Automated Pitch Convergence Improves Learning in a Social, Teachable Robot for Middle School Mathematics

Nichola Lubold[1], Erin Walker[1], Heather Pon-Barry[2], and Amy Ogan[3]

[1] Arizona State University, Tempe AZ 85281, USA
[2] Mount Holyoke College, South Hadley MA 01075
[3] Carnegie Mellon University, Pittsburgh PA 15213
nlubold@asu.edu, erin.a.walker@asu.edu
ponbarry@mtholyoke.edu
aeo@cs.cmu.edu

Abstract. Pedagogical agents have the potential to provide not only cognitive support to learners but socio-emotional support through social behavior. Socio-emotional support can be a critical element to a learner's success, influencing their self-efficacy and motivation. Several social behaviors have been explored with pedagogical agents including facial expressions, movement, and social dialogue; social dialogue has especially been shown to positively influence interactions. In this work, we explore the role of paraverbal social behavior or social behavior in the form of paraverbal cues such as tone of voice and intensity. To do this, we focus on the phenomenon of entrainment, where individuals adapt their paraverbal features of speech to one another. Paraverbal entrainment in human-human studies has been found to be correlated with rapport and learning. In a study with 72 middle school students, we evaluate the effects of entrainment with a teachable robot, a pedagogical agent that learners teach how to solve ratio problems. We explore how a teachable robot which entrains and introduces social dialogue influences rapport and learning; we compare with two baseline conditions: a social condition, in which the robot speaks socially, and a non-social condition, in which the robot neither entrains nor speaks socially. We find that a robot that does entrain and speaks socially results in significantly more learning.

Keywords: entrainment, convergence, pitch, teachable robot, rapport

1 Introduction

Pedagogical agents, including affect-sensitive tutors, emotionally responsive learning companions, and teachable agents, are becoming increasingly sophisticated. They facilitate learning through both cognitive feedback and socio-emotional support, using facial expressions, movement, and social dialogue strategies [1, 2, 3, 4, 5]. Social dialogue in particular has been explored in depth and found to influence engagement, motivation, and learning by drawing student attention to salient aspects of the problem domain while building rapport [6, 7]. We are interested in the effects of social behavior with a pedagogical agent in an area which remains largely unexplored—paraverbal behavior.

In human-human interactions, speakers often convey important social information to their listeners through paraverbal cues, by *how* they speak. For example, acoustic-prosodic **entrainment** is a phenomenon of speech where individuals adapt their paraverbal cues (such as their tone of voice or speaking rate) to that of their speaking partner while conversing. Correlated with **rapport** (a feeling of connection, harmony and friendship) as well as conversational flow, entrainment is thought to be a means of achieving social approval [8, 9]. It has been suggested that an individual on the receiving end of a high level of entrainment is likely to feel more rapport for their partner than if they were a receiver of low entrainment. A pedagogical agent which can model a learner's paraverbal cues and adapt to them might build a stronger social connection with the learner. In turn, a learner who feels more rapport for their agent may be more engaged and willing to evaluate misconceptions, leading to increased learning.

There are several challenges to implementing acoustic-prosodic entrainment in a pedagogical agent. To begin with, entrainment in human-human dialogue occurs on many features of speech and in many forms; it is unknown what the best method might be for automating entrainment to facilitate learning. Secondly, in exploratory work on implementing entrainment in agents, findings suggest social responses such as engagement [10] and likeability [11] may be enhanced by an agent that entrains. However, it is an open question whether automated entrainment will be powerful enough to influence outcomes like learning. In our own prior work, we explored whether a pedagogical agent which adjusted its pitch to match that of the learner could influence learning with college students [12]. While we found effects on social presence, there were no learning effects. There are several possible explanations for the lack of effect on learning. On the one hand, our implementation of entrainment may have been overly simple. On the other, many students were at ceiling on the posttest, and thus the domain content may have been too easy and prevented us from detecting effects.

In this paper, we iterate on our prior implementation of entrainment, deploy it as part of the interaction mechanisms of a teachable robot named Nico, and then explore its effects on learning with middle school students. Nico is a Nao robot that learners can teach how to solve math problems. It interacts with the learner using spoken dialogue and realistic gesture. More specifically, Nico uses social dialogue inspired by strategies successfully implemented in other AIED systems, such as praise [37], enthusiasm [3], and politeness [38]. Because dialogue of this sort has been shown to influence learning, and entrainment is a dialogue-based phenomenon, we investigate whether entrainment as a social behavior can enhance learning above and beyond this social dialogue.

We evaluate the influence of acoustic-prosodic entrainment with Nico using three conditions: a **social-entraining** condition, where Nico entrains and speaks socially, and two baseline conditions: a **social** baseline, where Nico speaks socially but does not entrain, and a **non-social** baseline, where Nico neither speaks socially nor entrains. We hypothesize that when Nico entrains and speaks socially, learners will report feeling more rapport and we will observe greater learning gains when compared to the social and non-social baselines. We also hypothesize that the social baseline will result in higher rapport and learning gains than the control. We further analyze how these three conditions differentially influence students of different genders, as recent research has shown that males and females respond to social behaviors from agents and robots differently, with females sometimes preferring social behavior more than males [21, 24, 42]. We believe it is possible we may observe gender differences with females responding to social behaviors more strongly than males.

In the next section, we review background on teachable agents, acoustic-prosodic entrainment, and gender differences. We then describe Nico and the implementation of acoustic-prosodic entrainment with social dialogue. The fourth section describes our evaluation study at two middle schools with 72 participants. The results of this study are given in Section 5; our discussion and conclusions are in the last section.

2 Background

2.1 Teachable Agents

We explore social behavior and effects on learning with a teachable robotic agent. By teaching, learners may attend more to the problem, reflect on their own misconceptions when correcting errors, and elaborate on their knowledge as they construct explanations [13], leading to learning. Teachable agents have demonstrated success in influencing learning [14, 15], and teachable robots have demonstrated similar positive effects [19, 20]. Indeed, due to their physical presence and rich channels of communication, robots have under some circumstances socially engaged users more than agents [22], and this may be the case with teachable agents as well.

It has been hypothesized that there is a social component to the success of teachable agents in influencing learning. Some research has shown that when learners feel rapport for their teachable agent [16] they are more likely to benefit. Others have demonstrated that learners can feel at once more responsible for their agent and believe the onus of failure belongs to the agent, easing the negative repercussions of failure [18]. Heightened feelings of responsibility for the agent can also lead to greater benefits from teaching the agent [17]. These responses may be enhanced by learners' feelings of rapport; within a teachable agent context, greater feelings of rapport may facilitate learning.

While social dialogue has not been extensively explored with teachable agents and acoustic-prosodic signals have received even less attention, there is reason to believe both will enhance learning and rapport. Social dialogue has been shown to build rapport [6, 30, 31] and in pedagogical agents it has been shown to increase learning [6]. In terms of acoustic-prosodic signals, there is some evidence that manipulating the agent's paraverbal behavior can positively influence social factors [23]. However, there is still little known about the potential of automating these signals to influence learning; we seek to provide more insight here.

2.2 Acoustic-Prosodic Entrainment

Entrainment, known also as accommodation, occurs when dialogue partners adapt to each other during an interaction. Acoustic-prosodic entrainment occurs when two people adapt their manner of speaking, including their speaking rate, intensity (i.e. loudness), or pitch (i.e. tone), to one another over a conversation. Explored in-depth in human-human dialogue, individuals can entrain in several ways. The most common forms are known as proximity, convergence, and synchrony. Proximity occurs when individuals match one another. Convergence occurs when speakers gradually grow closer. Synchrony is when speakers adapt in the same direction but do not match one another.

Explorations of automated entrainment are still in the early stages. Three studies have explored entrainment. Levitan and colleagues conducted a small pilot and found people unconsciously trusted an agent entraining proximally on speaking rate and intensity [11]. In our own work, we found college students felt more social presence for a teachable robot entraining proximally on pitch; learning was not affected [12]. Sadoughi and colleagues implemented a model of synchrony on pitch and intensity [10]. Varying whether the robot entrained in the first or second half of the interaction, children had higher engagement with the robot which began with entrainment

In contrast to this previous work, we implement entrainment as pitch convergence. In our own prior work, we implemented entrainment as proximity, matching the robot's pitch to the user's pitch. Both proximity and convergence on pitch have been found to be related to learning [25, 26] and rapport [27]. However, convergence rather than proximity may be more optimal for building rapport. Tickle-Degnen and Rosenthal suggest that the experience of rapport can be observed though behavioral correlates as they change over time [28]; for example, as rapport increases over time, coordination between partners also increases. Proximity does not have a temporal element but convergence does. Convergence is a form of increasing coordination as two speakers become more similar over time. An agent which converges may build more rapport and a partner who feels more rapport may learn more due to the social motivations in 2.1.

2.3 Gender

There is increasing evidence suggesting gender is an indicator of underlying individual differences which influence how individuals respond to social agents. Min and colleagues recently found that females were significantly more engaged with a narrative agent than males [24]. Other work has also suggested social behaviors are more favorable to females [21, 33]. This evidence suggests females respond more strongly to social interventions. There is, however, some work which has found no gender differences [34]. Additional analyses of gender are needed to understand these individual differences which may be indexed by gender. We include gender here, with the expectation that females may respond more strongly to social dialogue with entrainment.

3 Nico: An Entraining, Social Teachable Robot

Nico is an autonomous, social teachable Nao robot for middle school mathematics. Learners, using spoken dialogue and a tablet interface (MS Surface), teach Nico how to solve ratio word problems based on the Common Core Standards [29]. An example problem as displayed on the tablet interface is depicted in Figure 1. For each problem, Nico and the learner are given partial information; Nico requests the learner's help to solve for missing information. We describe the overall system design for Nico in the next section, followed by the social dialogue design, and the entrainment module.

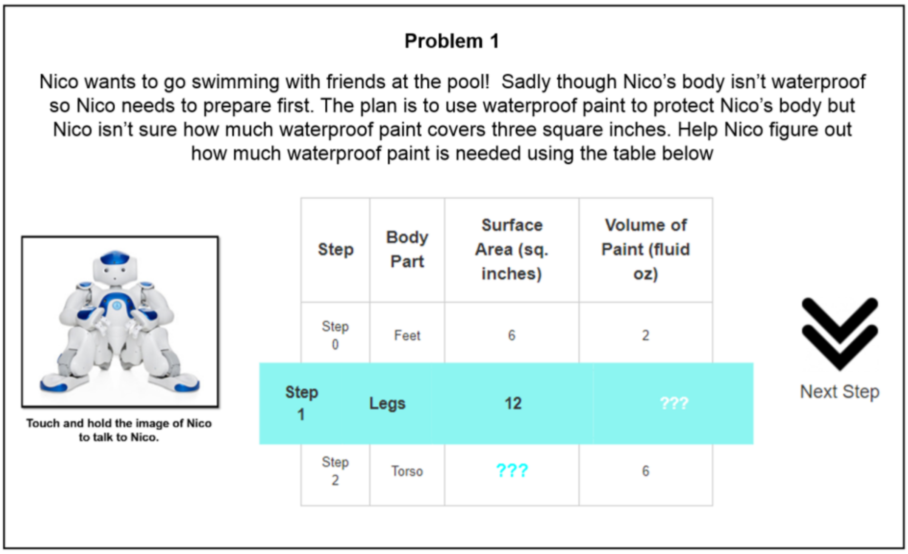

Fig. 1. Example problem in the tablet interface with current problem / step highlighted.

3.1 System

Learners are given a set of problems to teach Nico. The tablet interface displays visual progress to the learner as they move through the problems. Speech recognition is supported through the interface; to speak to Nico, the learner presses a button on the interface. After they are finished speaking, a notice appears on the interface indicating that Nico is 'thinking' while the system processes the input. Average response time is around four seconds. The interface tracks progress as the learner guides Nico through each problem step at their own pace, using buttons to advance forward with the current step highlighted and enlarged on the screen. When Nico 'answers' a step, the corresponding table cell is updated from question marks (see Figure 1) to the correct answer.

When a learner speaks, the dialogue system depicted in Figure 2 is engaged. The dialogue system processes the learner's speech, identifies a response, and transforms the response, depending on the condition. The interface captures the user's speech using the tablet's default microphone, and speech recognition is performed using the Google Speech API. The dialogue manager takes as input the recognized speech as well as the current problem and step. Utilizing the Artificial Intelligence Markup Language (AIML) [41], the dialogue manager identifies an appropriate response. Nico's responses are designed to encourage learning by prompting for explanations and providing feedback to encourage learners to feel that they are succeeding in helping Nico. This baseline dialogue is meant to foster learning by encouraging students to attend to the problem and elaborate on their knowledge [13]. An example of this dialogue is given in Table 1. Responses may then be transformed to entrain (see Section 3.3).

The dialogue system also introduces gesture. After an appropriate verbal response is identified, the gesture generator determines whether there is a corresponding gesture from a set of eight emblematic or easily recognizable gestures. These include waving

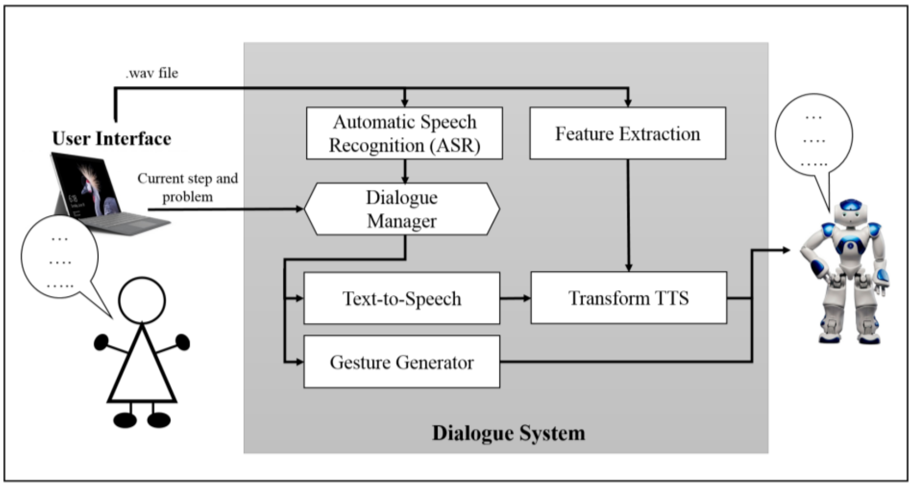

Fig. 2. Dialogue system for Nico

'hello,' nodding head as in 'yes,' shaking head as in 'no,' putting hands on hip to make a point, raising either hand, raising hands in celebration, and shrugging. The system identifies a gesture based on the content of Nico's utterance and times the behavior to the utterance. We also enabled "autonomous life", a default capability in the Nao robot which introduces a slight, swaying movement and listening behavior. Gesture and autonomous life are present in all three conditions and occur at the same rate.

3.2 Social Dialogue

By introducing social dialogue, we hope to replicate the effects of prior work demonstrating that social dialogue can positively influence learning and build rapport. We design Nico's dialogue to be social based on dialogue behaviors from prior work and on theories of rapport. One theory of rapport [30, 43] suggests an individual's use of linguistic politeness is a method for managing rapport. For example, if an individual praises their dialogue partner, this may positively enhance their partner's feelings towards them. If they are rude, this may introduce face-threat, hindering rapport. Following this theory, we utilize verbal behaviors indicative of linguistic politeness such as name usage, inclusive language, praise, and politeness. These verbal behaviors have been found to lead to more rapport in peer tutoring [31]. In addition, enthusiasm and off-task dialogue were found to lead to greater feelings of self-efficacy and learning [3, 6, 7]. We introduce social dialogue as using the learner's name, inclusive language, praise, enthusiasm, and off-task dialogue. Table 1 illustrates Nico's social and non-social dialogue responses to examples of potential student dialogue.

3.3 Acoustic-Prosodic Entrainment: Pitch Convergence

We implement a method for entrainment known as *local convergence* on a single acoustic-prosodic feature, pitch. Convergence occurs when individuals adapt over time to

Table 1. Example of non-social and social dialogues

	Non-Social Dialogue	Social Dialogue	Strategy
Nico:	Okay so I setup a ratio with 2.5 over 10. But then what do I do?	*Awesome*! So, *we* setup a ratio with 2.5 over 10. But then what do *we* do?	*Enthusiasm* *Inclusive*
Learner:	So, then you can reduce it or I guess simplify it to 1 over 4	So, then you can reduce it or I guess simplify it to 1 over 4	
Nico:	Oh, I simplify? How do you know I can do that?	*Sounds great*! So, *we* simplify? How do *we* know *we* can do that?	*Enthusiasm* *Inclusive*
Learner:	Because we know that 2.5 can go into 10 four times	Because we know that 2.5 can go into 10 four times	
Nico:	Oh, I get it. So, then I have 1 over 4?	Oh, I get it! *Thank you for explaining, [learner name]. You're a great tutor*. So, then *we* have 1 over 4?	*Politeness* *Praise* *Name usage*

their speaking partner; local convergence refers to this phenomenon happening on a local, turn-by-turn level. Individuals converge towards one another over a series of turns and then 'reset,' moving apart, typically when there is a change in topic or context.

We explore local convergence on pitch by gradually matching Nico's mean pitch to the learner's mean pitch over a series of turns. The mean pitch refers to the average pitch of a speaker's entire turn or utterance. Nico will speak with a mean pitch that is closer and closer to the learner's at each turn. When Nico and the learner move to a new problem, Nico will 'reset' and temporarily stop converging for one turn. Nico has a baseline pitch of approximately 230 Hz. To 'reset,' Nico speaks with a pitch at that baseline. Figure 3 depicts the changing mean pitch values as Nico converges and resets to the learner over a series of turns across two problems.

More specifically, our entrainment algorithm adapts the mean pitch of Nico's utterance utilizing a method which was previously found to produce high ratings of rapport and was perceived to be as natural as regular text-to-speech (TTS) [32]. This method involves generating the non-transformed TTS output and then shifting that output up or down such that the overall mean pitch of Nico's utterance matches a target value. That target value is calculated using the mean pitch of the learner's turn immediately prior. This work differs from previous work in the calculation of the target value. In our previous work, the robot mirrored the learner's pitch exactly, meaning the target value was the mean pitch from the learner's utterance, *target value = learner pitch* [12]. In this work, the calculation mimics local convergence by considering the number of turns which have passed, whether this is a new problem, and Nico's current mean pitch. Within a single problem context, the distance between Nico's mean pitch and the learner's mean pitch is gradually reduced. The target value to shift Nico's pitch is determined by the learner's pitch and the number of exchanges that have passed (one exchange = learner speaks, Nico speaks). Depending on the number of exchanges that have passed, Nico's pitch is shifted to be within a certain range of the learner's pitch (e.g., 0-1 exchanges: 50 Hz, 2 exchanges: 40 Hz, …, > 8 exchanges: 0 Hz). Thus, after 8 exchanges, Nico's mean pitch will equal the learner's mean pitch.

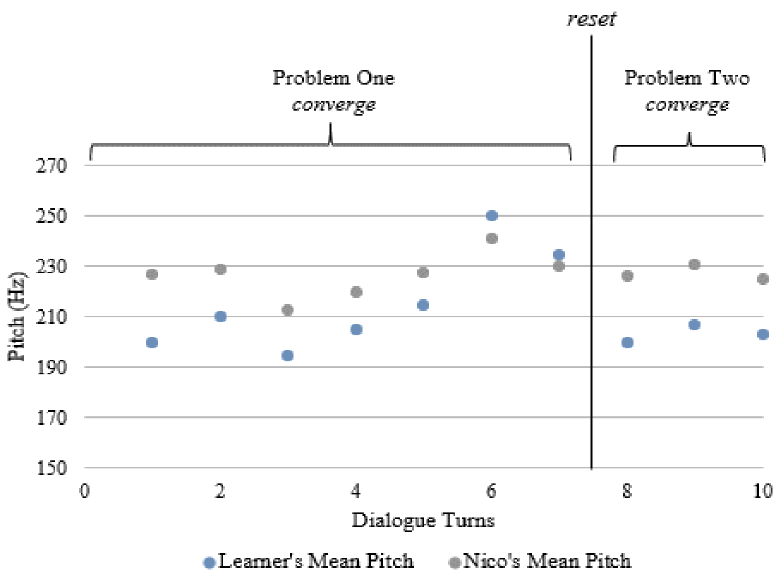

Fig. 3. Mean pitch values for a learner and Nico with entrainment

One additional restriction is placed on the adaptation. Nico will only adapt up to ±75 Hz, to reflect a realistic entrainment distance. Nico speaks with the same voice for both males and females, a version of the default Nao text-to-speech voice, with a baseline pitch of 230 Hz. This means Nico will adapt within the range of 155 Hz – 305 Hz. We validated the pitch convergence with four middle school students (2 female/2 male).

4 Study

We conducted a between-subjects experiment in which learners teach Nico how to solve ratio-based problems in one of three conditions: (1) **non-social**: Nico exhibits dialogue meant to foster a learning experience and does not introduce social dialogue or entrainment, (2) **social**: Nico encourages social interaction and rapport through social dialogue, and (3) **social + entrainment**: Nico introduces equivalent social dialogue and additionally entrains via convergence on pitch. Across all three conditions, the experimenter instructions and the content of the activity were held constant.

Participants were 72 middle-school students from two public middle schools in the Southwestern United States. 51% of the students were recruited from one school and 49% from the other, with a mean age of 11.25 (SD = 0.47). The gender breakdown is given in Table 2. Sessions lasted 60 minutes and took place at the participant's school. As shown in Figure 4, students sat a desk with a Surface Pro tablet in front of them. Nico stood on the desk next to the Surface Pro, to the right of the participant. Three participants experienced technical issues during the experiment and were excluded from the results. Thus, 22 participants remained in the non-social, 23 participants in the

Fig. 4. Students interacting with Nico at the two middle schools

social condition, and 24 participants in the social + entrainment condition.

Participants began with a 10-minute pretest and a short pre-survey to evaluate their initial self-efficacy towards math and tutoring. The participants were then given a few minutes to review the ratio problems and the worked-out solutions. After watching a short video depicting how to interact with Nico, students engaged in a teaching activity with Nico for 30 minutes. After the activity, they completed a 10-minute posttest and a short survey on self-efficacy, rapport, and their goals. Given the scope of this paper and the focus on learning and rapport, we do not explore effects of self-efficacy here.

We measured rapport with 12 questions that we designed and developed based on Tickle-Degnen and Rosenthal's understanding of rapport as being composed of three parts: attention, positivity, and coordination [28]. We developed questions for positivity and attention; we drew upon measures proposed by Sinha and Cassell [26] for coordination. Given our age group, we designed and iterated over the questions in a series of 14 pilot studies, adjusting the questions to target the desired measures and be understandable to middle schoolers. We finalized four questions assessing positivity, four questions measuring attention, and four questions for coordination[1]. We averaged the rapport questions to create a single representative construct with an acceptable internal reliability (Cronbach's $\alpha = 0.83$). To measure learning, we utilized a pretest-posttest design with an A and B form of the test. The two forms were isomorphic and counterbalanced within condition (half of the participants in each condition received test A as the pretest and test B as the posttest, and vice versa). The tests consisted of 10 procedural and conceptual questions around ratios. As with the rapport measures, we piloted and iterated on the design of the questions through 14 pilot studies. We calculated the

Table 2. Gender breakdown and dialogue statistics per session

	Females	Males	Total Turns *M (SD)*	Words per Turn *M (SD)*
control	13	11	141.7 (37.0)	8.13 (4.5)
social dialogue	13	11	124.9 (28.6)	8.56 (3.4)
social dialogue + entrainment	13	11	123.8 (26.5)	10.7 (4.9)

[1] Survey questions can be found at www.public.asu.edu/~nlubold/surveys/nico_rapport.pdf

normalized learning gains according to Hake [36]. If the posttest was lower than the pretest, we used equation (2):

$$gain = (posttest - pretest) / (1 - pretest) \quad (1)$$

$$gain = (posttest - pretest) / (pretest) \quad (2)$$

5 Results

In this section, we report the results for learning and rapport where individuals interacted with Nico, our teachable robot, in one of three conditions: a **social-entraining** condition where Nico was both social and entrained, a **social** condition where Nico was only social, and a **non-social** where Nico was neither social nor entrained. Studies were conducted across two schools. After analyzing differences between schools, there were no significant differences or interactions with school by condition or gender on learning or rapport. We therefore report the results without the additional factor of school.

5.1 Learning

With learning, we hypothesized that the social-entraining condition would result in greater learning than the social and non-social baselines. We verified this hypothesis by analyzing learning gains in a two-way analysis of variance (ANOVA) with condition and gender as the independent variables and gain as the dependent variable. The means and standard deviations for gain by condition and gender are in Table 3. The gain was significantly different across conditions, $F(2, 63) = 6.06$, $p = 0.004$. Partial eta squared was .16, meaning the effect size was medium. We did not find that gender to be significant, $F(1, 63) = .05$, $p = .82$ and the gender by condition interaction was not significant $F(2, 63) = 2.13$, $p = 0.12$. Tukey post-hoc analyses indicated significant pairwise differences for the social-entraining condition when compared to the non-social. The social-entraining condition was significantly higher than the non-social ($p=.005$). The social condition approached a significantly higher gain than the non-social ($p = .06$). The social-entraining condition was not significantly higher than the social condition ($p = .6$).

We had hypothesized that the social-entraining condition would result in higher learning gains. Our hypothesis was partially validated; we found that the addition of entrainment results in the highest learning gains and that it was significantly higher than the non-social control. Our hypothesis regarding gender was not verified. We did not find differences in regards learning by gender on the introduction of social behaviors.

Table 3. Means and standard deviations for learning gains and rapport across condition and by gender.

	Non-Social		Social		Social-Entraining	
	Males	**Females**	**Males**	**Females**	**Males**	**Females**
Learning Gain	-.11 (.5)	.02 (.34)	.13 (.17)	.17 (.24)	.34 (.2)	.13 (.12)
Rapport	4.0 (.7)	4.3 (.4)	4.2 (.7)	3.8 (.9)	4.1 (.7)	4.2 (.6)

5.2 Rapport

We had hypothesized that rapport might increase for the social and the social-entraining conditions. Given prior work, we expected that social dialogue and entrainment might influence rapport and that rapport may be related to learning. The self-reported means and standard deviations for rapport by condition and gender are also in Table 3. We first explored if rapport and gain were correlated but found they were not: Pearson's r = -.115, p = .347. We then explored if rapport differed by condition by analyzing rapport in an ANOVA with condition and gender as independent variables. We found our hypothesis was rejected. There were no significant differences in rapport across conditions, $F(2, 63) = .751$, $p = .48$, $\eta^2 = 0.02$. There was also no effect of gender, $F(1, 63) = .04$, $p = .84$, $\eta^2 = .001$ or gender by condition, $F(2, 63) = 1.49$, $p = .23$, $\eta^2 = 0.04$.

We found the lack of difference in self-reported rapport across conditions surprising, especially given that there were significant differences in learning. Prior work has suggested that the length of dialogue turns may play a role in learning and potentially rapport [39, 40]. Every learner interacted with Nico for thirty minutes regardless of condition but there may have been differences in number of turns and words per turn issued by each learner. Table 2 gives the means and standard deviations. We explored whether the total number of dialogue turns and average number of words per turn played any role in responses. We did not find any differences across conditions in the number of turns exchanged, $F(2, 63) = 1.22$, $p = .30$, or the number of words used, $F(2,63) = 1.7$, $p = .19$. We did not find a significant influence of turns or words on rapport or learning.

6 Discussion

Paraverbal manipulation is a less-explored modality in learning companions but it has potential for influencing learning. In this work, we explored paraverbal manipulation with our teachable robot, Nico. Nico adapted its pitch to that of the student as the student taught Nico. We were interested in effects of acoustic-prosodic adaptation on learning; we hypothesized that interactions with the social, entraining version of Nico would result in higher learning gains. Our hypothesis regarding learning was validated; the addition of entrainment with social dialogue significantly improved learning and the learning gains were significantly higher than when Nico did not speak socially and did not entrain. This is the first time that an implementation of acoustic-prosodic entrainment in an agent has shown positive effects on learning, and suggests that entrainment may be useful mechanism for enhancing learning interactions with agents.

We also hypothesized that measures of rapport would increase with the social and the social-entraining conditions as compared to the non-social baseline. Our hypothesis was not validated; self-reported rapport was reported at high and consistent levels across conditions. There are several possible explanations. One is that Nico was very successful in building rapport across all conditions and that our measure was at ceiling. An alternative possibility is that a single post-session survey does not capture the dynamic changes in rapport which occur during interaction and may be more influential to learning. Future work will include assessing behavioral rapport as it changed within interactions, and its relationship to learning across conditions.

Another possibility is that our measure for rapport was accurate, and our conditions simply did not influence social factors but instead influenced cognitive factors. Entrainment has been hypothesized to have social origins but alternative theories have suggested it has a cognitive function as well. The Interactive Alignment Model (IAM) [35] suggests entrainment is an outcome of individuals aligning on their understanding and knowledge of a situation. It is possible our implementation of entrainment facilitated the learners' convergence towards Nico and a deeper understanding of the problem.

For example, the learner might explain to Nico that Nico "*needs to multiply by two*." Nico will elaborate on this statement, recognizing the need to multiply by two because there are twice as many bags, "*Oh because we have two more bags? We have twice as many and multiply by two*?" Even though learners are given the worked-out solutions, they may not always have full domain knowledge and so Nico's queries lead to deeper understanding of the problem. With local convergence, learners are invited to converge to Nico and Nico's understanding. As Nico and the learner converge, this may facilitate deeper understanding of the domain content. There is evidence that this may occur in human-human peer tutoring. Sinha and Cassell [26] explored relationships of learning, convergence and rapport in dyads of peer tutors with a mean age of 13. They found that the individual who influences entrainment or induces the other speaker to entrain 'to' them, has higher learning gains. They suggest that a virtual peer that both converges to its human partner and invites convergence may be a more effective learning partner. Future work will include exploring the degree to which individuals entrained to Nico and how this may have led to learning considering the cognitive mechanism of IAM.

Finally, we did not observe any differences by gender. We had hypothesized that females might respond to the social behaviors more favorably. That we did not find differences is not unusual, but it may be due to our rapport measure or that we were unsuccessful in influencing the social factors which gave rise to gender differences in other work. In our future work, we intend to explore whether individuals responded differently on other dimensions which may be related to gender differences, including comfort-level in interacting with robots and interactional goals for teaching a robot.

7 Conclusions

In this paper, we explored how acoustic-prosodic entrainment influences learning and rapport in interactions with a social, teachable robot. This is the first evidence of automated entrainment in a pedagogical agent to find a significant effect on learning. In our future work, we plan on exploring how the cognitive mechanism of entrainment may be at work in these interactions and how other measures of rapport might provide insight into how entrainment and social dialogue influence rapport responses.

Acknowledgements

This work is supported by the National Robotics Initiative and the National Science Foundation, grant # CISE-IIS-1637809. We would also like to thank Samantha Baker and Ishrat Ahmed for their help in setting up this study and the data collection process.

References

1. Blanchard, E. G., Volfson, B., Hong, Y. J., Lajoie, S. P.: Affective artificial intelligence in education: from detection to adaptation. In: AIED, pp. 81-88. Springer, Heidelberg (2009).
2. Girard, S., et al.: Defining the behavior of an affective learning companion in the affective meta-tutor project. In: AIED, pp. 21-30. Springer, Heidelberg (2013).
3. Lane, H. C., et al.: The effects of a pedagogical agent for informal science education on learner behaviors and self-efficacy. In: AIED, pp. 309-318. Springer, Heidelberg (2015).
4. Afzal, S., Robinson, P.: A study of affect in intelligent tutoring. In: Workshop on Modeling and Scaffolding Affective Experiences to Impact Learning, vol. 57 (2006).
5. McDaniel, B.T., et al.: Facial features for affective state detection in learning environments. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 29 (2007).
6. Gulz, A., Haake, M., Silvervarg, A.: Extending a teachable agent with a social conversation module–effects on student experiences and learning. In: AIED, Springer, Heidelberg (2011).
7. Kumar, R., Ai, H., Beuth, J., Rosé, C.: Socially capable conversational tutors can be effective in collaborative learning situations. In: Aleven, V., Kay, J., Mostow, J. (eds.) *ITS 2010*. LNCS, vol. 6094, pp. 156–164. Springer, Heidelberg (2010).
8. Gallois, C., Giles, H.: Communication accommodation theory. *The international encyclopedia of language and social interaction* (2015).
9. Giles, H.: *A new theory of the dynamics of speech*. Diogenes, 106, 119-136 (1979).
10. 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. In: HRI, pp. 91-99. ACM (2017).
11. Levitan, R., et al.: Implementing acoustic-prosodic entrainment in a conversational avatar. In: *INTERSPEECH,* vol. 16, pp. 1166-1170. (2016).
12. Lubold, N., Walker, E., Pon-Barry, H.: Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion. In: HRI, pp. 255-262. IEEE (2016).
13. Roscoe, R. D., Chi, M. T.: Understanding tutor learning: knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research* 77(4), 534-574, (2007).
14. Leelawong, K., Biswas, G.: Designing learning by teaching agents: the betty's brain system. *International J. of Artificial Intelligence in Education* 18(3), 181–208 (2008).
15. Pareto, L., Arvemo, T., Dahl, Y., Haake, M., Gulz, A.: A teachable-agent arithmetic game's effects on mathematics understanding, attitude and self-efficacy. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) *AIED 2011*. LNCS, vol. 6738, pp. 247–255. Springer, (2011).
16. Ogan, A., et al.: Oh dear stacy!: social interaction, elaboration, and learning with teachable agents. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 39-48. ACM (2012).
17. Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., Roscoe, R.: Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology Enhanced Learning*, 5(02), 123-152, (2011).
18. Chase, C. C., Chin, D. B., Oppezzo, M. A., Schwartz, D. L.: Teachable agents and the protégé effect: increasing the effort towards learning. *Journal of Science Education and Technology*, 18(4), 334-352 (2009).
19. Hood, D., Lemaignan, S., Dillenbourg, P.: When children teach a robot to write: an autonomous teachable humanoid uses simulated handwriting. In: HRI, pp. 83-90. ACM (2015).
20. Tanaka, F., Matsuzoe, S.: 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) (2012).
21. Arroyo, Ivon, et al.: Affective gendered learning companions. In: AIED, pp. 21-48. (2009).

22. Liu, P., Glas, D. F., Kanda, T., Ishiguro, H., Hagita, N.: It's not polite to point: generating socially appropriate deictic behaviors towards people. In: HRI, pp. 267-274. IEEE (2013).
23. Westlund, K, et al.: Flat vs. expressive storytelling: young children's learning and retention of a social robot's narrative. *Frontiers in human neuroscience*, 11, 295 (2017).
24. Min, Wookhee, et al.: Thanks alisha, keep in touch: gender effects and engagement with virtual learning companions. In: AIED 2017, pp. 299. Springer, Heidelberg (2017).
25. Thomason, J., Nguyen, H. V., Litman, D.: Prosodic entrainment and tutoring dialogue success. In: AIED, pp. 750-753. Springer, Heidelberg (2013).
26. 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). ACM (2015).
27. 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* (pp. 5-12). ACM (2014).
28. Tickle-Degnen, L., Rosenthal, R.: The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4), 285-293 (1990).
29. Common Core State Standards Initiative. Common Core State Standards for Mathematics. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers (2010).
30. Spencer-Oatey, H.: (Im) Politeness, face and perceptions of rapport: unpackaging their bases and interrelationships. Politeness Research, 1(1)(Im), 95–119 (2005).
31. Bell, D., Arnold, H., Haddock, R.: Linguistic Politeness and Peer Tutoring. *Learning Assistance Review* 14.1, pp. 37-54, (2009).
32. Lubold, N., Walker, E., Pon-Barry, H.: Naturalness and rapport in a pitch adaptive learning companion. In: ASRU, pp. 103-110. IEEE (2015).
33. Burleson, W., Picard, R.: Gender-specific approaches to developing emotionally intelligent learning companions. *IEEE Intelligent Systems*, 22(4), (2007).
34. 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).
35. Pickering, M. J., Garrod, S.: An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329-347, (2013).
36. Hake, R. R.: Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pretest scores on mathematics and spatial visualization. In: Physics education research conference, pp. 30-45, (2002).
37. Maldonado, H., et al.: We learn better together: enhancing elearning with emotional characters. In: *CSCL*, pp. 408–417. International Society of the Learning Sciences (2005).
38. Wang, N., et al: the politeness effect. *Int. J. Human Computer Studies* 66, 96–112 (2008).
39. Litman, D. J., et al.: Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence in Education*, 16(2), 145-170, (2006).
40. Rosé, C. P., Bhembe, D., Siler, S., Srivastava, R., VanLehn, K. The role of why questions in effective human tutoring. In U. Hoppe, F. Verdejo & J. Kay (Eds.) AIED (2003).
41. Wallace, R.: The elements of AIML style. Alice AI Foundation (2003).
42. Kim, Y., Baylor, A., Shen, E.: Pedagogical agents as learning companions: the impact of agent emotion and gender. Journal of Computer Assisted Learning 23(3), 220–234 (2007).
43. Spencer-Oatey, H.: Managing rapport in talk: using rapport sensitive incidents to explore the motivational concerns underlying the management of relations. Journal of Pragmatics, 34(5), 529–545 (2002).