Talk to Me: Verbal Communication Improves Perceptions of Friendship and Social Presence in Human-Robot Interaction

Elena Corina Grigore[1], Andre Pereira[1], Ian Zhou[1], David Wang[2], and Brian Scassellati[1]

[1] Yale University, 51 Prospect Street, New Haven, CT, 06511, USA
{elena.corina.grigore, andre.pereira, ian.zhou, brian.scassellati}@yale.edu
[2] Amity Regional High School, Woodbridge, CT, 06525, USA
wangda16amity@amityschools.org

Abstract. The ability of social agents, be it virtually-embodied avatars or physically-embodied robots, to display social behavior and interact with their users in a natural way represents an important factor in how effective such agents are during interactions. In particular, endowing the agent with effective communicative abilities, well-suited for the target application or task, can make a significant difference in how users perceive the agent, especially when the agent needs to interact in complex social environments. In this work, we consider how two core input communication modalities present in human-robot interaction—speech recognition and touch-based selection—shape users' perceptions of the agent. We design a short interaction in order to gauge adolescents' reaction to the input communication modality employed by a robot intended as a long-term companion for motivating them to engage in daily physical activity. A study with n = 52 participants shows that adolescents perceive the robot as more of a friend and more socially present in the speech recognition condition than in the touch-based selection one. Our results highlight the advantages of using speech recognition as an input communication modality even when this represents the less robust choice, and the importance of investigating how to best do so.

1 Introduction

Artificial agents, such as virtually-embodied characters or physically-embodied robots, can take on the role of help givers, companions, teachers, coaches, and so on [11], [25]. In this work, we focus on a physically-embodied robot intended as a long-term companion that keeps adolescents motivated to engage in daily physical activity throughout time. The importance of developing methods for keeping adolescents engaged in physical activity is stressed by data revealing alarmingly low numbers of adolescents who engage in levels recommended by health guidelines [2]. Furthermore, this occurs at an age when physical activity holds essential benefits for healthy growth and development [6]. Given that the constant support needed for behavior change [16] is either not readily available

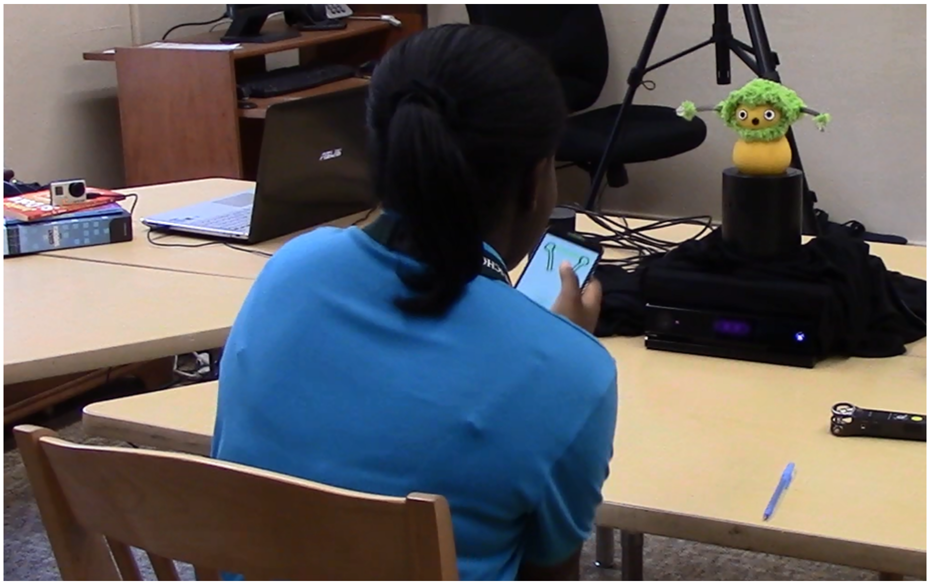

Fig. 1. Participant interacting with the robot using our smartphone interface.

or costly, employing a social agent as a companion would provide valuable social support to keep adolescents motivated to exercise routinely.

In order for such a companion to successfully provide its users with social support over time, enabling smooth and natural interactions through an effective communication modality is essential. We thus investigate the effects of the communication modality on users' perceptions of the agent. To this end, we conduct a study to observe adolescents' reactions to different ways of communicating with the robot (Figure 1 shows a participant during a session). Investigating how users' perceptions of the robot are affected by communication modality can guide us towards creating better and more effective agent interactions. In particular, we are interested in how users' perceptions of the agent change in terms of friendship and social presence factors. These factors are particularly relevant for creating engaging and compelling interactions, and are important to assess prior to the long-term deployment intended for our robot companion.

Given that verbal communication is at the core of how people interact with each other, one of the most natural ways of designing a human-robot communication interface is employing speech. Speech recognition provides a basic and natural way of conveying information to the robot. Even though it has seen serious advances in recent years, speech recognition technology still suffers from unreliable performance in noisy acoustic environments and when faced with speaker and language variability (especially when working with children [8]), as well as from difficulty handling free-style speech [4]. Due to the challenges faced by utilizing speech recognition reliably, many interfaces for communication with agents use touch-based inputs via tablet or smartphone devices [5], [18]. The use of such interfaces makes for more robust and dependable communication, but might take away from the benefits of having a more natural interaction.

Our work investigates how adolescents' perceptions of the robot change when speech recognition versus touch-input selection is used to communicate with the agent. Results reveal that, even though the former was more error-prone and less robust than the latter, users perceived the robot as more of a friend and more socially present in the speech recognition condition than in the touch-based selection one. This highlights the importance of employing a more natural way of communication, even when this represents the less robust choice.

2 Related Work

Speech recognition has advanced significantly in recent years, with numerous applications deployed on a large scale. Smartphones represent some of the most widely used devices that employ this capability, with 3.4 billion subscriptions in 2015 set to double by the year 2021 [7]. Most smartphones come equipped with intelligent personal assistants that help users complete day-to-day activities, from using social media platforms to searching for travel information. Such assistants take inputs in the form of voice, images, or contextual information to provide useful answers, typically in the form of natural language. Through the widespread use of these devices, owners have become acquainted with the power provided by speech recognition, making for smooth and natural interactions with their devices.

Speech recognition systems, however, still suffer from lack of robustness beyond constrained tasks and are not reliable when used in noisy environments. The highest performing systems in research struggle to obtain word error rates (WER) lower than 10%, and employ strategies that might prove to be unfeasible for most real-world applications [4]. Furthermore, data on the performance of commercially available speech recognition services show considerable variability with respect to performance metrics, with WERs ranging from 15.8% to 63.3% [14]. In this work, we wish to investigate whether employing speech recognition to enable communication between a robot and its users is of value even when this represents the less robust choice.

2.1 Speech Recognition Communication Interfaces

Speech interfaces have been used for communicating with social agents in different contexts and application domains, given the natural way of interaction such interfaces can provide. A salient research area in which speech recognition is used is that of creating multimodal human-robot interaction interfaces. Such interfaces include verbal communication and visual perception modules, among others, and employ speech recognition as an important part of the interface allowing users to communicate with the robot [9], [20].

In a study focusing on teaching a robot a complex task [24], the robot learns the task by both observing a human perform it and by interpreting the speech used by the person while doing so. The authors emphasize the utilization of speech recognition in human-robot interfaces to create natural and familiar ways for people to interact with robots, which could more quickly lead to their acceptance in human environments like homes and workspaces. Alongside providing a natural mode of communication, other advantages include more easily handling situations when a person's hands or eyes are occupied, benefits for enabling communication between robots and handicapped persons, and the use of the ubiquitous mobile devices discussed above that allow for two-way voice communication [23].

Although there exist a number of studies investigating the effect of output communication modalities on users' perceptions of agents (i.e. using different interfaces to convey information from the agent to the user such as text-to-speech, text, etc.) [22], [21], research on input communication interfaces is more scarce and is the area where we focus our current efforts.

2.2 Touch-based Selection Communication Interfaces

Faced with the challenges surrounding the reliable use of speech recognition, a growing number of studies have started utilizing touch-based selection input as the main form of interaction between humans and robots. In the past ten years, we can observe a clear trend in the increasing number of studies that make use of mobile devices or tablet touchscreens for this purpose. For example, the Nao or DragonBot robots have commonly been paired with touchscreen tablets to provide a context for human-robot interaction [5], [25].

Other studies have used mobile devices and touch interfaces to communicate with robots. In such a study, authors used a smartphone as a means of interacting with Nao and teaching it about new objects [18]. In a study exploring the concept of enjoyment in a human-robot interaction scenario with the elderly, authors used a touchscreen interface in place of a speech recognition system due to increased reliability and smoother interactions [11].

The research presented above gives us interesting insights into using different kinds of devices and interfaces for communication in human-robot interactions. However, researchers have not yet explored the tradeoff between employing a more natural versus a more robust interface for enabling communication between a user and a robot with respect to users' perceptions of the agent. Our work explores this tradeoff by investigating perceptions of friendship and social presence engendered by the use of different input communication modalities.

3 Methodology

3.1 Interaction Context

The interaction context for our study is that of a robot companion motivating adolescents to engage in daily physical activity. The companion is intended for long-term use, and the single-session study presented here constitutes the first interaction users would be going through during a longer-term study. During this session, the robot walks participants through its back-story and explains the different motivational strategies it would be using over time. The back-story consists of the agent taking on the role of a "robot-alien" whose space ship broke down on Earth and who needs help from the user in order to return home. The user can help the robot by exercising routinely and transferring "energy" to the robot by doing so. In a long-term scenario, this is accomplished by providing users with a wristband device that measures the physical activity level they engage in daily, which the robot can connect to in order to gain "energy points".

The back-story is created in order to build an engaging, compelling, and persuasive interaction by linking elements of the story (e.g., the fact that all inhabitants of the robot's planet have a high level of knowledge of physical activity) to ways in which the robot would help the user. In a long-term deployment, the agent can accomplish this by employing four different motivational strategies—cooperative persuasion, competitive persuasion, conveying information about physical activity through lessons and quizzes, and promoting self-reflection [1]. These motivational strategies are reinforced by providing users with a smartphone they can carry around with them during the day, while away from the physical robot. Participants also use the smartphone during their short, daily interactions with the physical robot in order to communicate with it.

3.2 Study Design

The user study we present herein examines the effects of speech recognition versus touch-based selection on users' friendship and social presence perceptions of the robot within the interaction context described in Subsection 3.1. Each session involves a participant interacting with the robot for six-to-nine minutes and being engaged in a dialogue on the topic of physical activity motivation for adolescents. The interaction is structured in the form of a dialogue controlled by the agent in which the robot either speaks to the participant or asks a question and waits for a response. Participants interact with the robot via a smartphone application that displays the appropriate interface based on whether the robot or the person is speaking, as well as based on the condition to which the participant is assigned. The conditions represent the use of two different input communication modalities: (1) speech recognition—when they are prompted through the smartphone interface, participants respond to the robot's questions by saying one of the answer choices displayed on the screen and their answers are evaluated through speech recognition, and (2) touch-based selection—participants respond to the robot's questions by using their finger to drag a button onto one of the answer choices displayed on the smartphone screen. We hypothesize that:

H1: Users would perceive the robot as more of a friend in the speech recognition than in the touch-based selection condition.

H2: Users would perceive the robot as more socially present in the speech recognition than in the touch-based selection condition.

3.3 Measures

The two main dependent variables (DVs) we examine are users' perceptions of the robot in terms of friendship and social presence. The two measures are based on standardized questionnaires measuring friendship [17] and social presence [10]. We examine subscales of both measures to investigate the effects of the input communication modality on users' perceptions of the agent. The two main DVs constitute important factors to consider in any human-robot interaction scenario in which we wish to create an engaging and compelling interaction and engender a positive rapport with an artificial agent.

Friendship. We employ the McGill Friendship Questionnaire (MFQ) [17] concerning a subject's assessment of the degree to which someone fulfills six friendship functions. Although friendships, like other relationships, vary in quality, research suggests it is possible to assess specific qualities, and the questionnaire used herein was created to define theoretically distinct friendship functions that distinguish between friends and non-friends, and that are associated with affection and satisfaction [17]. We believe that it is fundamental to engender the existence of such elements during interactions with social agents, whether they be long- or short-term. These factors are key to creating positive impressions of the agent, which is especially important during initial rapport development in relationships [29]. The more users perceive that the agent can fulfill friendship functions, the more likely they are to start building positive rapport with the agent.

The relevance of using friendship as a measure of the ability of an agent to establish meaningful relationships with its users has already been highlighted in a short-term study [15]. We employ this measure in a similar way to inform this ability of the agent for the intended long-term use of our robot companion. The MFQ consists of 30 zero-to-eight Likert items in total, with six subscales composed of five questions each. The six subscales included in the questionnaire are: stimulating companionship, help, intimacy, reliable alliance, self-validation, and emotional security. Although not straightforward to delineate, some friendship subscales such as intimacy might appear more relevant to long-term interactions. Nevertheless, these subscales could give us insights into how verbal communication might affect a long-term interaction. Additionally, including the full set of subscales allows for future comparisons with long-term interaction results.

Social presence. Social presence is another core aspect of human-human interactions that we should take into account when developing interactions between users and robots or social agents. Social presence was initially defined as "the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships" [26]. It is immediately apparent that users' perceived social presence of the robot is of high importance, given that we strive to develop robots that interact with their users in a convincing and social manner. An example of the power of creating strongly socially present robots is [19], where social presence is used to create believable and enjoyable board game opponents. It has been shown that an agent is more effective at being persuasive when perceived as socially present [27]. A robot companion that aims to keep adolescents motivated over time by employing persuasive motivational strategies would thus benefit greatly from being perceived as a strongly socially present party in the interaction.

The questionnaire used herein is based on [10], which consists of 36 zero-to-eight Likert items in total and defines social presence to include six important factors for the interaction. These factors of social presence represent the sub-scales used in the questionnaire: co-presence, attentional allocation, perceived message understanding, perceived affective understanding, perceived affective interdependence, and perceived behavioral interdependence. Social presence is directly applicable to short-term interactions and is routinely employed in their assessment [3], [28].

3.4 Apparatus and Experimental Setup

In our study, we use a modified (to make programmable) version of MyKeepon, a commercially available variant of the Keepon robot. Keepon is a non-mobile robot with four degrees of freedom, designed for interaction with children [13]. The modified MyKeepon can be seen in Figure 1, presenting a green “alien” hat with antennae, consistent with the robot’s back-story during the interaction.

Throughout the study, each participant sat at a table, with the robot placed on the table. A Kinect was positioned in front of the robot and covered with a black cloth to look integrated with the robot. The Kinect tracked the position of the user’s head so that the agent would turn left or right, as needed, in order to mimic placing its attention on the participant as he or she moved. The user was also given a smartphone device with a custom-built application displaying the interface based on the condition. Figure 1 shows the experimental setup during an interaction session with an adolescent.

3.5 Participants

We conducted a study with n = 52 participants, 26 male and 26 female, in local high schools. The subject population consisted of early adolescents and adolescents aged 14-to-16, with an average age of 15. Apart from the age, no other recruitment criteria were applied. Recruitment was managed with the help of the schools’ staff and participants did not receive monetary compensation. We obtained ethical approval from the Human Subjects Committee at our university.

3.6 Procedure

Participants were randomly assigned to either the speech recognition or the touch-based selection condition. The former included 26 participants, while the latter included the other 26. Users were given an assent form to read and sign prior to participation. Each participant was seated at a table, in front of the robot that was initially still and silent. We first gave the user a brief overview of the physical activity scenario and then handed him or her the smartphone.

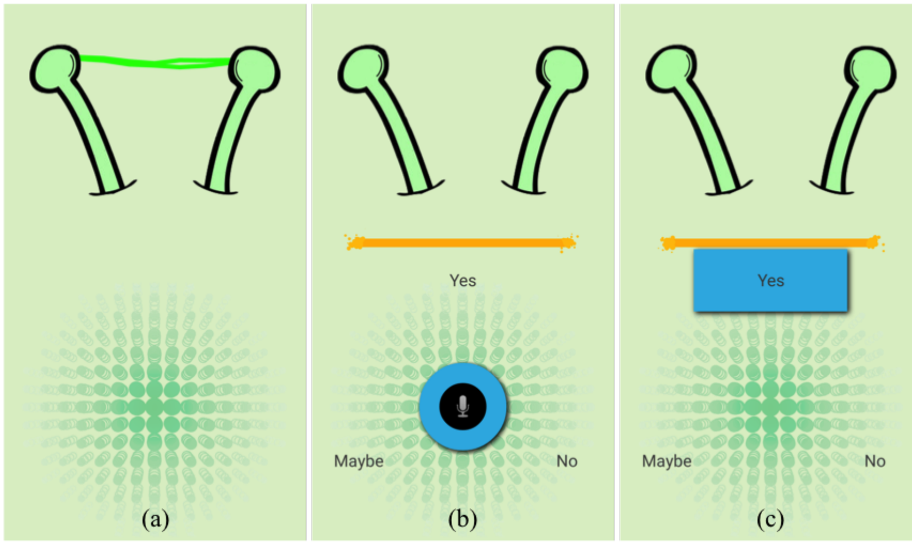

Fig. 2. Interface for communication with the robot. Figures show the interface displayed for: (a) the robot speaking during both conditions, (b) the speech recognition condition, and (c) the touch-based selection condition.

Each participant was given a brief explanation on how to use the interface to communicate with the robot. Figure 2(a) shows the interface displayed across conditions when the robot is speaking to the user. Figure 2(b) shows the interface displayed in the speech recognition condition. Users in this condition would say one of the options displayed on the screen to answer a question. We employed the Google Speech Recognition API [12]. The system could handle synonyms and expressions with similar meaning to the words shown (e.g., "yeah", "sure", and "of course" would all be recognized as "yes"). When the system could not correctly identify the input or the participant would not use the interface correctly, the robot would prompt the user with a brief instruction based on the particular mistake. Figure 2(c) shows the communication interface displayed in the touch-based selection condition. Participants in this condition would use their finger to drag the button displayed on the screen onto one of the answer choices in order to respond. As long as they kept the button pressed, participants could change their mind and move the button from one choice to another. In this condition too, users were prompted with a brief instruction if they would not utilize the interface correctly.

The interaction itself consisted of the robot unfolding its back-story and talking about physical activity motivation. The interface would change from Figure 2(a) to either Figure 2(b) or 2(c) (depending on the condition) when the robot would go from speaking to waiting for the participant's answer. The robot asked a total of six questions during each interaction, and users could choose among three simple entries to respond. The robot's replies to the different answer choices

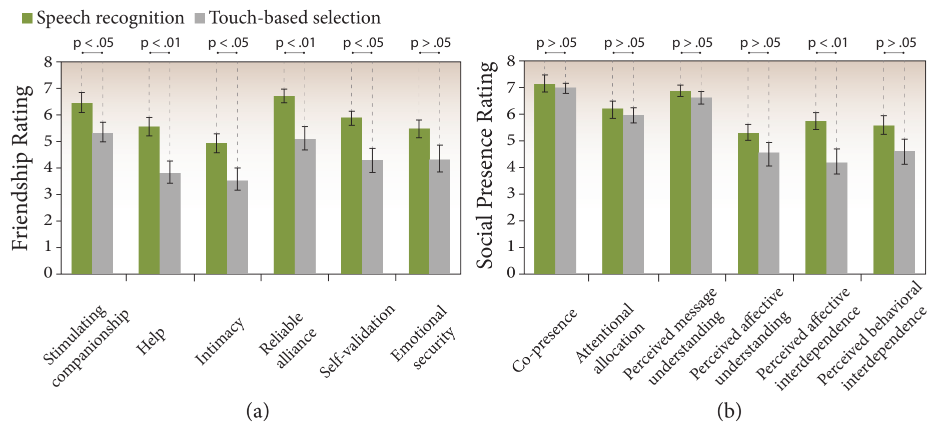

Fig. 3. Mean ratings for (a) friendship subscales (p-values based on the Mann-Whitney U test statistic, with means and error bars depicted to visualize results for this data), and (b) social presence subscales (p-values based on the independent-samples t-test statistic). We consider Bonferroni-adjusted α levels with $p < .008$ for significance and $p < .017$ for marginal significance. All error bars represent $\pm 1SE$.

differed slightly to give users feedback that their particular answer was understood, but otherwise the script was fixed. The interaction was approximately seven minutes long. At the end of the interaction, participants were asked to fill out the friendship and social presence questionnaires.

4 Results

This section presents the analysis of the friendship and social presence subscales, and discusses the robustness of the two conditions employed in the current study.

Friendship. We first evaluated the internal consistency for each of the six friendship subscales, resulting in highly reliable values. We thus computed the scores for the six subscales by averaging over values within each. The Shapiro-Wilk test of normality revealed our data is not normally distributed ($S - W = .93, df = 52, p = .005$). We thus employed the Mann-Whitney U test to compare users' ratings of the robot between the two conditions.

We performed the Mann-Whitney U test for each of the six subscales, using a Bonferroni-adjusted α level of .008 (.05/6) for significance and .017 (.1/6) for marginal significance. Participants rated the robot significantly higher in the speech input condition than in the touch input one for help ($U = 183, p = .004, z = -2.84$) and reliable alliance ($U = 186, p = .005, z = -2.79$), and marginally significantly higher for self-validation ($U = 200.5, p = .012, z = -2.52$). Based on the Bonferroni correction, there was no statistical significance in ratings for stimulating companionship ($U = 215, p = .024, z = -2.26$), intimacy ($U = 228.5, p = .045, z = -2.01$), or emotional security ($U = 246, p =$

$.094, z = -1.68$), although trends do suggest higher ratings in the speech recognition than in the touch-based selection condition. These results are highlighted in Figure 3(a) and they reveal the strong effect of employing speech recognition within the context of fostering positive rapport during human-robot interactions.

Social Presence. We computed social presence subscale scores in a similar manner to the friendship scores, with internal consistency evaluations resulting in reliable and highly reliable values for each subscale. The Shapiro-Wilk test of normality suggested that normality of data is a reasonable assumption for the social presence data ($S - W = .97, df = 52, p = .146$), and so we employed the independent-samples t-test to compare users' ratings between the two conditions.

We performed the independent-samples t-test for the six subscales, using a Bonferroni-adjusted α level of .008 (.05/6) for significance. Participants rated the robot significantly higher in terms of perceived affective interdependence in the speech recognition condition ($M = 5.72, SD = 1.58$) than in the touch-based selection one ($M = 4.17, SD = 2.31$), $t(50) = 2.83, p = .007$ (t-value reported for unequal variances), while the other five subscales (co-presence, attentional allocation, perceived message understanding, perceived affective understanding, and perceived behavioral interdependence) were not significant (Figure 3(b)). Although the friendship ratings reveal a stronger impact of the communication interface employed, social presence results also bolster the importance of employing a more natural and familiar mode of communication with robots.

Robustness. In order to assess the reliability of the two input communication modalities employed, we computed error rates for each. For both conditions, each session constitutes of the robot asking the user a total of six questions, for which the participant can choose among three simple, one-word entries to respond. In the speech recognition condition, we encountered two types of errors: (1) the system evaluates the user's speech as something other than the options available on the smartphone screen, and (2) there is no audible speech that can be evaluated by the system. Since each question response consists of one-word answers, we computed the word error rate (WER) for our system by dividing the total number of errors encountered (23) by the total number of questions asked (six questions per participant, with 26 participants in the speech recognition condition), and obtained a value of 14.74%. In the touch-based selection condition, the interface performed without errors, given the tailored nature of the interface design and the reliability of using a smartphone touch interface.

To ensure that the errors encountered in the speech recognition condition did not impact users' perceptions of the robot, we analyzed this data more in depth. Out of the 26 total participants in this condition, our system presented errors for 14 (with an average of 1.64 errors per user), and worked without errors for the remainder of 12. We thus applied the same statistical tests employed above for this condition only, using a binary coding scheme to divide the group into participants with and without errors. We again employed the Mann-Whitney U test for friendship subscales and the independent-samples t-test for social presence subscales, with the same significance levels used above. We obtained no significant effect for errors for either of the friendship or social presence subscales.

This shows that the difference in users' perceptions of the robot is not influenced by errors present in the speech recognition condition.

Compared to available commercial systems, our speech recognition WER represents a fairly low error rate, but this is due to the nature of the constrained task and simple answers employed during the interaction. To note, however, is the fact that the speech recognition condition constitutes the less robust choice out of our two input communication modalities, and that this condition still engenders higher perceptions of the robot for participants. These results stress that it is important to consider the tradeoff between utilizing a more natural versus a more robust interface when deciding which input communication modality to employ for interactions with robots, or artificial agents more broadly.

5 Discussion and Conclusions

This work examines the effects of employing speech recognition versus touch-based selection as an input communication modality on users' perception of an agent. We conducted a study to investigate how the two modalities shape users' friendship and social presence perceptions of the robot within the context of physical activity motivation for adolescents. Our intuition was that, as a more natural means of communication, speech recognition would prove to engender stronger perceptions of friendship and social presence than touch-based selection, even when this represents the less robust choice.

Hypothesis H1 predicted stronger friendship perceptions of the robot in the speech recognition than in the touch-based selection condition. We obtained statistically significant differences between conditions for help and reliable alliance, and marginally statistically significant for self-validation. Although we did not obtain statistically significant results for stimulating companionship, intimacy, and emotional security, we did notice a trend in the same direction. A possible explanation for why these subscales are only showing a trend could be due to their limited relevance to short-term interactions. We suspect that a longer exposure to our agent could result in significant results for these subscales as well, and consider this an interesting avenue for future research.

The second hypothesis, H2, is partly supported by our results. Co-presence, attentional allocation, and perceived message understanding engendered high perceptions of the agent for users across conditions, and so these subscales did not yield statistically significant differences. We did not obtain any significant differences for perceived message understanding and perceived behavioral interdependence. However, participants in the speech recognition condition perceived stronger affective interdependence than those in the touch-based selection one. Perceived affective interdependence is defined as "the extent to which the user's emotional and attitudinal state affects and is affected by the emotional and attitudinal states of the interactant." It represents an important facet of creating an engaging, persuasive, and compelling agent that aims to motivate adolescents. Having users perceive high affective interdependence when interacting with a social robot is fundamental to any type of agent, be it one that aims to entertain, motivate, help physically, or simply provide information.

The significant differences we observe in users' perceptions of the robot are within the context of employing a smartphone as part of the input communication modality. We do so deliberately, in order to design a robot that can realistically be used as a long-term companion in adolescents' homes and employ widespread, commercially available technology (smartphones) to this end. Although comparing our conditions to using speech recognition without a smartphone would help put the significant differences obtained into context, it would introduce another independent variable, i.e. the presence or absence of the device. Nevertheless, we believe this comparison to be an interesting future direction.

The results yielded by the current single-session study inform us that, for our long-term application scenario, using natural language for communication with a social agent is worthwhile despite the presence of speech recognition errors. Given the general pertinence of the two scales we employed, we believe that these results can be applied to many other human-agent interaction domains where social interactions are key.

References

1. Details omitted for double-blind reviewing
2. Physical activity guidelines for americans midcourse report: Strategies to increase physical activity among youth. Tech. rep., U. S. Department of Health and Human Services (2013)
3. Adalgeirsson, S.O., Breazeal, C.: Mebot: a robotic platform for socially embodied presence. In: Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction. pp. 15–22. IEEE Press (2010)
4. Barker, J., Marxer, R., Vincent, E., Watanabe, S.: The thirdchime'speech separation and recognition challenge: Dataset, task and baselines. In: 2015 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2015) (2015)
5. Baxter, P., Wood, R., Belpaeme, T.: A touchscreen-based 'sandtray' to facilitate, mediate and contextualise human-robot social interaction. In: Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction. pp. 105–106. HRI '12, ACM, New York, NY, USA (2012)
6. Boreham, C., Riddoch, C.: The physical activity, fitness and health of children. Journal of sports sciences 19(12), 915–929 (2001)
7. Ericsson: Ericsson mobility report (February 2016)
8. Gerosa, M., et al.: A review of asr technologies for children's speech. In: Proceedings of the 2nd Workshop on Child, Computer and Interaction. ACM (2009)
9. Gorostiza, J.F., et al.: Multimodal human-robot interaction framework for a personal robot. In: Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on. pp. 39–44. IEEE (2006)
10. Harms, C., Biocca, F.: Internal consistency and reliability of the networked minds measure of social presence (2004)
11. Heerink, M., et al.: Enjoyment intention to use and actual use of a conversational robot by elderly people. In: Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction. pp. 113–120. ACM, New York, USA (2008)

12. Inc., G.: Android speech (2014), available at http://developer.android.com/reference/android/speech/package-summary.html
13. Kozima, H., Michalowski, M.P., Nakagawa, C.: Keepon. International Journal of Social Robotics 1(1), 3–18 (2009)
14. Kudryavstev, A.: Automatic speech recognition services comparison (2016), available at http://blog.griddynamics.com/2016/01/automatic-speech-recognition-services.html
15. Leite, I., et al.: "why cant we be friends?" an empathic game companion for Long-Term interaction. In: Intelligent Virtual Agents, pp. 315–321. Lecture Notes in Computer Science, Springer Berlin Heidelberg (2010)
16. Marcus, B.H., Forsyth, L.: Motivating people to be physically active. Human Kinetics (2003)
17. Mendelson, M., Aboud, F.: Mcgill friendship questionnairerespondents affection (mfq-ra). Measurement Instrument Database for the Social Science (2012)
18. Oudeyer, P.Y., Rouanet, P., Filliat, D.: An integrated system for teaching new visually grounded words to a robot for non-expert users using a mobile device. In: IEEE-RAS International Conference on Humanoid Robots. Tsukuba, Japan (2009)
19. Pereira, A., Prada, R., Paiva, A.: Socially present board game opponents. In: Advances in Computer Entertainment, pp. 101–116. Springer (2012)
20. Perzanowski, D., Schultz, A.C., Adams, W., Marsh, E., Bugajska, M.: Building a multimodal human-robot interface. Intelligent Systems, IEEE 16(1), 16–21 (2001)
21. Qiu, L., Benbasat, I.: An investigation into the effects of text-to-speech voice and 3d avatars on the perception of presence and flow of live help in electronic commerce. ACM Transactions on Computer-Human Interaction (TOCHI) 12(4), 329–355 (2005)
22. Qiu, L., Benbasat, I.: Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. International journal of human-computer interaction 19(1), 75–94 (2005)
23. Roe, D.B., Wilpon, J.G., et al.: Voice communication between humans and machines. National Academies Press (1994)
24. Rybski, P.E., et al.: Interactive robot task training through dialog and demonstration. In: Proceedings of the ACM/IEEE International Conference on Human-robot Interaction. pp. 49–56. ACM, New York, USA (2007)
25. Short, E., et al.: How to train your dragonbot: Socially assistive robots for teaching children about nutrition through play. In: The 23rd International Symposium on Robot and Human Interactive Communication. pp. 924–929. IEEE (2014)
26. Short, J., Williams, E., Christie, B.: The social psychology of telecommunications (1976)
27. Skalski, P., Tamborini, R.: The role of social presence in interactive agent-based persuasion. Media psychology 10(3), 385–413 (2007)
28. Slater, M.: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society of London B: Biological Sciences 364(1535), 3549–3557 (2009)
29. Tickle-Degnen, L., Rosenthal, R.: The nature of rapport and its nonverbal correlates. Psychological inquiry 1(4), 285–293 (1990)