
STORYTELLING WITH A SOCIAL ROBOT

A PREPRINT

Nicole Goossens,
n.goossens_1@uvt.nl
CS&AI
Tilburg University
Tilburg, The Netherlands

Rian Aarts,
A.M.L.Aarts@uvt.nl
Culture Studies
Tilburg University
Tilburg, The Netherlands

Paul Vogt,
p.a.vogt@uvt.nl
CS&AI
Tilburg University
Tilburg, The Netherlands

March 1, 2019

ABSTRACT

Research on social robots in language education has become increasingly popular in recent years. However, the text-to-speech voices that robots have are not very expressive and lack many prosodic cues such as pitch and intonation. Yet, young children are very sensitive to prosodic cues and rely on them to comprehend spoken language. In this study the effect of the expressiveness and linguistic complexity of a robot's speech on language production and engagement in Dutch L2 children was investigated. In three reading sessions, children told stories from three picture books together with a robot that used either an expressive or inexpressive voice. The stories became more linguistically complex with each reading session. Results showed no effect of voice condition on either language production or engagement. Story complexity did have an effect on MLU.

Keywords second language learning · storytelling · child-robot interaction · engagement · synthetic speech

1 Introduction

In the last twenty years, technology has become increasingly important in language education. However, many existing technologies that are currently used for language learning, such as tablets and computers, lack social interactions which are essential for language learning [1, 2]. Social robots are able to provide these interactions, and many social robots have then also been applied for second language learning [3, 4, 5, 6]. Many studies in this field have used social robots to teach participants vocabulary in the second language using various methods, such as simple vocabulary tasks, games and, reading stories [7, 8, 9, 10]. However, nearly all these studies used synthetic text-to-speech (TTS) which might influence language learning in children. TSS lacks many prosodic cues that are important in learning and understanding language [11]. For example, prosodic cues, such as intonation and pauses, determine word and sentence boundaries, helping listeners to comprehend and identify words in a stream of speech easier [12]. Young children are also very sensitive to prosodic cues and rely on them to comprehend spoken language. Not much is known about the influence of synthetic speech in social robots on language learning. It is therefore interesting to investigate the effect of TTS on children's language learning.

In this study, we will investigate the effects of the expressiveness and linguistic complexity of a robot's speech when telling a story on children's language production and engagement. Language production will be measured by story length, expressive vocabulary and Mean Length of Utterance (MLU). Engagement is measured by both child-task and child-robot engagement. The design of this experiment is based on the insights of Dutch pre-school teachers that were interviewed previous to this study. In three sessions, a social robot will tell stories from picture books together with children that have Dutch as a second language. The robot will either use an inexpressive robotic voice or an expressive human voice. Furthermore, the linguistic complexity the robot uses will increase with each reading session.

2 Related work

As stated before, not much research has been done on the expressiveness of robot speech and how it effects human-robot interaction. One study showed that people seem to prefer interacting with a robot that sounds more human-like [13]. People also approach robots with a human voice closer compared to robots with a synthetic voice [14]. Furthermore, people feel psychologically closer, and antropomorphize robots more when they have the same gender and a human-like voice compared to robots that have the opposite gender and a synthetic voice [15]. However, these studies were done on adults.

In child-robot interaction, one study found that children enjoyed interacting more with a robot that had an expressive voice and used expressive gestures than with a robot that did not display this behaviour [16]. However, children did find the non-expressive robot easier to understand, possibly due to the lack of change in pitch [16]. In another study, they compared humans and robots telling stories in either a static or expressive way [17]. Children could recall more details of the story when the story was told in a behaviourally expressive manner. Furthermore, children's recall of the story from the expressive robot was equal to the expressive human, and higher than the static human [17]. Westlund et al.[7] also compared flat and expressive storytelling robots. Children told longer stories and were more engaged when they interacted with the expressive robot [7]. Children thus seem to enjoy interacting with expressive robots more than with non-expressive robots. Moreover, the more expressive storytelling robots appear to yield higher learning gains in children.

In this study, we investigate to what extent an inexpressive voice produced by the robot's TTS has an effect on children's storytelling and engagement compared to an expressive human voice. We expect that children produce more complex stories and are more engaged when interacting with an expressive robot. In addition, we investigate the effect that story complexity in three subsequent sessions has on children's storytelling and engagement. Here we expect children to produce more complex stories when the robot tells more complex stories, but that engagement will drop due to the novelty effect.

3 Methods

The study had a 2x3 mixed-design. In one between-subjects dimension, the robot's voice was either a flat TTS voice from the NAO robot (robotic voice condition) or a pre-recorded audio of a human voice, transformed to give the voice a robotic sense (human voice condition). In the other within-subjects dimension, there were three reading sessions in which the children interacted with the robot narrating increasingly complex stories.

3.1 Participants

We recruited 38 participants from two Dutch pre-schools. Only children that had Dutch as a second language and were between four and six years of age were eligible for participation. After dropout, 34 children remained (20 males, 14 females). Children had a mean age of 5 years and 4 months ($SD = 0.6$). There were 14 different native languages that were spoken by the children, which made the group very diverse. Participants were randomly divided into the two voice conditions.

3.2 Materials

For this experiment the NAO humanoid robot was used. A USB microphone was used to record the human voice. The audio files were altered by making the pitch of the human voice 18 - 20 % higher, making it sound more similar to the pitch of the text-to-speech voice and more "robot-like". Children heard the robot speak in the same voice every session. The movements the robot made were kept constant in both voice conditions and were very minimal. The robot moved its head towards the child and pointed at the screen every time it asked a question.

To control for the children's language proficiency in Dutch, the passive vocabulary sub task of the Toets Tweetaligheid (Dutch Bilingualism Test) was used [18]. The original version contains 60 items. However, due to time constraints, the sub task was split in half, resulting in two different versions with 30 items each. The children were only tested once, before the start of the reading sessions.

Three picture books were used: Boer Boris en de olifant by Ted van Lieshout and Philip Hopman, De verrassing by Sylvia van Ommen and Tim op de tegels by Tjibbe Veldkamp and Kees de Boer. The books were also used in that order. The number and complexity of selected target words and connectives increased with every book. Target words were words that are less commonly used in the classroom, as those were indicated to be difficult Dutch L2 children by the teachers that were interviewed prior to this study.

In each reading session the robot told a story from one of the picture books. A different picture book was used for every session. The picture books were shown on a laptop screen and the robot stood next to the child, turned towards the screen. The robot only narrated part of the story and asked several questions about events and characters of the story while telling the story. When the robot finished, it asked the child to finish the story using the pictures from the books as a guide. All sessions were videotaped and there were always two experimenters present in the room. One experimenter would interact with the child, while the other experimenter would keep more on the background and control the robot using a Wizard-of-Oz approach. This approach was chosen because of the poor performance of the automatic speech recognition (ASR) with children’s speech [19].

3.3 Procedure

The experiment consisted of four parts. The first part was a group introduction of the robot to the children, to become more acquainted with the robot before interacting with it one-on-one. During the introduction the robot performed two dances and played a short game with the children.

The second part was the Dutch as a second language test that all children took individually. This took place on the same day as the introduction. For each item, four pictures were shown on a laptop screen. Then, a recording of the word played and the child had to indicate the picture corresponding to the word they heard. When a child gave five consecutive wrong answers the test was terminated. When the test finished, the child was brought back to its class. The test took around five to ten minutes in total.

The third part are the reading sessions, which started one week after the language test. The child was taken to an experiment room in their school where the researchers first explained that the robot wanted to read the picture book together with them, but that it only knew part of the story and they had to finish the story using the pictures in the book. The robot first welcomed the child and asked the child if they knew the book that was shown on the screen. At some point during the story, the robot asked the child if they could finish the story by telling the robot what happened on the pictures. The robot had four standard encouragements in case the child did not respond. If the child still did not respond after these encouragements, one of the researchers intervened. After the child finished, the robot told the child that they did a good job and asked if they liked reading the book, after which it said goodbye. Then the child was taken back to the classroom. The reading sessions took between six to ten minutes of which the robot read for three to four minutes.

The fourth part took place when all children finished the reading sessions. Children said goodbye to the robot in small groups. The researchers gave all children a small reward for participating and the children were allowed to ask the researchers questions about the robot.

3.4 Analyses

All children’s stories were transcribed to assess their language production during storytelling. Language production was measured by counting the story length, the Mean Length of Utterance (MLU) and use of target words, all in number of words. Engagement was assessed by rating two-minute fragments of each session following an adapted version of the ZiKo coding scheme [20]. Two forms of engagement were assessed: child-task engagement (the extent to which the child was engaged with the task) and child-robot engagement (the extent to which the child was engaged with the robot).

4 Results

First we investigated if there were significant differences between the two voice conditions and schools on the language test. On average, children answered $M=21.97$ ($SD=5.19$) items correct on the Dutch language test. Children in the robotic voice condition answered significantly more items correct ($M = 23.60$; $SD = 4.35$) than children in the human voice condition ($M = 19.64$; $SD = 5.56$; $t(32) = 2.33$; $p = .026$). No significant differences were observed between the two schools ($t(32) = .31$; $p = .761$). Furthermore, test score had significant effects on story length, $F(1, 26) = 15.80$, $p < .001$, target word use, $F(1, 26) = 6.30$, $p = .02$ and child-task engagement, $F(1, 26) = 4.41$, $p = .05$. Children who had a higher test score, scored higher on all these measures.

A series of 2 (voice condition) \times 3 (story complexity) mixed ANOVA’s showed there was no significant difference between the two voice conditions on children’s story production, as measured by the length of the stories the children told, $F(1, 26) = .16$, $p = .69$, $MLUF(1, 26) = .22$, $p = .64$, $\eta^2 = .01$, the number of unique target words the children used, $F(1, 26) = 1.20$, $p = .28$, $\eta^2 = .04$. Results also did not show significant effects of story complexity on story length $F(2, 52) = 2.59$, $p = .09$, $\eta^2 = .09$ and target word use, $F(2, 52) = .08$, $p = .93$, $\eta^2 < .001$. Story complexity did have a significant effect on MLU, $F(2, 52) = 3.44$, $p = .04$, $\eta^2 = .12$.

Post-hoc tests showed that MLU was lower in the second session compared to the first ($MD = .52, SD = .20, p = .04$) and the third session ($MD = .88, SD = .17, p = < .001$).

The ANOVA's also did not reveal any significant effects of voice condition on child-task engagement, $F(1, 26) = .58, p = .46, \eta^2 = .02$, and child-robot engagement $F(1, 26) = 1.40, p = .25, \eta^2 = .05$. Story complexity also did not show a significant effect on child-task engagement, $F(2, 52) = .11, p = .89, \eta^2 < .01$ and child-robot engagement, $F(2, 52) = 1.88, p = .18, \eta^2 = .07$. This shows that, irrespective of the robot's voice, children remained engaged with the task and the robot over consecutive sessions with increasingly complex stories (cf. Fig. 1).

The ANOVA's only revealed a significant interaction effect of voice condition and story complexity on MLU, $F(2, 52) = 4.39, p = .02, \eta^2 = .26$. Planned contrasts showed that the change in MLU from the first to the third reading session, $F(1, 26) = 6.42, p = .02, \eta^2 = .20$, and from the second to the third reading session, $F(1, 26) = 6.96, p = .01, \eta^2 = .20$ were significantly larger for the robotic voice condition than for the human voice condition.

5 Discussion

The main goal of this study was to investigate the effect of the expressiveness and complexity of a storytelling robot's speech on children's language production and engagement. None of the language production measures were affected by voice condition. Children did not tell longer stories, did not have a higher MLU and did not use more target words in the human voice condition compared to in the robotic voice condition. Children in the human voice condition scored lower on the Dutch language test than the children in the robotic voice condition. These results should thus be interpreted with care. Voice condition also did not have a significant effect on either child-robot or child-task engagement. Children in the human voice condition were not more engaged with the task or the robot compared to children in the robotic voice condition.

In the study of Conti et al. [17] they did not find a difference between a expressive and inexpressive storytelling robot either. However, in their study the expressiveness of the robot was manipulated by implementing gestures and changing the colours of the eyes and not by changing the robot's speech. In contrast, Westlund et al. [7] found that children told longer stories when interacting with an expressive robot. However, this effect was only found in the delayed session and not immediately after hearing the story [7]. Westlund et al. [7] also found that children were more likely to use target words when interacting with an expressive robot. This was not the case in this study. An explanation could be that not all target words were repeated enough. The target words that children used the most were also usually most frequently repeated in the stories the robot told. In the study of Westlund et al. [7] children also had a stronger emotional engagement with the robot and the story, when the robot had an expressive voice. However, Westlund et al. [7] used Affdex, an emotion measurement software, to measure engagement. A recent study showed that the software has an accuracy of 55% when recognizing natural dynamic facial expressions and might thus not be the most accurate tool for measuring engagement [21].

When it came to the complexity of the story, there were no significant results found on story length, target word use, child-robot and child-task engagement. The lack of an effect on engagement may come as a surprise, as in many child-robot interaction studies, children lose interest in interacting with a robot in the long term [22, 23]. Although not significant, robot and task engagement increased over time when the stories became more complex, as is shown in Figure 1. The task from this study might thus be stimulating enough to keep children engaged, especially compared to other forms of tutoring that rely more on a tablet-based interaction, such as in [4].

A significant effect of story complexity was found on MLU. Children had a lower MLU in the second reading session compared to in the first and third reading sessions. An explanation for this might be the difference in the picture books that were used. The book used in the second reading session was a wordless picture book and the pictures were a lot less detailed. The story might thus have been too difficult, compared to the other stories that were used.

Concluding, this exploratory study aimed to provide more insight into the effect of the expressiveness and complexity of a robot's speech on children's language production and engagement. Storytelling with a social robot might be a great way for children to practise oral skills in a second language and to keep children engaged in interacting with a robot in the long term. Further research is necessary to explore the effects linguistic and emotional aspects in robots on child-robot interactions.

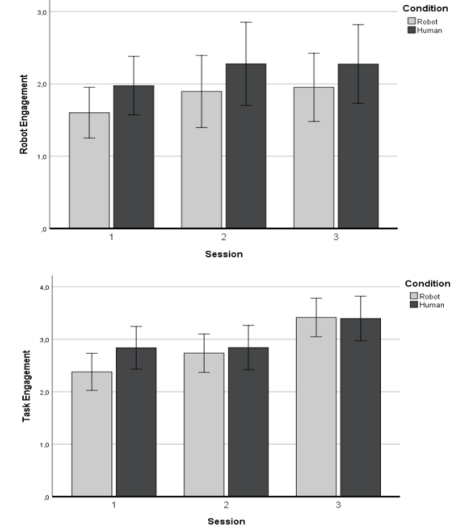


Figure 1: Mean child-robot and child-task engagement for each reading session.

References

- [1] Patricia K Kuhl, Feng-Ming Tsao, and Huei-Mei Liu. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, 100(15):9096–9101, 2003.
- [2] Patricia K Kuhl. Social mechanisms in early language acquisition: Understanding integrated brain systems supporting language. *The Oxford handbook of social neuroscience*, pages 649–667, 2011.
- [3] Junko Kanero, Vasfiye Geçkin, Cansu Oranç, Ezgi Mamus, Aylin C Küntay, and Tilbe Göksun. Social robots for early language learning: Current evidence and future directions. *Child Development Perspectives*, 2018.
- [4] Paul Vogt, Mirjam De Haas, Chiara De Jong, Peta Baxter, and Emiel Krahmer. Child-robot interactions for second language tutoring to preschool children. *Frontiers in human neuroscience*, 11:73, 2017.
- [5] James Kennedy, Paul Baxter, Emmanuel Senft, and Tony Belpaeme. Social robot tutoring for child second language learning. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pages 231–238. IEEE Press, 2016.
- [6] Tony Belpaeme, James Kennedy, Paul Baxter, Paul Vogt, Emiel EJ Krahmer, Stefan Kopp, Kirsten Bergmann, Paul Leseman, Aylin C Küntay, Tilbe Göksun, et al. L2tor-second language tutoring using social robots. In *Proceedings of the ICSR 2015 WONDER Workshop*, 2015.
- [7] Kory Westlund, M Jacqueline, Sooyeon Jeong, Hae W Park, Samuel Ronfard, Aradhana Adhikari, Paul L Harris, David DeSteno, and Cynthia L Breazeal. Flat vs. expressive storytelling: young children’s learning and retention of a social robot’s narrative. *Frontiers in human neuroscience*, 11:295, 2017.
- [8] Omar Mubin, Catherine J Stevens, Suleman Shahid, Abdullah Al Mahmud, and Jian-Jie Dong. A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, 1(209-0015):13, 2013.
- [9] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. Social robots for education: A review. *Science Robotics*, 3(21):eaat5954, 2018.
- [10] Marina Fridin. Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. *Computers & education*, 70:53–64, 2014.
- [11] Ellen Axmear, Joe Reichle, Maya Alamsaputra, Kathryn Kohnert, Kathryn Drager, and Kelli Sellnow. Synthesized speech intelligibility in sentences. *Language, Speech, and Hearing Services in Schools*, 2005.
- [12] Larry Vandergrift. Recent developments in second and foreign language listening comprehension research. *Language teaching*, 40(3):191–210, 2007.
- [13] Joe Crumpton and Cindy L Bethel. Validation of vocal prosody modifications to communicate emotion in robot speech. In *Collaboration Technologies and Systems (CTS), 2015 International Conference on*, pages 39–46. IEEE, 2015.
- [14] Michael L Walters, Dag Sverre Syrdal, Kheng Lee Koay, Kerstin Dautenhahn, and R Te Boekhorst. Human approach distances to a mechanical-looking robot with different robot voice styles. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pages 707–712. IEEE, 2008.
- [15] Friederike Eyssel, Laura De Ruiter, Dieta Kuchenbrandt, Simon Bobinger, and Frank Hegel. ‘if you sound like me, you must be more human’: On the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, pages 125–126. IEEE, 2012.
- [16] Myrthe Tielman, Mark Neerinx, John-Jules Meyer, and Rosemarijn Looije. Adaptive emotional expression in robot-child interaction. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 407–414. ACM, 2014.
- [17] Daniela Conti, Alessandro Di Nuovo, Carla Cirasa, and Santo Di Nuovo. A comparison of kindergarten storytelling by human and humanoid robot with different social behavior. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 97–98. ACM, 2017.
- [18] L Verhoeven, G Narain, G Extra, O. A. Konak, and R. Zerrouk. *Toets Tweetaligheid*. CITO, Arnhem, 1995.

- [19] J. Kennedy, S. Lemaignan, C. Montassier, P. Lavable, B. Irfan, F Papadopoulos, and T. Belpaeme. Child speech recognition in human-robot interaction: evaluations and recommendations. In *In Proceedings of the 2017 acm/ieee international conference on human-robot interaction*, pages 82–90. ACM/IEEE, 2017.
- [20] Ferdinand Laevers, Mieke Daems, Griet De Bruyckere, Bart Declercq, Kristien Silkens, Gerlinde Snoeck, Julia Moons, and Monique Van Kessel. Zelfevaluatie-instrument voor welbevinden en betrokkenheid van kinderen in de opvang (ziko). 2005.
- [21] Sabrina Stöckli, Michael Schulte-Mecklenbeck, Stefan Borer, and Andrea C Samson. Facial expression analysis with affdex and facet: A validation study. *Behavior research methods*, 50(4):1446–1460, 2018.
- [22] Iolanda Leite, Carlos Martinho, and Ana Paiva. Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, 5(2):291–308, 2013.
- [23] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. Interactive robots as social partners and peer tutors for children: A field trial. *Human–Computer Interaction*, 19(1-2):61–84, 2004.