Conversational Agents to Promote Children's Verbal Communication Skills

Fabio Catania[1(✉)], Micol Spitale[1,2], Giulia Cosentino[1], and Franca Garzotto[1]

[1] Politecnico di Milano, Milano, Italy
fabio.catania@polimi.it
[2] IBM Italy, Milan, Italy

Abstract. The fundamentals of verbal communication skills are developed during childhood, and existing studies pinpoint the benefits of stimulating language and expression skills from an early age. Our research is a preliminary evaluation of conversational technology to support this process. In this paper, we describe the design process of a speech-based conversational agent for children, which involved a Wizard-of-Oz empirical study with 20 primary school children aged 9–10 y.o. in order to identify the design guidelines for the automated version of the system. Our agent is called ISI, is integrated into a web application and exploits oral and visual interaction modes. ISI enables children to practice verbal skills related to the description of a person's physical characteristics. It provides opportunities for them to learn and use words and linguistic constructs. Also, ISI permits to develop their body awareness and self-expression (when describing their self) or the attention to "the other" (when describing someone else). ISI engages users in a speech-based conversational flow composed of two main repeated steps. It talks to the children and stimulates them with questions about a specific part of their body (e.g., "What color is your hair?"). When the users describe the required feature adequately, ISI provides a cheerful real-time visual representation of the answer; otherwise, it provides hints.

Keywords: Conversational technology · Natural language visualization · Children · Language learning · Learning

1 Introduction

Verbal communication is the foundation of relationships and is essential for learning, playing, and social interacting [17]. Early oral communication skills are developed during childhood [48]. Children learn how to convey information, needs, and feelings in a more effective way [13] by acquiring words and linguistic constructs of the language.

Previous studies proved that individuals differ in how they learn [53] and that various stimuli can support children in enhancing verbal communication capabilities and learning a language [10]. The VAK model [4] identifies three learning modalities:

A. Følstad et al. (Eds.): CONVERSATIONS 2020, LNCS 12604, pp. 158–172, 2021.
https://doi.org/10.1007/978-3-030-68288-0_11

- Visual learning, that exploits graphs, charts, maps, diagrams, pictures, paintings, and other kinds of visual stimulation;
- Auditory learning, that depends on listening and speaking;
- and Kinesthetic learning, that requires gestures, body movements, and object manipulation to process new information.

The strengths of each learning modality show up independently or in combination [3]. Generally, learners appear to benefit most from visual and mixed modality presentations, for instance, using both auditory and visual techniques. A review study [1] concluded that visual stimulation improves learners performance in the following areas:

- Retention. Learners remember and recall information better when it is represented and learned both visually and verbally;
- Reading comprehension. The use of visual stimulation helps to improve the reading comprehension of learners;
- Learning achievement. Learners with and without learning disabilities improve achievement across content areas and grade levels;
- Critical thinking. When learners use visual stimuli during learning, their critical thinking skills are enhanced.

In addition, according to [38], visual aids are used for various aspects in the teaching-learning process, and practicing teachers are often led to believe that "the more visuals, the better".

Our research concerns Computer-Aided Language Learning [47], i.e., language learning and communication skills training with the help of a machine. Our final goal is to investigate the use of an intelligent interface combining visual and auditory stimuli for children to improve their verbal communication skills.

We present ISI, an Italian speech-based conversational agent (CA) for children that exploits oral and visual communication modes. A conversational agent is a dialogue system able to interact with a human through natural language [14]. ISI enables children to practice verbal skills related to the description of a person's physical characteristics. In this way, ISI offers opportunities for children to learn and practice with words and linguistic constructs. Also, it permits to develop their body awareness and self-expression (when describing their self) or the attention to *the other* (when describing someone else).

The name ISI stands for "Io Sono Io", that is the Italian version of "I am me". This name takes inspiration from the German book "Das kleine Ich bin ich" [33] that supports children to answer the question "Who am I?" Furthermore, in Italian, ISI is pronounced in the same way as the English word "easy", that perfectly fits with the principles underlying the system:

- it is simple to use for children;
- it facilitates and trains communication skills and self-knowledge.

This paper not only contributes in exploring conversational technology merged with Natural Language Visualization for support learning, but also provides interesting design insights for conversational technology for children.

Indeed, in this paper, we describe the process of designing ISI: we implemented a Wizard-of-Oz version of the agent, and we conducted a preliminary empirical study with 20 primary school children aged 9–10 y.o. to identify effective design guidelines for the application.

Assessing the potential of ISI as a teaching tool is beyond the scope of this paper and will be analyzed in the future by exploiting an automated version of ISI. Here, we addressed the following research questions:

- "Is a system with the characteristics of ISI usable by children?"
- "Is a system like ISI engaging?"

2 Related Works

The rapid and continuous improvements in the field of Artificial intelligence are making spoken and written Conversational Technologies smarter, leading to new forms of collaboration with humans and new application areas [19]. Voice-based conversational agents (e.g., Apple's Siri, Amazon's Alexa, Google Assistant) are progressively getting more embedded into people's lives due to their intuitive, easy-to-use natural language human-computer interface: according to [24], 46% of United States adults use them in their daily routines.

From literature, we know some conversational agents specific for teaching and supporting the learning process. They are called Pedagogic Conversational Agent (PCA) and can be defined as smart systems that interact with the student in natural language, assuming the role of an instructor, a motivator, a student, or a companion. They are cheaper than human tutors and can exploit adaptive learning technology in order to meet the needs of each student [15].

There are PCAs for different targets – e.g., for children and adults – and for various topics that range from math to literature [27]. For children, one of the most common uses of PCAs is language teaching and practice. For example, [34] describes embodied agents offering language challenges to children. Also, Baldi is a tutor who guides students through a variety of exercises designed to teach vocabulary and grammar, to improve speech articulation, and to develop linguistic and phonological awareness [35]. In [11], an animated CA, named Marni, interacts with children to teach them to read and learn from the text. In [51], it is proposed the use of Pedagogic Conversational Agents to develop computational thinking in children. Hayashi [26] proposed multiple PCAs to support collaborative learning in children, and highlighted how multiple PCAs can implement roles yielding different types of suggestions. In [25], the authors implemented an intelligent virtual environment with many Embodied Conversational Agents, for improving speaking and listening skills of non-native English language learners. Finally, [50] reports about CAs to engage children in book reading activities and to create oral stories in a highly interactive manner.

Concerning design, there are several sets of established guidelines for Graphical User Interfaces (GUIs) (e.g., [40]) and Tangible User Interfaces (e.g., [54,55]) for children, but the ones for Conversational User Interfaces are few and not

universally accepted. Indeed, previous studies have already explored commercial and non-commercial CAs for children both for play [2,16,43] and learning [35,52], but without finding specific guidelines. Besides Nielsen, [41] who defined ten heuristics to test the usability of any interfaces, Moore and Arar [36] made the first attempt to suggest some guidelines for designing a conversational interaction experience for generic users and contexts. Also, Murad et al. [39] summarized some design guidelines to support researchers in solving issues related to usability and learning of hands-free speech interfaces.

To the best of our knowledge, so far, there is no conversational application teaching or enabling the practice of self-description and body parts learning for children with a Natural Language Visualization support system.

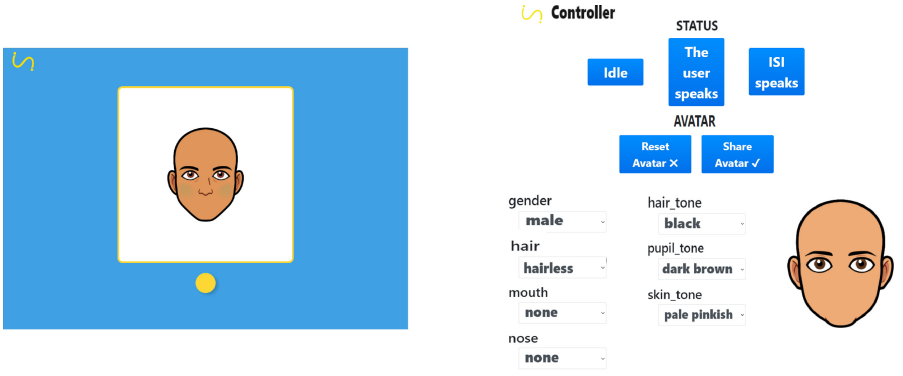
3 ISI

ISI is a tool for teachers and caregivers to make primary-school children practice with turn-taking and language constructs, and words related to the description of a person's physical characteristics, colors, and positions. Also, ISI allows children to develop their body awareness and self-expression (when describing themself) or the attention to *the other* (when describing someone else). It can be used autonomously, both in primary school and at home. Its design is grounded on the VAK model [4] for which the combination of visual and auditory stimuli improves the children learning.

ISI is a goal-oriented, domain restricted and Italian speaking conversational agent. It is integrated into a web application and enables both vocal and visual interaction with the user through the screen, the microphone, and the speakers of the device (both standalone and mobile). ISI engages children in a speech-based conversational flow composed of two main repeated steps. First, it talks and stimulates them with questions about a specific part of their or someone else's body. Questions can be about color (e.g., "What color is your hair?"), size (e.g., "How big is your nose?"), and position (e.g., "Where's the mole?"). Second, if users describe the required feature adequately, ISI shows a cheerful real-time visual representation of the answer; otherwise, it provides hints and feedback. The strength of ISI lies in exploiting an original Natural Language Visualization software module to associate expressions describing a person's physical appearance with an avatar produced in real-time using Bitmoji's API [49]. ISI provides real-time feedback, and support children through a dual visual and auditory stimulation, as defined in the VAK model [4].

ISI's GUI (see Fig. 1) is very basic since the screen represents just a support channel compared to the speech that is the primary interaction channel. The app shows a box in the middle of the screen where it visualizes in real-time the avatar as the user described it so far. During the experience, ISI provides the user of visual feedback about its status (idle - Fig. 1 -, listening, or speaking) to help her/him to handle the interaction and to understand the system better. Also, the system displays a digital button to be clicked by the user before and after speaking, respectively, to trigger the system and to let it stop listening

(i.e., to express the concepts "It is my turn" and "It is your turn"). The button was designed to be visible and intuitive to touch, using a color (yellow) in contrast with the background (light blue). Pushing the button implies a tangible interaction: its use is typical in GUI and requires the user to be able to point beyond pressing. Today's most popular CAs are triggered when they perceive a keyword or a short utterance spoken by the user; this phrase is universally known as a *wake word* (e.g., "OK Google" for Google Assistant, "Alexa" for Alexa, and "Hey Siri" for Siri). CAs stop listening when they recognize a pause that marks the end of the person's speech. This allows people to use CAs even when their hands and gaze are busy. We opted for pressing the button as *wake* and *sleep* action because we hypothesized that this method could promote the sense of agency and increase children's subjective awareness of being in control of the interaction [8]. Besides, this approach was already used for children (just as wake action) [6]. The motivations for having the same commands to wake up and put to sleep the system are the following: we have the vision that future conversational technologies will become more and more accessible and will be widely used even by children with special communication needs, and according to the theory of *partner-perceived communication* [12,30], the predictability and repetitiveness of the sequences makes it possible to better give meaning to them even for those children with complex communication needs.



(a) The GUI of ISI (b) The GUI of the Wizard

Fig. 1. On the left side the GUI of ISI waiting for the child to press the button (a); On the right side the GUI of the Wizard (b) (Texts are translated into English as a matter of paper readability) (Color figure online)

4 Wizard of Oz Experimentation

To study on the field the usability of ISI by children and their engagement while interacting with the agent, we conducted a Wizard of Oz experiment. Indeed,

the participants interacted with a prototype of ISI that they believed to be autonomous, but that actually spoke and performed thanks to an unseen human being in a second room (i.e., the Wizard).

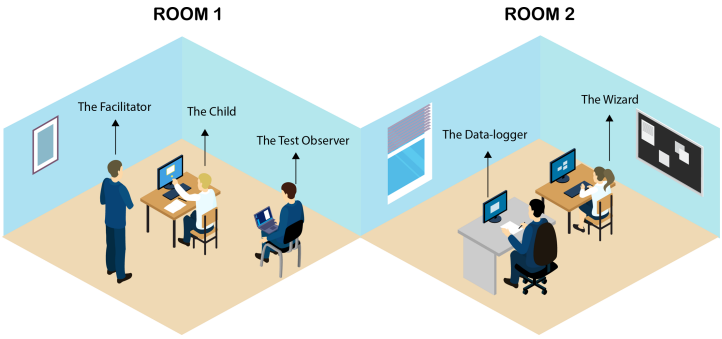


Fig. 2. Wizard of Oz experiment evaluation setting

4.1 Participants

We recruited 20 Italian speaking children aged 9 to 10 (11 girls and 9 boys) from the primary school of Cornaredo (MI, Italy). We have chosen this sample since they had already studied in school how to describe themselves, and consequently, we supposed they had the linguistic capabilities to accomplish in autonomy the self-description task interacting with the agent. From a survey by the teachers, we know that subjects were familiar with technology and computers through computer-classes at school. None of them said she/he had interacted with any conversational agents (e.g., Google, Siri, Alexa, and Cortana) previously. We are aware that all over the world many children deal with conversational technologies in their every-day life since conversational interfaces are spreading more and more in the people's homes [20]. However, this trend is slower in Italy (13% of penetration compared to other places - for example, to 24% of United States), and therefore our sample is closely representing the children of our country [9].

All participants provided us an informed consent signed by their parents.

4.2 Setting

The study was conducted during a summer campus in a primary school. This location was chosen to let children feel as comfortable as possible in a familiar space to them. The setting of the experiment was split into two separate rooms (see Fig. 2):

- Room 1. There was a touch-screen PC on a desk and a chair in front of it;
- Room 2. There was the research team, who could listen to the participant in the Room 1, speak to her/him, and observe her/his gestures through a connected laptop.

In addition to children, people involved in the study were:

- *The Facilitator*: A teacher who managed the experiment at the forefront in Room 1, introduced participants to the experience and helped them in case of requests for assistance (e.g., when they did not understand the task to complete);
- *The Test Observer*: A UX designer, who silently observed the experiment from the first room identifying problems and concerns during the interaction (e.g., in the use of the digital button) and taking qualitative notes;
- *The Data-logger*: A person collecting quantitative data (e.g., the number of errors and help requests of participants) from the second room;
- *The Wizard*: A trained person, who spoke with the children following a rigid script and simulating being a computer. This means that, for example, he did not respond when the user did not comply with the digital button interaction protocol (i.e., pressing the button before and after speaking). The Wizard controlled the system by using a web interface connected via the WebSocket API [37]. From the page shown in Fig. 1 the Wizard could remotely change the app status (idle, speaking, or listening) and the physical characteristics of the avatar displayed on the user's web page. The Wizard's voice was altered by using the Web Audio API to sounds more like a robot.

4.3 Procedure

Participants spontaneously showed up one at a time to play with ISI as single players. The Facilitator welcomed them in the classroom and invited them to sit down in front of the laptop. Afterward, she explained the participants of the coming experience and provided them with instructions about the wake and sleep action to use to trigger the system (i.e., pushing the digital button on the screen before and after speaking). When instructions were clear, the session started. The agent welcomed the children by saying, "Hello! My name is ISI. Today, we are going to create your avatar. Let's play together!" and asked her/him eight standardized questions: "What's your name?", "Are you a boy or a girl?", "What color is your skin?", "What color are your eyes?", "How big is your nose?", "How long is your hair?", "What color is your hair?", and "How big is your mouth?". Children described themselves, and every time they responded to a question, they saw the visual representation of the feature described directly on the avatar in the GUI. At every step, ISI prompted them, "Do you like your avatar as it is now?". If the answer was negative, ISI removed the last feature added and repeated the last question (allowing the children to change their avatar aspect).

During the whole experiment, the Test Observer and the Data-logger took note. If the participant did not understand what ISI said, she/he could directly

ask the system to repeat any time, or she/he could reach out to the Facilitator. At the end of the session, ISI showed the final avatar to the participant and thanked her/him for playing. The Facilitator asked the child to fill out a smile-o-meter with 5 levels about the likeability of the game. We opted for the smile-o-meter since it was proved to be a valid toolkit to measure children's engagement [46].

4.4 Data Collected

During the empirical study, the Data-logger took notes about the number of occurrences that each participant asked for help to the Facilitator about the interaction modality, and the number of errors made by each participant during the session. In the context of this study, an error is when children did not respect the interaction protocol, i.e., when they spoke before or without pressing the button, or when they pressed the button but did not speak. We calculated the variable n_{help} as the number of user's requests for help about the interaction mode during the session. n_{error} is defined as the number of errors committed by the participant while interacting with ISI divided by the number of dialogue exchanges. The Facilitator collected data about the likeability by the user with a 5 levels smile-o-meter. *liker* is the child's evaluation score voted in the questionnaire (from 1 to 5) at the end of the interaction.

Besides, we automatically stored the timestamps of some relevant events for each conversational step (i.e., when the user started/stopped speaking, when the user pushed the button, and when ISI started/stopped speaking). From the timestamps, we measured the following variables:

- the interval of time during which the child was speaking ($t_{durationspeak}$),
- the time difference between the end of ISI's speech and the user pressing the button on the screen ($t_{interact}$),
- the time difference between the child pressed the button, and she/he started to speak ($t_{startspeak}$),
- and the time elapsed between the last word spoken by ISI and the first word spoken by the child ($t_{turntaking} = t_{startspeak} + t_{interact}$).

In addition to quantitative data, during the empirical study, we also collected some qualitative data. The Observer looked at the child interacting with ISI and wrote down observations about comments aloud by participants and the Facilitator, questions from subjects to the Facilitator, requests for help to complete the self-description task, reactions, gestures [5], and behavior of participants during the experience, usability issues (e.g., difficulties interacting with the application due to the use of the button or to participant's pronunciation defects), breaking points in the conversation.

5 Results and Discussion

5.1 Results

In Table 1, there is a recap of the measured variables computed on the collected data. We are conscious that the homogeneity of our population in terms of age

(they were all 9 to 10 years old) leads to significant and reliable considerations about the target, but we are also aware that data analysis is less generalizable. We start reporting results from the analysis of the timing of turn-taking between children and ISI. From literature, we know that adults take split second between conversational turns (on average, between 0–200 ms). When it comes to having dialogues with young children, the turn-taking slows down [28]. Different children need different amounts of time to take a turn, but on average, they need 5–10 s [21]. We compared the timing data of our sample with the general population.

A single sample t-test was conducted to determine if a statistical significance exists between the time elapsed before speaking of our sample and the benchmark population in human-human conversations ($\mu_0 = 7.5$ s) [21]. Our null hypothesis is that the sample mean is equal to the population mean. Children interacting with ISI took much more time ($M = 5.09$ s, $SD = 3.90$ s) for starting the interaction (i.e., pushing the button) compared to the conventional children population, $t(19) = -2.78$, $p = .00$; on the other hand, they took much similar amount of time to start speaking ($M = 6.71$ s, $SD = 3.94$ s) compared to our benchmark value, $t(19) = -0.89$, $p = .18$.

Another single sample t-test was conducted to determine if a statistical significance exists between the duration of the child's turn of the empirical sample and the benchmark in human-human interactions ($\mu_0 = 2$ s) [21]. Children speaking with ISI takes less time to end their turn ($M = 1.6$ s, $SD = 0.4$ s) compared to the conventional children population, $t(19) = -4.65$, $p = .00$.

On average, children did not make many errors ($M = 0.15$, $SD = 0.37$), and they barely asked for help about the interaction mode ($M = 0.45$, $SD = 0.76$) during the session with ISI. Besides, they enjoyed playing with ISI as the questionnaire results revealed ($M = 4.2$, $SD = 0.83$).

The observations by the Observer are used as a starting point for identifying ISI's design insights described in the next section.

Table 1. Wizard of Oz experimental study: the variable obtained from the quantitative data about the children sample

Quantitative variable	Average (M)	Standard deviation (SD)
$t_{startspeak[s]}$	5.09	3.88
$t_{interact[s]}$	1.62	0.79
$t_{turntaking[s]}$	6.71	3.94
$t_{durationspeak[s]}$	1.60	0.62
n_{help}	0.45	0.76
n_{error}	0.15	0.37
$liker$	4.2	0.83

5.2 Discussion: Lessons Learned

From the results of the Wizard-of-Oz study, we elicited a set of lessons learned that could be applied to enhance the usability and engagement of ISI, and that may be useful for any conversational agent for children's learning.

The Ease of ISI's Multimodal Interaction – The time spent by children to start speaking with ISI during each conversational step is consistent with the literature's corresponding value in human-human interaction. This result suggests that the mixed interaction paradigm (speech, tangible, and visual) can be considered user-friendly and straightforward for this target group and could be applied to conversational agents for children in general. This result is also in line with what is described in the guidelines of Conversation design by Google [23], which claims that conversational interfaces are intrinsically multi-modal. Intuitiveness and ease of interaction are also supported by the low number of errors committed by children and the low number of times they asked for help about the interaction modality. As future work, pushing the button on the screen could be compared with other wake actions. A recent study [8] compared different actions to find out how effective they are as wake and sleep ones for children who want to interact with a conversational agent. Their results suggest that the physical button is the most appropriate solution for this target group, which opens new directions in the design of interaction affordances of CAs for children.

ISI as a Game with a Purpose – The smile-o-meter's results showed that children liked to play with ISI. Findings are in line with the positive feeling of the *Facilitator* and the *Test Observer* about the ISI-child conversational experience. We conclude that ISI has the potential to interact with children. Conversational agents need to be explored more for this population for different goals (e.g., learning, engagement, assistance). As a limitation of the study, we are aware that children tend to provide positive ratings to evaluation scales like smile-o-meter [46]. However, smile-o-meter is still an adequate tool for an easy and attractive method of scoring an opinion, especially with older children [45]. Future studies with ISI will exploit more general survey methods [45]. Also, the Again – Again table and the Fun Sorter will be used to rank specific features of the agent [45].

ISI's effectiveness for training their linguistic and communication skills is not verified, yet, and will be explored in the next study. We hypothesize that, if translated into various languages, ISI could help children to learn new languages interactively. Also, in a school environment, where socializing among peers is significant, ISI could enable the interaction with the classmates while playing in multiplayer mode.

ISI as a Self-explorative Tool for Elementary School Children – During the experimentation, the Observer took note of various requests for help by children to the *Facilitator* regarding their physical appearance in order to answer to ISI's questions. We drew two conclusions on this datum. On the one hand, the app is well-tailored around the target user, providing children an occasion to practice their self-description skills and to have a moment of reflection on their physical appearance. On the other hand, ISI lacks a tool to help children in the

self-description process since users had to ask the *Facilitator* for the information they were seeking. This consideration leads naturally to the necessity of providing the user of hints (visual and/or auditory) about the different options she/he can select from while describing her/his physical appearance; for example, the range of hair colors that can be opted. Additionally, it could be interesting to insert a *mirroring canvas* in the GUI of the application, centered on the exploration of the physical appearance by the user. Multi-modal support in interactive interfaces is suggested even in [42].

Speech-to-Text to Support the Interaction – During the experiment, the Wizard was able to understand the user even in non-optimal conditions (e.g., when the user was speaking looking away or when there was an external noise). While designing the automated version of ISI (and conversational agents in a broad sense), it would be important to take into account that Automatic Speech Recognition and Understanding would not be as straightforward as for humans. Indeed, there are many studies about speech-to-text performance evaluation both in typical laboratory settings (quiet environment, wideband, and read speech) [32] and in various non-standard and adverse situations and we know from those studies that machine performance degrades below that of humans in noisy situations – whatever noise we consider. For example, there are automatic transcription evaluations in noisy environments [29] (e.g., traffic, crowd), with foreign accents [18], with children's voice tone [44], with emotional speech [7], and with subjects presenting disfluencies in speech and with deaf and hard-of-hearing people [22].

To overcome this drop in performance, we recommend ISI to provide a real-time transcript of the conversation on the screen as evidence of what it understands. In this way, users would have a better understanding of the behavior by the system. Also, ISI should ask very targeted and scoped questions so that users could answer them briefly (ideally with one single word) and thus help the transcription by the system. Finally, for older children, ISI could also provide text-based interactions and prevent any speech-to-text misunderstanding.

The Potential of Natural Language conversations of ISI – Designing the GUI with the clickable button gave the children a sense of control [31], and they quickly understood how to interact with ISI. During the conversation, the graphical explanation of the system status was relevant to letting the children know exactly when it was time to listen and speak. We recommend conversational designers to take this functionality into account for every conversational agent for children, as system status visibility is also encouraged in Nielsen's 10 usability heuristics for user interface design [41]. We noticed that children were positively engaged in speaking with ISI, and they were surprised by the fact that ISI called them by their name. Moreover, we observed that they felt comfortable as the agent used a familiar tone with them. This suggests us to put children's needs, capabilities, and behavior first and attempt to design the conversation on how they react. And this applies to all conversational agents for children and not just to ISI. For example, in our case, it could be useful to add more positive feedback as "Good job!", to engage them to go on with the gamified experience and make

the conversation more natural and fluid. In this regard, it may be interesting to investigate the introduction of an avatar that reflects ISI and leads the children through the experience, examining children's perception and how they adapt the way they interact with the agent.

6 Conclusion and Future Works

This paper aims at exploring the usage of ISI, a speech-based conversational agent that enables children to practice verbal skills through natural language visualization. For this purpose, we ran a preliminary empirical study involving 20 primary school children. From this study, we learned the ease of multi-modal interaction for children and the potential of conversational technologies like ISI for this target group. The contribution of this paper is twofold:

- we described the design process to develop a conversational technology merged with a speech-to-image system that supports children in learning, and we tested its usability and the engagement produced;
- we reported the lessons we learned concerning our agent because we believe that they could be useful ever while designing other conversational agents for children to support the learning.

Our research brings up a few limitations. First, in our study, we involved just children within a restricted range of ages since we wanted them to be able to accomplish the self-description task easily. On one side, it was the right choice because all the children managed to complete the self-description task with the agent without any significant problems. On the other hand, this choice limits the generalization of our results to a broader children's age. To overcome this problem, in our future work, we will introduce different difficulty levels, and we will extend the research to a broader population. Second, so far, we tested only the Wizard-of-Oz version of ISI, and we reported the lessons we learned based on this first experience. However, we are conscious of the differences concerning the conversational skills of an automatic version of ISI compared to our human Wizard, and we know these differences could severely affect the conversational interaction with children. Unfortunately, we could not run an additional empirical study to validate the automated version of ISI because of the pandemic. We will do it as soon as it is possible.

Finally, this research opens up many questions that we want to address in our future studies. First, we will investigate whether a system like ISI that combines Conversational Technology and Natural Language Visualization can be a valid tool for children for improving communication skills – such as lexicon, expressions and sayings, observance of the dialogue timing, and prosody. Then, we would like to focus on the application of ISI as a tool for children to learn a foreign language. Also, we want to analyze if a technology like ISI can be an effective tool for children to improve their self-knowledge, self-awareness, and self-acceptance.

Acknowledgements. This work is partially funded by EIT Digital - Project LETSSAY "Conversational Technology for Speech and Language Therapy".

References

1. Graphic Organizers: A Review of Scientifically Based Research. The Institute for the Advancement of Research in Education at AEL (2003)
2. Al Moubayed, S., Lehman, J.: Toward better understanding of engagement in multiparty spoken interaction with children. In: Proceedings of the ACM ICMI 2015 (2015)
3. Barbe, W.B., Milone, M.N.: What we know about modality strengths (1981)
4. Barbe, W., et al.: Teaching Through Modality Strengths: Concepts and Practices (1979)
5. Begany, G., et al.: Factors affecting user perception of a spoken language vs. textual search interface: a content analysis. Interact. Comput. **28**, 170–180 (2015)
6. Benveniste, S., et al.: Designing improvisation for mediation in group music therapy with children suffering from behavioral disorders. In: Proceedings of the IDC 2009 (2009)
7. Catania, F., et al.: Automatic speech recognition: Do emotions matter? In: 2019 IEEE International Conference on Conversational Data and Knowledge Engineering (2019)
8. Catania, F., et al.: What is the best action for children to "wake up" and "put to sleep" a conversational agent? a multi-criteria decision analysis approach. In: Proceedings of the 2nd Conference on Conversational User Interfaces. ACM (2020)
9. Celi: Voice assistants: Celi's research reveals the habits of Italians (2019). shorturl.at/gvCS6
10. Coffield, F., et al.: Learning styles and pedagogy in post 16 education: a critical and systematic review (2004)
11. Cole, R., et al.: How marni teaches children to read. Educational Technology (2007)
12. Costantino, M.A.: Costruire libri e storie con la CAA: gli IN-Books per l'intervento precoce e linclusione. Erickson (2011)
13. Council, N.R., et al.: From Neurons to Neighborhoods: The Science of Early Childhood Development. National Academies Press, Washington, D.C. (2000)
14. DeepAI: Conversational agent (2019). shorturl.at/gsADL
15. Doswell, J.T.: Pedagogical embodied conversational agent. In: IEEE ICALT 2004 (2004)
16. Druga, S., et al.: Hey google is it ok if i eat you?: initial explorations in child-agent interaction. In: IDC Conference (2017)
17. Eğeci, İ.S., Gençöz, T.: Factors associated with relationship satisfaction: importance of communication skills. Contemp. Family Ther. **28**, 383–391 (2006)
18. Eskenazi, M.: Using automatic speech processing for foreign language pronunciation tutoring: some issues and a prototype. Lang. Learn. Technol. **2**(2), 62–76 (1999)
19. Forbes: How artificial intelligence is making chatbots better for businesses, May 2018. shorturl.at/koFU2
20. Gartner: 25% of customer service operations will use virtual customer assistants by 2020 (2018). www.gtnr.it/2MHVDG3
21. Garvey, C., Berninger, G.: Timing and turn taking in children's conversations. Discourse Process. **4**, 27–57 (1981)
22. Glasser, A.: Automatic speech recognition services: deaf and hard-of-hearing usability. In: The 2019 CHI Conference on Human Factors in Computing Systems (2019)
23. Google: What is conversation design? (2019). shorturl.at/byJL9

24. Hassani, K., Lee, W.S.: Visualizing natural language descriptions: a survey (2016). www.pewrsr.ch/2l4wQnr. Accessed 16 Oct 10 2019
25. Hassani, K., et al.: Design and implementation of an intelligent virtual environment for improving speaking and listening skills. ILE **24**, 252–271 (2016)
26. Hayashi, Y.: Multiple pedagogical conversational agents to support learner-learner collaborative learning: effects of splitting suggestion types. CSR **54**, 246–257 (2019)
27. Kerry, A., Ellis, R., Bull, S.: Conversational agents in e-learning. In: Allen, T., Ellis, R., Petridis, M. (eds.) Applications and Innovations in Intelligent Systems XVI, pp. 169–182. Springer, London (2009). https://doi.org/10.1007/978-1-84882-215-3_13
28. Levinson, S.C., Torreira, F.: Timing in turn-taking and its implications for processing models of language. Front. Psychol. **6**, 731 (2015)
29. Li, J., et al.: An overview of noise-robust automatic speech recognition. IEEE/ACM Trans. Audio Speech Lang. Process. **22**, 745–777 (2014)
30. Light, J., Drager, K.: AAC technologies for young children with complex communication needs: state of the science and future research directions. AAC **23**, 204–216 (2007)
31. Limerick, H., et al.: Empirical evidence for a diminished sense of agency in speech interfaces (2015)
32. Lippmann, R.P.: Speech recognition by machines and humans. Speech Commun. **22**, 1–15 (1997)
33. Lobe, M., Fritsch, R.: Das Kleine Ich Bin Ich Und Das Kleine Hokuspokus. Jumbo Neue Medien (2012)
34. Massaro, D.W.: Embodied agents in language learning for children with language challenges. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2006. LNCS, vol. 4061, pp. 809–816. Springer, Heidelberg (2006). https://doi.org/10.1007/11788713_118
35. Massaro, D., et al.: A multilingual embodied conversational agent. In: Proceedings of the 38th HICSS Annual Hawaii International Conference on System Sciences (2005)
36. Moore, R.J., et al.: Conversational UX design. In: ACM CHI 2017 (2017)
37. Mozilla: Websocket. shorturl.at/KMS27 (2019). Accessed 03 Feb 2020
38. Mueller, G.A.: Visual contextual cues and listening comprehension: an experiment. Mod. Lang. J. **64**, 335–340 (1980)
39. Murad, C., et al.: Design guidelines for hands-free speech interaction. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (2018)
40. Nielsen, J.: UX Design for Children (Ages 3–12). Nielsen Norman Group (2010)
41. Nielsen, J.: 10 usability heuristics for user interface design. NNG **1**(1) (1995)
42. Oviatt, S.: Ten myths of multimodal interaction. Commun. ACM **42**(11), 74–81 (1999)
43. Park, H.W., et al.: Telling stories to robots: the effect of backchanneling on a child's storytelling. In: HRI 2017 (2017)
44. Potamianos, A., et al.: Automatic speech recognition for children. In: Fifth European Conference on Speech Communication and Technology (1997)
45. Read, J.C., MacFarlane, S.: Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In: IDC 2006, New York, NY, USA (2006)
46. Read, J.C., et al.: Endurability, engagement and expectations: measuring children's fun (2002)

47. Sanjanaashree, P., et al.: Language learning for visual and auditory learners using scratch toolkit. In: CCI International Conference (2014)
48. Schmidt, C.R., Paris, S.G.: The development of verbal communicative skills in children. In: Advances in Child Development and Behavior, vol. 18. Elsevier (1984)
49. Snap: Libmoji (2018). wwww.github.com/matthewnau/libmoji
50. Sun, M., et al.: Collaborative storytelling between robot and child: a feasibility study. In: Proceedings of the 2017 Conference on Interaction Design and Children (2017)
51. Urrutia, E.K.M., et al.: A first proposal of pedagogic conversational agents to develop computational thinking in children. In: TEEM Conference (2017)
52. Wiggins, J., et al.: Conversational UX design for kids: toward learning companions. In: Proceedings of the Conversational UX Design CHI 2017 Workshop (2017)
53. Willingham, D.T., et al.: The scientific status of learning styles theories. Teach. Psychol. **42**, 266–271 (2015)
54. Xu, D.: Tangible user interface for children-an overview. In: Proceedings of the UCLAN Department of Computing Conference (2005)
55. Xu, D.: Design and evaluation of tangible interfaces for primary school children. In: Proceedings of the 6th International Conference on Interaction Design and Children (2007)