Exploratory Voice-Controlled Search for Young Users: Challenges & Potential Benefits ⋆

Michael Kotzyba * **Ingo Siegert** ** **Tatiana Gossen** *
Andreas Wendemuth ** **Andreas Nürnberger** *

* *Data and Knowledge Engineering Group, Faculty of Computer Science, Otto von Guericke University, Universitätsplatz 2, 39106 Magdeburg, Germany (e-mail: firstname.lastname@ovgu.de).*
** *Dept. of Electrical Engineering and Information Technology, Otto von Guericke University, Universitätsplatz 2, 39106 Magdeburg, Germany (e-mail: firstname.lastname@ovgu.de)*

Abstract: To date, children come more and more often into contact with information systems. Unfortunately, these systems are usually designed for adults and the increasing group of young users has been widely neglected. In order to create and support children's information seeking appropriately, we have to consider their specific state of development resulting into different needs and capabilities. In this paper, we discuss the design and the results of a user study to analyze exploratory voice-controlled search of young users. Exploratory search nicely reflects real-life information needs and voice-controlled interfaces have the advantage that children do not need to have good spelling skills. Hence the interaction can be more intuitive and motivating. We gained first insights into the children's search behavior in context of exploratory search and identified design issues for similar studies. Furthermore, we illustrate the challenges and potential benefits of interactive voice-controlled search systems for young users.

Keywords: Exploratory Search; Voice Control; Search User Interface; Young Users;

1. INTRODUCTION

The presence and impact of interactive technical systems in our daily life increases more and more. Also children come more often into contact especially with information systems at a very young age. Even though, a lot of studies about information seeking behavior have been conducted with adults, the increasing group of young users has been widely neglected. However, in order to support children's information seeking appropriately, we have to consider their specific state of development. In addition, the affective state often plays a large role in children's information seeking behavior (Nesset (2013)). In this work, we discuss the design and results of a user study to analyze exploratory voice-controlled search of young users. We use the construct of exploratory search as a class of information seeking processes to induce a natural search that reflects real-life information needs. A voice controlled search is motivated by the idea to support young users by compensating for their not well developed reading and writing abilities (see Gossen and Nürnberger (2013)). By using speech recognition, the user does not need to have good spelling skills. Besides, the interaction with a voice-controlled system can be more intuitive and hence more motivating for children as they do not have to learn the cumbersome interaction with mouse and keyboard. In addition acoustic characteristics can be used to uncover different (affective) states of the user.

The structure of the paper is as follows. First, we describe the preliminaries, our study setup, the procedure and technical requirements we have defined to achieve high quality user study data of different modalities like eye tracking data, voice recordings and video data. Afterwards, we discuss the collected data in form of questionnaires, interactions and acoustic characteristics – extracted from the user's voice – in order to derive typical interaction patterns during information seeking. Our goal is to reveal correlations between acoustics and interaction patterns with the interface during an exploratory search. Thereby, our research questions are:

(1) What kind of interactions do children use and how often do these interactions occur during the exploratory search process?
(2) How often do acoustic characteristics such as off-talk occur during an exploratory search and how it evolve over time?

On the basis of our findings we discuss current challenges and potential benefits of our results for voice-controlled search systems.

⋆ This paper is a revised and extended version of the workshop contribution "Kotzyba, M., Siegert, I., Gossen, T., Wendemuth, A. und Nürnberger A.: Exploratory Voice-Controlled Search for Young Users: Challenges & Potential Benefits, Proc. of 3. Interdisziplinärer Workshop Kognitive Systeme: Mensch, Teams, Systeme und Automaten, Magdeburg, March 25 - 27, 2014, Eds: Wendemuth, Jipp, Kluge und Söffker, Magdeburg, Germany, 2014.".

To date, there is only a few data material available for naturalistic human-machine interaction of children. One race dataset is presented in Batliner et al. (2004)), where children are interacting with an robot pet. Thus, it is not clear, which kind of interaction patterns and emotional statements are used by children that can serve as indicators for an affective system. In this paper, we develop a scenario of a longer lasting naturalistic interaction between a system and a child to investigate the occurring patterns and phenomena. This investigation can than be used to set up hypothesis that can be investigated in follow-up experiments.

2. RELATED WORK

Young users are naturally motivated to use the Web and interactively explore its content (Bilal and Kirby (2002)). However, they require support during information acquisition (Gossen et al. (2012)). In specific, children require emotional, language, cognitive, memory, interaction and relevance support. This means that children can easily get frustrated, have difficulties to formulate queries due to a small vocabulary and errors in spelling, have difficulties with thinking abstractly, can process less information than adults, have difficulties with complex interactions and have difficulties to judge the relevance of the retrieved documents to their information need. Hence it is important, to provide appropriate user interfaces and support mechanisms like voice control for children.

2.1 Search User Interfaces for Children

Information retrieval for children is a broad topic. A survey about information retrieval (IR) for children, where different components – both user interface and algorithms – of an IR system are covered, is given in Gossen and Nürnberger (2013). The survey explains specifics of young users, i.e., their cognitive skills, fine motor skills, knowledge, memory and emotional states and how they differ from those of adults. Previous user studies focused on children's information-seeking behavior, e.g. Bilal and Kirby (2002); Borgman et al. (1995); Kammerer and Bohnacker (2012), provided an overview of retrieval algorithms, e.g. Gyllstrom and Moens (2010), and search user interface concepts, e.g. Jansen et al. (2010), and described existing information retrieval systems for children, in specific web search engines and digital libraries, e.g. Hutchinson et al. (2006).

Children can get easily frustrated if they do not find relevant results, do not understand the search engine output or if a failure emerges (Bilal and Kirby (2002)). The fact that children also have difficulties to evaluate the relevance of retrieved documents to their information needs aggravates this (Jochmann-Mannak et al. (2010)). Furthermore, most children have difficulties with typing (Budiu and Nielsen (2010)). They are not able to type commands without looking at the keyboard (touch-typing). Instead they typically hunt-and-peck on the keyboard for correct keys. By looking at the keyboard while typing, children often do not spot spelling mistakes. In addition, some interaction techniques like scrolling or drag-and-drop are difficult for young users Budiu and Nielsen (2010). Therefore, young users would benefit from support mechanisms helping them to formulate their information need while interacting with a search user interface.

2.2 Voice Analysis

To improve search user interfaces for children, a voice-controlled search interface, that allows input and output voice interaction, can be used. By using speech recognition, the children do not need to have good spelling skills, the interaction can be more intuitive and motivating. But, it is known that speech does not only transmit the pure content of the message but also transmit further aspects, as relationship, feelings, or self-revelation (cf. Schulz von Thun (1981)). For an elaborated human-computer interaction, the analysis of the human behavior is needed. This includes the evaluation of affective behavior (cf. Picard (2000)). The goal is, to obtain a more human-like and more successful interaction with technical systems. Therefore, those systems have to be adaptable to the users' individual skills, preferences, and user characteristics. This includes both, the ability to understand the user's capabilities and a proper reaction towards him (cf. Wendemuth and Biundo (2012)). In this way, the utilization of speech-based search user interfaces allows us to evaluate the emotional and dispositional state of the children during the interaction with our system. This enables future systems to react onto the individual abilities and preferences of the child actually interacting. For this case, emotions are seen as a major source. Several researchers have presented methods, to automatically extract and evaluate the emotional and dispositional state of the user based on their acoustics (cf. Cullen and Harte (2012); Prylipko et al. (2014); Hassan et al. (2013); Schuller et al. (2011)). But only rarely children interacting with a technical system were in the focus (cf. Batliner et al. (2004)).

Generally, automatic emotion recognition make use of pattern recognition methods and measures input signals, extract characteristics and assign them to appropriate categories. The recognition of the user's affective state is still a challenging task. With ongoing research, the need for naturalistic databases grew. Especially Batliner et al. (2000) argued for the need of real-live material and pointed out the recognition difficulties with realistic material. Furthermore, in non-acted material the bandwidth of emotions is much broader and the a set of few emotions, like Ekman basic emotions, is not sufficient (cf. Siegert et al. (2011)). Additionally, the emotional expressiveness in realistic material is much lower than in acted material (cf. Grimm and Kroschel (2005)).

But, naturalistic data has the advantage that a complete interaction can be observed which is helpful for the needed annotation process, as the material lacks of a proper emotional labeling, cf. Siegert et al. (2013). Thus switching from short acted utterances to longer naturalistic interactions increases the reliability and variability of the occurring emotions but at the same time the number of variations is increased, such as acoustics, individuality, and recording conditions.

2.3 Previous Work

In general, affect detection during search is a promising approach especially in case of young users as they do not

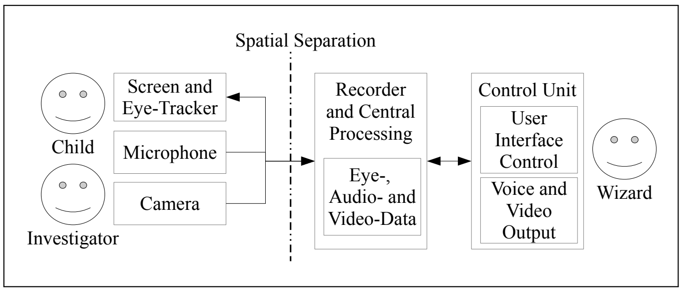

Figure 1. Spatial and technical setting of our study.

have the full developed ability to control their facial expressions (Ceschi and Scherer (2003)) and use impoliteness (insult) more frequently than adults when interacting with spoken dialog system (Yildirim et al. (2011)), thus making it easier to identify affective states.

In previous research, we already conducted a small experiment and found out that children are motivated to use voice-controlled search engines (Gossen et al. (2013a)). However, the quality of the recorded audio data was not good enough for an automatic speech and emotion recognition. Therefore, in this work, we use a high-tech equipment to overcome this issue. Furthermore, we study an exploratory search with a predefined task, whereas in previous work children used a free search.

3. STUDY DESIGN

For the study, we have created an environment, where a young user performs a natural search with the possibility to widen his or her knowledge about a certain domain and investigate new sub domains over several minutes without any external influence. To achieve this, the study needs adequate setting with respect to ambiance and installation, a well defined procedure, a child-friendly search engine and an adequate search task. In the following subsections we describe the used study setup, its technical details, the procedure and the different tasks for the young user and the investigators. The goal of this study are twofold: We like to gain insights into the search behavior of young users when using voice-controlled search. In addition this study should identify design issues of setting and procedure for a larger follow-up study.

3.1 Ambiance and Installation

To gather the necessary data of children's speech and interaction patterns, the study follows a so-called "Wizard-of-Oz-Experiment". Here the user assumes, that he or she interacts with a fully autonomous application, but actually it is controlled by a hidden operator (called Wizard). Figure 1 depicts the spatial and technical setting of our study. Child and Wizard are spatially separated. The child can notice only the different recording devices (eye tracker, headset microphones, cameras), the search user interface on the screen and the investigator. In Figure 2 the installation from the child's point of view is depicted. The room, where the study took place is designed to provide a comfortable, relaxed and thus natural feeling for both, the participant and the investigator. It has different interieurs like cupboards, frames and a sofa. The colors of the walls and interieurs are bright and the fanlight has a low neutral white luminous color (about 4000 K). By designing the study workspace for the children, one has to consider the large inter-individual differences in children's body size and hence provide adequate broad adjustments. To avoid (side) noises from outside the room, a sound-proof door is installed. For the study, we used the following recording devices:

(1) Eye Tracker: Tobii T60
(2) Two Sennheiser headset microphones: HSP 2-EW-3 (for child and investigator)
(3) Two AVT Pike F145C cameras with Tevidon objective 1,8/16.

The Wizard in the neighbor room is controlling the hole system. He or she controls the search user interface, sees the young user in the front (via the Pike cameras), hears the user and the investigator and has an overview over the hole scene in the study room. To synchronize the different video and audio streams, the audio interface Yamaha Steinberg MR 816X and its internal world clock trigger with 44 kHz was used. However, the used version of the eye tracker could not be triggered external and hence its data had to be synchronized afterwards manually.

3.2 Search User Interfaces

The young participants used the child-centered search engine "Knowledge Journey" (Gossen et al. (2012)). The search engine uses the metaphor of a treasure hunt, where a user takes a journey to gather relevant search results. In Figure 3 the interface of the Knowledge Journey and its elements are illustrated. We used the search user interface settings that were preferred by children according to Gossen et al. (2013b). That means we selected a predefined theme (here "space" instead of "pirates"), fonts, result illustration, etc. However, some settings were adopted to the current study design. Table 1 lists the used settings. In contrast to Gossen et al. (2013b) we did not activate the pie menu and the audio support. The pie menu suggest general categories to initiate a search and helps users to lead them into predefined domains. In our study the domain was given by the exploratory search task (see

Figure 2. Installation from the child's point of view.

next section “Search Task”) and a menu could distract the young user from the actual task. To hold the voice recordings from the children as clean as possible any kind of (side) noises from inside the room has to be avoided. For this reason, we deactivated the audio support, because a spoken audio tooltip from the system would be also recorded on the voice records.

Another deactivated but possible feature of the Knowledge Journey is the website preview. “Preview” provides a centered (pre-)view of a web site that is bigger than the summary in the coverflow but smaller than a full sized web site. Pictures and headlines are well readable in this view but the plain text of the web site may be pretty small. By using the preview functionality, two drawbacks for the eye tracker recordings may emerge. First, because of small plain text, the children tend to come more near the screen so that the eye tracker can not recognize the eyes. Second, if the participant uses the preview but not the full sized view of a web page, the gaze data is focused on the center of the screen and hence it is more difficult to identify crucial points at the website from the eye tracker data because of the reduced resolution (in comparison to a full sized view).

3.3 Search Task

Exploratory search tasks comply with the study requirements of a natural search, because they are designed to

Figure 3. Screenshot of the Knowledge Journey user interface: a guidance figure and a log book on the right hand side, query input elements on the top and a coverflow with search results in the middle.

Tab. 1. Settings for the Knowledge Journey

Parameter	Value
Theme	*Space*
Avatar	*Alien*
Result Visualization	*Coverflow*
Thumbnail	*true*
Keyword Highlighting	*true, yellow*
Website preview	*false*
URL	*true*
Result Page View	*In the same window*
Font Size	*14 pt*
Font Type	*Comic Sans MS*
Menu	*false*
Audio Support	*false*

reflect real-life information needs (White et al. (2007)). An exploratory search is a class of information-seeking and extends a standard lookup-search about learning and investigation activities (Marchionini (2006)). Here, the user has no clearly search goals and can formulate his or her information need only vaguely. A simple sequence of look-up search tasks generated by an investigator may have the disadvantage, that for the participant – depending on the found documents and content – the previous and the current search tasks have no relationship to each other. Hence a natural information seeking process can be interrupted.

A task to induce exploratory search for children needs to consider the following aspects: The user should not be too familiar with the search domain and the available search tools. With a sophisticated knowledge about the domain, the user is able to formulate concrete fact finding queries and thus the learning and investigation activities we are looking for may not occur. The task should be open ended. To “solve” the task, the user should spend a longer time on the task. Usual exploratory search can take hours, days or even weeks. To induce an exploratory search, we used the following task:

“Imagine you are a zoo director and you start your own small zoo. Your zoo shall provide different animal species a new home. To create an environment where all your animals are fine, you have to maintain an adequate animal housing. Use the Internet to read up on the needs of your animals. Use our Knowledge Journey to search for information and control it via speech”

This task compromises the domain issue because the needs of zoo animals are known to some extent by most children and should motivate them to execute a “longer” search. In contrast to that, only a few children should be able to name specific facts about adequate animal housing. For user studies, especially with children, the last constraint, addressing the search time, is difficult to fulfill. On the one hand, for a search of only a few minutes, it is almost impossible to produce learning and investigation activities. On the other hand, a search about hours is not realistic for children to conduct because after some time they get distracted. For our experiment we chose a 20 minutes search session time as a trade-off.

3.4 Procedure

The user study consists of the following four steps:

Pre-interview: We used a pre-interview to gather the user’s demographic information and their experience with computer systems and the Internet. Besides, we asked the participants about their experience with different input methods, i.e., keyboard, mouse, touch and voice control.

Introduction: In the next step, we gave the participants an introduction to the Knowledge Journey and how to use it. We briefly explained what the different elements are and what purpose they have. Intentionally, we gave only few information about how to use the interface per voice concretely. After that the investigator explained the exploratory search task that induce the search.

Tab. 2. Detected Types of Expressions

Types of Expressions	Example
Conducing soliloquy	"Why I wanted to say elephants [again]?"
Starting off-talk	"I also had guinea pigs."
Asking about the next search step	"Should I search for the next animal?"
Asking questions with regard to the content?	"What does Hz mean?"
Asking about user interface control	"How can I say to him, that he shall save this?"
Asking about user interface indicators	"Why is he writing Löwe (lion) with oe?"

Search Experiment: In this step, the actual search is performed. The young users could use the elements of the search user interface how they want, but only using voice commands. If a young user had no idea how he or she should use the system, an investigator gave some assistance, e.g., "Simply try to say the system what you want to do." or "You can try to say it in any other way.". If a child used an ambiguous command or the system (the Wizard) could not understand the command, the prepared synthesize female audio message "I could not understand you" was triggered.

Post-interview: The last step was a post-interview to evaluate the user's attitude towards the system and if they would use voice-controlled search user interface in the future. We also asked for recommendations to improve our search engine.

Each session with these four steps took about 45 minutes.

4. USER STUDY ACTORS

In this section we describe our participants and the problems we identified about the role and focus of the investigator and Wizard during the Wizard-of-Oz experiment.

4.1 Participants

Our user study was conducted in February 2014 at the Otto von Guericke University, Magdeburg, Germany. Our five participants were of age eight to nine (average 8.8 years), three female and two male. They were mostly third (4 children) and fourth (1 child) grade pupils and all of them spoke German. Thus the citation of the children – as given in the following – were translated as true to the original as possible. All children had experience with computer systems and the Internet. Three pupils use the Internet multiple times per week and two of them approximately once a week. The duration for using the Internet is between 10-30 min (2 pupils) and 30-60 min (2 pupils). Only one user stated an Internet usage between 1-2 hours (multiple times per week). Concerning the used devices and their input modalities (keyboard or touchscreen), no preferences have been identified. Generally all participants use the Internet without any supervision but on occasion relatives look at them or support on demand, for instance to write e-mails.

All children use the Internet to search information for them self or for school. Additionally two pupils write e-mails and hear music. In order to search for information, four participants use Google as default search engine and one pupil uses the German search engine for children *Blindekuh.de*. Other child centered search engine for children like *Fragfinn.de* or *Helles-Koepfchen.de* are known but used only seldom. No participant had ever used the Knowledge Journey.

4.2 Investigator

The investigator is confronted with different tasks within the four procedure steps but also beyond these steps. When the child (and the relatives) arrives at the study location, the investigator have to present and accustom them to the new surroundings. For the relative, we provided the possibility to monitor the whole experiment and stay beside the young user. But we also alluded that the presence of a relative may have additional influence to the experiment and we advise to wait outside of the room. For our study all relatives agreed to wait outside (or execute some other business) and thus the data is not influenced by the presence of relatives.

Before conducting the procedure steps, the investigator has to explain the individual study steps to the child and to lower his or her worries. This can be done by explaining the study goal and by the advise, that no performance is measured and the the user can not do anything wrong. In the pre-interview step, the investigator has to fill in the first part of the questionnaire, ask the child and if necessary to explain a question in more detail. The second step to introduce the usage of the system also includes the presentation of the search task and the calibration of the recording devices to the size of the user.

If the pupil processes the search task, the investigator should intervene as few as possible but also give assistance if necessary. In this situation the variance of investigators influence may be large. E.g. how to react, if a child says "Should I search for the next animal?". For this situation, a pre-defined list with behavior rules for the investigator may be helpful. Based on the experiment, we detected the following types of expressions from the (young) user. Table 2 lists the detected expressions and a respective example. These expressions can be considered and adequate types of reactions for the investigator should be pre-defined.

Of course the specific pre-defined reactions for the investigator depends on the concrete user study and its goals. E.g. in a study to investigate common used voice commands to control the interface in combination with a free search task, the investigator should not give example commands but he or she may explain some content related facts of a web site. The last step of the study procedure is similar to the first step. The investigator fills in the remaining questions and explain it in more detail when necessary.

4.3 Wizard

After the introduction step the Wizard starts the recording and imitates an autonomic system by reacting to participants' commands. From that point, similar to the investigator, the Wizard has also to decide to which commands from the user he or she has to react and how. Most of these decisions address the issue how tolerant the system (the Wizard) is in interpreting the commands. Here the degree of allowed interpretation is linked to the intelligence of the prospective system. In general, adequate reactions should be pre-defined for each individual study.

In our study, the participants were free to give any kind of voice command. The Wizard reacted to the commands of participants that contained descriptive terms for the user interface elements they wanted to use. For example, the expression "I want to search for ..." contains the term "search" that is descriptive for the search input field. This behavior of the Wizard gives much freedom for the participants to express their desired interactions. The Wizard also used the prepared audio message "I could not understand you". This message was triggered when:

- It was not clear if the participant still continued the off-talk/monologized or already gave a command to the system.
- The command of a participant did not contain any descriptive terms, e.g. "What do turtles eat?". However, if the user behaviour was persistent (after three command repetitions), the Wizard used the command to form a query and search for results. The idea behind that was to avoid user frustration.
- The command of a participant was not clear. For example, the command "image open" is not clearly defined in case there are several images on the screen.

The Wizard took into account the context and the user interaction sequence to interpret the user correctly. That means, participants could use the same command but the Wizard performed different actions depending on the current state, e.g. the command "continue" meant going to the next page in coverflow when looking at the result pages. It meant scrolling down if the web site is opened.

5. USER STUDY RESULTS

In this section we describe the study results focusing on interactions and acoustic characteristics related to the exploratory search. As suggested above, we recorded different modalities like eye tracking data, voice recordings and video data. The focus in this work lies on interaction and audio analysis. Thus, the video and eye tracking data are not analyzed. The transcripts, generated with the tools Exmaralda (Schmidt and Wörner (2009)) and Tobii Studio (eye tracker software), served as a basis for our analysis. These tools allow to analyze the interaction, conversation, and to perform discourse analysis.

5.1 Interaction Analysis

To answer the first research question – What kind of interactions do children use and how often do these interactions occur during the exploratory search process? – selected interactions of the children are discussed.

Due to the design of this study, the experiment can be seen as a conversation between three subjects, the children, the technical system and the investigator. All three of them make to some extend contributions to the interaction, but from the systems view only the children serve as a "commander". Therefore, we first identified two different kind of dialogues: the child talking to the machine ("command" also called interaction) and the child talking to itself or the investigator ("off-talk"). In our case, only statements, directly addressed to the system are regarded as "commands", all other statements are regarded as "off-talk". From a systems perspective, the distinction into commands that have to be processed and off-talk, that can be neglected but convey important information about the children's confidence with the task is of major importance.

The very first expression of participant 1 (female) was "habitat of penguins(?)". Since this expression was done during the search engine's start page it suggests itself to interpret a search query. The command terms indicted the user's wish to search for the expression and the system (the Wizard) adapted it as adequate search command. For all following search commands, the young participant uses the scheme "habitat space of [animal name]" again. If the system has not understood a command or the system was not fast enough, participant 1 shows a persistent behavior and she tried different alternatives for the same command, e.g. "save", "save this page", "system save" or she changed the pronunciation length, e.g. "save", "saaaaavvve".

Participant 2 (female) was a bit unsure concerning the user interface control. Before she tried a new command, often the investigator was asked, whether she can say this to the system or not. Furthermore she tried the same commands for different interface elements, e.g. "continue" for scrolling down in a web site but also to forward in the coverflow. Participant 2 used the text summary in the coverflow more extensive in contrast to the other pupils. After reading this summary text, she decided to enlarge or skip the search result. Another interesting fact is, that she continuously used the term "google" to formulate search querys, e.g. "I want to google for giraffe".

In comparison to the other participants, the third child (female) showed a dominant characteristic. Almost all commands were spoken pretty loud and clear in comparison to the off-talk. The same is true for participant 4 (male), all interactions are very clear articulated and pronounced quite confident. He initiated 18 search queries and all of them followed the same schema: "search [search terms]". The navigation-commands in the Knowledge Journey or in web sites also remained clear, articulate and almost unchanged. To navigate in the coverflow, he used "continue" and "back", to open a web page often "large", seldom "open" was used for navigating in web pages only "up" and "down" were used. His experience in using the internet, he stated in the questionnaire (multiple times per week, 30min-1h), was clearly noticeable. His most frequented web site during the experiment was Wikipedia (71%).

If a direct navigation to a headline was not possible, the user had to repeat a command (e.g. "down") multiple times. To avoid a large number of repetitions for the term "down", participant 4 tried to pronounce the term "down" lengthened. In Figure 4 the pronunciation of the term is illustrated by the pitch course. It was conspicuous, that during the whole search participant 4 was dominated by

Tab. 3. Detected Commands

Command Name	Description	Relative Freq.	Min[Participants]	Max[Participants]
P_{down}	Scroll down the web page	35,0%	11,11% [1]	46,15% [4]
R_{next}	Show next item in result list	16,0%	8,33% [1]	36,58% [5]
R_{search}	Search for new key word(s)	10,7%	9,89% [4]	14,63% [5]
R_{open}	Open the web page	10,7%	4,87% [5]	17,64% [3]
P_{close}	Close the web page	10,5%	4,87% [5]	17,64% [3]
P_{other}	Other (e.g. play a video)	4,0%	0,0% [1]	9,25% [2]
P_{up}	Scroll up the web page	3,8%	0,0% [2;3;5]	6,94% [1]
P_{ref}	Use a text hyper link	3,5%	0,0% [2;5]	15,27% [1]
R_{save}	Save web page shown in surrogate	2,3%	0,0% [2;4;5]	12,50% [1]
R_{prev}	Show previous item in result list	1,6%	0,0% [2;3;5]	8,33% [1]
R_{other}	Other (e.g. state a command error)	0,7%	0,0% [2;4;5]	2,77% [1]
P_{save}	Save web page	0,4%	0,0% [2;4;5]	1,47% [3]

straight forward interactions, i.e. commands like “back“ (in coverflow) or “up” on a website were very seldomly. If he wanted to enlarge a picture but also other pictures were on that side, he specified some characteristics of the picture to identify it: “‘picture of the world yellow enlarge”.

Participant 5 needed more adaptation of the system since he expressed questions without interpretable command terms or context. First, the pupil searched for turtles and opened a result page. Within the web page the next expression was “What do turtles eat?”. He repeated the question two times expecting the system to finally understand it. Here no command terms indicated the user’s wish to close the web site and start a new search. Thus the Wizard did no act. After that the term “crocodiles” was expressed, the Wizard did the first adaptation and searched for this term to avoid further user’s lonesomeness and frustration. Within a website about crocodiles the next expression was “What do crocodiles eat?”. Than the Wizard did the second adaptation and searched for “crocodiles”. The participants noticed that, was happy, smiled lightly and told the investigator, that “before he didn’t do it”.

In contrast to that example, participant 4 didn’t need any adaptation, but rather this one adapted his own commands to optimize the search: On the Wikipedia web pages this child enlarged pictures three times. The pictures were opened within the frame replacing the start page. After seeing a picture, the participant wished to return to the previous page, which was not possible directly due to technical reasons. Therefore, the system closed the frame, showed the result page and went back to the coverflow view. The participant tried different commands to achieve the desired result such as “go back‘“ and “a little bit back”. After three rounds, he understood the principle of the system and changed his strategy, accordingly. The pupil gave commands to close the current page and enlarge the same result page again to get to the start page.

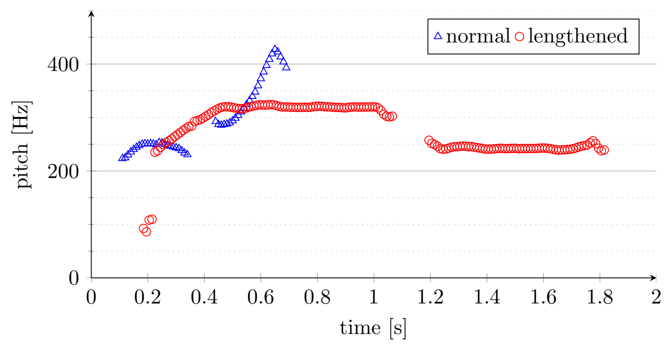

Fig. 4. Comparison of two pitch course for the same command (“runter”, engl. “down”), but with different speaking styles, normally spoken once again, lengthened word.

In the following a more detailed look to the occurrences of utterances implying an command is given. During the whole study 417 different utterances to apply an command in the search user interface were expressed by the pupils. Table 3 lists all expressed commands of all participants. The command name indicates the search user interface state (P = web page, R = result list, i.e. coverflow) and the succeeding interaction, for instance executing a new search in the result list R_{search}, scrolling down on a web page P_{down} or open a web page R_{open} (and change the state R to P). In the column “Relative Freq.” the probability for each detected command is given. In general slightly more commands were applied on state P (59,57%) in contrast to the state R (40,42%). A possible explanation is, that web pages are usually large and the user has to navigate/interact multiple times to get all desired information. Furthermore, web pages provide more possibilities to interact, like resize an image, play a video or click on a link.

The most applied command was P_{down} to scroll down on a web page. The characteristic behavior of the children here was to read a text passage, scroll down and continue to read. After scrolling forward within the search result page (16,0%), searching for new key word(s) (10,7%) is the third most frequent command. The frequent described loopy search behavior of children (e.g.: Bilal and Kirby (2002) and Gossen et al. (2011)) could not be confirmed in this study. The commands implying loopy behavior like P_{up} (3,8%) or R_{prev} (1,6%) have a low probability. One reason may lie in the type of search task. On the one hand, in exploratory search tasks, it is not necessary, to find specific facts, that may be missed in one of the last web pages. On the other hand (in contrast to a free search) the rough scheme of solving a task is given, that may prevent behavior where users stay or go back to distracting pages, that are no related to the topic. The commands to change between the states (R_{open} and P_{close}) are essential to switch between the main components of the search user interface and hence are necessarily used often. No child opened the logbook although the interactions was show in the introduction step. A possible explanation is, that the search focus was to solve the exploratory task and recent visited or saved pages do not contain new information to solve this task. Furthermore the save interaction was used only by participant 1 and 3.

In addition, the relative command frequencies for each participant were calculated separately. The last two columns in table 3 state the lowest and highest probabilities for each command, the corresponding participant(s) (in bracket) and hence indicate the range of variation. E.g. considering all commands of all pupils, the probability of R_{next} is 16,0%. A look at the individual frequencies reveals, that for participant 1 the probability of R_{next} is 8,33% (the minimum for all participants) and for participant 5 the probability of R_{next} is 36,58% (the maximum for all participants). Some of the minimum and maximum values need an additional explanation to avoid an overestimation of the range. To navigate on a web page, participant 1 used the commands P_{down} and P_{ref} with a similar frequency (11,11% and 15,27%). In contrast to this, the other pupils, used P_{ref} seldom. That is why, the range for P_{down} and P_{ref} is quite large. Furthermore due to technical reasons, some save commands R_{save} of participant 1 resulted into an error, which is why this commands was repeated several times and increase the maximum to 12,50%.

5.2 Audio Analysis

In the following section, we want to investigate the question – How often do acoustic characteristics such as off-talk occur during an exploratory search and how it evolve over time? As described in the previous section (5.1), we distinction two kind of statements: "commands" directly addressed to the system, all other statements are regarded as "off-talk". Thus, we are able to distinguish relevant human-machine interaction patterns from human-human interaction and furthermore estimate the amount of spoken content within this material. Table 4 lists the number and length of the labeled subject's utterances.

Tab. 4. Overview of the different children (P1 .. P5) utterances per experiment. Total denotes the total number of utterances including system responses and instructor. The recording time is given in mm:ss.

Subject	P1	P2	P3	P4	P5
Total	223	83	224	363	80
Command	72	54	68	182	41
Off-talk	103	09	131	71	26
Time	18:37	21:55	25:32	26:41	28:11

The total number of children utterances varied between 67 and 253, with a mean of 151.4 and a standard deviation of 83.8 turns. The average duration of turns was 1.82 seconds with an standard deviation of 1.58 seconds. The results in table 4 show that the children behave very individual. Participant 1 and 3 have used more off-talk than commands, the participants 2, 4, 5 behave exactly the other way around and have uttered more commands than off-talk. So one difficulty for a technical system would be to differentiate between commands, uttered to control the system, and off-talk uttered towards the instructor or to oneself, as both types could contain valid input to control the application. Therefore, these two kinds of utterances were analyzed phenomenologically based on the transcripts. Afterwards, an acoustic analysis was performed, which is described in the following.

As a first impression, we compared the length of utterances for both types of interaction on a inter-individually and a intra-individually base. The overall mean length for commands is 1.42s and for off-talk it is 2.31s. The duration difference between off-talk and command is significant in between all five subjects ($p < 0.001$) as well as intra-individually for participants 1, 3, 4, and 5 ($p < 0.001$). For participant 2, the duration of commands was quite similar to the off-talk duration ($\Delta = 0.37$) in comparison to the other subjects ($\Delta \approx 0.84$).

One reason, to choose children for further investigations is their higher motivation to make new experiences. They are not used to use a or talk to a technical systems in such a way adults are and thus see the interaction more as a game. This fact should be reflected in an increased verbosity, e.g. using politeness and personal pronouns. Furthermore, significant differences of verbosity should be seen between the beginning and end of the experiment.

Therefore, we calculated the average length of the ten first and the ten last commands to test this assumption. The inter-speaker analysis supports this assumption. The average command length at the beginning of the experiment is 1.71s (SD=0.83s) with differs significantly ($p < 0.01$) from the average length at the end of the experiment (M=1.25s, SD=0.56s). On the individual perspective, this observation is only true for the subjects 2 and 5, where a significant decrease of about 0.8s of the command length occurs during the experiment ($p < 0.025$, $p < 0.001$). For the other subjects we could not observe such a drastic decrease, which can be attributed to the fact, that these subjects used short commands right to the beginning of the experiment. Thus, these children should be encouraged to play with the voice interface. Additionally, the off-talk can be used for further analyses, as this does not change neither acoustically nor in duration during the experiment.

Unfortunately, the number and amount of material of this exploratory study is not sufficient for advanced automatic analysis. However, it can be noticed, that the energy level of both categories, command and off-talk does differ substantially. The talking style for off-talk is mostly like wispering, even if the investigator is asked. Thus, we assume that an automatically recognition of off-talk can be persuaded very easily. We furthermore noticed, that for controlling the system, mostly short often one-word commands are used, without any politeness word as "please" and "thanks". But we noticed an utilization of back-channel words as "hm" and "yes". These cues are exchanged among the interaction partners and used to signalize the progress of the dialogue (Allwood et al. (1992)). Hereby, especially the intonation of utterances transmits the communicative relation of the speakers and their attitude towards the current dialogue. As we found in previous studies (cf. Siegert et al. (2014)) that these cues are used as task-oriented signals, a further investigation and correlation with previous interaction pattern should be done.

5.3 User Preferences

We conducted a post-interview to evaluate the user's attitude towards the system and if they would use voice-controlled search user interface in the future. To the question whether the search was satisfying in general, four pupils answered yes. One child was not satisfied because not all information about the animal he needed were found. Since the learning component is an essential part of exploratory search tasks, we asked for newly learned facts. Four of the five pupils stated, that they had learned at least one new thing. Concerning the issues of conscious affects, two children stated, that they were delighted, as they saw animal pictures or videos (cheetah and giraffe). Their reasons were that these animals look pretty beautiful and by looking at videos it was "not necessary to read much". One child was surprised by learning that a crocodile swallows stones to dive.

All children stated, that they liked it very much to control the system via voice. As expected one of the main reasons was that the user does not have to write (difficult) words and thus the handling was faster. Another reason was the novelty to use such a system. All children also liked the ability of the system to speak (in case it was unable to understand a user's command). If the young users had used the bookmarking feature of the Knowledge Journey, than they grew to like it. The opinion was divided by asking, whether the voice controlled Knowledge Journey is better than the usual search engine or not. Two children stated that they like it more, two rated it both equally and one user preferred the common search engine more.

6. SUMMARY AND CONCLUSIONS

In this work we discussed the design and the results of a exploratory user study to analyze exploratory voice-controlled search of young users. The study provided insights into the search behavior of young users when using voice controlled search and helped to identify design issues of the setting. Due to the creativity of the young users, investigators have to define their role, tasks and freedom of action precisely. The discussed categories of expressions and situations give hints to create a framework of rules for the Wizard and investigator. The depicted interaction patters and acoustic characteristics can help to develop systems, that support the young user's interaction during information seeking. In future work, with more interaction data and a formalized interaction process, we will develop user models, that anticipate the next interaction steps of the user. Moreover, further analysis and user studies are necessary to investigate e.g. correlations between eye movements and interactions for children as well as build chains of interactions to increase the precision for the models. Furthermore the results can be used to develop classifier that recognize the users utterances and interpret the intended commands.

ACKNOWLEDGEMENTS

This work was supported partially by the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by the German Research Foundation (DFG).

REFERENCES

J. Allwood, J. N., and E. Ahlsn. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9(1):1–26, 1992.

A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. Desperately seeking emotions or: actors, wizards, and human beings. In *Proc. of the ISCA Workshop on Speech and Emotion*, pages 195–200, 2000.

A. Batliner, C. Hacker, S. Steidl, E. Nöth, M. Russell, and M. Wong. "You stupid tin box"- children interacting with the AIBO robot: A Cross-Linguistic Emotional Speech Corpus. In *Proc. of LREC*, pages 865–868, 2004.

D. Bilal and J. Kirby. Differences and similarities in information seeking: children and adults as web users. *Information processing & management*, 38(5):649–670, 2002.

C.L. Borgman, S.G. Hirsh, V.A. Walter, and A.L. Gallagher. Children's searching behavior on browsing and keyword online catalogs: the Science Library Catalog project. *JASIS*, 46(9):663–684, 1995.

R. Budiu and J. Nielsen. *Usability of Websites for Children: Design Guidelines for Targeting Users Aged 3–12 Years, 2nd edition.* Nielsen Norman Group Report, 2010.

G. Ceschi and K.R. Scherer. Children's ability to control the facial expression of laughter and smiling: Knowledge and behaviour. *Cognition & Emotion*, 17(3):385–411, 2003.

A. Cullen and N. Harte. Feature sets for automatic classification of dimensional affect. In *Proc. of the 23nd IET Irish Signals and Systems Conference*, pages 1–6, Maynooth, Ireland, 2012.

T. Gossen and A. Nürnberger. Specifics of information retrieval for young users: A survey. *Information Processing & Management*, 49(4):739 – 756, 2013.

T. Gossen, T. Low, and A. Nürnberger. What are the real differences of children's and adults' web search? In *Proc. of the 34th Int. ACM SIGIR Conf. on Research and development in Information*, pages 1115–1116. ACM, 2011.

T. Gossen, M. Nitsche, and A. Nürnberger. Knowledge journey: A web search interface for young users. In *Proc. of the Sixth Symposium on Human-Computer Interaction and Information Retrieval (HCIR 2012).* ACM, 2012.

T. Gossen, M. Kotzyba, S. Stober, and A. Nürnberger. Voice-controlled search user interfaces for young users. In *7th annual Symposium on Human-Computer Interaction and Information Retrieval*, 2013a.

T. Gossen, M. Nitsche, J. Vos, and A. Nürnberger. Adaptation of a search user interface towards user needs - a prototype study with children & adults. In *Symposium on Human-Computer Interaction and Information Retrieval (HCIR'13)*, 2013b.

M. Grimm and K. Kroschel. Evaluation of natural emotions using self assessment manikins. In *IEEE Workshop on Automatic Speech Recognition and Understanding*, pages 381–385, 2005.

K. Gyllstrom and M.F. Moens. Wisdom of the Ages: Toward Delivering the Children's Web with the Link-based AgeRank Algorithm. Proc. of the Int. Conf. in Information and Knowledge Management, 2010.

A Hassan, R. Damper, and M. Niranjan. On acoustic emotion recognition: Compensating for covariate shift. *IEEE Trans. Audio, Speech, Language Process.*, 21: 1458–1468, 07 2013.

H.B. Hutchinson, B.B. Bederson, and A. Druin. The evolution of the int. children's digital library searching and browsing interface. In *Proc. of the 2006 conf. on Interaction design and children*, pages 105–112. ACM, 2006.

M. Jansen, W. Bos, P. van der Vet, T. Huibers, and D. Hiemstra. TeddIR: tangible information retrieval for children. In *Proc. of the 9th Int. Conf. on Interaction Design and Children*, pages 282–285. ACM, 2010.

H. Jochmann-Mannak, T. Huibers, L. Lentz, and T. Sanders. Children searching information on the Internet: Performance on children's interfaces compared to Google. *SIGIR'10 Workshop on accessible search systems*, pages 27–35, July 2010.

Y. Kammerer and M. Bohnacker. Children's web search with google: the effectiveness of natural language queries. In *Proc. of the 11th Int. Conf. on Interaction Design and Children*, pages 184–187. ACM, 2012.

G. Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

V. Nesset. Two representations of the research process: The preparing, searching, and using (PSU) and the beginning, acting and telling (BAT) models. *Library & Information Science Research*, 35(2):97–106, 2013.

R. W. Picard. *Affective Computing*. MIT Press Cambridge, 2000.

D. Prylipko, D. Rösner, I. Siegert, S. Gnther, R. Friesen, M. Haase, B. Vlasenko, and A. Wendemuth. Analysis of significant dialog events in realistic humancomputer interaction. *Journal on Multimodal User Interfaces*, 8: 75–86, 2014.

T. Schmidt and K. Wörner. Exmaralda creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics*, 19(4):565–582, 2009.

B. Schuller, A. Batliner, S. Steidl, and D. Seppi. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Commun*, 53:1062–1087, 11 2011.

F. Schulz von Thun. *Miteinander reden 1 - Störungen und Klärungen*. Rowohlt, Reinbek, Germany, 1981.

I. Siegert, R. Böck, D. Philippou-Hübner, B. Vlasenko, and A. Wendemuth. Appropriate Emotional Labeling of Non-acted Speech Using Basic Emotions, Geneva Emotion Wheel and Self Assessment Manikins. In *Proc. of the 2011 IEEE Int. Conf. on Multimedia & Expo*, Barcelona, Spain, 2011. IEEE.

I. Siegert, R. Böck, and A. Wendemuth. Inter-Rater Reliability for Emotion Annotation in Human-Computer Interaction – Comparison and Methodological Improvements. *Journal of Multimodal User Interfaces*, page s.p., 2013. online.

I. Siegert, D. Prylipko, K. Hartmann, R. Böck, and A. Wendemuth. Investigating the form-function-relation of the discourse particle "hm" in a naturalistic human-computer interaction. In S. Bassis, A. Esposito, and F.C. Morabito, editors, *Recent Advances of Neural Network Models and Applications*, volume 26 of *Smart Innovation, Systems and Technologies*, pages 387–394. Springer, 2014.

A. Wendemuth and S. Biundo. A Companion Technology for Cognitive Technical Systems. In *Cognitive Behavioural Systems*, volume 7403 of *LNCS*, pages 89–103. Springer, Berlin, Heidelberg, 2012.

R.W. White, M. Bilenko, and S. Cucerzan. Studying the use of popular destinations to enhance web search interaction. In *Proc. of the 30th annual Int. ACM SIGIR Conf. on Research and development in information retrieval*, pages 159–166. ACM, 2007.

S. Yildirim, S. Narayanan, and A. Potamianos. Detecting emotional state of a child in a conversational computer game. *Computer Speech & Language*, 25(1):29–44, 2011.