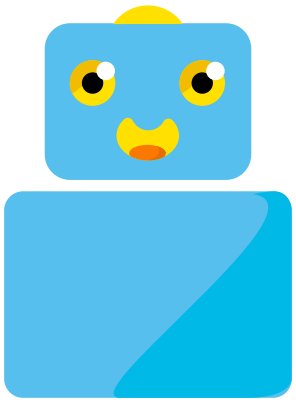
Smemo: A Multimodal Interface Promoting Children Creation Of Personal Conversional Agent



ABSTRACT

Popular Conversational Agents (CAs) are usually designed for adults, and for this reason they are focused on their needs and requests, but changing the target from the latter to children implies a change in the needs and the way they interact with them.

Smemo is a Multimodal interface for children from 6 to 9 years old that allows them to approach conversational technologies in an original way giving them the possibility to create their personal CA. Smemo allows users to approach vocal systems in an original way, creating and training it with any kind of content they prefer. This Intelligent Conversational Agent is also equipped with a graphical interface which allows children to have both a vocal and written/graphical response. As children typically concentrate for shorter periods of time, this interface is properly studied to be the most effective in enabling children to keep their interest level high to complete the activity.

Advanced User Interfaces
2019 - 2020

Claudia Maria Cutrupi
Department of Design, Milan Politecnico
claudiamaria.cutrupi@mail.polimi.it

Salvatore Fadda
Department of Electronics, Information and Bioengineering, Milan Politecnico
salvatore.fadda@mail.polimi.it

Giovanni Valcarenghi
Department of Electronics, Information and Bioengineering, Milan Politecnico
giovanni.valcarenghi@mail.polimi.it

Table of contents

Introduction

Our project could be thought as a new and innovative platform for learning. The child can teach what he has learnt and repeat it whenever he want.

It is also a new way to approach children to Conversational Agent that are more and more in our daily life. All of these in a completely safe and controlled platform because the only content that the child can see are the ones he has inserted.

All of these key points are enclosed in a very easy and user friendly extension for Google Assistant that could be accessible from all the devices (up to now only Android device).

1. The problem

1.1 Main Target Group

Children between the age of 6 and 9 would be the target group. They have been chosen because of the fundamental requirement to speak clearly and read at a certain level.

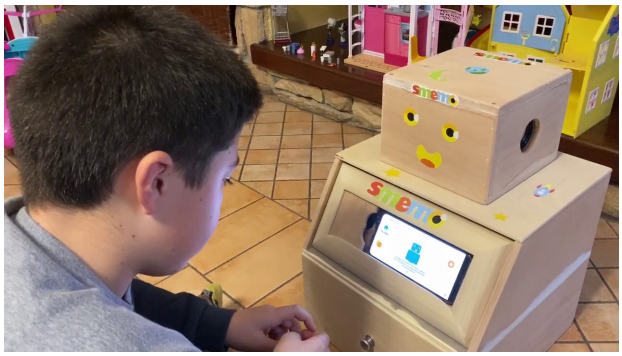
- Context
Children at home (or other places) that want to study in a different way or just have some fun with his Conversational Agent Smemo.

- Needs
One of the main needs is the curiosity that children have about Conversational Agent technology. We have studied a specific CA that can enhance children general knowledge on this technology.
Another important needs could be the desire of children to find different way to study (in a more involving way). Lastly is to spend some time playing and having fun.

- Constraints:
The only constraint is the capability of the children to read and speak properly.

- Goals of your project:

- To help children become involved in the Conversational Agent technology.
- To have fun while learning.
- To introduce children to the creative and logical process of Conversational Agents.
- To improve self-esteem and creativity.
- To assist in reviewing lessons previously learned.



1.2 State of the art

Recent research in conversational agent (CA) technologies have pushed the production of numerous prototypes and commercial applications around the world that provide natural spoken dialogue with machines. Usually the main target is the adult user population, but when we refer to conversational application for children it leads to a greater challenge. Children have different habits from adults and they approach CA technologies in a different way because of their linguistic features of speech, type of dialogues and needs. For this reason lot of researchers are interested in multimodal interfaces that combine speech with different input modalities such as visual, text or touch. As P. Cohen and colleagues suggest multiple modalities promote a more effective and natural interaction and improve the user experience. Nowadays commercial and non-commercial CAs for children are considered valuable for play and learning. For instance Sciuto described how children are increasingly experiencing CAs and they reported both parents and children having a positive opinion.

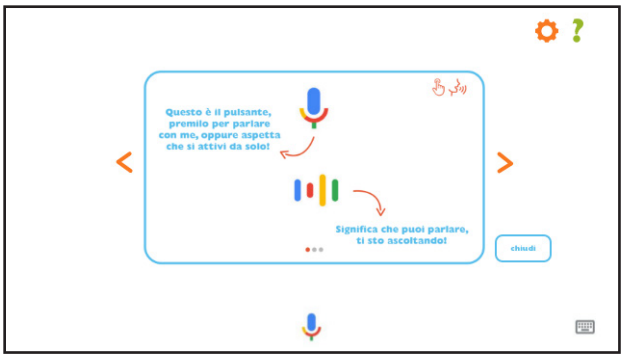
Given the above premises and stated that researches in CA for children is still in its infancy, this work sheds a light on the potential of a Multimodal interface that include vocal, and tangible interaction and allows children to create their own CA.

2. The solution

2.1 General approach

In order to better understand the characteristics to develop an effective and suitable multimodal interface for children, we have determined the guidelines for the realization of our project:

• A minimalist content, to avoid children distraction.
• A familiar shape and visual affordance.
• The creation of a catching and interesting history behind the product.
• A set of customizable assets both physical and digital.
• The system should be versatile and adaptable to each situation.
• The system must collect automatically any type of data given by the children.



2.2 Details of interaction and interfaces

We came up with a final idea called "Smemo", a multimodal interface composed by a CA supported by a graphical interface that provide visual stimuli useful to enforce the communication and maintain children's attention. To let children be more familiar with it, Smemo has also a real, and physical identity. He is a character with a story: after a short circuit it has lost all its memory and becomes child's job to teach to Smemo different and infinite information, of her/his choice. In this way the child enriches Smemo with personal information as a teacher would, but at the same time he could learn as a student asking Smemo question about what it learned. The interface provides a very simple graphic and language in order to allow children to approach independently, without the supervision or explanation by an adult. Smemo's strong point is therefore to interact with the child in different functions: as a friend, as a teacher, as a toy but also as a student himself.
Children could have different stimuli and could interact with Smemo in various ways:

1. Vocal: children can interact with the system using voice. Smemo responds in a familiar and direct way, recording child's name and choices and keeping the information of each game session.

2. Physical: the child can personalize the external part of Smemo with stickers and drawings to make it more personal.

3. Tangible: the child can interact with the interface not only with voice but also touching the screen. This makes the interface more usable and accessible.

2.3 Scenarios

The interface is divided mainly in two different session: "let's learn" and "let's talk". The first one gives the child the possibility to create his content, by following three simple steps:

- Choice of the topic (mathematics, animals, jokes etc ..)
- Insertion of the question
- Insertion of the answer

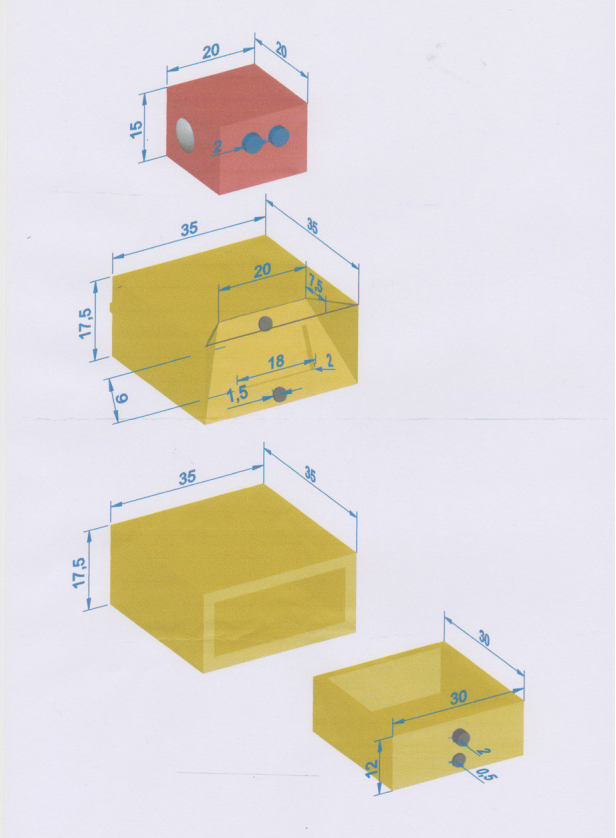
Once these steps have been completed, the child will be asked if he wants to continue teaching in the same topic enriching it with more information, otherwise he may end the session. Afterwards he/her could ask for the information that has taught, this time selecting the "let's talk" session.

Given the fact that during the codesign session teachers suggested us to let children systematize contents in categories, at the end of every game session, we insert an important feature that easily help the child to organize his data: "Cards". Every topics and interactions inserted in Smemo during the "let's learn" session will be saved and categorized into "cards". It could help the child with a logical structure to retrace his steps and to review progressions over time of his work. Another important aspect concerns

the settings' personalization which is available and editable in every moment, where the child could change text color, background theme and his own name.

2.4 The product

The physical structure was realized from technical drawing in AutoCAD, in order to facilitate the work we designed it in three different pieces, for a total height of 50cm. The final product is a manpower work made through the use of wood and plywood, and then it was painted with chalkboard paint.
The first component corresponds to the Smemo head, in which we decided to insert two speakers that have been connected to the smartphone, in the part below. The latter is a trapeze-shaped box composed in a frame in which the mobile phone is inserted. Also, inside this second box was insert the battery that power up the system. The last part is a drawer designed to allow children to keep their personal tools for playing and interacting with Smemo.



3. Implementation

- SW architecture:
The prototype of Smemo has been implemented as a web application to permit both vocal and visual interaction with the user, for example, through the screen, microphone and speakers of the running device. The agent is integrated into Google Assistant; consequently, it is accessible through Google Home, Android phones and other technologies. Once you have opened the Google assistant application you have just to say "talk to Smemo agent" in order to start your experience with Smemo. Dialogues are managed with Dialogflow, that is the Natural Language Understanding service 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. Besides, Dialogflow is integrated to a proprietary NodeJS webhook that allowed us to customize the behavior of the agent according to each recognized intention by the user. Once the user inserts the parameter for the intent creation thanks to the NodeJS webhook starts a method named "createIntent(request)" taken from the Google's Library "dialogflow. IntentsClient()" that is going to add the intent to the agent. The request that is sent as the parameter of the method contains an intent object where are stored the parameters that were taken before. The graphical interface is developed using Interactive Canvas. Finally, in order to host our web application and to store some real-time data that are used by the agent we have used respectively Firebase Hosting and Firebase Database.

- HW architecture:
To what concern the Hardware part there is not so much to say, the whole structure of the robot is entirely made of wood. It present two speaker embedded in the head of the robot and a device (placed on the belly) that contain the software.

4. Empirical evaluation

The tests were carried out on six children aged five to ten years old, one of whom with language disorders and one with learning disabilities, during the Christmas period in a familiar environment, which represented the primary scenario of product usage. Two different scenarios were assessed, initially without physical support and without tutorials and subsequently with physical support but always without tutorials.
We collected data through the analysis of children's behaviours while using the application, first without aid and then with aid.
The collected data showed us that the absence of the tutorial made using the application difficult to understand, especially for younger children. As far as the actual use of the application is concerned, no child had problems, in particular the child with learning disabilities proved to be the one with the most interest in the application and the child with speech disorders also had no problems too, when he was not understood, he was supported by the ability to click the buttons to go ahead. There was a substantial difference in children's interest in the application when it was used with the physical structure compared to when it was used simply by mobile phone. The children were much more interested in the first case, in fact they were fascinated and managed to give Smemo a real identity when the physical structure was used. We therefore decided to keep the teaching and request method as initially thought, similarly we kept the display of the topics taught and the customization of the graphic application unchanged. We therefore decided to add an introductory explanation part and tutorials, accessible at any time. At the end, in order to make the system more interactive and customizable we decided to cover the wood with blackboard paint.

5. Value proposition

The challenges we initially set are the following:
To help children learn in a complementary way to traditional school education, through an auto learning mechanism based on the use of a conversational agent.
To approach new applications that use a mix between classic graphic application and conversational application which in our case is based on Google technologies.
Creating a link between education and toys, which are seen within society as two completely different and very often contrasting fields.

Among the main difficulties encountered there was the connection of the graphic part to the conversational agent mechanism.
The technologies used by our application allow us to satisfy the needs of our customer segment and not only:
We allow students to learn through a teaching and request mechanism of the topics taught which allows the boy to teach SMEMO anything he wants and at the same time also allows to request what has been taught interactively.
We have developed a graphic application that captures the child's attention and can be personalized graphically, in such a way as to make it in his eyes unique, similarly the physical support can also be personalized by hand and made unique for the child who owns it.
On the hardware level, however, the application is memory-free, in fact it is contained in the cloud within the Google servers, similarly also the database to which it interfaces for data storage. It therefore allows it to be used by any device that contains the google assistant application and can be called and used directly from there.
Our competitors are few and use different approaches, in fact we are among the first to use this type of learning approach through the conversational agent's technologies.

6. Future work

We are already starting an exploratory study to understand children's reactions to the interface and received direct feedbacks from them. From a very first observation we notice that children are really engage during the session and they interact with the system in a natural way. We expected to experiment the entire system, with elementary schools' children from 6 to 9 years old, thanks to which we could evaluate the learning purposes and will give us complete feedback about usability and likeability.

Another future work that we have thought about, and in which we have given space within the design of the physical structure is the implementation of an interactive LED system managed through Arduino technology. Arduino will be programmed according to the voice commands that receives, showing different animations of the LEDs depending on the state of SMEMO. The positions to which we have left space are above the head and in the eyes, it would also be possible to implement a mechanical system of arms, always managed through Arduino via voice control that would allow the child to interact in a more real way with the physical structure.

The goal is to create a more vivid image of the robot, trying to make the smartphone installed as the "heart" of the physical structure and once inserted it activates and starts living.

Regarding the digital application we will add the multi-language feature and the multi-account mode.

7. Bibliography

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.

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.

Allen L Gorin, Giuseppe Riccardi, and Jeremy H Wright. 1997. How may I help you? Speech
communication 23, 1-2 (1997), 113–127.

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.

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

J Wiggins, B Mott, L Pezzullo, E Wiebe, K Boyer, and J Lester. 2017. Conversational UX design for kids: toward learning companions. In Proceedings of the Conversational UX Design CHI 2017 Workshop.

Ying Xu and Mark Warschauer. 2019. Young Children's Reading and Learning with Conversational Agents. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.
ACM, CS10.