A Child-Friendly Approach to Spoken Conversational Search

Thomas Beelen[1,*], Khiet P. Truong[1], Roeland Ordelman[1], Ella Velner[1], Vanessa Evers[2] and Theo Huibers[1]

[1]University of Twente, Drienerlolaan 5, 7522NB, Enschede, The Netherlands
[2]NTU Institute of Science and Technology for Humanity, 50 Nanyang Avenue, 639798, Singapore, Singapore

1. Introduction

Increasingly, search engines can be accessed via speech, often through Voice Assistants (VAs). As outlined before by Beelen et al. [1], children are insufficiently supported by technology during their search process, both by search engines on a computer, as well as by VAs. Children have more difficulty in formulating effective search queries that represent their information need well due to a smaller knowledge base and vocabulary [2]. Using speech does not inherently solve these obstacles. Furthermore, most VAs provide only a limited *question-answer* interaction style where the agent directly tries to find results based on the initial query. This interaction style causes several issues for children. Firstly, it is necessary to put all the required context into one query, which they struggle with [3, 4]. Secondly, usually it is not possible to ask follow-up questions, a functionality that children often expect [5, 3]. Lastly, children are typically not supported in formulating queries, for instance by query suggestions or clarifying questions [4, 6].

Many researchers have thus concluded that search tools should assist children in query formulation. With this goal in mind, our research project called CHATTERS (https://chatters-cri.github.io/), focuses on a robot for children's conversational search. We investigate if a *spoken conversational search* approach (see [7]) works better for children than the *question-answer* style interaction that current commercial VAs usually employ. The goal is to help children communicate their information need better via the conversational interaction with the system. We opt for a physically embodied robot to provide a natural conversation, and an engaging experience [8]. In this project we first develop an agent that uses a simplified conversational approach. The simplified conversational approach employs open-ended *elicitation questions* to prompt the child to talk more about their information need (for example: *"What do you want to know about animal species?"*). The elicitation questions are based on templates that do not require any domain knowledge. Instead, the templates extract keywords from the child's speech to generate questions. Subsequently extracted keywords are added to the memory until a threshold is reached and the agent moves on to result presentation (more on this in section 3). Secondly, we evaluate in experiments whether the simplified approach is able to elicit more keywords from children compared to the traditional question-answer paradigm. We also study whether such a robot is engaging to use because this is a precondition for potential long term use. To study these topics we conducted a pilot study with eleven children which will be described in section 4. This pilot is the basis for a main study that we describe in section 5. Here we also describe a different use case, which is to search an archive in a museum. In this case the physical embodiment is especially suited to draw visitors' attention.

2. Problem statement

The problem we address is that children are insufficiently supported in communicating their information needs to voice-based search systems. They are required to formulate queries that contain all necessary context in one statement, which is too complex. We study whether a robot that uses a back and forth conversation can help children communicate their need more effectively by leveraging multiple turns. Furthermore, we are interested if a conversational approach provides a more social and engaging experience. Our target audience are children between 10–12 years old.

3. Proposed system

We first develop a robot that uses *simplified conversational search*. In this interaction, the robot initiates the

MICROS'22: Mixed-Initiative ConveRsatiOnal Systems workshop at CIKM 2022, October 21, 2022, Atlanta, GA
*Corresponding author.
t.h.j.beelen@utwente.nl (T. Beelen); k.p.truong@utwente.nl (K. P. Truong); roeland.ordelman@utwente.nl (R. Ordelman); p.c.velner@utwente.nl (E. Velner); vanessa.evers@ntu.edu.sg (V. Evers); t.w.c.huibers@utwente.nl (T. Huibers)
https://thomasbeelen.com/ (T. Beelen)
0000-0002-5650-2830 (V. Evers); 0000-0002-9837-8639 (T. Huibers)

conversation with an introduction. It states its name, asks the child for theirs, and says it is pleased to meet them. Then the robot asks if the child wants help searching for information, and the child can ask a question. Since the child's speech may contain words that are not relevant to the search (such as "uhm, let's see"), keywords need to be extracted. In the pilot study described in section 4, keyword extraction was done by matching against a pre-programmed list of possible keywords. Detected keywords are added to the robot's memory. If the number of keywords is below a threshold, the robot will pose an *elicitation question.* This threshold will be optimized in the future. Too few keywords lead to an unspecific ranking of results, while trying to elicit too many keywords leads to a long interaction and possibly frustration. In the pilot study (section 4) this threshold was set to three keywords, which was an estimate based on the number of words in the search tasks. The elicitation questions are based on a simple pattern that includes the keywords that were extracted so far. The two question patterns are (translated from Dutch):

- What is it you want to know about `[recognised keywords]?`
- What don't you know yet about `[recognised keywords]?`

The word *"and"* is added to the list of keywords where necessary. The robot loops over the questions and adds new keywords until the threshold is met. After this phase, the robot will move on to present search results. Then the robot asks the child if it can be of any further assistance. Otherwise it goes to a closing interaction.

4. Pilot study

We conducted a pilot study as a first step to find out if the simplified conversational approach elicits more keywords (as described in the introduction), and how children experience such a robot. Furthermore, the pilot is a way to discover methodological issues that may still be corrected for the main study. In the pilot, we evaluated a robot that uses the simplified conversational search approach. As described in the evaluation framework by Landoni et al. [5], studies in IR for children can be described by the intended *search strategy*, for a particular *user group*, given a *task*, in a certain *context.* Our *user group* are children ages 10–12 years old. These children are in the final years of primary school in the Netherlands and are starting to work more on assignments such as presentations. We compare our conversational robot *strategy* to a traditional *question-answer* style interaction that is common with commercial VAs. The *task* in the experiments is searching information related to school subjects. The *context* of the searches is at school or in the home. In the future we change to a museum archive search task.

4.1. Method

Our study is a *within-subjects* comparison with two conditions. In one condition the robot uses the *simplified conversational* interaction style (see section 3), in the other it uses a *question-answer* interaction style. The order of the conditions was alternated between participants. The Furhat robot [9] and its software environment are used for both conditions. This keeps the two experimental conditions similar while the style of interaction is varied.

There are two search tasks, one for each condition (since condition order was alternated, each task was used on different conditions). Both tasks are factoid questions based on the work by Landoni et al. [5]. The first was to find out what hail is. The second was to find three endangered animal species. The tasks were presented in one sentence in Dutch on the task sheet. The search results that could be retrieved were pre-programmed in this pilot, and were the same in both conditions. When presenting results, the robot mentions the website where it found the information, the name of the article, and then reads aloud a snippet of the web page.

In the *question-answer* condition, the robot mimics the interaction of a commercial VA. This means it first waits for a *wake word*, in this case *"Hey robot"*, or *"Hey Furhat"*. Then the robot's LED ring lights up green to signal it is awake, and a question can be asked. The wake word and question can also be combined into one statement. The robot will then present results right away in the way described above. After the results the robot goes back to waiting for the wake word.

4.2. Measures

The measurements consist of observational notes, Likert scales using emojis (Smileyometers [10]), logs containing a raw transcript, and interview questions. The children were also asked about their current VA usage. Observations focused on how the children behaved and spoke to the robot. The Smileyometer was used to gauge the user experience of the interactions. The questions are based on [5], they include questions on: fun, ease of finding answers, intention to use again, kindness, and ease of conversing with. The logs are a transcript of how the robot interpreted the child and how it responded. This is used to evaluate whether the approach is able to elicit more keywords from children compared to the traditional paradigm. The goal of the interview at the end is to study children's perception of the differences and advantages of the two systems, and to add more qualitative data on their experience

4.3. Participants & procedure

Eleven children participated in the pilot at a local after school care over two days in June 2022. Children whose parents consented joined the experiments. Our user group are children ages 10-12, but the children that participated were mostly younger (mean age = 8.8, SD = 1.3). We decided to also let younger children participate due to the difficulty of recruiting children in our targeted age range.

The robot was set up on a table in a separate room with an open door to the main area. The researcher sat at the same table and height as the child and explained that they will be talking to two versions of the same robot. It was explained that the robots may not be fully functional yet, and that the children are helping to further develop them. The children were also explained that one of the robots will start talking right away, and the other requires a wake word. The children received a task sheet that also includes the questionnaire questions. The interview happened after interacting with both robots.

4.4. Results

The first day the speech recognition often did not understand children correctly due to background noise. Therefore, the second day a headset was used, and Wizard-of-Oz (undisclosed human operator) functionality was added. This way the researcher could take over when some responses were not understood correctly by the automatic speech recognition. Due to speech recogniser errors the logs were not usable for analysis.

Observation Many children resorted to reading the search task directly from the sheet. This caused the queries in both conditions to be similar. Mainly for younger users the task seemed complex and they required more input from the researcher. The children around the target age seemed more comfortable with the level of complexity, working more independently. In the conversational condition, the system entered the elicitation question loop in many cases. Sometimes, the child wrongfully assumed they already provided all the words in the search task, and got confused by the elicitation question. Especially younger children tended to look at the researcher when they were unsure how to continue the interaction. This also happened at the elicitation questions where the researcher had to give a hint. In other cases the child kept the conversation going and answered the elicitation question naturally.

Smileyometers The robots scored relatively similar in the Smileyometers. The results suggest that the *conversational* robot may be more enjoyable to use and easier to find answers with. Children also indicated being slightly more likely to use it again. The robots were seen as roughly equally friendly and easy to talk to. The results are shown in figure 1.

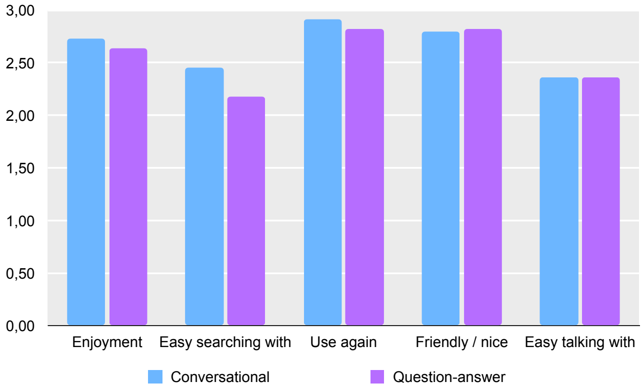

Figure 1: Smileyometer outcomes pilot study

Table 1
Children's statements on the two robots

Robot	Statements
Conversational	*It speaks directly (it is awake and thus faster).* *It is more fun because there is more talking.* *It is easier to use.* *It requires less talking.*
Question-answer	*I have to wake it up first.* *It is better because I have to talk less. It's faster.* *I like that it turns green* [after wake word].

Interview Four children preferred the *question-answer* condition, while five preferred the *conversational.* Two children had no preference. Some of the interesting statements on the robots are in table 1. The statements indicate there is a potential trade-off between efficiency and fun. There are also clear individual differences, as some children seemed to enjoy talking more, while others preferred a fast interaction. Concerning participants' VA usage, five children had no experience, three children used them a few times, and three used them frequently. The frequent users ask VAs about the weather, jokes, and finding information. No effects of prior VA usage on the outcome could be deterined in this pilot. Some children gave tips to improve the robot. These tips were: an easier to understand voice, and a touch screen on the robot's face to be able to select search results visually as well.

4.5. Conclusion and limitations

Based on the pilot, the number of elicited keywords could not yet be compared due to errors and task design. Children rated the robots quite similar but possibly find the conversational system more fun and easier to find answers with. More children preferred the conversational system. They perceived the robots as nearly equally friendly. The interview answers shed light on the individual preferences regarding the amount of conversation during the search process.

The pilot study also gave insights that influence the method of the main study. Firstly, the small sample size means the current findings have low confidence. In the main study we will increase the sample size by relying on cooperation with partners such as museums. Secondly, background noise led to errors and required the addition of Wizard-of-Oz controls. Another limitation is that most participants were younger than our target audience. A few years can have a significant developmental difference in children, therefore the interaction may be less complex for children in the target age. However, the complexity of the interaction seemed suitable for the older participants around the target age. Finally, the search tasks were mostly read aloud from the task sheet, which likely affects the naturalness of children's queries. In the next section we describe how a different task in the museum context may address this.

5. Future work

The next step in creating the conversational robot is to connect it the API of the Netherlands Institute for Sound and Vision[1], containing Dutch public broadcasting media. In line with the tip by one of the participants, this use case will also introduce a display for multi media search results. The API connection will enable us to study more natural search tasks and move away from pre-programmed search results. The tasks that were used in the pilot are fact finding and stated directly on the task sheet. The API connection would allow children to search for TV fragments that they come up with themselves, which is a more open search task than fact finding. This enables us to study children in a more natural setting, where their query formulation process more closely reflects a realistic scenario instead of reading from a task sheet. Elicitation questions may become more useful in this case. A more advanced keyword extraction from speech method will need to be implemented as well, such as the one by Habibi and Popescu-Belis [11]. The API connected robot will be tested in a similar method as the pilot study described above. The method compares the style of interaction without changing other aspects about the robot between conditions. The within-subjects setup allowed children to reflect on differences between the systems and their preference. The Smileyometers worked well even for participants that are younger than the target audience. With our pilot findings we can account for some important methodological issues. We look forward to learning more about children's conversational search process in our main study.

[1]https://www.beeldengeluid.nl

Acknowledgments

This research is supported by the Dutch SIDN fund https://www.sidn.nl/ and TKI CLICKNL funding of the Dutch Ministry of Economic Affairs https://www.clicknl.nl/. We would also like to thank the staff at BSO Partou de Vlinder for their time and cooperation.

References

[1] T. Beelen, E. Velner, R. Ordelman, K. P. Truong, V. Evers, T. Huibers, Does your robot know? Enhancing children's information retrieval through spoken conversation with responsible robots, arXiv:2106.07931 [cs] (2021). URL: http://arxiv.org/abs/2106.07931, arXiv: 2106.07931.

[2] H. Hutchinson, A. Druin, B. B. Bederson, K. Reuter, A. Rose, A. C. Weeks, How do I find blue books about dogs? The errors and frustrations of young digital library users, Proceedings of HCII 2005 (2005) 22–27. Publisher: Citeseer.

[3] S. B. Lovato, A. M. Piper, E. A. Wartella, Hey Google, Do Unicorns Exist? Conversational Agents as a Path to Answers to Children's Questions, in: Proceedings of the 18th ACM International Conference on Interaction Design and Children, IDC '19, Association for Computing Machinery, Boise, ID, USA, 2019, pp. 301–313. URL: https://doi.org/10.1145/3311927.3323150. doi:10.1145/3311927.3323150.

[4] S. Yarosh, S. Thompson, K. Watson, A. Chase, A. Senthilkumar, Y. Yuan, A. J. B. Brush, Children asking questions: speech interface reformulations and personification preferences, in: Proceedings of the 17th ACM Conference on Interaction Design and Children, IDC '18, Association for Computing Machinery, Trondheim, Norway, 2018, pp. 300–312. URL: https://doi.org/10.1145/3202185.3202207. doi:10.1145/3202185.3202207.

[5] M. Landoni, D. Matteri, E. Murgia, T. Huibers, M. S. Pera, Sonny, Cerca! Evaluating the Impact of Using a Vocal Assistant to Search at School, in: F. Crestani, M. Braschler, J. Savoy, A. Rauber, H. Müller, D. E. Losada, G. Heinatz Bürki, L. Cappellato, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2019, pp. 101–113. doi:10.1007/978-3-030-28577-7_6.

[6] S. Druga, R. Williams, C. Breazeal, M. Resnick, "Hey Google is it OK if I eat you?": Initial Explo-

rations in Child-Agent Interaction, in: Proceedings of the 2017 Conference on Interaction Design and Children, ACM, Stanford California USA, 2017, pp. 595–600. URL: https://dl.acm.org/doi/10.1145/3078072.3084330. doi:10.1145/3078072.3084330.

[7] H. Zamani, J. R. Trippas, J. Dalton, F. Radlinski, Conversational Information Seeking, arXiv preprint arXiv:2201.08808 (2022).

[8] J. Li, The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents, International Journal of Human-Computer Studies 77 (2015) 23–37. URL: http://www.sciencedirect.com/science/article/pii/S107158191500004X. doi:10.1016/j.ijhcs.2015.01.001.

[9] S. Al Moubayed, J. Beskow, G. Skantze, B. Granström, Furhat : A Back-projected Human-like Robot Head for Multiparty Human-Machine Interaction, Springer Berlin/Heidelberg, 2012, pp. 114–130. URL: http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-105606.

[10] J. C. Read, S. MacFarlane, Using the fun toolkit and other survey methods to gather opinions in child computer interaction, in: Proceedings of the 2006 conference on Interaction design and children, IDC '06, Association for Computing Machinery, New York, NY, USA, 2006, pp. 81–88. URL: http://doi.org/10.1145/1139073.1139096. doi:10.1145/1139073.1139096.

[11] M. Habibi, A. Popescu-Belis, Diverse keyword extraction from conversations, in: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2013, pp. 651–657.