsuzoe 2012). In this work, students teach the robot how to solve literal math

equations; the robot can introduce verbal rapport-building social dialog and simulate entrainment by adapting its pitch to a human user.

The teachable robot platform we introduce here presents a significant opportunity for exploring the design and outcomes of multi-channel social behavior. The system is unique in its ability to incorporate entrainment alongside other rapport-building verbal behavior, and we present the first complete description of that system. Additionally, teachable agent interactions are thought to benefit from increased social engagement. Student tutors who feel more invested or feel more rapport for their agent have been found to learn more (Leelawong and Biswas 2008; Ogan et al. 2012) and human–human dyads who exhibit higher rapport tend to have greater success in peer-tutoring interactions. It is possible that a teachable robot that builds rapport may facilitate increased learning. The conclusions we can draw regarding building rapport for a teachable robot through adaptation provide hopeful implications for similar adaptive behaviors in other human–agent interactions where social engagement and rapport are important.

We conducted a study with 48 undergraduate students who interacted with one of three versions of the teachable robot: (1) Non-social. The robot did not behave socially, (2) Social. The robot spoke socially but did not entrain, (3) Social plus entraining. The robot spoke socially and entrained by adapting prosodically. We collected subjective self-reported rapport measures and coded for observable behaviors related to rapport. In the remainder of this paper, we begin by summarizing related work on entrainment, rapport, and gender. We then describe the social, teachable robot we developed and the technical implementations of its behavior in Sect. 3. The study and measures are described in Sect. 4. The results and analyses are reported in Sect. 5. We conclude with a final discussion on the implications for the development of social, adaptive robots in Sect. 6.

2 Related work

2.1 Prosodic adaptation as entrainment

Entrainment is prevalent in human–human interactions. People have been found to mimic their conversational partners, adapting their facial expressions, their body language, the content of their speech, and their prosodic cues such as pitch, intensity, and speaking rate (Hess and Blairy 2001; Lakin and Chartrand 2003; Nenkova et al. 2008; Levitan et al. 2012). For this work, we are interested in entrainment on prosodic cues. Individuals entrain prosodically along two timescales: locally and globally. Local entrainment is measured on a turn-by-turn basis while global entrainment is measured across the course of a conversation, such as by comparing the beginning to the end. There are several types of local and global entrainment including synchrony (individuals exhibit similar rhythmic qualities and coordination), convergence (individuals become increasingly similar on their dialog features over time), and proximity (individuals match or mirror one another). Synchrony is typically only measured locally, while convergence and proximity can occur both globally and locally.

In the prior work exploring prosodic entrainment between people, there appears to be a significant relationship between entrainment and different kinds of social constructs including trust, engagement, positivity, and rapport (Gravano et al. 2014; Scissors et al. 2009; Levitan et al. 2012, 2015; Bone et al. 2013; Bonin et al. 2013; Gregory et al. 1997; Levitan and Hirschberg 2011; Natale 1975; Sinha and Cassell 2015; Mushin et al. 2003; Sidaras 2011; Benuš et al. 2012, 2014; Benuš 2009; Lee et al. 2011a, b; Lubold and Pon-barry 2014; Babel 2010; Babel and Bulatov 2012; Babel 2011; Kousidis and Dorran 2009; Lubold et al. 2015; Gweon et al. 2013; Schweitzer and Lewandowski 2013; Vaughan 2011). The results of these explorations are summarized in Table 1. Many explored proximity and/or convergence and found positive relationships. It is possible that the overwhelmingly positive findings may be due to a lack of reporting non-significant results. In any case, developing a model of automated entrainment in a conversational agent beginning with a local form of proximity or a form of local convergence may have a good chance of fostering social responses. We use proximity as the starting point in this work.

Automated models of entrainment are still in the early stages. Sadoughi et al. (2017) built a system for a social, human-like robot which adapts on a turn-by-turn basis to a child's pitch and intensity. In their approach, they utilized a Bayesian network to select the most appropriate verbal response along with its prosodic manipulation at run-time. This resulted in a limited number of possible adaptations. Sadoughi and colleagues evaluated engagement with the robot by varying whether the robot entrained in the first or second half of the interaction. Children had higher engagement with the robot which began the interaction by entraining. Sadoughi and colleagues did not explore the effects of manipulating the prosodic features real-time, instead using pre-recorded audio, and it remains unclear whether pitch or intensity or both resulted in the positive effects on engagement. Levitan et al. (2016) explored the effects of adapting intensity and speaking rate in a turn-by-turn manipulation on perceptions of a virtual agent's likability and reliability. In pilot evaluations, they found positive effects for English speakers. Their approach demonstrated the potential of real-time adaptations with intensity and speaking rate. Our work builds on these implementations by focusing on a real-time adaptation of pitch and providing insight into the impact adaptation has on feelings of rapport as well as the role of individual differences as indicated by gender.

2.2 Prior work on enabling robots to be social

A popular way of enabling robots to be social and build rapport is by introducing gesture and dialog to suggest a social connection; for example, a gesture which conveys 'friendliness' such as waving when one says hello or dialog such as polite language. These kinds of behaviors have been shown to increase rapport and learning when employed by robotic tutors, which are robots that can teach or tutor students. Kanda et al. (2004) conducted a 2-month trial in an elementary school with a social robot for learning English. The robot, called Robovie, could express various social behaviors, such as calling children by name. The social behaviors engaged the students and the students who interacted with Robovie longer learned more.

Table 1 Summary of prior work exploring relationships between prosodic entrainment and communicative behaviors or affective factors with connections to rapport

Type of Entrainment		Prosodic features explored	Direction of relationship between entrainment and social factor [Reference]
Global	Proximity	Pitch contour	↗ Engagement [Gravano et al. 2014]
		Positive words	↗ Trust [Scissors et al. 2009]
		Negative words	↘ Trust [Scissors et al. 2009]
		Pitch, intensity…	↗ Trying to be liked [Levitan et al. 2012]
		Pitch, intensity	↘ Conversational quality (ASD) [Bone et al. 2013]
	Convergence	Pitch, intensity	_ Agreement [Bonin et al. 2013]
		Pitch	↗ Solidarity [Gregory et al. 1997]
		Pitch, intensity…	↗ Backchannel preceding cues [Levitan and Hirschberg 2011]
		Intensity	↗ Social desirability [Natale 1975]
		Speaking rate	↗ Rapport [Sinha and Cassell 2015]
		Boundary tone	↗ Common ground [Mushin et al. 2003]
		Speaking rate	↗ Bias (vocal expectations) [Sidaras 2011]
Local	Proximity	Pitch, intensity…	↘ Latency [Levitan et al. 2015]
		Pitch, intensity…	↗ Backchannel [Levitan et al. 2015]
		Intensity	↗ Filled pauses [Benuš 2012; Benuš 2009]
		Pitch, energy…	↗ Positive/negative affect [Lee et al. 2011a, b]
		Pitch…	↗ Rapport [Lubold and Pon-Barry 2014]
		Pitch, intensity…	↗ Trying to be liked [Levitan et al. 2012]
		Intensity, CF	↗ Favorable voting [Benuš et al. 2014]
		Phonetic, pitch	↗ Positive bias [Babel 2010, 2011]
		Overlaps	↗ Engagement [Kousidis and Dorran 2009]
	Convergence	Pitch, intensity…	_ Backchannels [Lubold et al. 2015]
		Pitch	↗ Grounding [Lubold et al. 2015]
		Pitch, intensity…	↗ Transactive contributions [Gweon et al. 2013]
		Speaking rate	↗ Liking [Schweitzer and Lewandowski 2013]
	Synchrony	Pitch, intensity…	_ Rapport [Lubold and Pon-Barry 2014]
		Pitch, intensity…	_ Agreement [Vaughan 2011]

The ellipsis (…) indicates that there were additional acoustic-prosodic features explored in the prior work referenced

Saerbeck et al. 2010 investigated how a socially supportive robotic tutor (iCat) influenced the task of language learning. iCat exhibited a variety of rapport-building, social behaviors which were both verbal and non-verbal, including dialog, gaze, and facial expressions; they found students had higher learning performance with the socially supportive tutor. Kasap and Magnenat-Thalmann introduced a humanoid robot that combined affective facial expressions with supportive dialog and the ability to remember previous interactions (Kasap and Magnenat-Thalmann 2010, 2012). With the robot acting as a tutor, they found that how individuals perceived the robot was significantly influenced by both the robot's ability to remember and its use of socially supportive expressions, with some suggestion that the combination of social dialog and affective facial expressions can lead to more positive impressions of personality. Finally, Westlund et al. 2016 introduced Tega, an affect-sensitive robotic tutor which pre-school children interacted with to learn vocabulary. Tega demonstrated that adaptation can increase positive valence (Gordon et al. 2016). In one of the few explorations of prosodic manipulations, Tega was also used to explore engaging preschoolers in active reading (Kory-Westlund and Breazeal 2019). In this exploration, the robot's voice was either expressive, including a wide range of intonation and emotion, or flat, like a classic TTS engine. Their findings suggested an expressive robot is more beneficial.

Social dialog has been explored as a rapport-building behavior particularly in human–agent literature. Bickmore and Cassell (2001) demonstrated that social dialog such as small talk can have rapport-building effects, significantly enhancing feelings of trust in interactions with a virtual real estate agent. In later work, they also found that removing non-verbal cues available through facial expression and gesture negatively influenced the effects of social dialog, and that individual differences indicated by personality played a role in these effects, with individuals preferring and trusting an embodied conversational agent which matched their own personality more (Bickmore and Cassell 2001). Gulz et al. (2011) demonstrated that students who interacted with a teachable agent which engaged in social dialog in the form of 'off-task' dialog reported having a more positive experience and learned more. In work with a virtual agent tutor for multi-party dialogs, Kumar et al. (2010) showed that a tutor which uses social dialog to show solidarity, trigger tension release, and exhibit an agreeable attitude can significantly influence learning. Bickmore et al. (2013) showed that virtual agents which exhibited solidarity through common ground and self-disclosure, empathy, and humor improved engagement. For our work, we design social dialog in line with this prior work and use this dialog as a baseline to explore effects of social dialog plus entrainment on rapport with a teachable robot.

We hypothesize that the addition of prosodic entrainment to social dialog should positively influence rapport. This hypothesis is motivated by the Communication Accommodation Theory (CAT), which proposes that individuals will entrain as a means of achieving solidarity with their interaction partner. A socio-psychological theory explaining entrainment based on CAT argues that the phenomenon is driven by the need to achieve certain social effects and is based on the idea of similarity-attraction. The similarity-attraction theory posits that, "The more similar our attitudes and beliefs are to those of others, the more likely it is for them to be attracted

to us." (Giles and Smith 1979). Individuals use entrainment to obtain social approval from their interlocutor. This theory suggests that an individual on the receiving end of a high level of accommodation is likely to develop a greater sense of self-esteem and satisfaction and to feel more rapport for their speaking partner than if they were a receiver of low accommodation. With the addition of social dialog, a robot which also adapts its prosody should increase feelings of rapport.

2.3 Measuring rapport: self-reported, behavioral, and rapport state

We can measure feelings of rapport in several ways including as self-reported rapport through questionnaires (i.e., "I felt a connection with the robot"), as behavioral rapport where the user's behaviors are used as assessment of their rapport (i.e., does the individual smile, do they use rapport-building language), and as perceptual rapport through third party perceptions where an individual observes and rates an interaction for rapport (Pantic et al. 2007; De Carolis et al. 2015; Foster et al. 2013; Bechade et al. 2015). We are interested in the first two approaches which are rooted in the user themselves. As a self-reported measure, rapport has been assessed as general rapport related to feelings of connection and harmony (Novick and Gris 2014) and, particularly with agents and robots, as social presence (Huang et al. 2011). Social presence has been described as the "level of awareness of the co-presence of another human, being, or intelligence" and as the "feeling that one has some level of access or insight into the other's intentional, cognitive, or affective states" (Biocca and Nowak 2001). With robots and agents, social presence may be an important measure of rapport given the inherently remote properties typically associated with these technologies. For this work, we assess self-reported rapport as both general rapport and as social presence.

To measure behavioral rapport, rapport theory suggests that behaviors which are indicative of politeness may provide insight into a user's feelings of rapport. Spencer-Oatey (2005) suggests an individual's use of politeness is an example of how individuals manage rapport. For example, if an individual praises their conversational partner, this may positively enhance their partner's feelings toward them. If an individual is rude to their conversational partner by calling them a name, this may introduce face-threat, hindering rapport. Bell et al. (2009) performed an analysis of linguistic politeness and interpretation of its meaning in peer-tutoring scenarios. Based on the dialog of two pairs of tutors and tutees, the authors analyzed different politeness strategies on the part of the tutor based on verbal behaviors such as inclusive language, praise, and humor that were suited to the peer-tutoring domain. In first-time sessions, the tutors appeared to be reluctant to utilize positive politeness behaviors such as inclusive language and praise; over the course of multiple sessions, these behaviors increased and aligned to building rapport. Similar behaviors such as praise, inclusive language, name usage, and formal politeness have been found to be associated with positive rapport in other prior works (Ogan et al. 2012; Wheldall and Mettem 1985). Exploring how students utilize similar linguistic strategies when tutoring a robotic learning companion may provide insight into their level

of engagement; their use of different language strategies may provide insight into the degree of rapport they are feeling for the robot.

We hypothesize that both the behavioral measures and self-reported measures will indicate similar overall responses, with students who exhibit more behavioral rapport also self-reporting higher feelings of general rapport. Based on prior work and theory of rapport, behavioral and self-reported rapport should reflect a similar underlying rapport state. If an individual exhibits behavior that according to theory and analysis of human–human interactions is reflective of higher rapport, it seems logical their self-reported rapport should reflect this as well. Given this, we also measure rapport by exploring how behavioral rapport reflects an individual's underlying rapport state.

To measure an individual's underlying rapport state, hidden Markov models have historically been applied for understanding hidden states such as emotions, tutoring modes, and learner engagement (Nwe et al. 2003; Boyer et al. 2010; Beal et al. 2007). Once a model has been created, the frequency counts of the estimated hidden states can be used to understand the relationship between the hidden state (i.e., tutorial mode or learner engagement) and desired outcomes (i.e., learning). For example, Boyer and colleagues utilized an HMM to model effective tutoring modes based on observed dialog acts. Correlating the estimated frequency counts of the different tutoring mode states with learning, they found significant learning gains associated with state sequences. Beal, Mitra, and Cohen modeled learner engagement; relating the hidden state of learner engagement to learning, they identified learner engagement trajectories which directly related to learning gains. Bergner and colleagues explored how tutors assist tutees when tutees make a mistake (Bergner et al. 2017). Utilizing an IOHMM, they compared the assistance value of different tutor inputs in helping the tutee correct a mistaken step and found successful as well as deleterious patterns in collaborative learning. An IOHMM may reveal whether there is a hidden state associated with an individual's behavioral rapport, whether we can consider that state to be 'social' or reflective of an individual's rapport, and how that state was affected by Quinn's behavior. By exploring an individual's underlying rapport state, we can confirm the outcome of our hypothesis regarding how behavioral measures and self-reported measures align.

2.4 The role of gender

Explorations of gendered responses in the human–robot literature are limited; as of 2014, only 21 of 190 HRI papers published from 2006 to 2013 provided any form of gender-based analysis (Wang and Gratch 2009). However, there is evidence which suggests males and females might respond differently to rapport-building behaviors from a robot. Strait et al. (2015) found females were more sensitive to verbal communication, while males were more sensitive to multiple behaviors and consistency. If we examine prior work on gender differences in human–agent interactions, females tended to respond more positively to social behavior from virtual agents, while males tended to respond negatively (Burleson and Picard 2007; Vail et al. 2015; Arroyo et al. 2013; Lutfi et al. 2013;

Jokinen and Hurtig 2006; Krämer et al. 2016). Burleson and Picard introduced a multimodal, real-time affective agent which exhibited emotional intelligence in response to a user's affect. The agent's behaviors included speaking, nodding, smiling, fidgeting, and shifting its posture forward or backward; these behaviors were adapted to mirror the user and to give evidence of 'active' listening. In analyzing responses from 76 girls and boys aged 11 to 13, girls responded much more positively to the affective tutor, expressing a stronger social bond, persevering longer, and exhibiting higher gains in meta-affective skills. Boys responded more positively on these measures to the non-social tutor. In other work, Vail and colleagues explored gender responses to an agent which exhibited cognitive and affective support through verbal feedback; females reported significantly more engagement and less frustration with an agent which exhibited motivational and engaging support. Arroyo and colleagues supported these findings with an extensive analysis of an affective pedagogical agent deployed in several public schools; female students had significantly lower frustration, more excitement, higher self-efficacy in mathematics, and greater liking of mathematics when interacting with the affective agent. These agents were generally designed to exhibit rapport through dialog and physical gesture.

We hypothesize that females will respond with more rapport to the social, entraining robot than males given this prior work suggesting females tend to respond more positively to social behaviors from agents and robots. As we are interested in assessing rapport with both self-reported and behavioral measures, we also hypothesize males and females will differ in their use of behavioral rapport. Measuring behavioral rapport as linguistic politeness, males and females have been found to differ in how they exhibit rapport-building language like linguistic politeness. Empirical studies by Holmes (1995), Coates (2015), Tannen (1994), and Hong and Hwang (2012) point to women using conversation to establish, nurture, and develop relationships, while men are more likely to see conversation as a tool for obtaining and conveying information. Bell et al. (2009) did not report seeing a large difference in politeness between the male and female tutors they analyzed; however, women have been found to be politer in general, often being more likely to give praise and engage in formal politeness (Chalupnik et al. 2017; Brown 1980). It has been suggested that differences in linguistic strategies may be a result of an individual's experience and their role in the conversation as either a peer, expert, or sub-ordinate. If we observe differences in linguistic strategies between males and females, this could be due to males and females using politeness for different purposes (i.e., as a rapport builder vs. information conveyer) or it could be indicative of differences in their interpretation of their role as a peer versus an expert.

3 A social, entraining robotic learning companion

We designed and built Quinn, a social, robotic learning companion, for this research. We first describe Quinn and the dialog system. We then describe the design and implementation of the social dialog and the voice adaptation.

3.1 Quinn

Quinn is a teachable robot for literal equations, consisting of a LEGO Mindstorms base with an iPod Touch mounted on top to represent its face. Students taught Quinn how to solve literal equations (i.e., "Solve $bx + gy = 14by + 6x$ for x") using spoken language and a web application presented on a Windows Surface Pro tablet. The web application contained materials to guide the students in their teaching of Quinn with the worked-out solutions for each literal equation provided on the interface. Up to eight literal equation problems and quizzes were available in the application. The application presented one problem at a time and included the worked-out steps to reach a solution. The problems were ordered in increasing difficulty. New concepts were introduced every two problems; concepts included how to handle multi-step equations, re-arranging formulas, and factoring. Students walked Quinn through the worked-out problems using spoken language, explaining each step. Quinn responded using spoken language; the robot's facial expressions are animated when speaking, and neutral otherwise. At the end of each problem, a follow-up quiz was provided. Students asked Quinn to solve the quiz, step by step. Quinn solved the quiz independent of the student. Figure 1 depicts Quinn and a sample problem.

The speech interaction was real-time, and the dialog was recorded via the microphone on the tablet interface as the student spoke. After explaining each step, the students were instructed to pause, giving Quinn a chance to respond, and students would see a gif depicting that Quinn was "thinking" would appear on the screen to indicate that Quinn was occupied while the system processed the speech and a response was generated. The typical processing time was 3.7 s on average. Once a student was done speaking, their audio passed into the dialog system, which is described in more detail in the next section.

3.2 Dialog system

The dialog system developed was capable of both entrainment and social dialog for the purposes of exploring rapport. The overall structure follows that of typical dialog systems. The user's speech was recorded via a microphone on the tablet interface and once they were done speaking, the user's audio was passed into the dialog system which consisted of four main modules: (1) an automatic speech recognizer,

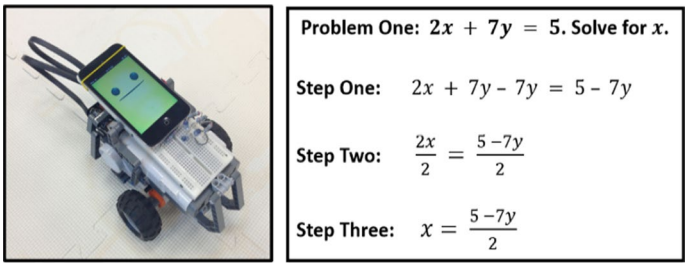

Fig. 1 Quinn and a sample problem

(2) a dialog manager, (3) an acoustic-prosodic feature extractor, and (4) a module for prosodic manipulation and text-to-speech generation. The overall dialog system infrastructure is shown in Fig. 2. For the automatic speech recognizer, we utilized the HTML5 Speech API available in Chrome ("Web Speech API" 2018). After the speech recognition, the dialog manager would identify an appropriate response for Quinn; we provide more detail about the dialog manager in the next section. For the acoustic-prosodic feature extraction, the student's mean pitch was extracted using Praat's pitch estimation algorithm which performs acoustic periodicity detection based on autocorrelation (Boersma 2006). Minimum and maximum fundamental frequencies for pitch estimation were set based on the gender of the speaker (i.e., for males, 75–300 Hz and for females 100–500 Hz). Once the student's features were extracted and a response identified, the prosodic adaptation and text-to-speech generation were performed with Praat and the Microsoft Speech API; we describe this module in more detail in Sect. 3.2.2.

3.2.1 Dialog manager

The text of the student's utterance was sent to the dialog manager to identify an appropriate response. The dialog manager utilized a modified version of the rule-based chatbot design found in chatbots like ELIZA and ALICE (Weizenbaum 1966; Shawar and Atwell 2002; ALICE 2002). Chatbot systems, first introduced in the 1960's, have increasingly been applied to practical applications (Shawar and Atwell 2007). In educational systems, chatbot frameworks have been developed that combine more traditional elements from task-oriented dialog with the shallow dialog moves chatbots were originally developed to produce (Gulz et al. 2011). This combination enables a system to produce social dialog opportunities while still maintaining domain knowledge representation and acceptable dialog responses. Inspired by Gulz and colleagues, we utilized a chatbot system based on the AIML framework (Wallace 2003) and introduced additional modifications unique to the domain and task-oriented dialog of the learning-by-teaching interaction.

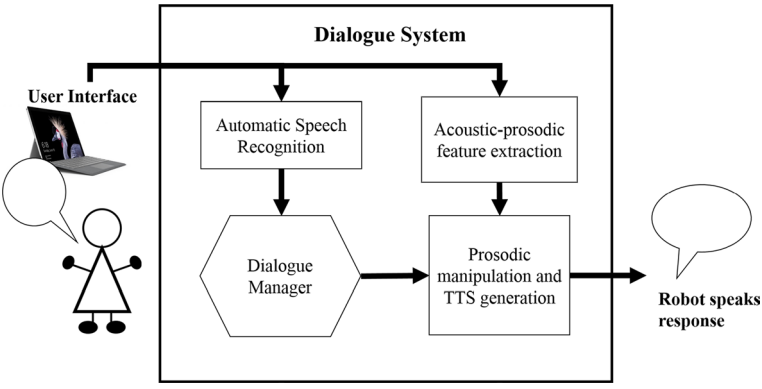

Fig. 2 Dialog system structure to enable entrainment in a social, robotic learning companion

To identify Quinn's responses, the dialog manager implemented the AIML process of linking keywords to pattern/transform rules. For example, if we consider the problem given in Fig. 1, the user may begin by explaining to Quinn that, "You need to *subtract* 7*y* from each side." Because the user's utterance contains the keyword *'subtract,'* this utterance would match the following rule/transform associated for this specific problem:

(∗ subtract ∗) → (Okay we subtract! Can you explain a little bit more about why we subtract?)

The system would then issue the response, "Okay we subtract! Can you explain a little bit more about why we subtract?" The simplest responses mapped the keyword back into a response that followed from the limited domain of the problem set. More complex responses included additional transformations or captured more explicit content. For example, an additional keyword mapping could capture that the learner said to "subtract 7*y*". The rule/transform would be:

(∗ subtract 7*y* ∗) → (Okay we subtract 7*y*! Can you explain a little bit more about why we subtract 7*y*?)

In this case, the system would issue the response "Okay I subtract 7*y*! Can you explain a little bit more about why I subtract 7*y*?" All keywords were given a rank; a higher rank increased the likelihood of a keyword being matched. The content-full response associated with "7*y*" would be given a higher priority based on the higher specificity.

To reduce the effects of ASR errors and enable more content-full responses, we incorporated additional information from the tablet interface that learners used to interact with the system. This information included the current problem and step. We considered each individual problem-step combination as a separate 'topic' with unique keywords, phrases, and associated pattern/transform rules. The learning companion would initiate the dialog whenever a new problem or a new step was started. This would set the 'topic' to that problem and step. Keywords belonging to the current problem and step were given the highest rank; general keywords and phrases not tied to the current problem and step were ranked lower and were therefore less likely to be matched first. If a student's speech could not be matched to a specific keyword, a response was selected from a set of 'generic' utterances. This set contained two types of responses: requests for clarification (i.e., "can you please repeat that?"), and general acknowledgements (i.e., "ok sounds good").

Within a problem-step, certain keywords when matched might initiate short two to three turn dialog trees where the system would then listen for keywords associated with the system's prior utterances. An example of a short dialog with dialog tree is given in Fig. 3. The sample dialog begins with Quinn initiating the conversation at the start of the step. The learner then began an explanation telling Quinn to subtract. This initiates a dialog tree based on the keyword subtraction. Keywords and phrases corresponding to subtraction and Quinn's prior utterances are then given the highest priority.

Finally, the ability to include social dialog was also a part of the dialog manager. To include social dialog, we modified the rules/transforms to include social

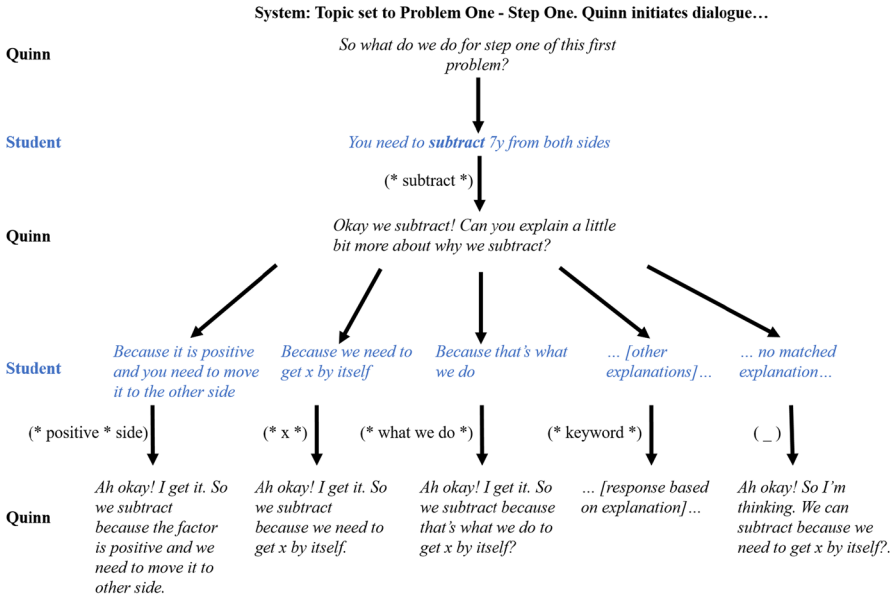

Fig. 3 Sample dialog tree initiated by a student's explanation to subtract

content in addition to the domain-based content. We describe the design of the social dialog more in the next section. The overall, high-level architecture describing how responses were identified in the dialog manager is summarized in Fig. 4.

3.2.2 Social dialog

Quinn's social dialog responses were motivated and designed based on the social interaction strategy proposed by Bales (1950) and utilized by Kumar and colleagues with their virtual tutoring agent (2010). We chose this framework since Kumar and colleagues demonstrated that their designed social utterances had positive effects on collaborating students' communication and because similar social behaviors were

```
function QUINN_DM(user utterance, current problem) returns response

prior←Get Quinn's last prior statement from memory
Find the word/phrase w in utterance that has the highest keyword rank in problem given prior
        if w exists
                Choose the highest ranked rule r for w that matches utterance
                response←Apply the pattern/transform in r to utterance
                push response on to memory
        else
                response←Apply the transform for the NONE keyword to sentence
                push response on to memory
return (response)
```

Fig. 4 Dialog manager architecture

found to have positive effects by Gulz et al. (2011) and Bickmore et al. (2013). Bales defined three main categories of positive socio-emotional behaviors: showing solidarity, showing tension release, and agreeing. Examples of social responses Quinn could give in each category are given in Table 2. Bales based his process on observations of group interactions; however, these responses and categories are also supported by human–human peer-tutoring dialog analysis which has shown that peer tutors can engage in behaviors which indicate solidarity (i.e., praise and encouragement, "come on, we can do this"), tension release (i.e., off-topic conversation such as "so what do you do for fun?"), and agreeing (i.e., comprehension, "yes, okay, you are right") (Bell et al. 2009). All social content that was built into the dialog was based on dialog moves that have been validated as being socially oriented in prior work.

The social dialog was generated by creating two parallel dialog options; a social dialog and a non-social dialog. This means that there were two identical sets of potential dialog responses where one set had 'social' moves included with the problem-solving dialog, as shown in Table 2, and one set had only problem-solving dialog. A 'social' dialog response was selected in the social and social plus entraining conditions 15–20% of the time. This frequency mirrored results from analysis of human–human social responses in collaborative dialogs (Lubold and Pon-Barry 2014; Kumar et al. 2010). The trigger for using a 'social' response rather than a 'non-social' response was uniformly randomized to ensure that at least 15% and no more than 20% of Quinn's responses to each participant in the social and social-entraining conditions were social and that these social responses were not all given at once but were distributed across the interaction. Given that the social content was combined with the problem content, the social responses still made contextual sense as the responses were generated based off the same keywords in either condition, but the social dialog conditions swapped out additional problem content for more socially oriented content. In the case that a social dialog response might trigger a social response from the student, this response would be handled socially in the social and social plus entraining conditions. For example, if the robot asked if the student enjoyed math and the student responded 'yes', the robot would exhibit cheerfulness and solidarity with the response "That's cool! Me too."

3.2.3 Prosodic adaptation

The pitch adaptation was based on the form of entrainment known as local proximity or turn-by-turn accommodation, a form of entrainment highly correlated with rapport, learning and task success (Lubold and Pon-Barry 2014; Thomason et al. 2013). There are multiple approaches to implement local proximity on pitch. Described in detail in Lubold et al. (2015), we identified three potential methods for adaptation inspired by observations of how human conversation partners entrain. To identify the most rapport-like manipulation which still maintained levels of naturalness equivalent to regular text-to-speech, we utilized Amazon Mechanical Turk (AMT), a popular resource for crowdsourcing research tasks including annotations, transcripts, and subjective analysis (Buhrmester et al. 2011). We found that a method of pitch adaptation that maintains the contour of the original TTS but shifts it up

Table 2 Categories and descriptions for social responses; examples of social and non-social responses

Category	Description	Social response	Non-social response
Solidarity	Compliments	Ok so we add *x*. You're a really great teacher!	Ok so we add *x*. I get that we are adding *x* here
Tension release	Being cheerful	Ok so we add *x*. I'm so happy to be working with you	Ok so we add *x*. It makes sense that we would add *x* here
	Off-topic	Ok so we add *x*. Do you like math?	Ok so we add *x*. I get adding
Agreeing	Comprehension	I hear what you're saying. You're saying we add *x*	We add *x*. It makes sense that we would add *x* here

or down to match the mean pitch of the speaker resulted in higher rapport ratings and was as natural as basic text-to-speech. Based on interviews, the pitch adaptation did not appear to influence the affective perception of the voice, in the sense that individuals did not perceive a pitch-adaptive agent as 'happier' or 'sadder'. Figure 5 depicts the adaptation.

The pitch adaptation was applied to the text-to-speech output; the TTS output was generated by the Microsoft Speech API. The gender of the TTS output was matched to the gender of the speaker; the female voice of "Zira" was used to adapt to female speakers, and the male voice of "David" was used to adapt to male speakers. In the non-adaptive interactions, females heard a non-adapted version of "Zira" and males heard a non-adapted version of "David." To adapt the pitch of the text-to-speech output to the user, the system utilized the estimated pitch values extracted from the user's previous turn. This occurred in the 'Feature Extraction' module depicted in Fig. 2.

To adapt the TTS, we extracted the mean pitch of the non-adapted TTS output using Praat's pitch estimation algorithm. The TTS output was then adapted by shifting all the frequencies of the TTS output by the difference between the mean pitch of the user and the mean pitch of the non-adapted TTS output. With the new frequencies, re-synthesis of the modified TTS output was performed with Praat's version of Time-Domain Pitch-Synchronous Overlap-and-Add (TD-PSOLA). Re-synthesis with TD-PSOLA has the potential to introduce some distortion based on speaker characteristics (Longster 2003), which can lead to potential attenuation of some frequency values and reverberation. To identify if there would be issues regarding intelligibility, we reviewed differences in the values of the vowel formants produced pre-adaptation and post-adaptation in the 80 min of dialog described in Lubold et al. (2015). Formants correspond to resonances in the vocal tract; vowels are identifiable based on formant ranges, and there is a clear link between perceived vowel quality and the first two formant frequencies. For each of the adaptations, the resulting formants stayed consistently within expected ranges for intelligibility. We compared the pre-adaptation and post-adaptation values for the contour shift adaptation applied in this work. A paired *t* test indicated no significant differences between the

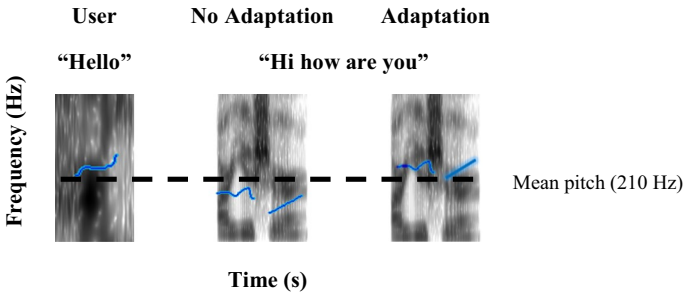

Fig. 5 Spectrograms of waveforms with pitch contour shown in blue. The contour is shifted up to match the user's mean pitch in the chosen adaptation. (Color figure online)

pre- and post-F1 ($t = -.11$, $p = .91$) values and the pre- and post-F2 formant values ($t = .06$, $p = .95$), indicating the adaptation was acceptable.

4 Study

We conducted a between subjects' experiment with three conditions: (1) a social plus entraining condition in which Quinn introduced social statements and adapted its pitch in a form of entrainment to the participant (2) a social condition in which Quinn only introduced social statements, and (3) a non-social condition in which Quinn did not speak socially, staying purely on task and did not adapt. A total of 48 individuals participated, 16 in each condition consisting of 8 females and 8 males. Despite the small sample size, we chose a between-subjects design rather than a within-subjects design because we anticipated that students would form a different mental model of the robot and its characteristics when it behaved socially when it did not, and thus a within-subjects design would be vulnerable to order effects.

Participants were undergraduate students between the ages of 18 and 30, and all were native English speakers. Individuals were randomly assigned to a condition, and all sessions lasted for 90 min. Participants were compensated $15 upon completion.

4.1 Procedure

Participants began by completing a 10-min pretest on literal equations. They were then given a practice exercise consisting of two worked-out examples of literal equations. The participants were asked to explain the problems and how to solve each step out loud. This exercise was to help participants adjust to the tutoring task and encourage them to think about how they might explain the content to another before having them attempt to explain it to Quinn. After this exercise, participants watched a short video which introduced Quinn and described the task.

Participants were told the task consisted of helping Quinn learn how to solve literal equations; they should walk Quinn through the steps to solve six problems, and they would have the opportunity to test Quinn's understanding through quizzes. Participants were also informed they could 'reteach' Quinn if Quinn struggled on a quiz by moving back to the previous problem. After teaching Quinn all six problems, participants were given a 10-min posttest and a short questionnaire assessing their attitudes toward Quinn. If a participant had additional availability, meaning they could stay longer than the 90 min of allotted time, the experimenter asked them a few focused interview questions. Outside of availability, no other criteria were used to determine which participants were interviewed. The interviews were approximately distributed across gender (11 female, 9 male) and condition (6 control, 6 social, 8 social plus entraining).

4.2 Measuring behavioral and self-reported rapport

We collected self-reported rapport and verbal, behavioral rapport for analyzing social responses to a social, robotic learning companion. For self-reported rapport, we were interested in general rapport related to feelings of understanding and connection, and social presence, related to feeling that one's partner is real, present and attentive. For general rapport, we based our questions on work by Huang and colleagues (Huang et al. 2011) and Gratch and colleagues (Gratch et al. 2007) who developed a rapport scale over several iterations. We adapted their questions to create a nine Likert-scale questionnaire to capture feelings of connectedness, coordination, and understanding (see "Appendix"). Cronbach's alpha for the nine questions was .72. We averaged these nine questions to create one representative construct for general rapport, referred to simply as 'rapport' in the results.

To measure social presence, we utilized eight Likert-scale questions from the attentional allocation portion of the Networked Minds Social Presence Inventory (Biocca and Harms 2002, "Appendix"). We utilized the attentional allocation portion of the survey because attention has been identified as a critical element of both social presence and rapport (Tickle-Degnen and Rosenthal 1990). Cronbach's alpha for the eight questions was .69. We averaged the eight questions and refer to this measure as 'social presence' in the results.

For the verbal behaviors of rapport, we assessed elements of linguistic politeness, including praise, formal politeness, inclusivity, and name usage; examples of the behaviors can be found in Table 3. The detailed coding scheme is given in the "Appendix". In deciding on a coding scheme for linguistic politeness, we considered that different situations create unique interpretations for which linguistic structures are positively polite and may build rapport versus hinder rapport. Distinctions were made for any situations in which these behaviors may have been used to express negativity. This was rare; in those cases, the behavior was not included. Two individuals each independently coded for these behaviors. The average Cohen's kappa for these behaviors was .88. Individual kappas are reported in Table 2 along with the overall means and standard deviations for each condition. To assess how these behaviors differed across conditions, we aggregated them into a single representative construct of linguistic rapport where Cronbach's alpha was .70.

Table 3 Means, standard deviations, and kappa ratings for linguistic politeness behaviors across all conditions

Behavior	Example dialog	*M*	SD	*k*
Praise	"Great job", "Good answer"	2.63	4.01	.79
Politeness	"thank you", "you're welcome"	2.98	5.49	.83
Inclusive	'we' or 'lets'	26.0	19.5	.98
Name	"That's right, Quinn", "okay Quinn"	23.5	17.8	.95

4.3 Measuring rapport state

To assess rapport state, we used an input–output hidden Markov model (IOHMM), a special type of hidden Markov model, to explore how an individual's rapport state can be predicted from their use of linguistic rapport and Quinn's own social dialog. Hidden Markov models are the simplest form of a dynamic Bayesian network. In an HMM, the states are unobserved (i.e., hidden), making the HMM a useful model for estimating internal conditions such as social state which is only hinted at only by observable social cues. HMMs utilize the Markov property and assume the probability of the current state depends only on the prior state. We utilize input–output HMMs because they include one additional dependency, where the current state depends not only on the probability of the prior hidden state but also on the preceding input (for example, whether Quinn is social or not).

A complete description for how IOHMMs operate can be found by Bengio and Frasconi (1995). Like an HMM, the joint probability distribution of a given sequence of observations ($O_{1:T}$) and hidden states ($S_{1:T}$) is based on the Markov property. The distribution is given in Eq. (1) for a sequence of length T.

$$P\left(O_{1:T}, S_{1:T}\right) = P\left(O_1\right)P(S_1|O_1)\prod_{t=2}^{T} P(S_t|S_{t-1})P(O_t|S_t) \tag{1}$$

With IOHMMs, the hidden state at time t, S_t, is dependent on both the prior hidden state S_{t-1} and the prior input I_{t-1}. This primarily affects the transition probability, or the probability of a particular state given what has already occurred. The transition probability can be described by the input (I) and the prior state as shown in Eq. (2). Given the total number of input types (K) and the total number of state types (N), the transition probabilities can be broken into K separate $N \times N$ transition matrices, one for each input type. When we report the rapport states of individuals, we report the transition probabilities as K separate $N \times N$ transition matrices.

$$P(S_t|S_{t-1}, I_{t-1}) \tag{2}$$

A model of the network based on the general form of an IOHMM is given in Fig. 6. We aggregated Quinn's social dialog so that we had two input types ($K = 2$) consisting of whether Quinn speaks socially. We then analyzed the IOHMMs across the three conditions. For the hidden state, we were interested in a state indicative of whether the student was responding socially. We proposed two hidden states ($N = 2$), corresponding to whether the student was 'socially engaged' or not. We utilized two states because we did not want to overcomplicate the representation, and the measures we used are more interpretable with fewer states. In addition, a model with two states resulted in an acceptably high likelihood while keeping the number of parameters suitably smaller than the dataset. For observations, we labeled a turn as rapport building if the student used any one of the four behaviors, giving us two possible observations ($O = 2$), either linguistic rapport was present, or it was not.

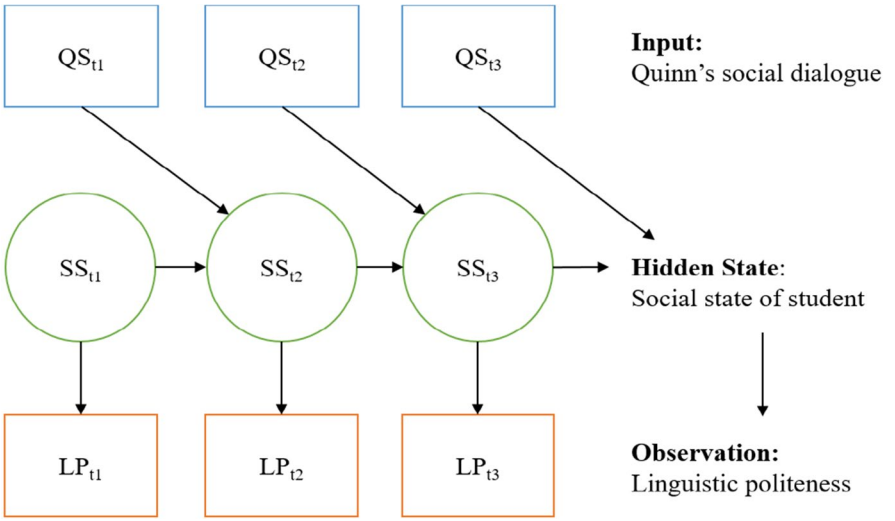

Fig. 6 Visual depiction of the IOHMM for exploring the rapport state of students

We trained the HMM for each condition (non-social, social, social plus entraining) on sequences composed of each students' turn-by-turn dialog with Quinn. Each student had taught Quinn six problems. A single sequence consisted of a single student's turn-by-turn exchange with Quinn on one problem. This resulted in 319 sequences with 2545-time slices; each time slice consisted of an observed input and output. All parameter learning was carried out using Murphy's Bayes Net Toolbox for Matlab (Murphy 2001), which uses a variation of the expectation–maximization (EM) algorithm. The likelihood manifold has local maxima, so we used multiple restarts of EM from different initial values. Using 300 restarts, we found the ten best runs, in terms of log-likelihood, resulted in values consistently within a narrow range. Additionally, we ran models for males and females across conditions considering how the underlying rapport state of genders might differ.

4.4 Measuring learning

While our interest was primarily in effects on rapport, we measured learning gain as well. Learning gains were assessed with a pretest–posttest design with an A and a B form of the test. The two forms were isomorphic and counter-balanced within condition (half of the participants in each condition received test A as the pretest with test B as the posttest and vice versa). The tests consisted of 10 questions related to literal equations, mostly procedural with some conceptual. We computed normalized learning gains according to Hake (2001) using (3) to account for prior knowledge. If the posttest scores were lower than the pretest scores, we used (4).

$$\text{gain} = (\text{posttest} - \text{pretest})/(1 - \text{pretest}) \tag{3}$$

$$\text{gain} = (\text{posttest} - \text{pretest})/\text{pretest} \tag{4}$$

4.5 Pilot evaluations

Prior to running the full study, we ran a set of 7 pilot evaluations to test the system and validate the credibility of the dialog. The pilot procedure and participants mirrored that for the full study. One of the challenges in our approach for the design of the dialog system was that it requires identifying appropriate keywords to map to responses. With the pilot studies and the dialog generated from them, we were able to validate that we had successfully mapped a sufficient number of potential keywords to responses and connected these within two to three turn dialog trees such that the dialog flow was not perceived to be unnatural and participants felt that Quinn understood them. We were also able to use the pilot studies to verify common ASR issues (i.e., 'add' might be heard as 'at') and mitigate these with context specific substitutions. Pilot participants reported that they "enjoyed teaching Quinn" and "felt like Quinn learned" suggesting that the basic goals of the interaction were being achieved. We did not explicitly ask pilot participants whether they perceived the social dialog moves to be social; however, participants who interacted with the social content tended to report in the post-interviews that Quinn was "amusing" and "fun" indicating a social perception, versus participants who interacted with the non-social version of Quinn.

5 Results

We were interested in three questions: (1) how does an entraining, social robot build rapport? (2) How do these effects differ based on the measure of rapport used? And (3) how are these effects mediated by the gender of the user? To answer these questions, we evaluated how individuals responded to a teachable robot as they interacted with the robot in one of three conditions—the robot was social, engaging in social dialog, and entrained by adapting its pitch (condition = social plus entraining), the robot was social, engaging in social dialog, but did not entrain (condition = social) or the teachable robot did not entrain and was not social (condition = non-social). In the next section, we summarize our findings from Lubold et al. (2015), which explored the effects of the robot in these three conditions on self-reported measures of rapport. In Sect. 5.2, we describe how we expanded on these initial findings by exploring the effects on behavioral rapport measured as linguistic politeness. We include an analysis of how gender mediated an individual's use of these rapport-building behaviors. We then compared the effects of an entraining robot depending on whether responses were measured as self-reported or behavioral.

5.1 Self-reported rapport

As reported in Lubold, Walker, and Pon-Barry, we utilized multivariate analysis of variance (MANOVA) to explore how gender-mediated self-reported measures to a

social, pitch-entraining teachable robot. A two-way MANOVA examining rapport and social presence as dependent variables and gender and condition as independent variables revealed significant main effects for condition (Wilks' $\lambda = .80$, $F = 4.41$, $p = .02$) and gender (Wilks' $\lambda = .77$, $F = 2.54$, $p = .04$, partial eta squared = .124). The interaction between gender and condition was not significant (Wilks' $\lambda = .85$, $F = 1.52$, $p = .21$, partial eta squared = .124). The means and standard deviations are given in Table 4.

Performing an analysis of univariate effects to understand the effect of condition, we found individuals reported significantly less social presence when Quinn was social but did not adapt its pitch, F (2, 42) = 4.0, $p = .02$. Simple pairwise comparisons of gender and social presence indicated that for males, the social condition differed significantly from both the social plus entraining ($p = .001$) and the non-social ($p = .01$) conditions, with males reporting significantly *less* social presence in the social condition. This suggests that it was the males who were driving the overall condition difference.

Analysis of the univariate effects of gender revealed that females felt significantly more rapport overall for the teachable robot than males, F (2, 42) = 8.86, $p = .006$. Regardless of the robot's behavior within conditions, females expressed more rapport for Quinn. The effect size η^2 for this difference is large at .18. In considering social presence, the difference between males and females approached significance, F (2, 42) = 3.76, $p = .06$, with males reporting less rapport overall.

To summarize, we found individuals self-reported significantly less social presence in the social condition. These findings suggest individuals found social dialog without the presence of prosodic entrainment on pitch to be less engaging, indicating consistency and balance of design is critical when incorporating social behaviors to build rapport. In the next section, we explore whether these findings are supported by behavioral measures of rapport and whether the alignment of social dialog with entrainment appeared to facilitate more rapport-building behaviors in the social plus entraining condition than in the social condition. In addition, we found females responded with significantly more rapport overall to the robot than males and that males drove the low social presence response. With the exploration of behavioral measures, we are interested in how gender-mediated behavioral responses and whether we find a further support for the positive response of females to the robot and to the robot's social behaviors.

Table 4 Means and standard deviations for social presence and rapport across genders and condition

Condition	Social presence			Rapport		
	Males	Females	All	Males	Females	All
Non-social	4.63 (.35)	4.75 (.68)	4.69 (.52)	4.58 (.61)	5.35 (1.11)	5.04 (1.06)
Social	4.05 (.74)$^{+}$	4.49 (.51)	4.27 (.65)	4.90 (.75)	5.58 (.58)	5.27 (.91)
Social plus entraining	4.88 (.29)	4.71 (.79)	4.75 (.60)	4.92 (1.36)	5.59 (.78)	5.36 (1.2)
All conditions	5.18 (.79)	5.55 (.70)	4.57 (.62)*	4.70 (.97)	5.60 (.71)**	5.22 (1.05)

$^{+}$Indicates approaching significance, *Indicates significance at $p < .05$, **Indicates significance at $p < .01$

5.2 Behavioral rapport

We explored behavioral rapport as verbal rapport-building dialog. Specifically, we looked at how individuals used four rapport-building linguistic politeness behaviors: name usage, inclusive language, praise, and formal politeness (for example, "please" or "you're welcome") while interacting with Quinn. We combined all four linguistic rapport behaviors into a single construct of linguistic rapport. We expected individuals' use of linguistic rapport to reflect similar findings as with the self-reported rapport. We conducted a two-way ANCOVA to examine the effect of condition and gender on use of linguistic rapport while controlling for dialog length. There was a statistically significant interaction between the effects of gender and condition on the presence of linguistic rapport, $F\ (2,\ 42) = 5.45$, $p = .008$. In terms of main effects, there was not a statistically significant difference in linguistic rapport for the different conditions, $F\ (2,\ 42) = 1.26$, $p = .29$. However, we did observe significant differences by gender, $F\ (1,\ 42) = 10.6$, $p = .002$. The means and standard deviations are given in Table 5.

We explored the significant interaction effect; simple main effects analysis showed that females used on average significantly fewer linguistic rapport behaviors in the social plus entraining condition as compared to both the non-social ($p = .03$) and social conditions ($p = .002$). Males, however, did not change significantly in the number of linguistic rapport behaviors they used between conditions. In addition, females used significantly more linguistic behaviors than males in the non-social ($p = .03$) and social conditions ($p = .001$).

To summarize, these findings indicated that females utilized rapport-building behaviors significantly more in the social and non-social conditions when compared to males and when compared to themselves in the social plus entraining condition. This suggests that the robot's social behavior did influence individuals' use of these behaviors but that it was mediated by gender and potentially that these behaviors may be more informative of female responses. If these behaviors are positively related to self-reported rapport, this may indicate that the entraining mechanism failed for females because they used fewer rapport-building behaviors in the social plus entraining condition. However, if linguistic rapport is negatively related to self-reported rapport, then these findings may suggest that there is a mismatch between how females self-report rapport versus their behavior. We investigated the relationship between self-reported and linguistic rapport in the next section.

Table 5 Descriptive statistics for linguistic rapport with Quinn

Linguistic rapport	Non-social	Social	Social + entraining
Females	57.9 (29.9)	72.3 (31.0)	28.3 (24.7)
Males	29.1 (17.3)	19.6 (16.2)	36.3 (32.7)
Overall	43.5 (27.9)	45.9 (36.2)	32.2 (28.3)

5.3 Relating self-reported rapport and behavioral rapport

We utilized the Pearson product-moment correlation coefficient to explore whether there was a relationship between self-reported general rapport, social presence, and the measure of linguistic rapport. We had hypothesized a relatively simple, positive relationship between self-reported measures and linguistic measures. The correlations indicate that social presence was significantly, negatively correlated with linguistic rapport, $r\,(46) = -.44$, $p = .002$; self-reported general rapport was not significantly correlated with the linguistic rapport, although it approaches a positive relationship, $r\,(46) = .23$, $p = .10$.

Looking at the correlations by gender, we found that females may be the driving force behind the significant negative correlation between social presence and linguistic rapport, $r\,(22) = -.49$, $p = .001$. For males, social presence and linguistic politeness were not correlated, $r\,(22) = .05$, $p = .80$. Males, however, did approach a significant positive relationship between general rapport and linguistic rapport, $r\,(22) = .36$, $p = .08$. Females exhibited no relationship between general rapport and linguistic politeness, $r\,(22) = -.12$, $p = .57$. Figure 7 summarizes these findings.

Gender appears to be a significant indicator of individual differences in how self-reported rapport and behavioral rapport are related. We find that when males used more praise and politeness, they self-reported feeling more rapport. Females on the other hand reported lower feelings of social presence and rapport when they used more inclusive language and Quinn's name. These findings suggest that there are individual differences present in how these verbal behaviors reflect an individual's internal rapport state and potentially how these behaviors are influenced by the robot's use of social dialog and entrainment.

5.4 Assessing rapport state

Using the IOHMM to assess rapport state, the final models suggest that there is a distinguishable hidden state associated with observing linguistic politeness. When linguistic politeness was observed, this was associated with a distinct state (S_2) which was separate from when linguistic politeness was not observed (S_1). These states were observable from the observation probabilities given in Table 6 broken out by gender. The state associated with observing linguistic politeness (S_2) was also clearly related to Quinn's social dialog and adaptation as suggested by the results of the transition probabilities, which were broken out by males and females across conditions and presented as 2 ($K = 2$), 2×2 ($N \times N$) matrices in Table 7.

Gender differences were present in how the hidden state associated with linguistic politeness manifested, particularly when Quinn was social. In the non-social when Quinn was not social nor entraining, males and females responded similarly. If they were in a state which was associated with politeness, they stayed in that state. However, in the social condition, when Quinn exhibited social dialog but did not adapt, we began to see a difference in how males and females responded. For males, if Quinn was not social, males would either move to a non-linguistic-politeness state or they would stay

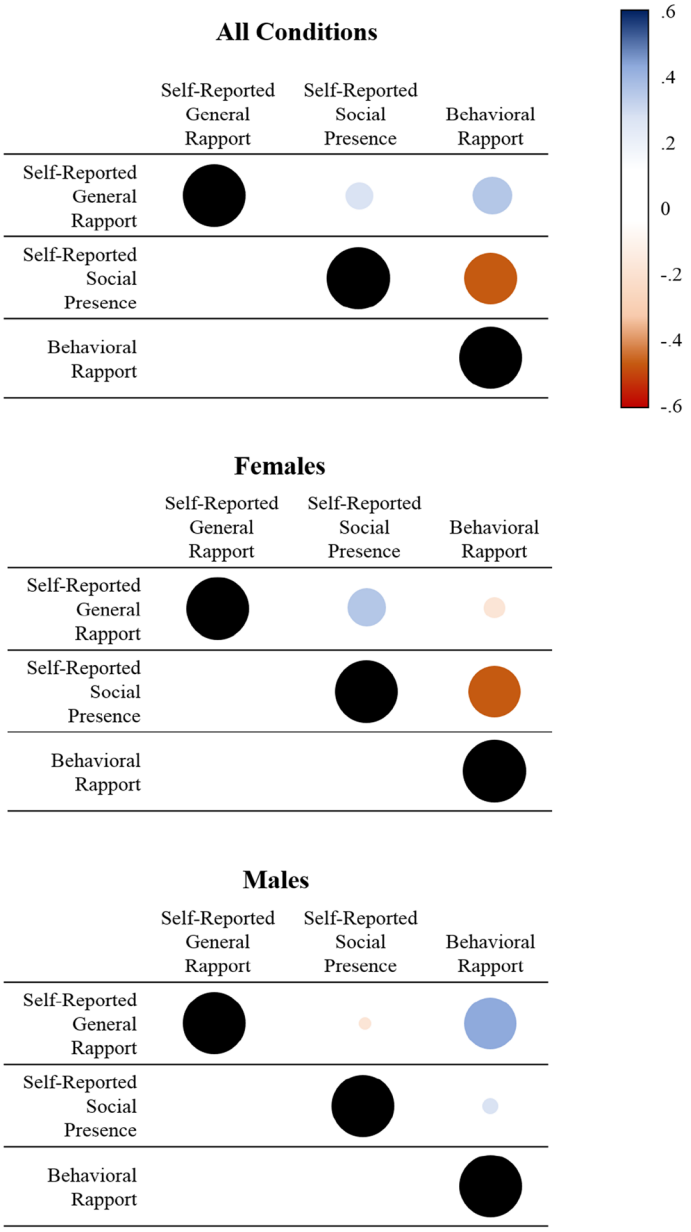

Fig. 7 Correlations of self-reported and behavioral rapport as linguistic politeness

in a non-linguistic-politeness state. If Quinn was social and the male student was already being polite, the male student was likely to remain polite. If they were not exhibiting politeness already and Quinn was social, a male student had a 50–50 chance of moving to a state associated with politeness. Females on the other hand were more likely to move

Table 6 Observation probabilities for the IOHMM

Student is…	Non-social		Social		Social + entraining	
	S_1	S_2	S_1	S_2	S_1	S_2
Males						
Not exhibiting rapport	**.92**	.13	**.95**	.11	**.75**	.12
Exhibiting rapport	.08	**.87**	.05	**.89**	.25	**.88**
Females						
Not exhibiting rapport	**.77**	.06	**.88**	.18	**.92**	.12
Exhibiting rapport	.23	**.94**	.12	**.82**	.08	**.88**

Bolded values indicate the dominant transition probability for transitioning to that state

S_1 and S_2 represent the two hidden states

Table 7 Transition matrices for IOHMM

Student is…	Quinn is…	State	Non-social		Social		Social + entraining	
			S_1	S_2	S_1	S_2	S_1	S_2
Male	Not social ($K=1$)	S_1	**.95**	.15	**.95**	**.93**	**.96**	**.92**
		S_2	.05	**.85**	.05	.07	.04	.08
	Social ($K=2$)	S_1	–	–	.45	.12	.05	.08
		S_2	–	–	.55	**.88**	**.95**	**.92**
Female	Not social ($K=1$)	S_1	**.95**	.01	.12	.15	.22	.26
		S_2	.05	**.90**	**.88**	**.85**	**.78**	**.74**
	Social ($K=2$)	S_1	–	–	**.66**	**.86**	**.92**	**.99**
		S_2	–	–	.34	.14	.08	.01

Bolded values indicate the dominant transition probability for transitioning to that state

S_1 and S_2 represent the two hidden states

to a state characterized by linguistic politeness when Quinn was NOT social. If Quinn was social, female students were more likely to move to a non-linguistic-politeness state. We saw these patterns intensify when Quinn was social and adapted. Males were more likely to move to a state associated with observed linguistic politeness when Quinn was social and adapted. Female students were more likely to move to a state which was NOT associated with linguistic politeness when Quinn was social and adapted. These results suggest that (1) a hidden state exists which is associated with linguistic rapport and is clearly influenced by Quinn's social behaviors, (2) how this hidden state manifested and the effects of Quinn's social behaviors on it are strongly mediated by gender and (3) males and females entered this underlying state on different social triggers.

5.5 Learning gains

Analyzing learning as gain, we found 10 individuals at zero gain, 10 individuals who gained in a normal distribution, and 23 hitting full gain. The descriptive statistics are given in Table 8. We determined analysis would be better served by grouping the learners into three groups—no gain, some gain and all gain. Having grouped the students into three learning groups, we analyzed the learning gains in terms of a multinomial logistic regression. However, even with this adjustment, the overall model in the analysis including both condition and gender was not significant, $X^2(6) = 6.86$, $p = .33$, and we found that none of the individual predictors are significant.

We assessed social presence and rapport in terms of learning. Running a multinomial regression with rapport and social presence, we found the model was not significant, $X^2(4) = 4.68$, $p = .32$. However, in viewing the individual coefficients, social presence does approach a significant effect on learning ($p = .06$). For those individuals who gained but did not hit ceiling on their gain, social presence is 1.38 times higher than for those individuals who did not gain.

6 Discussion and conclusions

6.1 Discussion

Forty-eight college students interacted with the teachable robot Quinn in one of three conditions, a social condition where the robot utilized social dialog, a social plus entraining condition where the robot spoke socially and entrained using the pitch adaptation, and a non-social condition where the robot neither spoke socially nor entrained. We were interested in three research questions:

RQ1 How does an entraining, social robot build rapport?

RQ2 How do these effects differ based on the measure of rapport used?

RQ3 How are these effects mediated by the gender of the user?

Table 8 Descriptive statistics for learning gains

Condition	Learning		
	Males	Females	All
Non-social	.72 (.44)	.83 (.41)	.81 (.33)
Social	.34 (.48)	.73 (.48)	.50 (.54)
Social plus entraining	.53 (.51)	.60 (.43)	.56 (.45)
All conditions	.53 (.48)	.71 (.43)	.62 (.46)

*Significant at $p < .05$, **Significant at $p < .01$

This work demonstrated that entrainment can be modeled as a form of turn-by-turn pitch adaptation and that an entraining, social robot can have a positive impact on rapport. We provided one of the first in-depth descriptions of a system that implements adaptation on pitch based on the phenomenon of entrainment, and we performed an analysis of the effects on spoken interaction with a robot. People entrain on many features in addition to pitch and in many different ways besides turn-by-turn. Our introduction of adapting to users based on the entrainment phenomenon paves the way for future implementations and explorations of adaptation mechanisms based on entrainment. Several interesting observations emerge from the exploration of pitch proximity with a social, robotic learning companion which tie back to our research questions regarding how a social robot that adapts its pitch might build rapport, how different measures of rapport reflect different responses to social behavior, and the role of gender in those responses.

First, it appears that entrainment as pitch adaptation improved perceptions of social dialog. When social dialog was present without the pitch adaptation, individuals perceived the robot as significantly less socially present. Our results suggest that prosodic manipulation as a form of entrainment may have served to enhance the positive perception of social dialog while social dialog without prosodic manipulation decreased perceptions of Quinn's social presence. In prior works, social dialog has been shown to build rapport but, in this study, social dialog unexpectedly produced the lowest responses, even lower than no social behavior at all. Individuals reported significantly lower social presence in the social condition, and we found individuals increased linguistic rapport behaviors negatively correlated with social presence in the social condition while individuals in the social + entraining condition reported the highest feelings of social presence and rapport. We designed the social dialog based on dialog moves validated as social in prior work; however, the social dialog did not have the desired outcome. These findings suggest a single channel of designed social behavior can fail where two channels can succeed, and strongly supports critical design considerations when incorporating multiple channels of social behavior to facilitate rapport. Other work has indicated that the misalignment of behavior can harm perceptions (Meena et al. 2012; Lubold et al. 2018). The results we have found here suggest that facilitating alignment through pitch proximity can potentially improve social responses. While we did not observe that pitch proximity was able to improve social responses beyond the non-social control in this work, there is evidence in other work that alternative forms of pitch adaptation can improve social responses above and beyond a non-social control (Lubold 2018). These findings on entrainment and social behavior demonstrate the promise of entrainment and the potential adapting to users based on the entrainment phenomenon might have for enhancing social responses.

Regarding our second and third research questions, the interaction results with Quinn provide interesting insights into human–human and human–agent interactions, particularly regarding gender and the degree to which gender indicates individual differences present in social responses. We observed that behavioral rapport measured as linguistic rapport was negatively correlated with perceptions of social presence, particularly for females, and females engaged in significantly *more* linguistic rapport in the social condition. These results suggest that perceptions and

behaviors reflecting rapport are not always aligned, but this relationship between different measures may be mediated by individual differences as indicated by gender. Our findings also indicate that Quinn's social dialog influenced how individuals engaged in linguistic rapport as mediated by gender. It is not surprising that males and females might respond differently to social behavior from a robot and exhibit different linguistic rapport when we consider that females have been found to respond more positively to social behavior from a robots and agents and that females in general tend to use linguistic rapport (Burleson and Picard 2007; Chalupnik et al. 2017; Brown 1980). Compared to the males, females felt significantly more rapport for the robot overall. They also changed how they used linguistic rapport across the different conditions, using more linguistic behaviors associated with rapport in the non-social and social conditions. We found females used these verbal behaviors when they felt less rapport as opposed to more rapport for Quinn, and they were more likely to stay or move to a social state (i.e., states not associated with linguistic rapport) when Quinn was not engaging in social dialog. It is possible females in the study increased linguistic rapport behaviors when they felt less rapport because they were attempting to increase rapport and build a relationship where they currently did not sense one. According to prior work, women are more likely to see conversation as a means for building rapport (Tannen 1994; Hong and Hwang 2012). While females used these verbal behaviors to build a relationship, males may have utilized these verbal behaviors as relationship indicators, engaging in linguistic rapport only once a positive relationship had been initiated, confirmed, and pushed by their conversational partner. This would suggest that for males, linguistic rapport emerged because they felt rapport, not because they were trying to build rapport. We observed in analyses of rapport state that verbal behavioral measures were not indicative of the same underlying rapport state for all individuals and this rapport state manifested differently in response to social triggers. This interpretation suggests that we may want to assess responses differently when observing linguistic behavior from different individuals.

An alternative interpretation is that males and females had different initial social inclinations toward Quinn, and these initial inclinations resulted in different rapport responses. We measured the linguistic rapport behaviors based on theories of rapport and politeness. In these theories, politeness to build rapport is more commonly associated with initial encounters with strangers. As individuals get to know one another, rapport increases and politeness decreases. The longer people know each other the less polite they tend to be and the more rapport they tend to feel. We observed females used fewer of the rapport behaviors in the social plus entraining condition. Females may have been more comfortable with viewing Quinn as a 'teachable' entity that could learn, being more likely to anthropomorphize Quinn and expect Quinn to be social. As a result, when Quinn engaged in social dialog and entrained, females were more likely to accept Quinn's social behavior as genuine and treat Quinn as a friend, dropping the social niceties of linguistic politeness typically used with strangers. In contrast, males followed a more traditional path. Feeling less rapport in general, males were less comfortable with Quinn. Quinn was a 'stranger' that they could potentially develop rapport with, but they did not feel as if Quinn was their 'friend.' We observed some evidence of these attitudes in the post-interviews, where females were more likely to refer to Quinn

as "my friend" and "we're best friends now," and males were more likely to describe Quinn as "an interesting robot" and "decently complex." This suggests that individuals who are more prone to social interaction with a robot will respond with more familiarity to a robot's social behaviors; they will be more inclined to increased rapport overall, and this will impact their behavioral responses accordingly. If this interpretation is accurate, this has implications for the design of social interaction for different individuals—for those who are more inclined to social behavior, social interaction models may move to more quickly to familiar behavior than for users who are less inclined to social interaction.

6.2 Limitations

Limitations of our work include the subject pool which consisted of college-aged, native English speakers; this potentially limits our findings to interactions within this group though we believe our findings can inform and be used to explore prosodic adaptation with other groups. In addition, the ceiling effects on the posttest make it difficult to generalize our learning results to other contexts.

Another limitation is that we explore adaptation based on entrainment for only one feature, pitch; however, individuals often entrain on more than one feature and in more ways than the simple approach modeled here. This limits our conclusions to what can be said about entrainment on pitch specifically versus entrainment more broadly. Implementing acoustic-prosodic entrainment is still highly novel in dialog systems, with only a few other examples (Aneja et al. 2019; Kory-Westlund and Breazeal 2019; Levitan et al. 2016) and thus the roadmap for how best to do so is not clear. We chose pitch as an initial step toward implementing more human-like forms of entrainment because it has been shown to correlate with a variety of outcomes of interest, including engagement (Gravano et al. 2014), solidarity (Gregory et al. 1997), rapport (Lubold and Pon-barry 2014), and positive affect (Lee et al. 2012). In addition, our early explorations of pitch-based entrainment suggested that we can implement entrainment that is both perceived as natural and rapport-building (Lubold et al. 2015). We hope our work lays the foundation for the study and implementation of more complex forms of entrainment.

A final limitation of our work is our use of a rather simple robot. Quinn was constructed using a digital face displayed on an iPod Touch and mounted on a physical base. Comparisons of the use of a robot to a virtual agent that behaves in the exact same way have found that robots promote increased attention (Looije et al. 2012), time on task [Hsu], enjoyment of the interaction (Kose-Bagci et al. 2009), positive attitudes toward the robot (Powers et al. 2007), and empathy for the robot (Seo et al. 2015), simply because of their physical presence. Given this related work, we hypothesized that because Quinn had a physical body, and it might enhance the rapport and responsibility students' felt for their agent, increasing the effects of our social manipulations. Future work could test this hypothesis in a controlled study. Nevertheless, because Quinn did not gesture or move around over the course of the study, our results may not generalize to robots that have more channels of communication available to them.

6.3 Conclusions

The overall results of this work highlight the complexities inherent in measuring rapport, the benefits of using both self-reported rapport and observable rapport, and the potential advantages of modeling a user's rapport state based on their observable behavior. We found males and females responded very similarly to the conditions, but this was not immediately obvious from their self-reported scores. The work suggests that self-reported scores may be more informative for some individuals than for others and the addition of verbal behaviors can provide more insight into those individuals for whom the self-report is less informative. In addition, self-reported rapport measures were not aligned in the same direction as the verbal rapport behaviors we collected, particularly for females. The results emphasize the importance of assessing social responses like rapport from multiple dimensions and that when using verbal behaviors to gain insight into an individual's underlying feelings, individual differences such as gender should be considered because the underlying state indicated by their verbal behavior may manifest differently.

We have established that a social, entraining robot can produce complex self-reported and linguistic responses rooted in individual differences. Combining social dialog and prosodic entrainment can lead to higher self-reported measures of rapport, but individual differences mediate both self-reported and linguistic responses. This is important to consider in how we might 'stereotype' users based on their individual differences, from implementing social behavior that is adaptive to users to measuring how users respond. Future work will focus on exploring whether the interpretations suggested here regarding the individual differences indicated by gender are due more to differences how language is used to build relationships or if it is due to the degree to which individuals are comfortable with social interaction from a robot. Additional analyses of other combinations of multiple behaviors will also be explored to expand on the possibilities of multi-channel behavior to build rapport.

Funding This work was supported by the National Robotics Initiative and the National Science Foundation, Grant # CISE-IIS-1637809.

Compliance with ethical standards

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

Ethical approval This paper is void of plagiarism or self-plagiarism as defined by the Committee on Publication Ethics and Springer Guidelines.

Appendix

Rapport Measures: Participants responded on a Likert scale from 1 to 5

I felt I had a connection with Quinn
I felt I was able to engage Quinn
I think that Quinn and I understood each other

I felt that Quinn was interested in what I had to say
I felt that Quinn was warm and caring
I felt that Quinn was intensely involved in the interaction
I felt that Quinn seemed to find the interaction stimulating
I felt that Quinn was respectful to me
I felt that Quinn showed enthusiasm while talking to me

Social Presence Measures: Participants responded on a Likert scale from 1 to 7
Quinn was easily distracted
I was easily distracted
Quinn tended to ignore me
I tended to ignore Quinn
I sometimes pretend to pay attention to Quinn
Quinn sometimes pretended to pay attention to me
Quinn paid close attention to me

Coding Scheme:

Politeness: "P" is polite to Quinn, follows conversational niceties (like saying hello)

Ex 1: Thank you, Quinn
Ex 2: ah step four please

Complimenting or praising: "P" praises Quinn

Ex 1: good job Quinn
Ex 2: great! Now I factor out the two
Ex 3: nice!

Name usage: "P" uses Quinn's name

Ex 1: Nice job Quinn (this would contain checks in both the praise column and the name column)

Inclusive: "P" includes Quinn, for example by using 'inclusive' language such as "us," "we,", "together", "let's"

Ex 1: Let's do problem one!

References

ALICE: A.L.I.C.E AI Foundation (2002). http://www.alicebot.org/. Accessed 21 Apr 2015
Aneja, D., Hoegen, R., McDuff, D., Czerwinski, M.: Designing style matching conversational agents. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM (2019)

Arroyo, I., Burleson, W., Tai, M., Muldner, K., Woolf, B.P.: Gender differences in the use and benefit of advanced learning technologies for mathematics. J. Educ. Psychol. **105**(4), 957–969 (2013). https://doi.org/10.1037/a0032748

Babel, M.: Dialect divergence and convergence in New Zealand English. Lang. Soc. **39**, 437–456 (2010)

Babel, M.: Evidence for phonetic and social selectivity in spontaneous phonetic imitation. J. Phon. **40**, 177–189 (2011)

Babel, M., Bulatov, D.: The role of fundamental frequency in phonetic accommodation. Lang. Speech **55**(2), 231–248 (2012)

Bales, R.F.: A set of categories for the analysis of small group interaction. Am. Sociol. Rev. **15**(2), 257–263 (1950)

Beal, C., Mitra, S., Cohen, P.: Modeling learning patterns of students with a tutoring system using Hidden Markov Model. In: Luckin, R., et al. (eds.) Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED), pp. 238–245. Marina del Rey (2007)

Bechade, L., Dubuisson Duplessis, G., Sehili, M., Devillers, L.: Behavioral and emotional spoken cues related to mental states in human–robot social interaction. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, November 2015, pp. 347–350. ACM (2015)

Bell, L., Gustafson, J., Heldner, M.: Prosodic adaptation in human–computer interaction. In: Proceedings of ICPHS, vol. 3 (2003)

Bell, D., Arnold, H., Haddock, R.: Linguistic politeness and peer tutoring. Learn. Assist. Rev. **14**(1), 37–54 (2009)

Bengio, Y., Frasconi, P.: An input output HMM architecture. In: Advances in Neural Information Processing Systems, pp. 427–434 (1995)

Beňuš, Š., Levitan, R., Hirschberg, J., Gravano, A., Darjaa, S.: Entrainment in Slovak collaborative dialogues. In: 2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom), November 2014, pp. 309–313. IEEE (2014)

Beňuš, Š., Levitan, R., Hirschberg, J.: Entrainment in spontaneous speech: the case of filled pauses in Supreme Court hearings. In: 3rd IEEE Conference on Cognitive Infocommunications, Kosice, Slovakia (2012)

Beňuš, Š.: Are we in sync': turn-taking in collaborative dialogues. In: Tenth Annual Conference of the International Speech Communication Association (2009)

Bergner, Y., Walker, E., Ogan, A.: Dynamic Bayesian network models for peer tutoring interactions. In: Innovative Assessment of Collaboration, pp. 249–268. Springer, Cham (2017)

Bickmore, T.W.: Relational Agents: Effecting Change Through Human-Computer Relationships (Doctoral dissertation, Massachusetts Institute of Technology) (2003)

Bickmore, T., Cassell, J.: Relational agents: a model and implementation of building user trust. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 396–403 (2001)

Bickmore, T.W., Picard, R.W.: Establishing and maintaining long-term human-computer relationships. ACM Trans. Comput.-Human. Interact. (TOCHI) **12**(2), 293–327 (2005)

Bickmore, T.W., Vardoulakis, L.M.P., Schulman, D.: Tinker: a relational agent museum guide. Auton. Agent. Multi-Agent Syst. **27**(2), 254–276 (2013)

Biocca, F., Harms, C.: Defining and measuring social presence: contribution to the networked minds theory and measure. In: Proceedings of PRESENCE, 2002, pp. 1–36 (2002)

Boersma, P.: Praat: doing phonetics by computer. http://www.praat.org/ (2006). Accessed 31 Mar 2014

Bone, D., Lee, C.C., Chaspari, T., Black, M.P., Williams, M.E., Lee, S., Levitt, P., Narayanan, S.: Acoustic-prosodic, turn-taking, and language cues in child–psychologist interactions for varying social demand. In: INTERSPEECH-2013, pp. 2400–2404 (2013)

Bonin, F., De Looze, C., Ghosh, S., Gilmartin, E., Vogel, C., Polychroniou, A., Salamin, H., Vinciarelli, A., Campbell, N.: Investigating fine temporal dynamics of prosodic and lexical accommodation. In: INTERSPEECH-2013, pp. 539–543 (2013)

Borrie, S.A., Lubold, N., Pon-Barry, H.: Disordered speech disrupts conversational entrainment: a study of acoustic-prosodic entrainment and communicative success in populations with communication challenges. Front. Psychol. **6**, 1187 (2015)

Boyer, K.E., Phillips, R., Ingram, A., Ha, E.Y., Wallis, M., Vouk, M., Lester, J.: Characterizing the effectiveness of tutorial dialogue with hidden markov models. In: International Conference on Intelligent Tutoring Systems, June 2010, pp. 55–64. Springer, Berlin, Heidelberg (2010)

Breazeal, C.: Emotion and sociable humanoid robots. Int. J. Human-Comput. Stud. **59**(1–2), 119–155 (2003)

Brown, P.: How and why are women more polite: some evidence from a Mayan community. In: McConnell-Ginet, S., Borker, R., Furman, N. (eds.) Women and Language in Literature and Society, pp. 111–136. Praeger, New York (1980)

Brown, P., Levinson, S.: Politeness. Some Universals in Language Usage. CUP, Cambridge (1987). Originally published as Universals in language usage: politeness phenomenon. In: Goody, E. (ed.) Questions and Politeness: Strategies in Social Interaction. CUP, Cambridge (1978)

Burleson, W., Picard, R.W.: Gender-specific approaches to developing emotionally intelligent learning companions. IEEE Intell. Syst. **22**(4), 62–69 (2007). https://doi.org/10.1109/MIS.2007.69

Cassell, J., Bickmore, T.: Negotiated collusion: modeling social language and its relationship effects in intelligent agents. User Model. User-Adap. Inter. **13**(1–2), 89–132 (2003)

Chalupnik, M., Christie, C., Mullany, L.: (Im)politeness and gender. In: Culpeper, J., Haugh, M., Kádár, D. (eds.) The Palgrave Handbook of Linguistic (Im)politeness, pp. 517–537. Palgrave Macmillan, London (2017)

Chidambaram, V., Chiang, Y.-H., Mutlu, B.: Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In: 7th Annual ACM/IEEE International Conference on Human–Robot Interaction (HRI'12), pp. 293–300 (2012). http://doi.org/10.1145/2157689.2157798

Csapo, A., Gilmartin, E., Grizou, J., Han, J., Meena, R., Anastasiou, D., Jokinen, K., Wilcock, G.: Multimodal conversational interaction with a humanoid robot. In: 2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom), pp. 667–672. IEEE (2012)

Coates, J.: Women, Men and Language: A Sociolinguistic Account of Gender Differences in Language. Routledge, Abingdon (2015)

Darves, C., Oviatt, S.: Adaptation of users' spoken dialogue patterns in a conversational interface. In: Seventh International Conference on Spoken Language Processing (2002)

De Carolis, B., Ferilli, S., Palestra, G., Carofiglio, V.: Modeling and simulating empathic behavior in social assistive robots. In: Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter, September 2015, pp. 110–117. ACM (2015)

Drolet, A.L., Morris, M.W.: Rapport in conflict resolution: accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. J. Exp. Soc. Psychol. **36**(1), 26–50 (2000)

Foster, M.E., Gaschler, A., Giuliani, M.: How can I help you': comparing engagement classification strategies for a robot bartender. In: Proceedings of the 15th ACM on International Conference on Multimodal Interaction, December 2013, pp. 255–262. ACM (2013)

Giles, H., Smith, P.: Accommodation theory: optimal levels of convergence. In: Howard, G., Robert, N. (eds.) Language and Social Psychology, pp. 45–65. St Clair (1979)

Gordon, G., Spaulding, S., Westlund, J.K., Lee, J.J., Plummer, L., Martinez, M., Breazeal, C., et al.: Affective personalization of a social robot tutor for children's second language skills. In: Thirtieth AAAI Conference on Artificial Intelligence, March 2016 (2016)

Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R.: Creating rapport with virtual agents. In: International Workshop on Intelligent Virtual Agents, September 2007, pp. 125–138. Springer, Berlin, Heidelberg (2007)

Gravano, A., Benus, S., Levitan, R., Hirschberg, J.: Three ToBI-based measures of prosodic entrainment and their correlations with speaker engagement. In: 2014 IEEE Spoken Language Technology Workshop (SLT), December 2014, pp. 578–583. IEEE (2014)

Gregory, S.W., Dagan, K., Webster, S.: Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. J. Nonverbal Behav. **21**(1), 23–43 (1997)

Gulz, A., Haake, M., Silvervarg, A.: Extending a teachable agent with a social conversation module—effects on student experiences and learning. In: International Conference on Artificial Intelligence in Education, June 2011, pp. 106–114. Springer, Berlin, Heidelberg (2011)

Gweon, G., Jain, M., McDonough, J., Raj, B., Rosé, C.P.: Measuring prevalence of other-oriented transactive contributions using an automated measure of speech style accommodation. Int. J. Comput. Support. Collab. Learn. **8**(2), 245–265 (2013)

Hess, U., Blairy, S.: Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. Int. J. Psychophysiol. **40**(2), 129–141 (2001)

Holmes, J.: Women, Men and Politeness. Longman, London (1995)

Hong, J.-C., Hwang, M.-Y.: Gender differences in help-seeking and supportive dialogue during online game. Procedia Soc. Behav. Sci. **64**(2007), 342–351 (2012). https://doi.org/10.1016/j.sbspro.2012.11.041

Huang, L., Morency, L.P., Gratch, J.: Virtual rapport 2.0. In: International Workshop on Intelligent Virtual Agents, September 2011, pp. 68–79. Springer, Berlin, Heidelberg (2011)
Jacq, A., Lemaignan, S., Garcia, F., Dillenbourg, P., Paiva, A.: Building successful long child–robot interactions in a learning context. In: 2016 11th ACM/IEEE International Conference on Human–Robot Interaction (HRI), March 2016, pp. 239–246. IEEE (2016)
Jokinen, K., Hurtig, T.: User expectations and real experience on a multimodal interactive system. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, January 2006, vol. 2, pp. 1049–1052 (2006)
Kanda, T., Hirano, T., Eaton, D., Ishiguro, H.: Interactive robots as social partners and peer tutors for children: a field trial. Hum. Comput. Interact. **19**(1), 61–84 (2004)
Kasap, Z., Magnenat-Thalmann, N.: Building long-term relationships with virtual and robotic characters: the role of remembering. Vis. Comput. **28**(1), 87–97 (2012)
Kasap, Z., Magnenat-Thalmann, N.: Towards episodic memory-based long-term affective interaction with a human-like robot. In: 19th International Symposium in Robot and Human Interactive Communication, September 2010, pp. 452–457. IEEE (2010)
Kory-Westlund, J.M., Breazeal, C.: Exploring the effects of a social robot's speech entrainment and backstory on young children's emotion, rapport, relationship, and learning. Front. Robot. AI **6**, 54 (2019)
Kose-Bagci, H., Ferrari, E., Dautenhahn, K., Syrdal, D.S., Nehaniv, C.L.: Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot. Adv. Robot. **23**(14), 1951 (2009). https://doi.org/10.1163/016918609X12518783330360
Kousidis, S., Dorran, D.: Monitoring convergence of temporal features in spontaneous dialogue speech. In: First Young Researchers Workshop on Speech Technology, Dublin, Ireland, January 2009 (2009)
Krämer, N.C., Karacora, B., Lucas, G., Dehghani, M., Rüther, G., Gratch, J.: Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction. Comput. Educ. **99**, 1–13 (2016)
Kumar, R., Ai, H., Beuth, J.L., Rosé, C.P.: Socially capable conversational tutors can be effective in collaborative learning situations. In: Aleven, V., Kay, J., Mostow, J. (eds.) Intelligent Tutoring Systems, pp. 156–164. Springer, Berlin (2010)
Lakin, J.L., Chartrand, T.L.: Using nonconscious behavioral mimicry to create affiliation and rapport. Psychol. Sci. **14**(4), 334–339 (2003)
Lakens, D., Stel, M.: If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity. Soc. Cognit. **29**(1), 1–14 (2011)
Lee, C.C., Katsamanis, A., Black, M.P., Baucom, B.R., Georgiou, P.G., Narayanan, S.S.: An analysis of PCA-based vocal entrainment measures in married couples' affective spoken interactions. In: Proceedings of Interspeech, Florence, Italy (2011)
Lee, M.K., Forlizzi, J., Kiesler, S., Rybski, P., Antanitis, J., Savetsila, S.: Personalization in HRI: a longitudinal field experiment. In: 2012 7th ACM/IEEE International Conference on Human–Robot Interaction (HRI), March 2012, pp. 319–326. IEEE (2012)
Lee, N., Shin, H., Sundar, S.S.: Utilitarian vs. hedonic robots: role of parasocial tendency and anthropomorphism in shaping user attitudes. In: Proceedings of the 6th International Conference on Human–Robot Interaction, March 2011, pp. 183–184 (2011)
Leelawong, K., Biswas, G.: Designing learning by teaching agents: the Betty's Brain system. Int. J. Artif. Intell. Educ. **18**(3), 181–208 (2008)
Levitan, R., Hirschberg, J.: Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In: Interspeech 2011 (2011)
Levitan, R., Benus, S., Galvez, R.H., Gravano, A., Savoretti, F., Trnka, M., Hirschberg, J.: Implementing acoustic-prosodic entrainment in a conversational avatar. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 08–12 September 2016, pp. 1166–1170 (2016). http://doi.org/10.21437/Interspeech.2016-985
Levitan, R., Beňuš, Š., Gravano, A., Hirschberg, J.: Acoustic-prosodic entrainment in Slovak, Spanish, English and Chinese: a cross-linguistic comparison. In: Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 325–334 (2015)
Levitan, R., Gravano, A., Willson, L., Benus, S., Hirschberg, J., Nenkova, A.: Acoustic-prosodic entrainment and social behavior. In: Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, June 2012, pp. 11–19. Association for Computational Linguistics (2012)

Longster, J.: Concatenative speech synthesis: a framework for reducing perceived distortion when using the TD-PSOLA algorithm. Dissertation, Bournemouth University (2003)

Looije, R., van der Zalm, A., Neerincx, M.A., Beun, R.J.: Help, I need somebody: the effect of embodiment on playful learning, pp. 718–724. IEEE (2012). http://dx.doi.org/10.1109/ROMAN.2012.6343836

Lubold, N., Pon-Barry, H., Walker, E.: Naturalness and rapport in a pitch adaptive learning companion. In: 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), December 2015, pp. 103–110. IEEE (2015)

Lubold, N., Pon-Barry, H.: Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. In: Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge, November 2014, pp. 5–12 (2014)

Lubold, N., Walker, E., Pon-Barry, H., Ogan, A.: Automated pitch convergence improves learning in a social, teachable robot for middle school mathematics. In: International Conference on Artificial Intelligence in Education, pp. 282–296. Springer, Cham (2018)

Lubold, N.: Producing acoustic-prosodic entrainment in a robotic learning companion to build learner rapport. Doctoral dissertation, Arizona State University (2018)

Lutfi, S., Fernández-Martínez, F., Lorenzo-Trueba, J., Barra-Chicote, R., Montero, J.: I feel you: the design and evaluation of a domotic affect-sensitive spoken conversational agent. Sensors **13**(8), 10519–10538 (2013)

Meena, R., Jokinen, K., Wilcock, G.: Integration of gestures and speech in human–robot interaction. In: 2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom), December 2012, pp. 673–678. IEEE (2012)

Mehrabian, A.: Nonverbal Communication. Transaction Publishers, Piscataway (1972)

Murphy, K.: The Bayes net toolbox for Matlab. Comput. Sci. Stat. **33**(2), 1024–1034 (2001)

Mushin, I., Stirling, L., Fletcher, J., Wales, R.: Discourse structure, grounding, and prosody in task-oriented dialogue. Discourse Process. **35**(1), 1–31 (2003)

Natale, M.: Convergence of mean vocal intensity in dyadic communication as a function of social desirability. J. Pers. Soc. Psychol. **32**(5), 790 (1975)

Nenkova, A., Gravano, A., Hirschberg, J.: High frequency word entrainment in spoken dialogue. In: Proceedings of ACL-08: HLT, Short Papers, pp. 169–172 (2008)

Novick, D., Gris, I.: Building rapport between human and ECA: a pilot study. In: International Conference on Human–Computer Interaction, June 2014, pp. 472–480. Springer, Cham (2014)

Nwe, T.L., Foo, S.W., De Silva, L.C.: Speech emotion recognition using hidden Markov models. Speech Commun. **41**(4), 603–623 (2003)

Ogan, A., Finkelstein, S., Walker, E., Carlson, R., Cassell, J.: Rudeness and rapport: insults and learning gains in peer tutoring. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) Intelligent Tutoring Systems, pp. 11–21. Springer, Berlin (2012)

Pantic, M., Pentland, A., Nijholt, A., Huang, T.S.: Human computing and machine understanding of human behavior: a survey. In: Artificial Intelligence for Human Computing, pp. 47–71. Springer, Berlin, Heidelberg (2007)

Powers, A., Kiesler, S., Fussell, S., Torrey, C.: Comparing a computer agent with a humanoid robot. In: Proceedings of the ACM/IEEE International Conference on Human–Robot Interaction, March 2007, pp. 145–152 (2007)

Sadoughi, N., Pereira, A., Jain, R., Leite, I., Lehman, J.F.: Creating prosodic synchrony for a robot co-player in a speech-controlled game for children, pp. 91–99 (2017). https://doi.org/10.1145/2909824.3020244

Saerbeck, M., Schut, T., Bartneck, C., Janse, M.D.: Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1613–1622. ACM (2010)

Saint-Aimé, S., Le-Pevedic, B., Duhaut, D., Shibata, T.: EmotiRob: companion robot project. In: RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication, pp. 919–924. IEEE (2007)

Schermerhorn, P., Scheutz, M., Crowell, C.R.: Robot social presence and gender: do females view robots differently than males?. In: Proceedings of the 3rd ACM/IEEE International Conference on Human–Robot Interaction, March 2008, pp. 263–270 (2008)

Schweitzer, A., Lewandowski, N.: Convergence of articulation rate in spontaneous speech. In: INTERSPEECH, August 2013, pp. 525–529 (2013)

Scissors, L.E., Gill, A.J., Geraghty, K., Gergle, D.: In CMC we trust: the role of similarity. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, April 2009, pp. 527–536. ACM (2009)
Seo, S.H., Geiskkovitch, D., Nakane, M., King, C., Young, J.E.: Poor thing! Would you feel sorry for a simulated robot?. In: Proceedings of International Conference on Human–Robot Interaction—HRI'15, pp. 125–132. ACM (2015). https://doi.org/10.1145/2696454.2696471
Shawar, B.A., Atwell, E.: A comparison between Alice and Elizabeth chatbot systems. University of Leeds, School of Computing research report 2002.19 (2002)
Shawar, B.A., Atwell, E.: Chatbots: are they really useful?. In: Ldv Forum, January 2007, vol. 22, no. 1, pp. 29–49 (2007)
Sidaras, S.K.: Hearing what you expect to hear: the interaction of social and cognitive mechanisms underlying vocal accommodation. Doctoral dissertation, Emory University (2011)
Sinha, T., Cassell, J.: We click, we align, we learn: impact of influence and convergence processes on student learning and rapport building. In: Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence, pp. 13–20 (2015). http://doi.org/10.1145/2823513.2823516
Spencer-Oatey, H.: (Im)Politeness, face and perceptions of rapport: unpackaging their bases and interrelationships. Polit. Res. **1**(1), 95–119 (2005). https://doi.org/10.1515/jplr.2005.1.1.95
Stewart, M., Brown, J.B., Boon, H., Galajda, J., Meredith, L., Sangster, M.: Evidence on patient–doctor communication. Cancer **25**(1999), 30 (1999)
Strait, M., Briggs, P., Scheutz, M.: Gender, more so than age, modulates positive perceptions of language-based human–robot interactions. In: 4th International Symposium on New Frontiers in Human Robot Interaction, April 2015, pp. 21–22 (2015)
Tanaka, F., Matsuzoe, S.: Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. J. Hum. Robot Interact. **1**(1), 78–95 (2012). https://doi.org/10.5898/JHRI.1.1.Tanaka
Tannen, D.: Gender and Discourse. Oxford University Press, Oxford (1994)
Thomason, J., Nguyen, H.V., Litman, D.: Prosodic entrainment and tutoring dialogue success. In: International Conference on Artificial Intelligence in Education, July 2013, pp. 750–753. Springer, Berlin, Heidelberg (2013)
Tickle-Degnen, L., Rosenthal, R.: The nature of rapport and its nonverbal correlates. Psychol. Inq. **1**(4), 285–293 (1990)
Vail, A.K., Boyer, K.E., Wiebe, E.N., Lester, J.C.: The mars and venus effect: the influence of user gender on the effectiveness of adaptive task support. In: International Conference on User Modeling, Adaptation, and Personalization, June 2015, pp. 265–276. Springer, Cham (2015)
Vaughan, B.: Prosodic synchrony in co-operative task-based dialogues: a measure of agreement and disagreement. In: Twelfth Annual Conference of the International Speech Communication Association (2011)
Walker, E., Girotto, V., Kim, Y., Muldner, K.: The effects of physical form and embodied action in a teachable robot for geometry learning. In: 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT), July 2016, pp. 381–385. IEEE (2016)
Wallace, R.: The elements of AIML style. Alice AI Foundation (2003)
Wang, N., Gratch, J.: Rapport and facial expression. In: 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, September 2009, pp. 1–6. IEEE (2009)
Web Speech API: https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API. Accessed 26 Oct 2018
Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM **9**(1), 36–45 (1966)
Westlund, J.K., Lee, J.J., Plummer, L., Faridi, F., Gray, J., Berlin, M., Dos Santos, K., et al.: Tega: a social robot. In: The Eleventh ACM/IEEE International Conference on Human Robot Interaction, March 2016, pp. 561–561. IEEE Press (2016)
Wheldall, K., Mettem, P.: Behavioural peer tutoring: training 16-year-old tutors to employ the 'pause, prompt and praise'method with 12-year-old remedial readers. Educ. Psychol. **5**(1), 27–44 (1985)

Nichola Lubold is a research scientist in the Human Centered Systems group at Honeywell Laboratories and research staff at Arizona State University in the School of Electrical, Computer, and Energy Engineering. She received a Ph.D. in computer science from Arizona State University in 2018 and holds a B.S. in computer engineering from the University of Notre Dame. Her research focuses on bridging the gap between applications of intelligent technology and what we know about human interaction to create natural and productive interactions. This includes exploring behavioral phenomena and identifying the contributions and complications resulting from critical individual and group differences that emerge in interactions with intelligent technology.

Erin Walker is an Associate Professor in the School of Computing and Information and the Learning Research and Development Center at the University of Pittsburgh. Previously, Walker was a faculty member in the School of Computing, Informatics, and Decision Systems Engineering at Arizona State University. She was awarded a Computing Innovations Postdoctoral Fellowship from the Computing Research Association, also at Arizona State. She completed her PhD in 2010 at Carnegie Mellon University in Human–Computer Interaction. Her research uses interdisciplinary methods to improve the design and implementation of educational technology, and then to understand when and why it is effective. She is currently working on projects that incorporate social and contextual adaptation into learning technologies, including implementing a teachable robot for mathematics learning and reimagining the design of the digital textbook. Walker's work has resulted in over ten journal articles and thirty peer-reviewed full conference papers, including a best paper award at Creativity and Cognition, best young researcher's track paper award at AIED, best paper nominations at ITS and AIED, and a best technology design nomination at CSCL.

Heather Pon-Barry is an Associate Professor of Computer Science at Mount Holyoke College. She received a Ph.D. in computer science from Harvard University in 2013 and holds B.S. and M.S. degrees in Symbolic Systems from Stanford University. She is a recipient of the NSF CAREER award and a Google CS Capacity Award. Her research examines spoken language processing and dialog in the context of educational robots.