Effects of Voice-Adaptation and Social Dialogue on Perceptions of a Robotic Learning Companion

Nichola Lubold and Erin Walker
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University, Tempe, AZ, USA
nichola.lubold@asu.edu, erin.a.walker@asu.edu

Heather Pon-Barry
Department of Computer Science
Mount Holyoke College
South Hadley, MA, USA
ponbarry@mtholyoke.edu

Abstract— **With a growing number of applications involving social human-robot interactions, there is an increasingly important role for socially responsive speech interfaces that can effectively engage the user. For example, learning companions provide both task-related feedback and motivational support for students with the goal of improving learning. As a learning companion's ability to be socially responsive increases, so do learning outcomes. This paper presents a socially responsive speech interface for an embodied, robotic learning companion. We explore two methods of social responsiveness. The first method introduces social responses into the dialogue, while the second method augments these responses with voice-adaptation based on acoustic-prosodic entrainment. We evaluate the effect of a social, voice-adaptive robotic learning companion on social variables such as social presence and rapport, and we compare this to a companion with only social dialogue and one with neither social dialogue nor voice-adaptions. We contrast the effects against those of individual factors, such as gender. We find (1) that social presence is significantly higher with a social voice-adaptive speech interface than with purely social dialogue, and (2) that females feel significantly more rapport and are significantly more persistent in interactions with a robotic learning companion than males.**

Keywords—adaptive learning companion, spoken dialogue, acoustic-prosodic entrainment, social presence, rapport

I. Introduction

As robots become increasingly pervasive, filling every aspect of life, at home, at work, at school, they can offer continued and individualized support in cases where it is not always possible to have a constant human companion. With this growing number of applications involving social human-robot interactions, there is a growing need for adaptive, socially responsive speech interfaces. In human-robot interactions, people tend to consciously treat robots as non-living mechanics but unconsciously, they engage robots in fundamentally social ways [1, 2]. When it comes to speech interfaces, people react similarly, applying automatic and unconscious social responses [3]. This work presents a socially responsive speech interface for human-robot interaction motivated by our understanding of how people interact in spoken human-human interactions.

In spoken HRI interactions, one approach to enhance a robot's social responsiveness is to add social lexical content to a dialogue [4]. In human-human interaction, people also communicate social information through their *manner* of speaking; they may speak fast or slow, loudly or softly, high or low. Modifying the vocal prosody of a robot can convey emotional states such as anger, happiness, and sadness [5]. We are interested in creating a robot that conveys social information in its manner of speaking, *adaptively*, to enhance the perceived social responsiveness. To create this voice adaptation, we focus on a phenomenon known as acoustic-prosodic entrainment. Acoustic-prosodic entrainment occurs when two speakers adapt their acoustic-prosodic features including tone, intensity, and speaking rate to mirror one another. In human-human interaction, entrainment is correlated with social factors including communicative success [6], conversational flow [7], and rapport [8]. A robot that can entrain has the potential to improve interactions by enhancing these factors.

We present the design and preliminary evaluation of a socially responsive, voice-adaptive speech interface for an embodied, robotic learning companion comprised of a LEGO® Mindstorms® NXT robot and an iPod-Touch that displays facial expressions and outputs speech. The primary goal of the learning companion is to facilitate student learning by providing both task-related feedback and motivational support. Learning companions, based on the theory that learning is influenced by social interactions [9], require social sensitivity to influence students' socio-motivational factors and increase student learning [10]. We explore the learning companion as a robotic teachable agent. A robotic teachable agent applies the concept of peer tutoring within a robotic learning companion framework. With the advantage of physical embodiment, the learning companion as a robotic teachable agent has the potential to create more social engagement with the activity, enhance motivation, and promote learning. Recent work on robotic teachable agents has shown students respond positively to these interventions [11, 12].

To explore and evaluate the effect of voice-adaption in this environment, we analyze the platform under three conditions: a *social* condition with social dialogue content in addition to the educational content, a *voice adaptive plus social* condition with the addition of both social dialogue and voice adaptation, and a *control* condition with neither social dialogue nor voice adaptation. Prior work on entrainment and human-robot interaction suggests males and females respond very differently to entrainment and to robotic interventions [13]. Given this prior work, we propose the following three research questions:

1. How do social variables like social presence and rapport differ depending on condition and gender?
2. How does persistence in the interaction differ depending on condition and gender?
3. How is learning affected by condition and gender?

We explore these questions in a 3 (condition) x 2 (gender) experiment. We find social presence differs significantly between conditions, with the voice plus social condition resulting in the highest average social presence. In addition, females report significantly more rapport than males; we do not find an effect between rapport and condition. We also find females persisted longer in the interaction than males but we observe no differences in learning gains by condition or gender.

In the next section, we provide background on the learning companion as a robotic teachable agent, prior findings on acoustic-prosodic entrainment, and relationships to social presence and rapport. We then outline the platform introduced for this analysis, discuss the study and procedure, and give an overview of the results. We conclude with a discussion of the results and directions for future work.

II. Background

A. *Learning Companion as a Robotic Teachable Agent*

While learning companions can be implemented in various forms, we focus on the learning companion as a robotic teachable agent [13]. Teachable agents are based on the observed benefits of peer tutoring [14]. When students teach other students, they are driven to reflect and elaborate on their knowledge and identify misconceptions [15]. Studies with teachable agents show that students are highly motivated to teach their agents and the agents can be highly beneficial to the student [16].

By utilizing the advantages of the teachable agent framework in a robotic learning environment, we can leverage the principles of both to create more social engagement with the activity, enhance motivation, and promote learning. This approach is supported by prior work; for example, Saerbeck, et al. [17] used the iCat robot to investigate how a socially supportive robotic cat influenced the task of language learning. Users who interacted with the socially supportive robotic cat were more motivated and learned more than those who interacted with a socially neutral robotic cat. Hood, et al. [11] introduced a teachable NAO robot which children could teach to write, and they validated this interaction paradigm for both engaging and educating students. We believe influencing social variables in the teachable agent framework has a high potential for resulting in improved learning because increased feelings of social presence and rapport will positively affect student motivation. We expect our findings will generalize to other learning companion environments.

B. *Social Dialogue*

In HRI as well as spoken dialogue research, the introduction of social content to otherwise non-social dialogue improves the interaction. Kanda, et al. [18] conducted a two-month trial in an elementary school with a social robot called Robovie, who could express various social behaviors, such as calling children by name. The social behaviors engaged the students; the students who interacted with Robovie longer learned more. Breazeal [19] increased social engagement by detecting emotion and providing expressive, emotive responses. Tapus and Matarie [20] demonstrate users prefer interacting with robots that display similar personalities via dialogue and vocal adaptations. While they did not introduce social content as we interpret it here, Lee et al. [4] improved rapport, cooperation, and engagement with a service robot that personalized its interactions and dialogue. Kumar, et al. [10] found an intelligent tutoring system which introduced text-based social dialogue by giving encouragement and promoting cohesion increased learning gains. Bickmore and Cassell [21] found adding social content to spoken dialogue can have a significant impact on a user's trust of an embodied real-estate agent engaging a user in real-time dialogue. In this work, we explore the effect of social dialogue on rapport and social presence in spoken interactions with a robotic teachable agent, where students are *teaching* the robot rather than the robot acting as a tutor.

C. *Acoustic-Prosodic Entrainment*

Entrainment, known also as accommodation, occurs when dialogue partners adapt their behavior to each other during an interaction. Entrainment can be gestural, via gaze or facial expressions [22], word-based or lexical [23], or speech-based [24]. Acoustic-prosodic entrainment occurs when two people adapt their manner of speaking, such as their tone, speaking rate, or pitch, to one another. Acoustic-prosodic entrainment is correlated with communicative success, conversational flow, and social factors like rapport [6, 24, 25]. Explored in-depth in human-human conversation, entrainment has been found to be both continuous and dynamic. Speakers will entrain over the course of a conversation, growing more similar over time, and they will also fluctuate in similarity dynamically within the conversation, growing closer and then resetting. Entrainment has been measured and analyzed both globally, at the conversation level, and locally, at the turn level, in human-human corpora. In prior analysis of turn-level entrainment, we found individuals entraining on pitch on a turn-by-turn basis had higher measures of communicative success [26] and rapport [25].

Exploration of entrainment with computer systems has shown people will entrain to a computer [27] and that individuals prefer computer voices which are similar to their own. For example, Nass [3] found that users who were extroverts preferred a computer voice which displayed extroverted speech features such as increased intensity and speaking rate. Levitan [28] found that people unconsciously trusted a virtual avatar which adapted to the user's speaking rate and intensity more than one that did not.

In this work, we provide further insight into human-computer entrainment by exploring a voice-adaptive speech interface similar to Levitan [28]. Instead of intensity and speaking rate, we focus on adapting to pitch. Given analysis of entrainment on pitch in human-human dialogues, adaptation on pitch is likely to affect communicative success, conversational flow, and social factors like rapport. In our prior work [29], we found a speech interface which adapted to a user's pitch could achieve higher 3rd party perceptual ratings of naturalness and rapport over other pitch adaptions. We also found that while the pitch adaptation resulted in more rapport and naturalness, it was not significantly better than normal text-to-speech output.

D. Gender

It is well established that stereotypes, especially gender stereotypes, can play a significant role in influencing human-human interactions, and there is evidence to suggest this effect applies to human-robot interactions as well. Utilizing a security robot, Tay, et al. [30] showed users applied gender stereotypes to a security robot, with users perceiving a security robot which a male gender overtones more useful and acceptable. Schermerhorn, Scheutz, and Crowell [13] found females viewed robots as more machine-like and responded less socially. These findings suggest we are likely to see a gender effect in our study given the robotic interaction; however, whether we will see the same effect is uncertain if we take into account prior work on virtual learning companions and prior work on entrainment. Prior work on socially responsive virtual learning companions has shown females tend to respond more positively to the interaction than males and that females also tend to persist in the interaction longer for virtual learning companions [31]. These prior works suggest we will see males and females respond differently to a voice adaptive, robotic learning companion, but how they will differ is a nuanced question.

E. Hypotheses

To evaluate the effect of a social voice adaptive robotic learning companion, we report on two social variables: rapport and social presence. Rapport is a complex phenomenon characteristic of many successful interactions. We base our approach to rapport on Tickle-Degnen and Rosenthal's [32] theory of positive emotions, mutual attentiveness, and coordination, and we utilize a rapport scale developed by Gratch, et al. [33] to assess rapport. Social presence is also a complicated social factor with multiple interpretations. We utilize the human-computer interpretation for our work, construing social presence as the "perceptual illusion of non-mediation." In prior work, social presence is correlated with increased satisfaction, enjoyment, and motivation [34], implying that the more people feel like a mediated condition is not mediated, the more successful the interaction can be.

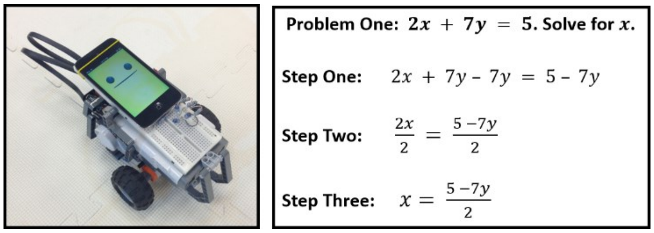
With these interpretations for rapport and social presence and the background work on entrainment, gender, and human-robot interactions, we propose two primary sets of hypotheses, one for gender and one for condition, for each of our research questions, as depicted in Figure 1. Given the effects of voice adaption and gender on rapport and social presence, we hypothesize for research question one that the voice adaptive plus social condition will result in the highest ratings of rapport and social presence, followed by the social condition, and finally the control, and that females will feel more rapport and social presence than males. If we find our first hypotheses are validated, then we believe based on theories of motivation that there will be increased motivation with increased rapport [35]. This leads us to hypothesize for our second research question that there will be greater persistence in teaching in the voice adaptive plus social condition and females will persist in the interaction longer. For our third research question, if we find increased rapport, social presence, and persistence for females and the voice adaptive plus social condition, the teachable agent framework suggests we will also find greater learning. We hypothesize for research question three that the voice adaptive plus social condition will lead to greater learning, and that females will learn more than males, since we expect them to experience greater rapport and persist in the interaction longer.

III. Designing a social voice-adaptive Robotic Learning Companion

Drawing on previous work involving social dialogue and pitch-adaptations, we designed and built Quinn, a social voice-adaptive teachable robot. Quinn is a teachable robot for the math domain; students teaching Quinn how to solve variable equations. Quinn consists of a LEGO Mindstorms base with an iPod mounted on top of it representing its face. Quinn's facial expressions are animated when speaking, and neutral otherwise. Students interact with Quinn using spoken language and a web application. The web application contains materials to guide the students in their teaching of Quinn: there are six variable equation problems (i.e. "Solve $bx + gy = 14by + 6x$ for x"), and six quizzes. The application presents one problem at a time and includes the worked-out steps to reach a solution. The problems are ordered in increasing order of difficulty with particular concepts introduced in each problem and follow-up quiz. Students walk Quinn through the worked-out problems using spoken language, explaining each step. Quinn responds using spoken language. At the end of each problem, students ask

Fig. 1: Research questions and hypotheses; the hypotheses from research questions one and two drive the hypothesis for research question three

RQ #1: How do social presence and rapport differ by condition and gender?	RQ #2: How does persistence in the interaction differ depending on condition and gender?	RQ #3: How is learning affected by condition and gender?
Condition: *voice adaptive plus social* will have **more social presence** and **more rapport** than the *social* and *control*	Condition: *voice adaptive plus social* will result in **greater persistence** than the *social* and *control*	Condition: *voice adaptive plus social* will result in **greater learning** than the *social* and *control*
Gender: *females* will feel **more social presence** and **more rapport** for the teachable robot than *males*	Gender: *females* will **persist longer** in the interaction than do *males*	Gender: *females* **will learn more** than will *males*



Problem One: $2x + 7y = 5$**. Solve for** x**.**

Step One: $2x + 7y - 7y = 5 - 7y$

Step Two: $\frac{2x}{2} = \frac{5-7y}{2}$

Step Three: $x = \frac{5-7y}{2}$

Fig. 2: The teachable robotic agent, Quinn, and a sample problem

Quinn to solve the quiz, step by step. Figure 2 shows an image of Quinn and a sample problem.

The spoken dialogue system consists of a speech recognition module that utilizes the Web Speech API specification, a pattern-matching dialogue manager written using the XML-compliant Artificial Intelligence Markup Language [36], and a text-to-speech synthesis module that utilizes the Microsoft Speech API. Our voice adaptation module takes a synthesized waveform as input and uses Praat [37] to alter the voice and output a new waveform. An advantage of this approach is that the voice adaptation module can be introduced independently into other dialogue systems.

A. *Voice Adaptation*

We adopt a method for voice adaptation that manipulates a single acoustic-prosodic feature—pitch. The pitch adaptation method preserves the pitch contour of the original text-to-speech output but shifts the pitch up or down to match the mean pitch of the previous speaker turn. This method, described in detail in Lubold, Pon-Barry, and Walker [27], was found to have the highest ratings of perceived naturalness and rapport among three different methods of automated pitch adaptations. In the voice-adaptive condition, every turn spoken by Quinn is adapted to the mean pitch of the user using this method. In human-robot interactions the effect of adapting on pitch mean has been less explored. This work contributes insight into how pitch mean can be utilized as a voice adaptation in a human-robot platform.

B. *Social Responses*

Quinn's social dialogue responses are motivated by the social interaction strategy proposed by Kumar et al. [10] based on Bales' socio-emotional interaction categories [38]. There are three main categories: showing **solidarity**, showing **tension release**, and **agreeing**. Examples of social responses Quinn might give in each category are given in Table 1. These responses are supported by human-human dialogue analysis which categorizes social responses as including positive dialogue moves, such as compliments [38]. Table 1 includes non-social response for contrast. Quinn's social and non-social responses were aligned to the same number of syllables as much as possible; both versions included Quinn's acknowledgement of the student's dialogue. In the social condition, Quinn selects a social response 15–20% of the time, in line with analysis of human-human social responses in peer tutoring and collaborative dialogues [8, 10].

IV. Study

A. *Conditions*

We conducted a between subjects experiment with three conditions: **control**, **social**, and voice-adaptive plus social, referred to as **voice plus social**. In each condition, there were 16 participants: 8 females and 8 males. In the control condition, Quinn had no social responses and no pitch adaptation. In the social condition, Quinn introduced statements of a social nature as described in the prior section. In the voice plus social condition, Quinn introduced social dialogue *and* adapted the pitch of its voice based on the student's voice. The gender of the synthesized voice was pre-set to match the gender of the participant. Experimenter instructions and the content of the activity were held constant for all conditions.

B. *Participants*

We recruited 48 undergraduate students for the experiment (24 female, 24 male). All students were native English speakers between ages 18 and 30 and were randomly assigned a condition. Sessions lasted 90 minutes and students were compensated $15 upon completion. Students sat a desk with a Surface Pro tablet in front of them. Quinn sat on the desk next to the Surface Pro, to the right of them. 5 participants were excluded for scoring 100% on the pretest, and thus having too much prior knowledge for the study. Thus, 16 students remained in the voice plus social condition, 14 students in the social condition, and 13 students remained in the control condition.

C. *Study Design & Procedures*

Students began the study by completing a 10 minute pretest. Next, each student was given a practice exercise which contained two worked-out examples of variable equation problems. The student was asked to explain the problems and the steps described out loud. After this practice exercise, the students watched a 4-minute video introducing Quinn and

TABLE 1: 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.

describing the task. Students were told they should help Quinn learn how to solve variable equations by walking Quinn through the six example problems and quizzing Quinn after each problem to assess Quinn's knowledge. Students were also informed they have the option to re-teach Quinn if Quinn struggles on a quiz. Students worked with Quinn through all six problems and quizzes. They then completed the post-test, which took around 10 minutes. They were then given a questionnaire assessing their motivation during the study and attitudes towards Quinn. Given time, they were asked some final interview questions. In total, the study took 90 minutes.

D. Measures

For measuring rapport and social presence, the follow-up questionnaire adopted nine Likert-scale questions from prior literature to assess rapport [33] and eight Likert-scale questions for social presence (8 questions). The social presence questions were adopted from the attentional allocation portion of the Networked Minds Social Presence Inventory [39]. Attentional allocation is a critical element of social presence [40] and is the most applicable within our robotic teachable agent scenario. We average the rapport and social presence questions to create three representative constructs with an acceptable internal reliability (Cronbach's $\alpha \geq 0.70$).

We assess a measure regarding persistence in the interaction by collecting the number of times a student retaught Quinn. Quinn was pre-programmed to get the wrong answer on two of the quizzes. This re-teaching metric was calculated as the total number of times the student retaught Quinn, with four possible values observed: 0, 1, 2, or 3.

Learning gains were assessed with the pretest and posttest scores. We computed normalized learning gains according to [41] using (1) to account for prior knowledge. If the posttest scores were lower than the pretest scores, we used (2).

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

$$gain = (posttest - pretest)/pretest \tag{2}$$

After removing the five participants who scored 100% on the pretest, we found of the 43 participants remaining, 23 hit a ceiling on their learning gains (scoring 100% on the posttest). With 10 individuals at zero gain, 10 individuals who gained in a normal distribution, and 23 hitting full gain, we determined analysis would be better served by grouping the learners into three groups – no gain, some gain and all gain. The results on learning gains are analyzed in this context.

V. RESULTS

A. Social Presence and Rapport

We compare a social voice-adaptive robotic learning companion (condition = *voice + social*) to a social (condition = *social*) and to a non-social, non-voice adaptive (condition = *control*) robotic learning companion. Our first research question, introduced in the Introduction, was: "How do social variables like social presence and rapport differ depending on condition and gender?" We answer this question with an initial two-way MANOVA examining social presence and rapport as dependent variables and gender and condition as independent variables. The means and standard deviations by gender and condition are in Table 2. The MANOVA analysis reveals a significant multivariate main effect for condition, Wilks' λ = .80, F = 4.41, p = .02, partial eta squared = .197 and a significant multivariate main effect for gender, Wilks' λ = .77, F = 2.54, p = .04, partial eta squared = .124. However, the interaction between condition and gender is not significant, Wilks' λ = .85, F = 1.52, p = .21, partial eta squared = .124. Given the significance of the multivariate main effects, we examine univariate main effects for condition and gender on social presence and rapport. We also report the univariate results of the interactions.

Univariate analyses for the effect of condition indicate significant differences related to social presence, F(2, 42) = 4.0, p = .02. The η^2 effect size is 0.17, meaning the condition explained 14% of the total variability in social presence scores, which is large effect by conventional standards (Cohen 1988). Analyzing the pairwise differences for condition on social presence, significant pairwise differences are found between the voice plus social condition and the social condition. The voice plus social condition results in significantly higher ratings of social presence than the social condition (p = 0.02), but the voice plus social and control are not significantly different. We do not see an effect of condition on rapport, F(2, 42) = .16, p = .86.

While this result partially validates our hypothesis (as the voice plus social condition has higher social presence than the social condition), we expected the social condition to score higher. One possibility is the percentage of social turns within Quinn's dialogue moderated the effect on social presence. However, we did not find a significant effect comparing the percentage of social turns in the two social conditions, F(2, 42) = 2.51, p = 0.09.

Univariate analyses for gender reveal significant differences between males and females in regards to rapport, F(2, 42) = 8.86, p = 0.006. The η^2 effect size is 0.18, meaning gender explains

TABLE 2. MEANS AND STANDARD DEVIATIONS FOR GENDER AND CONDITION ON SOCIAL PRESENCE (LIKERT SCALE 1 – 7), RAPPORT (LIKERT SCALE 1 – 7), PERSISTENCE (0 – 3), AND LEARNING GAINS (0 – 1). * INDICATES SIGNIFICANCE AT P < 0.05, ** INDICATES SIGNIFICANCE AT P < 0.01

	Social Presence		**Rapport**		**Persistence**		**Learning Gain**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control	5.54	.67	5.04	.84	1.6	1.1	.81	.33
Social	4.90	.74	5.21	.84	1.1	1.2	.50	.54
Voice + Social	5.57*	.75	5.30	1.1	1.4	1.2	.56	.45
Males	5.18	.79	4.70	.97	1.0	1.1	.53	.48
Females	5.55	.70	5.60**	.71	1.7*	1.1	.71	.43

18% of the total variability in the degree of rapport, which is large effect by conventional standards [42]. Analyzing the differences between the genders, females feel more rapport for Quinn than males, with the total average rapport score of females, ignoring condition, equal to 5.59 and males equal to 4.78. In regards to social presence, the difference between males and females approaches significance, F(2, 42) = 3.76, p = 0.06.

This result, while partially validating our hypothesis, is also somewhat surprising given prior work has found females often do not respond warmly to human-robot interactions [13], liking the robot less than their male counterparts. One possibility for the difference in our result may be the larger number of females we have from technical majors, including engineering and math. Their technical background and experience may make them more inclined to feel more rapport for Quinn. We run a follow-up 2-way ANOVA with major and gender as independent variables and rapport as a dependent variable. However, we find no significant relationship between how males and females feel about Quinn and their majors (p = 0.16), and there is no significant relationship between major and rapport (p = 0.33).

Finally, univariate analyses of the interaction between condition and gender on rapport is not significant, F(2, 42) = .18, p = .84, as we would expect. However, the interaction between condition and gender on social presence is approaching significance, F(2, 42) = 3.02, p = 0.06. We examined the differences in social presence among the conditions separately for males and females. The male simple effect test indicated statistically significant differences among the means, F(2, 37) = 6.73, p = 0.003, $\eta^2 = .27$, whereas the female simple effect test was non-significant, F(2, 37) = .31, p = .74, $\eta^2 = .02$. Simple pairwise comparisons among the male means indicated the social condition differed from both the voice plus social (p = 0.001) and the control (p = 0.01) conditions. This indicates males may be the driving force behind the differences we see between the conditions on social presence.

B. *Persistence*

We utilize the re-teaching metric described earlier to answer our second research question: "How does persistence in the interaction differ depending on condition and gender?" The means and standard deviations for persistence by gender and condition are shown in Table 2. We utilize multinomial logistic regression to estimate the influence of condition and gender on persistence in the interaction, given that we measure persistence in terms of total re-teaching. In our analysis, the overall model including both condition and gender was not significant, $X^2(9)$ =12.35, p = 0.19. Looking at the predictors individually, gender is significant when controlling for condition. The likelihood of a female persisting in the interaction and re-teaching Quinn was 2.13 times more likely than a male, p = 0.03.

C. *Learning Gains*

Having grouped the students into three learning groups, we analyze 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 is not significant, $X^2(6)$ =6.86, p = 0.33, and we find that none of the individual predictors are significant.

Given the significance of re-teaching in relation to gender, we then explore whether re-teaching is related to learning. We run Pearson's chi-squared correlation on the categorical learning gains described above. We find that there is a significant correlation between re-teaching and the categorical learning gains, with $X^2(6)$ =17.9, p = 0.006.

We also assess social presence and rapport in terms of learning. Running a multinomial regression with rapport and social presence, we find the model is not significant, $X^2(4)$ =4.68, p = 0.32. However, in viewing the individual coefficients, social presence is approaching a significant effect on learning. 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, p = 0.06.

VI. DISCUSSION

We find our two hypotheses regarding research question one regarding how social presence and rapport are effected by condition and gender partially validated. The social voice-adaptive robotic learning companion has higher social presence than the social condition and females feel more rapport in general. We do not see an effect of condition on rapport and we find the effect of gender on social presence merely suggestive.

We explore one potential explanation for these results in the potential speech recognition errors made by the dialogue system. The Web Speech API we utilized for speech recognition uses Google's voice recognition service, which has been publicized as having an error rate of only 8%. To analyze the effect of speech recognition errors on rapport and social presence, we focus on the output of the dialogue manager (DM). The DM selected responses based on pattern matching, keywords, and context [35]. If the DM could not match the student's words to a particular pattern or response, the DM would return one of two types of responses, either a request for clarification (i.e. "can you please repeat that?") or a general acknowledgement (i.e. "ok sounds good"). Classifying the number of generic responses Quinn returned when Quinn could not match an exact pattern to a precise response, we ran an ANCOVA with gender and condition as independent variables and social presence and rapport as covariates, with the percentage of turns where Quinn requested clarification or gave a general acknowledgement as the dependent variable. We found this did not have a statistically significant effect on the differences reported by gender and condition on social presence and rapport, with F=1.1, p=0.41.

To gain some qualitative insight into the results, we explore the interview responses collected as a part of the experimental procedure. Participants were interviewed as time allowed, resulting in a total of 20 interviews. The interviews were approximately distributed across gender (11 female, 9 male) and condition (6 control, 6 social, 8 voice adaptive plus social).

Analyzing social presence, we are surprised the social condition scored lower than both the control and the voice adaptive plus social condition. Given prior work on social dialogue, we presumed the social condition would have a positive effect when compared to the control. Turning to the interview data, we asked participants how they felt about Quinn's responses. We find individuals in different conditions responded with very different views. Participants in the voice plus social condition said they “liked Quinn’s responses”, Quinn responded like “a normal every day person,” and “sometimes Quinn was kind of sassy!” Participants in the social condition tended to feel Quinn was less focused, “not on the same page,” “didn’t really seem to listen,” and Quinn’s responses “sometimes felt like they came from out of nowhere.” The interview responses from those in the control condition, who heard no social dialogue, suggest the absence of social dialogue led to different expectations of Quinn – Quinn was “a robot so of course it’s not going to respond like a human would.”

Comparing these responses, there are several possible theories as to why those in the social condition may have felt differently about Quinn’s responses. One possibility is the pitch adaptation is counter-balancing the adverse effect of the social dialogue when the social dialogue is presented without non-verbal cues. Prior work suggests users are sensitive to social dialogue timing and non-verbal cues which accompany social dialogue, such as facial expressions. When non-verbal cues do not accompany social dialogue, there can be a mismatch in expectations leading to dichotomous results. Another possibility is that individuals are responding differently as a result of the different social responses Quinn is capable of giving.

In assessing rapport, females felt significantly more rapport towards the robotic learning companion than males. We review the interview responses of the males and females to identify any relevant clues as to how or why females and males are reacting differently. In the interviews when asked how they would describe Quinn, three of the females described Quinn as “so cute” or “very cute.” The males described Quinn literally, as a robot made of Lego Mindstorms. When asked how they felt about Quinn, seven of the females replied “yeah I really like Quinn” or “I liked Quinn!” When asked why they liked Quinn, females would explain with “we’re best friends now,” “I feel like Quinn is now my friend,” or “teaching her felt like a connection.” Four of the male interviewees, when asked how they felt about Quinn, also replied with “I liked Quinn.” When asked why, they explained they liked Quinn because he was an “interesting robot,” “decently complex,” or a “quick learner.”

These responses give some insight into the differences we are observing between males and females. The females appear to be suspending disbelief more readily and easily than the males. Aligned with the statistical findings on rapport, this suggests females are viewing Quinn as more of a social embodied companion than as a robot. Taking into account background work with virtual learning companions showing that females respond with more rapport in these types of domains, these findings suggest domain may play an important role in how participants of different genders will respond socially in robotic interactions. These responses also help illuminate possible reasons for why males are viewing Quinn as significantly less socially present in the social condition. While the interaction between gender and condition was not significant, the male gender differs significantly on measures of social presence in the social and voice plus social conditions. If males are viewing Quinn from a more technical, robotic viewpoint, they may perceive Quinn’s social prompts differently. This suggests when designing social robots, awareness of gender is vital in how we apply social mechanisms because we cannot assume that males and females will respond similarly simply because it is a social mechanism.

In reviewing persistence in terms of persistence, we are disappointed to find condition does not appear to be having an effect but given rapport is also not significant by condition, this is not surprising. We do find females, who report significantly higher measures of rapport, are also persisting in the interaction significantly more. This finding supports learning theories on motivation which suggest rapport can have a positive effect on motivation [43]. In reviewing the interview data from those females who retaught Quinn, many gave motivational reasons in line prior work on teachable agents for why they persisted in teaching Quinn. For example, one interviewee responded “I felt responsible for Quinn failing the quiz” and another replied she retaught Quinn because she “wanted Quinn to succeed and I felt like it was my fault she wasn’t.” These responses validate persistence as an intrinsic measure of motivation. Examining these responses in association with females who also reported higher rapport, we find that those who said Quinn was now their friend also commented about being motivated to reteach Quinn, further supporting the relationship between rapport and motivation. There did not appear to a major difference between males and females who did not reteach Quinn.

VII. Conclusions

As a part of this work, we introduced a platform for performing voice adaptation and we explored the effect of adapting to one promising acoustic-prosodic feature, pitch, incorporating analyses of both gender and contextual social dialogue into our exploration. We find the voice adaptation with the addition of social dialogue is significantly higher than pure social dialogue alone. We also find females react with more rapport to the interaction and are more persistent in teaching the robotic learning companion. Interestingly, males have a lower social response to Quinn, and this is validated by interview responses. We are limited in our results regarding learning gains; due to a large number of participants earning 100% on the posttest; we fail to detect effects of learning. However, we believe the nature of the interactions can still provide interesting insights. To further understand the potential design repercussions, future work includes a deeper process analysis of the social and voice adaptive conditions to assess the types of social responses Quinn is evincing and how students are responding to Quinn’s social responses. In addition, our implementation of pitch adaptation is naïve – we entrain to every turn of the user to the absolute mean of the user every time, which is not realistically representative of the fine-grained and nuanced phenomenon in human-human conversations. Recent analysis of human-human data is providing ideas for how we might operationalize entrainment in future work [43]. We plan on exploring more nuanced forms of pitch adaptation as well as other acoustic-prosodic features for voice adaptation, including intensity, speaking rate, and vocal quality.

Acknowledgements

This work is supported in part by the Ira A. Fulton Schools of Engineering through a Dean's Fellowship and the Google Anita Borg Memorial Scholarship. We would also like to thank Tyler Robbins for his help in the data collection process.

REFERENCES

[1] Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*(3), 143-166.

[2] Leyzberg, D., Avrunin, E., Liu, J., & Scassellati, B. (2011). Robots that express emotion elicit better human teaching. In *Proceedings of the 6th International Conference on Human-Robot Interaction* (pp. 347-354).

[3] Nass, C. I., & Brave, S. (2005). *Wired for Speech: How voice activates and advances the human-computer relationship*. Cambridge: MIT press.

[4] Lee, M. K., Forlizzi, J., Kiesler, S., Rybski, P., Antanitis, J., & Savetsila, S. (2012, March). Personalization in HRI: A longitudinal field experiment. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on* (pp. 319-326). IEEE.

[5] Crumpton, J & Bethel, C. (2015). Validation of vocal prosody modifications to communicate emotion in robot speech. In *International Conference on Collaboration Technologies and Systems* (pp.39-46).

[6] Borrie, S. A., & Liss, J. M. (2014). Rhythm as a coordinating device: entrainment with disordered speech. *Journal of Speech, Language, and Hearing Research*, *57*(3), 815-824.

[7] Porzel, R., Scheffler, A., & Malaka, R. (2006, January). How entrainment increases dialogical effectiveness. In *Proceedings of the IUI* (Vol. 6).

[8] Lubold, N., & Pon-Barry, H. (2014, November). Acoustic-Prosodic Entrainment and Rapport in Collaborative Learning Dialogues. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (pp. 5-12). ACM.

[9] Vygotsky, L. S. (1978). Mind and Society: the Development of Higher Mental Processes,(edited by Cole, M., et. al).

[10] Kumar, R., Ai, H., Beuth, J. L., & Rosé, C. P. (2010). Socially capable conversational tutors can be effective in collaborative learning situations. In *Intelligent tutoring systems* (pp. 156-164). Springer Berlin Heidelberg.

[11] Hood, D., Lemaignan, S., & Dillenbourg, P. (2015). When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 83-90).

[12] Tanaka, F., & Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*, *1*(1).

[13] Schermerhorn, P., Scheutz, M., & Crowell, C. R. (2008). 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* (pp. 263-270). ACM.

[14] Ploetzner, R., Dillenbourg, P., Preier, M., & Traum, D. (1999). Learning by explaining to oneself and to others. *Collaborative learning: Cognitive and computational approaches*, 103-121.

[15] Roscoe, R. D., & Chi, M. T. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, *77*(4), 534-574.

[16] Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, *19*(3-4), 363-392.

[17] Saerbeck, M., Schut, T., Bartneck, C., & Janse, M. D. (2010). 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.

[18] Kanda, T., Sato, R., Saiwaki, N., & Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human–robot interaction. *Robotics, IEEE Transactions on*, *23*(5), 962-971.

[19] Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*(1), 119-155.

[20] Tapus, A., & Mataric, M. J. (2008). Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance. In *AAAI Spring Symposium: Emotion, Personality, and Social Behavior* (pp. 133-140).

[21] Bickmore, T., & Cassell, J. (2005). Social dialongue with embodied conversational agents. In *Advances in natural multimodal dialogue systems* (pp. 23-54). Springer Netherlands.

[22] Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of nonverbal behavior*, *27*(3), 145-162.

[23] Friedberg, H., Litman, D., & Paletz, S. B. (2012). Lexical entrainment and success in student engineering groups. In *Spoken Language Technology Workshop (SLT), 2012 IEEE* (pp. 404-409). IEEE.

[24] Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive science*, *35*(4), 587-637.

[25] Lubold, N., & Pon-Barry, H. (2014). Acoustic-Prosodic Entrainment and Rapport in Collaborative Learning Dialogues. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (pp. 5-12). ACM.

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

[27] Coulston, R., Oviatt, S., & Darves, C. (2002). Amplitude convergence in children's conversational speech with animated personas. In *Proceedings of the 7th International Conference on SLP* (pp. 2689-2692).

[28] Levitan, R. (2013). Entrainment in Spoken Dialogue Systems: Adopting, Predicting and Influencing User Behavior. In *HLT-NAACL* (pp. 84-90).

[29] Lubold, N., Pon-Barry, H., & Walker, E. (2015). Naturalness and Rapport in a Pitch-Adaptive Learning Companion. In *Automatic Speech Recognition and Understanding (ASRU), 2015 IEEE Workshop*. IEEE.

[30] Tay, B. T. C., Park, T., Jung, Y., Tan, Y. K., & Wong, A. H. Y. (2013). When stereotypes meet robots: The effect of gender stereotypes on people's acceptance of a security robot. In *Engineering psychology and cognitive ergonomics. Understanding human cognition* (pp. 261-270). Springer Berlin Heidelberg.

[31] Burleson, W., & Picard, R. W. (2007). Gender-specific approaches to developing emotionally intelligent learning companions. *Intelligent Systems, IEEE*, *22*(4), 62-69.

[32] Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological inquiry*, *1*(4), 285-293.

[33] Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. In *Intelligent Virtual Agents* (pp. 125-138). Springer Berlin Heidelberg.

[34] Heerink, M., Ben, K., Evers, V., & Wielinga, B. (2008). The influence of social presence on acceptance of a companion robot by older people. *Journal of Physical Agents*, *2*(2), 33-40.

[35] Murphy, E., & Rodríguez-Manzanares, M. A. (2009). Teachers' perspectives on motivation in high-school distance education. *International Journal of E-Learning & Distance Education*, *23*(3), 1-24.

[36] Wallace, R. (2003). The elements of AIML style. *Alice AI Foundation*.

[37] Boersma, Paul & Weenink, David. Praat: doing phonetics by computer [Computer program]. V. 5.4.12, retrieved 07/10/15 http://www.praat.org/

[38] Bales, R. F. (1950). Interaction process analysis; a method for the study of small groups. Cambridge: Addison-Wesley.

[39] Biocca, F., & Harms, C. (2002). Defining and measuring social presence: Contribution to the networked minds theory and measure. *Proceedings of PRESENCE*, *2002*, 1-36.

[40] Baron-Cohen, S., & Swettenham, J. (1996). The relationship between SAM and ToMM: Two hypotheses. In Theories of theories of mind (pp. 158-169). New York: Cambridge University Press.

[41] Hake, R. R. (2002). Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pretest scores on mathematics and spatial visualization. In *submitted to the Physics Education Research Conference,* Boise, ID.

[42] Cohen, J. (1988). Statistical power analysis: A computer program. Routledge.

[43] R. Levitan, S. Benus, A. Gravano, & J. Hirschberg. (2015). "Entrainment and turn-taking in human-human dialogue." In *AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction*.