A Multimodal Approach to Adaptive Dialogue Interaction for Learning Companion Robots

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

1 Mount Holyoke College
2 Arizona State University
3 Carnegie Mellon University

*** ponbarry@mtholyoke.edu**

Abstract

Learning companion robots that support spoken natural language dialogue present exciting opportunities for adaptive and personalized interaction. We hypothesize that adaptive robot behaviors, for example choosing between a straightforward versus a social dialogue move, can have a positive effect on *student motivation*. Supporting these adaptive behaviors in real-time requires that learning companion robots construct and dynamically update models of student motivational state. Our project examines the utility of *speech prosody* in contributing to a dynamic model of student motivational state in human-robot peer tutoring interaction. The project involves collecting spoken dialogue data with a learning companion robot, measuring the motivational state of students, and modeling the relationship between speech prosody and student motivation.

Introduction

Learning companion robots for STEM domains, especially socially-capable robots, have the potential to engage populations who are not currently engaged in STEM activities. This can be achieved by designing personalized, or adaptive, robot behaviors that can sense a student's motivational state and adjust its behaviors during the interaction to foster greater engagement and motivation [10]. Furthermore, recent studies suggest that personalization behaviors such as animated gestures, partner-directed gaze, and using social dialogue have a positive impact on student learning [1]. To foster engagement and motivation, especially in long-term interactions, a current challenge is the question of what kind of sensory input will inform a personalized robot's adaptive decision making.

We have developed a robot learning companion for middle school mathematics that interacts with students using spoken natural language. Human tutors are often able to pick up on social cues to remain informed on how effective the learning experience is, and adapt their approach based on this information [3]. Past research suggests that feelings of closeness or rapport are perceived through a person's prosody — the intonational and rhythmic patterns of speech [6]. Robots that adapt their prosody to match their human partner's prosody have been shown to foster social presence in a teachable robot setting [7] and greater engagement in a game co-player setting [9]. This current work aims to understand how multimodal input, particularly prosodic information from the speech signal, might be used to construct a real-time model of student motivational state during spoken dialogue interaction with a learning companion robot.

Learning Companion Robot

The platform for this project is a NAO humanoid robot, called *Nico*, that interacts with students as a learning companion. Learning companions that are capable of social interaction have the potential to influence student motivation and increase student learning [4]. Our learning companion robot builds upon past work examining prosodic entrainment, rapport, and social dialogue with a non-humanoid learning companion robot [7]. In this current work, students interact with Nico in a learning-by-teaching style [8]. The students are told that their goal is to help Nico solve a set of mathematics problems. Prior to the interaction, the students are provided with worked-out problem solutions and time to prepare. During the interaction, Nico takes initiative in leading the dialogue, asking students for help in how to approach the problem sub-parts (e.g., *"How do I figure out how much paint to mix?"*). Students respond by explaining their reasoning to Nico (e.g., *"We want to have six cans of green paint so we mix three cans of yellow paint and three cans of blue paint because..."*). Nico can respond with actions such as entering numbers in a tablet interface, gestures such as scratching its head, and dialogue. Figure 1 shows a student teaching Nico in our lab.

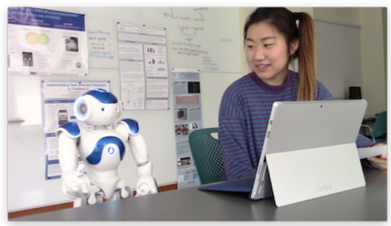

Figure 1. **Nico, a learning companion robot, being taught by a student**.

Spoken Dialogue Interaction Experiments

We have recently completed two human-robot spoken dialogue interaction experiments: the first in a lab setting with Wizard-of-Oz robot control, the second in a middle school setting with autonomous robot control. In both settings, students engaged with Nico to solve a collection of mathematics problems.

Mathematics domain. The problem domain for this work is middle-school mathematics, emphasizing understanding ratios and solving word problems. The same set of mathematics word problems was used in both experiment settings. A tablet touchscreen interface displays the problem text as well as a table with missing information that Nico needs to complete. Figure 2 shows this interface.

Wizard-of-Oz setting. In this experiment setting, 20 student participants (average age 20.0.) engaged in problem-solving dialogues with Nico in a research lab on a college campus. Nico was controlled behind the scenes by a human in a Wizard-of-Oz experiment setup [2]. The Wizard had a selection of pre-programmed dialogue and gesture moves and phrases at their disposal, as well as the ability to input additional phrases when necessary. Each session consisted of four ratio problems, as well as pre- and post-surveys.

Middle school setting. In this experiment setting, 72 student participants in the sixth grade (average age 11.25) engaged in problem-solving dialogues in a classroom at

Figure 2. A ratio word problem displayed in the tablet touchscreen interface. The current problem step is highlighted.

their middle school. Nico interacted with students autonomously, with a dialogue system that we have recently developed [5]. Each session consisted of four or more ratio problems, as well as pre- and post-surveys.

Approach

To investigate whether prosody can be used to model student motivation, we are initially focusing on three prosodic features that can be automatically estimated from the speech signal: pitch, intensity, and speaking rate. We hypothesize that within a single interaction session, increasing values in pitch, intensity and speaking rate will indicate greater motivation, while decreasing values will indicate lower motivation.

We sample prosodic feature values four times per session, at the start of each problem. In the Wizard-of-Oz setting, we use 30-second samples. In the middle school setting, we use the automatically segmented dialogue turns, which vary in length from 13.2 to 21.1 seconds.

To gauge student motivation, we analyze self-reported attitudes from post-session surveys. From each survey, we identify three statements that measure attention to Nico, teaching motivation, and teaching efficacy. Table 1 shows the average agreement and variance for these three statements, for each experiment setting.

Table 1. Average agreement and variance for student attitudes of attention to Nico, teaching motivation, and teaching efficacy. Agreement scale is from 1=strongly disagree to 5=strongly agree.

	Avg. Agreement	Std. Dev.
Wizard-of-Oz ($N = 20$)		
Paid attention to Nico	4.25	0.94
Motivation to teach Nico	4.14	0.88
Self-efficacy in teaching a peer	3.46	0.71
Middle school ($N = 72$)		
Paid attention to Nico	4.31	1.16
Motivation to teach Nico	3.83	1.30
Self-efficacy in teaching a peer	3.42	1.19

Preliminary Results The data analysis is currently in-progress. Below, we present examples of prosodic data from two participants. Figure 3 shows the prosodic feature trends for a participant in the Wizard-of-Oz setting whose self-reported motivation (aggregated) was among the *bottom* 20% of participants. Figure 4 shows the prosodic feature trends for a participant in the middle school setting whose self-reported motivation (aggregated) was among the *upper* 20% of participants.

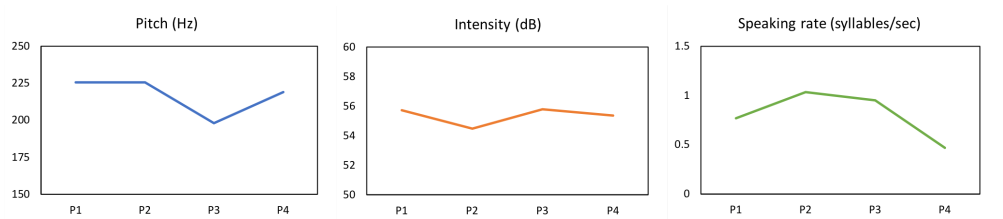

Figure 3. Prosody data from a low-engagement participant in the Wizard-of-Oz experiment. For the four math problems in the dialogue, P1-P4, the participant's average pitch, intensity, and speaking rate during the first 30 seconds of each problem sub-dialogue are shown.

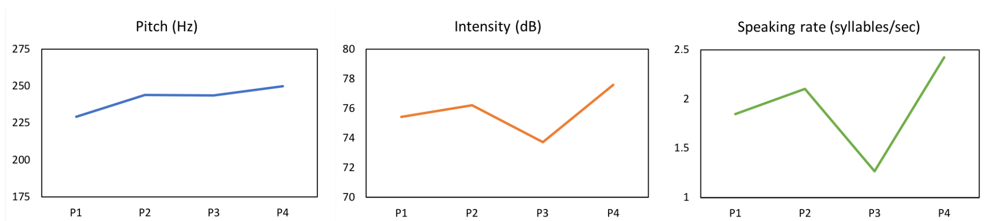

Figure 4. Prosody data from a high-engagement participant in the middle school experiment. For the four math problems in the dialogue, P1-P4, the participant's average pitch, intensity, and speaking rate during the first dialogue turn of each problem sub-dialogue are shown.

Discussion

The preliminary analysis of the low-motivation and high-motivation participants leads us to believe that prosody may be a factor worth considering in constructing a model of student motivation. Completing the analysis of all the participants will provide more information. One limitation of our analysis is that student attitudes related to motivation are recorded at a single point of time, in the post-session survey; we do not have real-time measures of engagement during the interaction.

To support robust adaptive dialogue interaction, we do not intend to model student motivation based on prosody alone; rather, we intend to use features of the student's problem-solving progress in combination with prosody. Exploring ways of combining these multimodal inputs is an important direction of future research.

Acknowledgments

The authors would like to thank Tricia Chaffey, Hyeji Kim, Emilia Nobrega, Yuliana Flores, Sam Baker, and Corey Bird. This work was supported by the National Science Foundation under Grant No. IIS-1637947.

References

1. P. Baxter, E. Ashurst, R. Read, J. Kennedy, and T. Belpaeme. Robot education peers in a situated primary school study: Personalisation promotes child learning. *PLOS ONE*, 12(5):1–23, May 2017.

2. N. Dahlbäck, A. Jönsson, and L. Ahrenberg. Wizard of Oz studies—why and how. *Knowledge-based systems*, 6(4):258–266, 1993.

3. J. Kerssen-Griep, A. R. Trees, and J. A. Hess. Attentive facework during instructional feedback: Key to perceiving mentorship and an optimal learning environment. *Communication Education*, 57(3):312–332, 2008.

4. R. Kumar, H. Ai, J. L. Beuth, and C. P. Rose. Socially capable conversational tutors can be effective in collaborative learning situations. In *Proceedings of Intelligent Tutoring Systems*, pages 156–164, 2010.

5. N. Lubold. Building rapport through dynamic models of acoustic-prosodic entrainment. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 297–300. ACM, 2017.

6. N. Lubold and H. Pon-Barry. Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, pages 5–12, 2014.

7. N. Lubold, E. Walker, and H. Pon-Barry. Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction*, pages 255–262, 2016.

8. R. Ploetzner, P. Dillenbourg, M. Praier, and D. Traum. Learning by explaining to oneself and to others. *Collaborative-learning: Cognitive and computational approaches*, pages 103–121, 1999.

9. N. Sadoughi, A. Pereira, R. Jain, I. Leite, and J. F. Lehman. Creating prosodic synchrony for a robot co-player in a speech-controlled game for children. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17, pages 91–99, 2017.

10. D. Szafir and B. Mutlu. Pay attention!: Designing adaptive agents that monitor and improve user engagement. In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, CHI '12*, pages 11–20, 2012.