Smemo: a Multi-modal Interface Promoting Children's Creation of Personal Conversational Agents

Claudia Maria Cutrupi*, Salvatore Fadda*, Giovanni Valcarenghi*, Giulia Cosentino, Fabio Catania, Micol Spitale, Franca Garzotto
Politecnico di Milano
Milano, Italy
<name>.<surname>@polimi.it,<name>.<surname>@mail.polimi.it*

Figure 1: Three children interacting with Smemo (vocal, physical, and touch mode) - Agent's height: 50cm, Material: wood

ABSTRACT

Recent improvements in natural language processing pushed the development of a wide range of products and applications that allow natural spoken interaction for children. Previous studies explored conversational agents for children and found out that they are valuable both for playing and learning. In addition, recent researches highlighted that, when we want to develop a conversational tool for children, we must consider that they have different modes from adults and they approach conversational technologies in a different way because of their linguistic skills, their way of conducting dialogues and, more generally, their interaction attitude. In this paper, we describe Smemo, a multi-modal interface for children aged 6 to 9 years old. This tool permits to approach conversational technologies in an original and multi-modal way giving the user the possibility to customize and personalize the agent. The main goal of Smemo is to promote an innovative learning method by putting children at the center of the process.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; *User centered design*; *Interaction techniques*; *Graphical user interfaces*; Sound-based input / output.

KEYWORDS

Conversational Technology, Natural Language Visualization, Children, Language Learning, Learning

CUI '20, July 22–24, 2020, Bilbao, Spain

ACM ISBN 978-1-4503-7544-3/20/07.
https://doi.org/10.1145/3405755.3406162

ACM Reference Format:
Claudia Maria Cutrupi*, Salvatore Fadda*, Giovanni Valcarenghi*, Giulia Cosentino, Fabio Catania, Micol Spitale, Franca Garzotto. 2020. Smemo: a Multi-modal Interface Promoting Children's Creation of Personal Conversational Agents. In *2nd Conference on Conversational User Interfaces (CUI '20), July 22–24, 2020, Bilbao, Spain.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3405755.3406162

1 INTRODUCTION AND RELATED WORK

In the last few years, researchers in speech and multimedia technology developed several prototypes and commercial applications that enable natural interaction using the spoken language. These tools are mainly targeted to the adult user population [2, 5, 6]

Building speech-based applications for children generates greater challenges for designers and engineers because children speak and interact differently. In fact, the main differences are in the linguistic characteristics of speech, dialog interaction strategies, problem solving skills, and user preferences[8]. Previous studies suggest that the use of multi-modal interfaces combining speech with different input modes such as visual, text or touch improve the user experience, especially when we speak about children [3, 10].

Purposes of a conversational agent for children can be many: researchers already investigated the use of voice-based interfaces for playing [1, 4] and learning [7, 11]. Both children and caregivers seem to like these innovative interfaces: Sciuto reported how children are increasingly experiencing conversational agents (CAs) and described both parents and children having a positive opinion of them [9].

Given the above premises, we present Smemo, a web application with a wooden embodiment that allows children to play with it training it with any kind of content they prefer. Smemo was designed to be used at school or at home, for play and learning. This work sheds light on the potential of conversational agents that learn from children while they play and interact with them in many different and combined ways.

2 DESIGN

We involved in the design process ten primary-school teachers. We interviewed them to understand their routine with children and their teaching approaches and techniques. They gave us some valuable advice that helped us to come up with some guidelines for the realization of a dialogue system for children:

- create a familiar shape and visual affordance, to attract and engage children.
- design a minimalist content, to prevent them to distract.
- invent a catching and interesting storytelling, to stimulate their imagination.
- define a set of customizable assets both physical and digital, to keep their attention up.
- develop a system that is versatile and adaptable to each situation to be used for different purposes.

3 SMEMO

3.1 Concept

"Smemo" is a multimodal interface composed by a conversational agent with a physical identity supported by a graphical interface that provides visual stimuli exploiting the avatar of the robot represented on the screen (Figure 1 on the left). The graphical interface is minimal and simple and dialogues use a basic language to allow children to use the system autonomously, even without the caregiver's supervision. The main function of Smemo is its ability to listen and memorize any data the children want to teach it, giving them full access to all the informations it had memorized, in every moment. Smemo's strength is that it interacts with children playing different roles at the same time: it is a companion, teacher, and a student itself. Indeed, children can teach the agent with information as a teacher would do, but at the same time they could learn as students, asking Smemo questions about its knowledge so far. At the beginning of the first session, Smemo tells its story to the users: due to a short-circuit, the agent lost its memory and it is children's job to teach it whatever they want.

3.2 Interaction

Children can receive different stimuli by Smemo and interact with it in various ways:

(1) Vocal: they can speak with Smemo using voice and it responds in a familiar and direct way, it remembers about child's name and preferences and saves the information of each game session.
(2) Physical: children can personalize the external structure of Smemo with stickers and drawings to make it more personal.
(3) Touch: little users can also communicate without using their voice by touching the buttons on the screen and writing input messages with the digital keyboard (see Figure 1). This makes the interface more usable and accessible.

The UX of Smemo is divided in two different sections: "Let's learn" and "Let's talk" (Figure 2) and Smemo's avatar guides the children in both phases. "Let's learn" section gives the child the possibility to teach his/her content to Smemo, by following three simple steps:

Figure 2: Smemo graphical user interface representing the "Lest's learn" and "Let's talk" sections.

- The Choice of a topic (e.g."mathematics")
- The Insertion of the question (e.g."what is 2 times 5")
- The Insertion of the answer (e.g."10")

Given the fact that teachers suggested us to divide topics in categories to let children find their contents, we insert a feature named "cards". Every topics inserted in Smemo during the "Let's learn" section will be saved in a card, helping the child to retrace his/her steps and allowing him/her to change over time his/her work. Selecting the "Let's talk" section, he/she may ask for the information he/her taught previously. To make the user experience even more engaging, we decided to provide to the user the possibility to customize the avatar color and the background theme.

3.3 Technology

Inside smemo's head, there is a tablet that runs a web app, that allows both vocal and visual interaction with the user. Dialogues are managed with Dialogflow by Google. This software module can handle unstructured inputs that are governed by poorly defined and flexible rules and can convert them into a structured form that a machine can understand and act upon. In addition, Dialogflow can extract parametric data from the input text by the user and permits to the software engine to use this data for future processing. We connected Dialogflow to a proprietary NodeJS webhook that processes the recognized intention by the user and produces a consistent output. In fact, every time the user wants to teach something to the agent (i.e. to link a specific input from the user to a specific system output), the webhook uses the Dialogflow API to pull Smemo to recognize the new input and respond with the new output.

4 CONCLUSION AND FUTURE WORK

In this work we have highlighted the possibility of creating a conversational agent for children that allows them to customize and train it as they prefer to have the opportunity to learn from it. We have just started a first exploratory study that will last a year, with primary-school children from 6 to 9 years old. Our next steps are to evaluate Smemo's usability and likeability. Moreover, we want to analyze its learning effectiveness, the way in which children interact with it and how much they get engaged. Since Smemo has various and promising applications, we would expand our future work by creating a multi-player and multi-language platform that could add value to our contribution.

REFERENCES

[1] Samer Al Moubayed and Jill Lehman. 2015. Toward better understanding of engagement in multiparty spoken interaction with children. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 211–218.

[2] Hyunji Chung, Michaela Iorga, Jeffrey Voas, and Sangjin Lee. 2017. Alexa, can I trust you? *Computer* 50, 9 (2017), 100–104.

[3] Phil R Cohen, Michael Johnston, David McGee, Sharon L Oviatt, Joshua Clow, and Ira Smith. 1998. The efficiency of multimodal interaction: A case study. In *Fifth International Conference on Spoken Language Processing*.

[4] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. Hey Google is it OK if I eat you?: Initial explorations in child-agent interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children*. ACM, 595–600.

[5] Matthew B Hoy. 2018. Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Medical reference services quarterly* 37, 1 (2018), 81–88.

[6] Gustavo López, Luis Quesada, and Luis A Guerrero. 2017. Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 241–250.

[7] Koichi Mori, Rafael Ballagas, Glenda Revelle, Hayes Raffle, Hiroshi Horii, and Mirjana Spasojevic. 2011. Interactive rich reading: enhanced book reading experience with a conversational agent. In *Proceedings of the 19th ACM international conference on Multimedia*. 825–826.

[8] Shrikanth Narayanan and Alexandros Potamianos. 2002. Creating conversational interfaces for children. *IEEE Transactions on Speech and Audio Processing* 10, 2 (2002), 65–78.

[9] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. 2018. Hey Alexa, What's Up?: A mixed-methods studies of in-home conversational agent usage. In *Proceedings of the 2018 Designing Interactive Systems Conference*. ACM, 857–868.

[10] Toshiyuki Takezawa and Tsuyoshi Morimoto. 1998. A multimodal-input multimedia-output guidance system: MMGS. In *Fifth International Conference on Spoken Language Processing*.

[11] Elizabeth K Morales Urrutia, José Miguel Ocaña, Diana Pérez-Marín, and Silvia Tamayo. 2017. A first proposal of Pedagogic Conversational Agents to develop Computational Thinking in children. In *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality*. ACM, 2.