Help-Giving Robot Behaviors in Child-Robot Games: Exploring Semantic Free Utterances

Cristina Zaga, Roelof A.J. De Vries, Sem J. Spenkelink, Khiet P. Truong, Vanessa Evers
Human Media Interaction, University of Twente, The Netherlands
{c.zaga, r.a.j.devries, k.p.truong, v.evers}@utwente.nl

***Abstract*—We present initial findings from an experiment where we used Semantic Free Utterances — vocalizations and sounds without semantic content — as an alternative to Natural Language in a child-robot collaborative game. We tested (i) if two types of Semantic Free Utterances could be accurately recognized by the children; (ii) what effect the type of Semantic Free Utterances had as part of help-giving behaviors with *in situ* child-robot interaction. We discuss the potential benefits and pitfalls of Semantic Free Utterances for child-robot interaction.**

***Index Terms*—child-robot interaction, semantic free utterances**

I. Introduction

An alternative to using Natural Language in child-robot interaction (cHRI) is the use of Semantic Free Utterances (SFUs)[6]. Using the terminology of [6], a distinction can be made between Gibberish Speech (GS), i.e., vocalizations of meaningless strings of speech sounds that resemble human speech, and Non Linguistic Utterances (NLU), i.e., nonvocal re-created sounds that show no resemblance to natural language or speech (e.g., beeps, whirrs). However, existing work provides limited information about: (i) which types of SFU (i.e., GS or NLU) are more effective in real-life scenarios, where SFUs are coupled with robot's actions, and (ii) the effect SFUs have on *in situ* child-robot interaction, as existing work only focuses on video-based evaluations [2], [3]. To explore the aforementioned gaps, we conducted an exploratory study to evaluate the effects of GS and NLU in a child-robot collaborative game. We focused on a widely studied aspect of peer collaboration i.e., help-giving [5]. Our aim was twofold: (i) to investigate if SFUs can be accurately recognized by the children in a real-life help-giving context; and (ii) to investigate how two types of SFUs (i.e., GS and NLU) impact the perceived help-giving of robot behaviors in the game.

II. Method

Setup and procedure. Eighteen Dutch children ($N = 18$; age: $M = 9.16$, $SD = 0.38$; 8 males, 10 females) in same-sex couples played a tangible 4x4 pair-matching memory game (see Fig. 1) with a non-anthropomorphic robot, i.e., Festo Robotino[1], which was remotely controlled. To support ecological validity, the experiment was carried out in the gym of the school of the children. The goal of the game is to match all pairs of identical objects hidden under crates in a sequence of rounds. In each round two crates can be lifted to see if the objects match. The first rounds usually consist of making guesses — there is no information about the hidden objects yet — but after a few of rounds enough information should be available to make a match, if the objects are remembered. The robot — a help-giving peer [7] — assisted to find pairs of objects in the game. The knowledge of the robot was always equal to maximum possible knowledge of the children. The robot gave help by providing game move suggestions when the possibility of matching a pair of objects was possible from the knowledge of the game, but otherwise it would follow and watch the children making guesses.

Help-giving behavior design. We designed the help-giving robot behavior movements to suggest a good game move by nudging the crate hiding an object that could be matched. In the NLU and GS condition, we paired these movements with either an NLU or GS sound matching the time of the movements, i.e., both 2.68 seconds. For the NLUs we used synthesized *metaphoric auditory icons* [6]. For the GS, we inspected prosodic characteristics of children's suggestive speech in a Dutch child-child collaboration corpus [1] and tried to replicate those prosodics with a sustained */m/*-sound.

Conditions. We designed the study as a between-subjects experiment with three conditions, each with six participants: the two types of SFUs, Gibberish Speech and Non-Linguistic Utterances and a no-sound control condition (NS). The manipulation was applied only to the sounds, the robot movements were equal across conditions. We hypothesized that: (H1) SFU sounds will be perceived as being more help-giving than no sound, and in particular, (H1a) GS will be perceived as being more help-giving than NLU and no sound, because GS conveys more paralinguistic cues [4] and NLUs can be more prone to ambiguity [3]. For the latter reason, (H2) the recognition rate of the intention of the sound (help-giving suggestion) for GS will be higher than for NLU.

Evaluation. To evaluate if the children recognized the intention of the sounds (H2), we did an *ex-post* recognition test in the GS and NLU conditions. The children had to match the sounds — replayed during this time — to four cartoon-like images depicting the intentions of the sounds (i.e., suggestion, doubt, error, reward). To evaluate the effect of our manipulation on the perceived help-giving of the robot behaviors on the game (H1, H1a), we used observational methods. Two coders observed compliance to suggestion rate (number of times complied to help-giving suggestions divided

[1] http://www.festo-didactic.com/

978-1-4673-8370-7/16/$31.00 © 2016 IEEE

Fig. 1. The setup of the tangible 4x4 pair-matching memory game.

by total number of help-giving suggestions) and average game-round duration (i.e., total game duration divided by number of rounds in the game). The rationale behind these measures is to have an objective overview of the children's perception of the help-giving behavior in a collaborative game context. We operationalized the robot being perceived as help-giving when there is: high compliance to help-giving suggestion rate, resulting in a low average round game duration. We think that if there is perceived help-giving, compliance will be high, the children will be decisive and therefore the rounds will be short. Consequently, if there is no or less perceived help-giving, the compliance will be low, the children are more indecisive, and the game rounds will take more time.

III. Results and Discussion

The recognition of the help-giving sounds was higher in the GS condition (85%) than in the NLU condition (20%). A three-way ANOVA with post-hoc Bonferroni correction showed that the participants took less time on average each round ($F(2, 15) = 4.202$, $p = .036$) in the NLU ($M = 17.76$, $SD = 3.41$) than in the NS ($M = 28.36$, $SD = 10.62$) condition, but no differences between GS ($M = 19.92$, $SD = 3.16$) and NLU or NS were found. The same trend emerged for the rate of compliance ($F(2, 15) = 8.152$, $p = .004$), where the rate was lower in NLU ($M = 0.69$, $SD = 0.02$) than in NS ($M = 0.90$, $SD = 0.22$), and again no differences between GS ($M = 0.78$, $SD = 0.20$) and NLU or NS were found. Our findings provide insights on (i) the recognition of help-giving SFUs in a game context and (ii) how the interpretation of SFUs affects perceived help-giving of the robot on the game.

Quantitative analysis. As expected (H2), the children were able to match the GS to its intention with a better recognition rate than the NLU. We assume that the stronger paralinguistic cues carried by the GS helped the children in the recognition [4]. Additionally, the results of the recognition of the NLUs show that they lead to more ambiguity in the interpretation of the intention. This is partly in line with the literature on NLUs [2]. Without the SFUs, the intentions of the actions were clear to the children, as the high compliance in the control condition (NS) shows. Therefore, the actions by themselves were not ambiguous, but when combined with the NLU, they were misleading to the children. We expected SFU sounds to be perceived as more help-giving than no sounds (H1), specifically for GS (H1a). We hypothesized that perceived help-giving would result in a high compliance to suggestion rate and therefore a fast round time. However, for NS, we observed a significant *higher* compliance rate and *slower* round time compared to NLU. We think that NLU behavior caused more indifference to robot suggestions as the NLU sound part was not well recognizable. But, we expected that a lower compliance rate for NLU would result in *slower*, not faster round time.

Qualitative analysis. To explain *why* in the NS condition the children had a higher rate of compliance, but were slower in finishing the rounds compared to the NLU condition, we performed qualitative analyses on the video recordings. We observed that the children in NS were rather confused by the absence of sounds and this led them to play the game very passively and await robot suggestions. Conversely, the children seemed more engaged with the game in the NLU condition, which can be substantiated by the slower game time. The children in the GS condition seemed more engaged with the game *and* the robot, which is in line with the high recognition of sound intention (help-giving 85%). Here, we also observed more 'collaborative' behavior (e.g., joint-actions) with the robot, which was lacking in the NLU condition where the participants played the game ignoring robot help-giving.

Conclusion. Our preliminary results suggest that when designing help-giving SFU robot behaviors for cHRI, it is important to not only investigate the recognition of the behaviors, but also how the recognition of the behaviors will affect the robot interaction; as demonstrated by the passive and compliant behavior of the children in the control condition, the excluding behavior, but fast rounds in the NLU condition, and the collaborative and recognizable behavior in the GS condition.

Future Work. We will run a bigger study combining our results to investigate the interaction between SFU recognition and children's behavior. We will incorporate our annotation scheme with some of the qualitative interaction features that proved to be informative.

Acknowledgment

This work is funded by EU FP7 SQUIRREL (Clearing Clutter Bit by Bit) project (grant 610532)

References

[1] J. Kim et al. Vocal turn-taking patterns in groups of children performing collaborative tasks: an exploratory study. In: *Proceedings of Interspeech*, pp. 1645-1649. Dresden, Germany. 2015.

[2] R. Read and T. Belpaeme, How to use non-linguistic utterances to convey emotion in child-robot interaction, In: *HRI* pp. 219-220. Boston, MA, U.S.A.: ACM/IEEE, 2012

[3] R. Read and T. Belpaeme, Situational context directs how people affectively interpret robotic non-linguistic utterances. In: *HRI* pp. 41-48. Bielefeld, Germany: ACM/IEEE, 2014.

[4] R. E. Remez, P. E. Rubin, D. B. Pisoni and T. D. Carrell, Speech perception without traditional speech cues. In: *Science*, 212, 947-950, 1981.

[5] E. Walker, M. Rummel, P. K. Koedinger, Modeling Helping Behavior in an Intelligent Tutor for Peer Tutoring. In: *AIED*, pp. 341-348, 2009.

[6] S. Yilmazyildiz, R. Read, T. Belpaeme and W. Verhelst, Review of Semantic Free Utterances in Social Human-Robot Interaction, In: *International Journal of Human-Computer Interaction*, 2015

[7] C. Zaga, M. Lohse, K.P. Truong, and V. Evers *The Effect of a Robot's Social Character on Childrens Task Engagement: Peer Versus Tutor.* In Proceedings of The International Conference of Social Robotics (ICSR'15) pp. 704-713, Springer International Publishing, 2015