

SPECIAL ISSUE ARTICLE

Misalignment in Semantic User Model Elicitation via Conversational Agents: A case study in navigation support for visually impaired people

Jakub Berka^a, Jan Balata^a, Catholijn M. Jonker^{b,c}, Zdenek Mikovec^a, M. Birna van Riemsdijk^d, Myrthe L. Tielman^b

^aCzech Technical University, Prague, Czech Republic; ^bDelft University of Technology, Delft, the Netherlands; ^cLeiden University, Leiden, the Netherlands; ^dUniversity of Twente, Enschede, the Netherlands

ARTICLE HISTORY

Compiled January 28, 2022

ABSTRACT

Disabled people can benefit greatly from assistive digital technologies. However, this increased human-machine symbiosis makes it important that systems are personalized and transparent to users. Existing work often uses data-oriented approaches. However, these approaches lack in transparency and make it hard to influence the system's behaviour. In this paper, we use knowledge-based techniques for personalization, introducing the concept of Semantic User Models for representing the behavior, values and capabilities of users. To allow the system to construct such a user model, we investigate the use of a conversational agent which can elicit the relevant information from users through dialogue. A conversational interface is essential for our case study of navigation support for visually impaired people, but in general has the potential to enhance transparency as users know what the system represents about them. For such a dialogue to be effective, it is crucial that the user understands what the conversational agent is asking, i.e., that misalignments which decrease the transparency are avoided or resolved. In this paper, we investigate whether we can use a conversational agent for Semantic User Model elicitation, which types of misalignments can occur in this process and how they are related, and how misalignments can be reduced. We investigate this in two (iterative) qualitative studies (n=7 & n=8) with visually impaired people in which a personalized user model for navigation support is elicited via a dialogue with a conversational agent. Our results show four hierarchically structured levels of human-agent misalignment. We identify several design solutions for reducing misalignments, which point to the need for restricting the generic user model to what is needed in the domain under consideration. With this research we lay a foundation for conversational agents capable of eliciting Semantic User Models.

KEYWORDS

Conversational agent ; User model ; Values ; Semantic user model misalignment ; Visually impaired ; Navigation support system

1. Introduction

Computers are increasingly getting smarter, leading to an increase in their prevalence in and influence on our daily lives. This increased symbiosis between people and these

systems also gives rise to the need for systems which understand their specific users (i.e., are personalized), and which are in turn understandable to these users (i.e. are transparent) (Stephanidis et al. (2019)). We need the first so that the system can make decisions which truly fit with what the individual wants, and the second so the individual can understand, assess, and influence these decisions. Whenever a system gives advice or makes decisions for users, explainability and transparency are important to allow for responsible use of such systems (IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2017)).

Existing approaches for user modelling and personalization often use data-oriented approaches in which user preferences are inferred from users' previous behaviour (see, e.g., (Georgiou and Demiris (2017))). Data-oriented approaches, however, typically lack in transparency due to the complex relation between the input data and a model's output (Dignum (2017)). This lack makes it not only difficult for users to understand the system, but also to influence its behaviour. Moreover, by its very nature data-oriented approaches are designed for majority groups, ignoring the outliers and the individual uniqueness of minority groups (e.g., disabled people Wald (2021)). In the context of navigation support, specifying user route preferences based on generic low-level characteristics such as route length, type of crossings, etc., does not always work well, since preferences for route alternatives are influenced by the context of route segments (Balata, Mikovec, and Slavik (2018)). Thus, it should be possible to specify higher-level route choices. Moreover, data typically lacks the "why", i.e., the underlying values such as independence or safety, behind the decision, which means that users cannot interact with the system about the reasons behind their choices. Finally, the data is about the past, which means it does not suffice if people want the system to support them in changing their future behaviour.

Therefore, in our work we investigate a different approach to personalization which makes use of knowledge-based techniques (Brachman and Levesque (2004)). In this approach, we explicitly model the user's (desired) behavior such as the location they want to go to and the various route options, their values, and user capabilities.¹ This approach facilitates transparency and explainability, because the system can make its reasoning explicit (Harbers (2011)). For example, a navigation app offering advice would be able to explain that reasoning to a user when asked (e.g., I would go right here, as this route has fewer crossings and is therefore safer).

We refer to these types of models as *Semantic User Models* since they explicitly capture the meaning of the represented concepts through their interrelations, as in the work on ontologies for knowledge representation and semantic web (Lüttich, Mossakowski, and Krieg-Brückner (2004)). For example, a Semantic User Model can capture the relation between a route option and the supported value such as safety. In previous work, we have developed the formal and semantic foundations of these models (Kließ, Stoelinga, and van Riemsdijk (2019); Tielman, Jonker, and van Riemsdijk (2018)). In this paper, we investigate for the first time the elicitation of these models for users, and the misalignments between the system and users that might occur when users interact with the system about what should be represented.

As the process of eliciting necessary information from the user for constructing a Semantic User Model leads to the need of increasing the accuracy in an iterative way, a *conversational agent* seems to be an appropriate choice for user interface implementation. The use of a conversational agent to ask what it wants to know in order

¹Eventually data-oriented and knowledge-based approaches may be combined, for example by establishing a baseline through a data-oriented approach as the starting point for creating a knowledge-based user model. However, in this paper we focus on the knowledge-based aspects.

to model the user in the system also has a potential benefit regarding transparency. It allows users to immediately get an understanding of what knowledge the system is basing its decisions on. User model elicitation is a unique opportunity to already give the user insight into what is important to a system. However, this only truly works if the communication between the user and the system during the elicitation is free of misunderstandings. Being open about what you wish to know as a system will only lead to true transparency if the user also understands you. Therefore, to achieve transparent communication, the detection and avoidance of misalignment is crucial.

Thus, we focus on improving our understanding of which *misalignments* can occur in the conversation between a user and agent to endanger transparency. Moreover, we wish to better understand whether these misalignments are related to potentially shed light on how we could address them.

In summary, we address the following research questions:

RQ1: Which types of human-agent misalignment occur when using a conversational agent to elicit a Semantic User Model describing the navigation behaviour options and corresponding capabilities and values of a traveller?

RQ1.1: Is there a relationship between these different types of misalignments?

RQ2: How can a conversational agent be designed to reduce human-agent misalignments when eliciting a Semantic User Model?

A specific dialogue strategy was created and implemented in the conversational agent for eliciting the Semantic User Model. The model and dialogue strategy are generic frameworks for representing and eliciting user behaviour, facilitating the application of this work to other behaviour support domains in follow-up research as part of our broader research aims (van Riemsdijk, Jonker, and Lesser (2015)). We use the conversational agent to perform a qualitative user study in two iterations in which we let the participants have a conversation with our agent to explore if a Semantic User Model can be elicited in this way, and which misalignments occur in the process. In the first iteration (running in Czech language), an experimenter acts as a speech-to-text subsystem (Wizard of Oz technique) of the conversational agent by transcribing the utterances of the participant into the dialogue system. The main reason of the usage of Wizard of Oz technique was the low reliability of the speech-to-text systems for Czech language available. In the second iteration running in English, we removed the Wizard and used an automated speech-to-text subsystem. In addition, in the second iteration, we made a number of adjustments to the conversational agent based on insights from the first iteration.

Section 2 introduces the case study of navigation support for visually impaired people, the assumptions this work is based on, and their grounding in the literature. Section 3 describes the combined methods of the two experiments that were done, as well as the formal structure of the user model and the dialogue that is based on it. Section 4 presents the results of the first study, Section 5 outlines what changes were made based on this first iteration to improve the dialogue, and Section 6 outlines the results from the second study. Together, these studies show that a conversational agent can be used to elicit Semantic User Models, but that misalignments can be found in distinct places, and that these types of misalignment are indeed connected. The paper ends with a discussion and our conclusions in Section 7.

2. Background and Motivation

In this section, we give more background on the case study we use in this paper, and we provide background and motivation for the assumptions underlying this paper.

2.1. Case Study: Navigation Support for Visually Impaired People

A suitable case study we were looking for had to fulfill several conditions. The user’s problem that should be solved must be complex with a high need of personalization. The target group should be represented by frequent users of ICT and speech interfaces in particular to avoid the unfamiliarity of the conversational agent. The use case chosen should be well covered by previous research studies describing behavior patterns, user needs and desires, capabilities and limitations to see if the model created is in alignment with the reality.

From our knowledge and previous experience, the case study ”Navigation Support for visually impaired people” can suit well to above mentioned conditions. For visually impaired people, the technology plays a very important role in allowing them to navigate more independently (Balata et al. (2018)) as *visual impairment* limits travelling and mobility capabilities (Golledge (1993); Wycherley and Nicklin (1970)). So visually impaired people either do not travel at all (White and Grant (2009)) or travel mostly on well known routes (Golledge (1999)), even though they often wish to travel more on unknown ones (Tuttle and Tuttle (2004)). *Navigation aids* have the potential to improve this situation. An important aspect of developing such navigation aids in an effective manner, is the incorporation of personalization in the route calculation and presentation. Personalization needs to be done regarding new parameters specific for this setting, such as safety, difficulty of crossings, etc. (Bujacz, Baranski, Moranski, Strumillo, and Materka (2008); Faria, Lopes, Fernandes, Martins, and Barroso (2010); Völkel and Weber (2008)).

Visually impaired people use ICT daily (especially smartphones) for various activities such as magnifying objects, visual search, text recognition, pedestrian navigation, etc. They are used to speech user interface and advanced speech assistants like Siri on iOS.

Personalization is important for visually impaired people because of large differences among them regarding navigation behaviour, capabilities, and personal values (Ahmetovic, Guerreiro, Ohn-Bar, Kitani, and Asakawa (2019); Guerreiro, Ohn-Bar, Ahmetovic, Kitani, and Asakawa (2018); Ohn-Bar, Guerreiro, Kitani, and Asakawa (2018)). For instance, unlike people navigating with a guiding dog, only those navigating with a white cane need specific information to find a crossing. Moreover, in unfamiliar places, the level of detail of environment description needs to be higher to lower the stress level, while in familiar places just brief information is sufficient to save time and preserve the efficiency of navigation. Some people would cross the street only at crossings with traffic lights, while others will never cross the street where there is tram traffic.

2.2. Assumptions

Modelling Values

The first assumption we make is that it is desirable to model user values, and that values can be modelled in a formal user model. The choice for values is based on

the premise that to properly personalize any type of technological support, one first needs to understand what is important to users (van Riemsdijk et al. (2015)). Values are useful to achieve this understanding, as they are concepts that represent the criteria used by people to choose what to do, and to evaluate people and events (Friedman, Jr., and Borning (2006); van de Poel (2015)). A first step towards understanding the values of visually impaired people was taken by Azenkot et al. (Azenkot et al. (2011)), who for instance identified independence, confidence and safety as important. Moreover, values are increasingly also represented explicitly in technology itself, for instance to make people aware of their environmental behavior (Haller et al. (2017)), to decide between norms (Kayal, Brinkman, Neerincx, and Riemsdijk (2018); Serramia et al. (2018)), or to choose behavior plans for agents (Cranefield, Winikoff, Dignum, and Dignum (2017)). In order to model values such that they can be used by a system for run-time personalisation, they need to be connected to possible behaviour choices. In the context of navigation these choices are travelling activities, for instance different possible routes. Pasotti, Jonker, and van Riemsdijk (2017) present a framework modelling activities in hierarchies, based on how people conceptualize them themselves. Pasotti, van Riemsdijk, and Jonker (2016) and Tielman et al. (2018) show how values can be included in this framework. Through linking values to activities in a hierarchy, they specify what values are promoted or demoted by a certain choice. Such hierarchies of activities with values form the base of our formal user model for visually impaired travellers.

User Model Elicitation

Our second assumption is that explicit information elicited from users should be used to build the user model. One important reason for this assumption, is that personal values cannot be learned from current behavior. After all, if the user’s behavior would represent exactly the behavior most aligned with their values, they would not require support from a system to do additional or different things.

Moreover, values are abstract concepts which do not always carry the same meaning to all people. This is the reason why many methods using values involve directly talking to prospective users in the design phase (van de Poel (2013)). Another important reason to elicit information from users is to ensure the transparency of the system. The user should know and be able to control what the system knows (IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2017)), which is realized by having explicit conversations about its knowledge.

Human-Agent Alignment

The third assumption that we make, is that building a user model for a complex task through conversation with a system is difficult and that to succeed, the user and system need to be aligned in their understanding of their conversation. Conversational agents or intelligent assistants are increasingly deployed in households and are present in our smartphones (Emarketer (2017)). However, these systems still have many limitations to be solved (Laranjo et al. (2018)). Human-machine voice interaction is fundamentally different from human-human interaction, demonstrably so in the ways in which responses from the device do not necessarily coherently follow the input (Porcheron, Fischer, Reeves, and Sharples (2018)). And agents typically lack knowledge of the surrounding environment (Sciuto, Saini, Forlizzi, and Hong (2018)). Moreover, not many conversational agents can support users in complex tasks (Vtyurina (2019)). It is es-

pecially in such complex tasks, that all involved need to have a shared mental model of the environment, task and their role (Converse, Cannon-Bowers, and Salas (1991)).

This concept of shared mental models is useful in modelling any task where multiple parties need to collaborate (Fan and Yen (2010); Scheutz, DeLoach, and Adams (2017)) including complex dialogues (Abdulrahman, Richards, Ranjartabar, and Mascarenhas (2019); Faulkner, Niekum, and Thomaz (2018)). At its core, having a shared mental model in conversation means that the user and the agent are talking about the same things, i.e. that there is no misalignment. Avoiding misalignment is also crucial to truly achieve transparency, as there is a need not just for the system to share what it wishes to know, but also for the user to understand this. To achieve this transparency, the first step is to better understand where and how possible misalignment between the conversational agent and user can occur.

3. Experiments

Two studies were done to investigate whether we can use a conversational agent for eliciting Semantic User Models, and to better understand where misalignment could occur.

The task participants were asked to perform was to have such a conversation with an agent, which asked them about their navigation behaviour, capabilities, and underlying values. Considering the novelty of the task of eliciting Semantic User Models via a conversational agent, we did not know what shape the misalignments would take. Therefore, both qualitative measures (interview) and data about the final user model and usability were gathered. Two studies were done, the first building on the second, so that initial findings could be incorporated in the design of the second. This approach facilitated an initial exploration of design solutions for addressing misalignments.

Both studies followed a very similar methodology, so Section 3.1 presents the methods for both studies. For the second study, changes were made to the user model and dialogue structure in particular, to resolve some of the misalignment issues from the first study. Whenever changes were made from the first to the second study, this is noted explicitly. The experimental methods are presented first, the structure of the mental models (and therefore the dialogues) are presented second. Both studies were approved by the Ethics Committee of Delft University of Technology (nr. 423), and included an informed consent procedure.

3.1. Methods

3.1.1. Participants

All participants had a visual impairment, were native Czech speakers and were recruited via e-mail leaflet. Sample sizes in qualitative studies are often around 10 participants. In our case, we settled for a minimum of 7 per study, as we prioritized recruiting participants in our specific user group (so with a visual impairment) above a larger sample size. The first study had seven participants (1 female, 6 male), age 32 to 69 ($mean = 41.57, SD = 13.00$). The second study had eight participants (3 female, 5 male), age 20 to 68 ($mean = 40.38, SD = 10.03$). In the first study, four participants had category 5 visual impairment (VI)² and three participants had cate-

²See WHO (2009) for definitions of categories of visual impairment

gory 4 VI. Three participants were late blind, and four participants were congenitally blind. In the second study, five participants had category 5 VI; two participants had category 4 VI, 1 participant had category 3 VI. Four participants were late blind and four participants were congenitally blind.

3.1.2. Measures

A number of different measures were used to identify misalignments between the agent and the user. In general, these measures can be split into observations from the experimenter and observations/opinions from the participant. We included both as the experiment leader might not always realize when the participants were confused, and participants (due to not knowing what the agent expected) might not always realize they misunderstood something. Finally, we also looked at the resulting information in the system, as this might also uncover hidden misalignment.

The way these measures were taken was as follows. Firstly, the experiment leader noted whenever misunderstandings occurred during the conversation between agent and user, by keeping track of all situations in which confusion arose or was expressed; or when the user gave an unexpected answer. And after the elicitation process, the participants were asked in an interview to 1. indicate whether they understood the concepts the agent asked them about (see Section 3.2) and 2. if they felt the final information represented by the system was correct. Some statistics were gathered about the final user models to provide extra insight (for instance, in some cases (e.g. for values) the user could say *none* as answer, how often this was done and where). Finally, some measures regarding usability were gathered, as the user understanding the system is a major part in usability. Low usability, therefore, could indicate more misalignment. To measure general usability, we choose to use system usability scale (SUS) (Brooke (1996)) as it can generate reliable results even on small sample sizes, and we were able to compare our results with the baseline (i.e. with other systems). Additionally, the participants filled two questions about the level of naturalness and confusion about the dialogue on the 5 point Likert scale (1 to 5). These questions were added as the SUS is not about dialogue in particular. The experiment leader also noted anything that the participant said or what was noticed regarding the usability during the sessions.

3.1.3. Procedure

Both studies lasted about 1.5 hour and started with a welcome and brief instructions. The consent form was read out loud and consent was recorded on audio with participant name and date. The session itself consisted of two phases. The first phase was the dialogue with the conversational agent. For the first study this phase lasted a maximum of 45 minutes, in the second there was no time limit as changes in the dialogue structure made the limit obsolete (see Table 1 for details). In both studies, the second phase of the experiment consisted of firstly answering the SUS and additional questions, and then an interview where participants were asked whether they understood the concepts, and whether the information in the final user model was correct.

3.1.4. Data preparation and qualitative analysis

The qualitative remarks written down by the experimenter were analyzed keeping in mind the concept of misalignment. We wanted to find all situations in which the user and agent did not understand each-other, replied unexpectedly to the other or were

not talking about the same thing. With a group of 3 researchers, different categories of misalignment were established based on the written down initial observations from the experimenter. During this process, the 3 researchers first each individually came up with their own categories. These results were then compared, and refined until agreement was reached on what final categories to use. As a next step, the researchers then considered all remarks which indicated a misunderstanding and misalignment between the system and user, and individually categorized the remarks of all participants using the previously established categories. This categorization was performed by two people for the first and three for the second study. After the individual coding, the categories of misalignment were further refined and redefined based on that coding. During this process, all inconsistencies between coders were discussed and the categories were re-defined based on these discussions. Finally, all remarks were re-categorized by the individual coders, and as a last step any remaining differences in opinion were discussed and resolved between the coders. This process resulted in a list of remarks for indicating the different types of misalignment.

3.2. User model & Dialogue structure

Creating an agent that elicits a user model via dialogue requires two main components: a structure to represent the user model, and the dialogue for eliciting that user model. Both of these two main components are presented in this section (Section 3.2.1), as well as the way they were implemented (Section 3.2.2).

3.2.1. Modelling

The user model has three main components: hierarchically structured actions representing user behaviour options, related user values that are promoted or demoted by these actions, and user capability regarding the execution of actions.

The core of the user model is a hierarchical tree structure representing the user’s activities by means of two types of relationships following the work presented in Kließ et al. (2019); Pasotti et al. (2017, 2016): (i) a relationship where one action is a more concrete or specific way of doing another action, for instance *going to work by bus* is a more concrete way of *going to work*, and (ii) one action being a smaller part of doing the other action, for instance *get bus ticket* is a part of *going to work by bus*. All actions are nodes in such a hierarchy, and a node always only has one type of children, so either ways-of or parts-of. Moreover, the layers of the tree alternate between ways-of and part-of, starting with ways-of. The choice to alternate was made to simplify the structure, so people could learn what to expect. So if the root is *going to work*, its children are ways of going to work, for instance *going to work by foot* and *going to work by bus*. And the children of those actions are parts of, for instance *get bus ticket* for *going to work by bus* and *cross street* for *going to work by foot*.

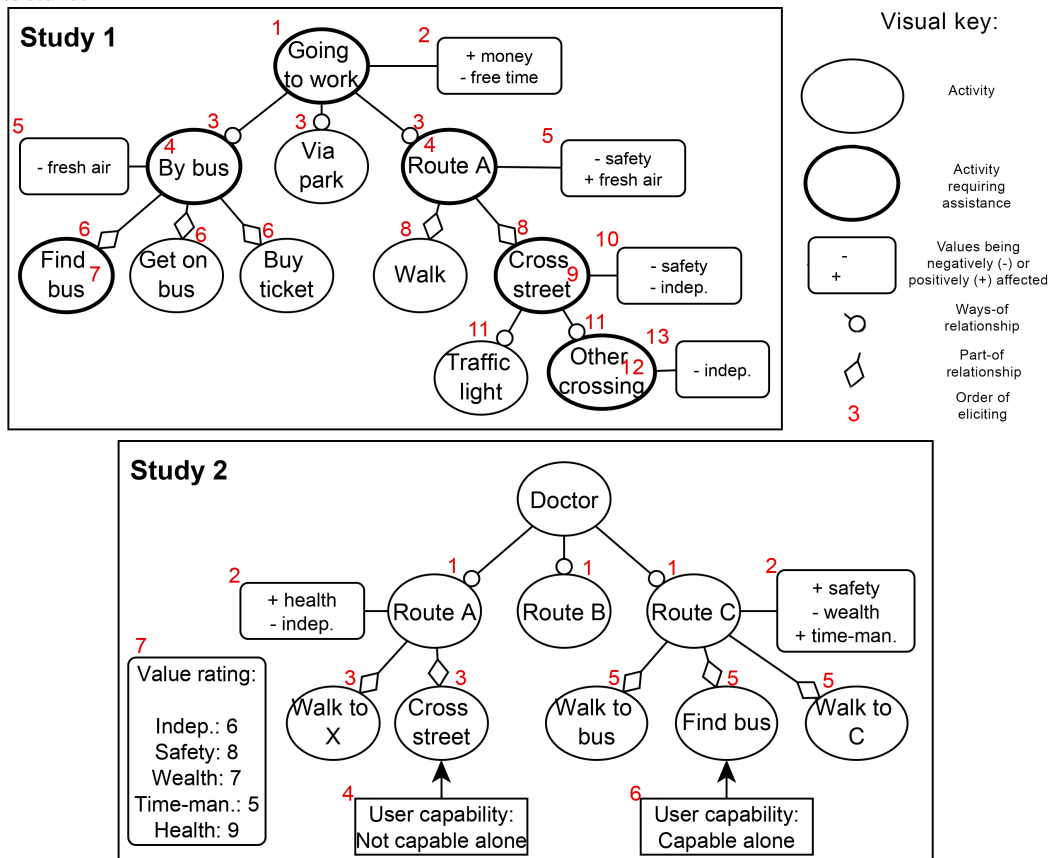
Personal values are included via their relation to actions as presented in Pasotti et al. (2016) and Tielman et al. (2018): so one or more values relate to an action. This relation indicates that an action either demotes or promotes a value. An action can be related to none, one or multiple values. The goal of identifying these relationships is that the agent has a better notion of which of the user’s possible actions are most in line with their values.

Finally, the user model contains information on the user’s capability, as it is relevant for a support agent to know where that support is most needed.

This basic structure is the same for both studies. The dialogue follows the structure

of the action tree. That is, it starts with a top action (the root) and elicits the children of this root, then values of the children, then the next layer, etc. Although the main concepts and structure remained consistent, some changes were made to the model after the first study based on its results. The motivation for the changes between study 1 and 2 is discussed in section 5. The specifics of the user model for both studies (including their differences) is shown in Table 1. Figure 1 shows a graphical representation of an example tree structure of the user model for both studies, including the exact order in which the information is elicited during the dialogue.

Figure 1. A graphical representation of possible tree structures (user models), and the order of elicitation for both studies. Ovals represent activities, rounded off cornered boxes the information on values, and sharp cornered boxes additional information on user capability. Arrows with a circle indicate a ways-of relationship, arrows with a diamond a part-of relationship. For the first study, activities could be marked as requiring assistance, here shown with bold edges. For the second study, only those part-of activities requiring assistance were elicited, so those can automatically be marked as such. The red numbers represent the order in which the information was elicited. In case of a number inside a box, this represents the question what activities require assistance.



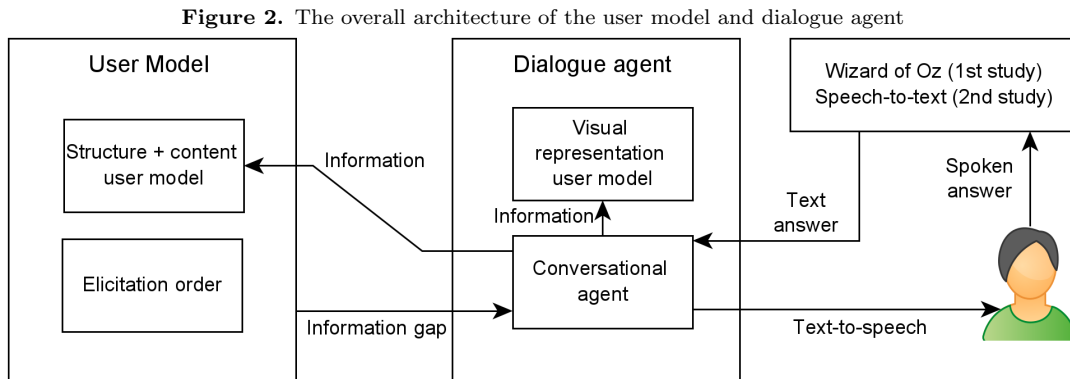
3.2.2. Implementation

The conversational system was implemented with a dual structure, one part implementing the data structure of the user model, the other the dialogue agent for eliciting the information in the user model from people. Figure 2 shows the overall architecture

Table 1. Table describing the specifics of the user model and dialogue in both studies.

| | Study 1 | Study 2 |
|--------------------|--|---|
| User model | | |
| Assistance | Both part-of and ways-of are either marked as 'requiring assistance' or not | Implicit in the model, only parts-of requiring assistance are elicited. |
| User capability | Only included in terms of 'requiring assistance' | Specific user capability marked when the (future) navigation system can't offer full assistance |
| Value-activity | Values for all activities requiring assistance | Values are elicited for all ways-of only |
| Value options | Anything can be a value, the term the user gives is put into the system | Predefined list of 7 items (see section 5) |
| Value rating | Not included | Values are rated from 1 to 10 on importance |
| Dialogue | | |
| Answer input | Participant speaks answer, Wizard types answers into the system | Speech-to-text recognizes answers of the participants |
| Top activity | User is asked for 'travel related activity where assistance is helpful'. This is the top. | Two activities are given, going to the doctor and to a restaurant. For both a tree is elicited. |
| Tree depth | Not specified. The user determines when to stop, or when time (45min) runs out. | 1 layer of way-of activities (routes) are elicited, and then for every route the parts-of. |
| Next layer in tree | The next layer is asked only for those activities where assistance would be helpful | The next layer is asked for all activities in present layer. |
| Activity input | The user is asked to list all ways of or parts of simultaneously. | The user is asked to give the first way/part of, then the second, etc. |
| Value input | The user is asked to list all relevant values. Then per value to indicate if it is positive or negative. | The user is asked to give one value, then if it is positive or negative. Then the second, etc. |
| Tutorial | No | Yes |

of the system. The code can be found on GitHub³.



The structure of the user model was implemented in the agent-language GOAL⁴. This part also kept track of what information needed to be elicited next given the current state of the user model. This ‘information gap’ was then communicated to the dialogue agent.

The dialogue agent visualized the information in the user model for the experimenter in their interface, transformed the information gap to a natural language question and dealt with conversational aspects such as repetitions, grounding and extra examples.

As a grounding strategy, the system presented summaries to the user of how the system interpreted the activities, values mentioned by the user, and provided more specific descriptions (*concretisations*). This strategy was used to confirm whether the agent understood the answer correctly, and to present the possibility for the user to correct it if not. In the second study as the dialogue structure changed grounding strategy was slightly changed as well, see the excerpt of the dialogue from first study in Table 2 and from second study in Table 3. Whenever the dialogue agent had the final answer from the user about a new part of the user model, it would be sent back to the corresponding module. The dialogue agent was implemented in Java 8 SE. The answers were sent to the IBM Watson Assistant⁵ service, which is used for natural language understanding. The voice of the dialogue agent was generated with Mac OS VoiceOver functions.

For the first experiment, conducted in Czech language, the system was designed to work with a Wizard of Oz set-up, where an experimenter (the wizard) was acting as a perfect speech-to-text system, listening to the user, and transcribing the utterances into the dialogue agent without any modifications. Therefore, the dialogue agent interface also included space for this text input. The experimenter (wizard) was in the same room as the participant. The main reason for usage of Wizard of Oz setup was to avoid problems with very low reliability of speech-to-text systems for Czech language available. The second experiment was conducted in English, and the participants’ utterances were recognized using automated speech-to-text system (STT) so the participant could directly talk to the system. For transcription of the user speech to text, we used a Google Cloud STT API.

³<https://github.com/CoreSEAP>

⁴<https://goalapl.atlassian.net/wiki/spaces/GOAL/overview?mode=global>

⁵<https://www.ibm.com/watson/ai-assistant/>

Table 2. Example of the dialog between the user (U) and the conversational agent (A) from first study, specifically when the agent is asking for values.

Agent Which values does the activity searching for a tram stop promote or demote? I will ask later if they are negative or positive. First you can just mention the values.

User Independence.

Agent Is the value independence influenced by doing the activity searching for a tram stop?

User Yes

Agent Great. Now I will ask for each value, if it is positively or negatively influenced by doing of the activity searching for a tram stop. Does doing the activity searching for a tram stop. influence the value independence positively or negatively?

User Positively

Agent Is the value independence influenced positively?

User Yes

Agent To be sure, I'll repeat the values influenced by doing the activity searching for a tram stop. The value independence is influenced positively. Is that right?

User Yes

Table 3. Example of the dialog between the user (U) and the conversational agent (A) from second study, specifically when the agent is asking for values.

Agent Now I want to know, which of your values would be positively and also negatively supported when you choose route: by tram to your favourite restaurant. You can choose values in this list: comfort, health, independence, safety, socializing, time-management, and wealth ,or say 'no value' if it's none of them. What is the first value that is positively supported when you choose route: by tram to your favourite restaurant?

User Time-management

Agent Ok, time-management is positively supported. Is it correct?

User Yes.

Agent Great, you can say: that's all, when you are done. What is the next promoted value?

User That's all.

Agent To be sure, I will repeat it. Promoted values is: time-management, is it correct?

User Yes

Agent Great. What is the first value that is negatively supported when you choose route: by tram to your favourite restaurant?

User Repeat the list of values

Agent You can choose values in this list: comfort, health, independence, safety, socializing, time-management, and wealth. You can also say none or no value.

4. Results Experiment 1

In this section, the results from the first experiment are discussed. In the next section, we will discuss the changes that were made based on these results for Experiment 2, and the results of Experiment 2. Throughout the presentation of results, we will summarize our findings through numbered and named misalignments. This summary will make it easier to get an overview of our findings, and connect them with design solutions we propose for Experiment 2.

4.1. Misalignment

During the qualitative analysis, three different categories of misalignment arose. Misunderstandings occurred because users did not understand the general structure of the dialogue and user model; because users were confused about concepts, and when users misunderstood how to talk to the system. In the following subsections, we present these three categories in more detail. All mentioned participant and experimenter comments can be found in Berka et al. (2020), the Px references refer to participant numbers.

4.1.1. Misunderstanding the general structure

This category of misalignment between system and user sprung from the user not understanding the overall structure of dialogue and user model. For instance, some users indicated losing context within the structure (e.g. P6), or confusion in how concepts relate (e.g. values to actions P5, or how many child actions could exist P4). Some of this misalignment can also be found in the statistics about the user models, as seen in Berka et al. (2020). Firstly, there is a high number of nodes with one child (i.e. so called ‘1 branches’), which is unexpected. If an action only has one way of (or part of) child, further specification usually does not make sense because it does not add extra information and the branch should stop instead of a single child being added. That these ‘1 branches’ occurred might be because people felt forced to answer, or did not understand the hierarchy could end. Adding to this confusion might be that people often started with very specific actions, which meant that the questions asking for even more specifics stopped making sense very early on. This is also reflected in the relative shallowness of the structures.

Misalignment 1. [Relations between concepts in the user model] There were misunderstandings about the relations between values and actions, and it was difficult to keep track of the relations between actions (context).

Misalignment 2. [Abstraction level of actions] There were misunderstandings about the abstraction level at which actions in the user model should and could be specified.

4.1.2. Misunderstanding the concepts

The second point of misunderstanding lay in the concepts that were used. The conversational agent talked about *values*, and about more specific description *concretisations* and *parts-of* regarding the actions. Regarding the actions, the concretisation question was particularly difficult to understand for many participants. They also regularly answered with a part of an activity, instead of a way of. The other main point of confusion was understanding the concept of values. Many participants had difficulties to grasp exactly what was meant by the word ‘values’, and how they could be promoted or

demoted by actions. This lack of understanding of values might have also contributed to the frequent occurrence of an inverse relationship, where values influenced actions. For instance, when ‘presence of a friend’ would be named as a value positively influenced by the action ‘swimming’, whereas from the dialogue analysis it emerged that the person probably meant that the action ‘swimming’ was positively influenced by a friend being present. This type of mix-up occurred more than once, indicating that it is very probable that participants indeed meant an inverse relationship, which might be due to the way the question would have been phrased in Czech. Another inverse relation occurred, where a positive/negative confusion existed. In these cases, the value itself was a negative thing (e.g. ‘fear’), which is contrary to how values are normally thought of, as positive motivators. This meant a double negation occurred.

Misalignment 3. [Concretisations and parts-of actions] There were misunderstandings about meaning of and differences between the notions of action concretisations and parts-of relations.

Misalignment 4. [Values] There were misunderstandings about the meaning of the term ‘value’ and how values can be promoted or demoted by actions (cf. Misalignment 1).

4.1.3. Not knowing how to answer

Finally, confusion arose when people did not know how to talk to the system. For instance, the question where assistance would be helpful required the participant to re-iterate all relevant activities, they could not just say ‘all of them’ (P7). Similarly, participants could indicate that no values were relevant, but some did not understand that this was possible right away (P4). The Wizard-of-Oz design probably did influence this category as well. A bias was observed where participants used long utterances during the sessions as they could expect higher ability of speech recognition. However the Wizard-of-Oz was not altering their utterances and transcribed them as they were pronounced. This influence of the design was one of the factors which contributed to the choice for speech-to-text in the second study.

Misalignment 5. [Communication options] There were misunderstandings about which communication options were available and which communication style to use in order to communicate effectively with the conversational agent.

4.1.4. Misunderstandings because of misunderstandings

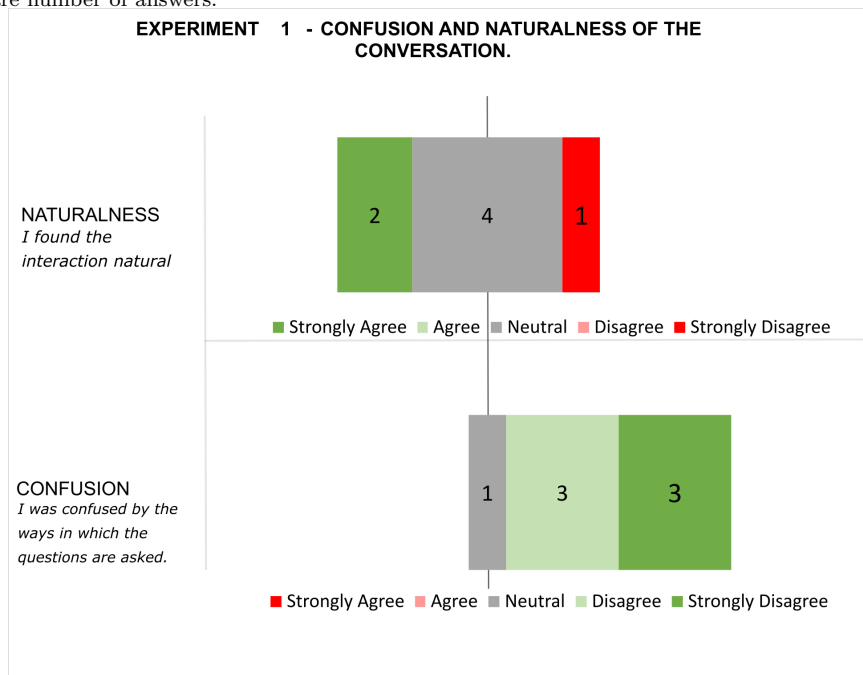
These three categories in themselves shed light on where misalignment between the system and the user can occur. However, perhaps even more interesting was the observation that these misunderstandings can also cause each other. Based on the qualitative analysis, several comments showed how one category of confusion led to another. Confusion about the general structure could lead to confusion about the concepts (e.g. P6 did not understand the concept of *concretisation* and parts-of because of misunderstanding the structure). Confusion about the general structure also sometimes led to confusion about how to answer (e.g. P6 did not understand values could be either positively or negatively related and therefore tried to answer ‘both’). Finally, confusion about the concepts also led to confusion about how to answer (e.g. P7 did not get the concept of assistance and then answered in a way the system could not understand). These examples show how one misalignment can lead to the next. This

also means, however, that some misunderstandings could be prevented by reducing misunderstandings on other levels. We also have to note that misunderstandings on how to answer did not lead to further misunderstandings.

4.2. Usability

Two questions were asked on the naturalness (*I found the interaction natural.*) and confusion (*I was confused by the ways in which the questions are asked.*) of the conversation. The mean score for *naturalness* was 1.17, $SD = 1.38$ (high is more natural), Regarding the *level of confusion*, the mean scores was 0.71, $SD = 0.76$ (high is more confusing). See the figure 3 for more detailed results.

Figure 3. Results of subjective judgements from first study. Green color indicates desired answers. Numbers in chart are number of answers.



The resulting SUS scores were $max = 60, min = 10, mean = 38.21, SD = 20, 14$, which indicates according to Bangor, Kortum, and Miller (2009) poor adjective rating or F grade, which means not acceptable usability (Bangor, Kortum, and Miller (2008)). The main source of problems could rise from the grounding strategy used, which was perceived as slow and annoying.

5. Experiment 2 - goal and changes

This second experiment had three main goals. Firstly, it attempted to improve the structure of the dialogue to reduce misunderstandings in concepts and how to answer. The hypothesis here is that less misunderstandings in the structure and concepts should also lead to less misunderstandings about the concepts and how to answer. Secondly, it was meant to see if the categories of misalignment from the first study would be reproducible. And thirdly, the Wizard was omitted to study if further levels

of misalignment would be observed if there was no experimenter in the loop. Below we discuss the specific changes we made to address these points. We will summarize these changes through numbered and named ‘design solutions’. This summary will make it easier to get an overview of our proposed solutions, and to refer to them in the discussion.

The main change in the dialogue was the inclusion of STT instead of the Wizard transcribing the answers. This inclusion was enabled by switching to English as the communication language because the STT systems for English language demonstrate satisfactory level of reliability. In such case we could get closer to the future real system. The other changes in the dialogue and formalism were made to improve the usability, and understandability of the concepts, so the structure was simplified in the following ways. Firstly, the system asked for *routes* instead of *concretisations* when talking about ways of going somewhere, as that concept is easier to understand. This change was possible because the depth of the action hierarchy was restricted to one destination as root, one layer of ways of getting there (routes), and one layer of parts of each of those routes. This restriction was done to simplify the structure, and made sense given the relatively small trees of the first study. Additionally, values were only asked for the routes, as they represent reasons to choose one route over the other. For the parts of, the only possible choice would be on whether to include the action as part of the route, which makes less sense as people typically only give parts of that are essential anyway.

Design solution 1. [Domain specific action terms] The generic term ‘action concretisations’ that was used by the conversational agent was replaced by the domain specific term ‘routes’. This change was aimed at addressing Misalignment 3.

Design solution 2. [Restricted action hierarchy] The generic action hierarchy that allows arbitrarily deep trees and usage of concretisations and parts-of at any level was replaced by a restricted tree with a fixed number of layers and fixed usage of action types. This change was aimed at addressing Misalignments 1, 2, and 3.

When eliciting the values, a pre-defined list of possible values was given based on the answers from the first study and Azenkot et al. (2011), namely: comfort, health, independence, safety, socializing, time-management and wealth. While in the first study any answer was possible to the values question, in the second participants could choose from this list or answer ‘no value’ if none from this list were applicable. The goal of letting people choose from a pre-defined list was twofold. Firstly, to help the speech-to-text recognize the values. Secondly and most importantly, to give participants a mental picture of what type of answers was expected, and simultaneously give examples.

From the results of the first study, we found that the concept of values proved to be the most tricky to understand. We see this result both in the qualitative results and in the list of values which was given, which included things like ‘company of a friend’ and ‘personal freedom’. Though clearly related to values, values as described by Schwartz (1992) are more abstract. The answers from the first study were manually linked to more abstract values by the experimenters (e.g. ‘company of a friend’ to socializing and ‘personal freedom’ to independence) to investigate whether a pre-defined list could be made. This resulted in 8 categories, and the only category which was eventually left out of our list was ‘information access’ (for instance for answers such as ‘route knowledge’ or ‘find connection’), as for visually impaired people this need for information typically comes down to another value such as time-management (getting there quicker) or safety (getting there safer). All other categories were included, as they also largely

corresponded to those found by Azenkot et al. (2011) in their study into the values of visually impaired travellers.

This study gives further evidence that a relatively small set of values lie at the core of travelling choices of visually impaired people. Although we cannot exclude that sometimes participants would have values not included in the list, the need for more clarity for the participants and this evidence supported the choice to have people choose from a pre-defined set of options. After talking about the actions, the users would also be asked to rate the importance each of the values had for them as this importance might also differ per person. This importance rating was left out of the first study as the conversation would otherwise become too long, and was made possible now because of the simpler structure.

Design solution 3. [Provide a pre-defined list of values] Instead of allowing participants to specify arbitrary values that could be promoted or demoted by the specified actions, a pre-defined list of values was used. The inclusion of this list was aimed at addressing Misalignments 1, 4, and 5.

A final change that was made after the first experiment is that instead of asking whether help would be useful for each action, capability was only included for the part-of actions. The routes are more high level, while the parts of are more often the concrete actions (e.g. *cross the street*) that a system might actually be able to help with. This simplification also gave room to ask for capability more specifically, namely to also check whether assistance would be crucial or just helpful.

Design solution 4. [Restricted use of capabilities] Instead of asking participants about their capability for any specified action, it was only asked for part-of actions. This was a simplification inspired by the before-mentioned design solutions.

The exact differences in the formalism and dialogue order between the first and second study can be seen in Figure 1 and Table 1.

6. Results Experiment 2

6.1. Misalignment

All qualitative remarks were analyzed the same way as in the first experiment. During analysis, a fourth category of misalignment arose, namely the system misunderstanding the user. The sections below describe the findings for each of the four categories.

6.1.1. Understanding the general structure

Generally, less confusion was expressed about the general structure than in the first experiment. The beginning of the elicitation was still sometimes problematic in the sense that confusion was often expressed here, despite the fact that users went through the tutorial before. One typical mistake was that some participants gave the location of the destination instead of possible ways to get there (e.g. P3, P6). This mistake can be viewed as a variant of Misalignment 3 regarding the understanding of concretisations. However, the grounding strategies helped people understand their mistake and learn about the structure, and level of understanding increased during the dialogue, as indicated by less confusion expressed later on.

In the statistics about the user models, it is good to look at the amount of informa-

tion that was present in the user model in the end. The structure was simplified and could only hold two destinations in total. It is interesting to note, therefore, that the number of actions present in the hierarchies at the end is similar to the first experiment (see additional files Berka et al. (2020)). This finding indicates that the reduction of the possible number of levels in the hierarchy (Design solution 2) did not cause people to express less knowledge. Although there was some confusion at the start of the experiment about the overall idea of the structure, Misalignments 1 and 2 did not surface as in the first experiment.

6.1.2. Understanding the concepts

The confusion about the difference between ways-of and parts-of (Misalignment 3) was much less during this study, probably due to the re-framing of the question about *concretisations* as routes (Design solution 1). The concept of values was still difficult to understand for some participants (Misalignment 4), but the addition of a predefined list (Design solution 3) did seem to help as it provided them with an answer space. Several participants tried to answer with an option that was not in the list, but the answer they gave was typically a synonym of one of the values in the list and, therefore, recognized without problem.

6.1.3. Not knowing how to answer

There was still confusion when people did not know how to talk to the system (Misalignment 5). For instance, some users gave their answers all at once, which the system did not accept (for instance for parts of, where the user was asked to name one at a time). Others did not know exactly what the system would accept, specifically how brief or concrete the answers should be (P6). Some people also did not know how to ask for a repetition of the question or explanation, for instance when the system was asking to choose from the list of values. Not knowing this became a problem mostly as it was sometimes hard to remember the list of predefined values (P6), or when it was not clear what to say if there is no value positively or negatively supported (P5).

6.1.4. System misunderstanding

The new category of system misunderstanding describes situations where the system did not understand the user's words. These misunderstandings can be divided into two subcategories, namely STT limitations, and errors in recognizing the intent. The STT limitations occurred quite often during the sessions, despite the fact that we provided the STT with a list of words and phrases as hints for the speech recognition. For instance, words like 'on foot' or 'by foot' were wrongly transcribed to 'on food' and 'buy food', respectively. Unfortunately this type of wrong transcription led to some false confirmations by users and, therefore, false information stored in the user model (P1, P2). When eliciting values e.g. 'health' was transcribed as 'house' three times in a row by the system, but it was not stored in the model as the system could only accept the predefined values, but it did cause frustration to the user.

Misalignment 6. [Recognition of words] The speech recognition technology was unable to recognize some words accurately, which lead to the user model containing inaccurate content, and other content being erroneously rejected by the system.

Errors in intent recognition were often caused by answers which were too long, for

instance when eliciting routes (P1, P3). The system also did not recognize activities which were not in, and not even similar to activities in the intent example list of the conversational agent, e.g. activities which described what happened at the destination and were, therefore, not from the travelling domain (P8 - looking for free seats in the restaurant). Longer and more human-like utterances were also problematic to recognize, e.g. utterances starting with 'Well maybe' followed by the activity name, resulted in recognizing 'well maybe' as an activity. Fortunately this mistake was not accepted by the user during grounding (P7).

Misalignment 7. [Recognition of intent] The system was sometimes unable to recognize which activity a user intended to capture in the user model.

6.1.5. Misalignment model

As with the previous study, several examples were found where one misunderstanding led to another. All the relationships between the three previously identified levels of misalignment were found in this study as well, solidifying the evidence for this model. Additionally, this study showed that the user not understanding how to answer often led to the system not understanding the user. Given this additional relationship, we present the following model of our levels of misalignment.

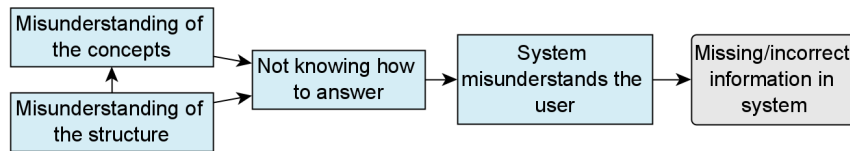


Figure 4. Model representing the four levels of misalignment and how one category can lead to another

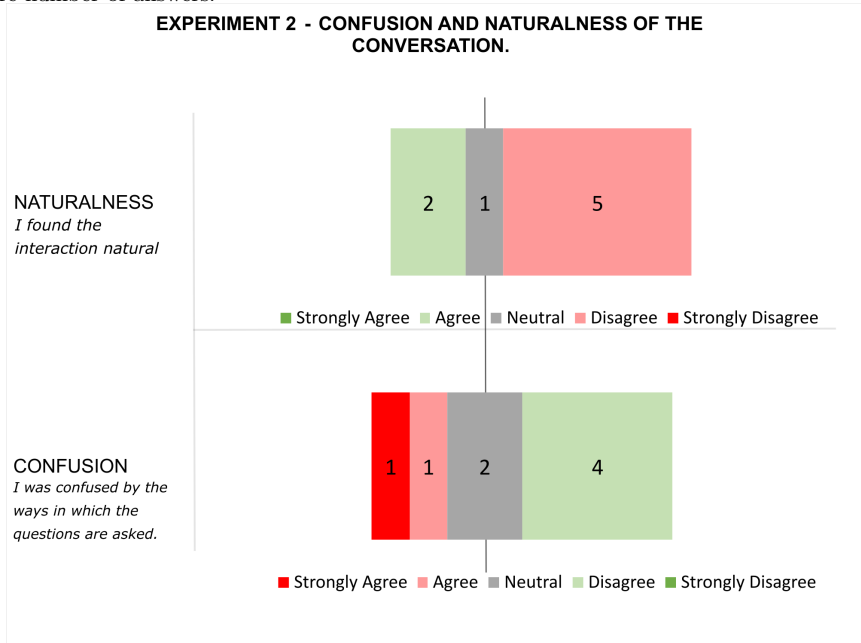
As can be seen in this model, a final box has been added which depicts the user model containing faulty information. In several cases, the fact that the system misunderstood the user led to this mistaken information being stored in the final user model. Although the grounding strategy caught and corrected some of these misunderstandings, in some cases the participants also wrongly confirmed the misunderstanding. Either because they did not notice the mistake, or because they did not want to bother correcting the agent.

6.2. Usability

Similarly to the first study, usability was measured with the SUS and additional questions on naturalness and confusion of the conversation. The mean score for *naturalness* was 2.38, $SD = 0.92$. Regarding the *level of confusion*, the mean scores was 1.88, $SD = 1.13$. Both of these scores were very similar to the first study. See the figure 5 for more detailed results.

The resulting SUS scores were $max = 72.5$, $min = 35$, $mean = 59.7$, $SD = 12.6$, which according to Bangor et al. (2009) indicate an OK adjective rating or D grade, which means marginal acceptable usability (Bangor et al. (2008)). Although still on the low side, this score was significantly higher than for the first experiment $t(9.81) = -2.44$, $p < 0.05$, despite the fact that the Wizard was replaced by STT.

Figure 5. Results of subjective judgements from second study. Green color indicates desired answers. Numbers in chart are number of answers.



7. Discussion

The aim of this work was to better understand possible misalignment between a conversational agent and user in the context of user-model elicitation for visually impaired travellers. The ultimate goals are to increase both personalisation and transparency. Both these goals are hindered by misalignments. Personalisation because misalignments in the conversation can lead to the wrong information being stored in the user model. Wrong information in the model would in turn lead to the system personalising in the wrong way (e.g. advising a user to go the quick route instead of the safe one while the user actually prefers safety). And transparency because a user can only learn more about what the system wants to know through elicitation conversations if that user also actually understands what the system wishes to know. Transparency in communication cannot just be achieved by the system sharing what it wishes to know, it also requires the user understanding the system, which is hindered by misalignment. Therefore, we specifically wished to know which types of misalignment would occur, and if these were in some way related to each-other, as knowledge about both is crucial for solving or preventing misalignment.

From our results we see that it is possible to use a conversational agent to elicit Semantic User Models for navigation support for visually impaired people, but, as expected, this process still comes with several types of human-agent misalignment. We can characterize the types of misalignment in four ‘levels’ (RQ1); the user’s misunderstanding of the structure of the user model and dialogue; the user’s misunderstanding of the concepts in the user model; the user’s not understanding how to talk to the system, and the system’s misunderstanding of the user’s utterances. These four levels range from misunderstanding abstract structures, to misunderstanding concrete utterances.

Additionally, we find that there were indeed relationships between the different

types of misalignment (RQ1.1). Specifically, we can structure the types in an hierarchical way, where the more abstract levels of misunderstanding can give rise to the more practical misunderstandings. This hierarchy means that we can also draw some conclusions on how to prevent misalignments (RQ2). Practically, our findings imply that if a user understands the general structure of the user model and the concepts, this will also lead to fewer misunderstandings in how to talk to the system, fewer occurrences of the system misunderstanding the user, and eventually to a more correct user model. This hierarchy starting with a user understanding the structure and concepts shows the importance of having an understandable and explainable user model whenever it is elicited in direct conversation with a user. Moreover, the hierarchy provides a strategy for making improvements in the development of such conversational agents; working from the user model up to the conversational aspects. In this work we, therefore, aimed our design solutions towards the model and concepts in particular, simplifying the structure by making it more domain specific, offering a tutorial, and rephrasing and re-framing the concepts. Although more work is needed to study the exact effects of these improvements on different levels, the preliminary results from our second study do indicate that high-level improvements reduced misunderstandings. For instance, we found that usability was higher for the second study after the implementation of the identified design solutions, despite the addition of STT which can often lead to lower usability.

7.1. *Fundamental vs. methodological issues*

We identify four different types of misalignment in our work. In general, when misalignment and confusion happens between a human and artificial agent during a conversation, we can distinguish between issues which are fundamental or methodological. Methodological issues arise from the specifics of the conversation, for instance the topic of the conversation, the implementation of the agent or the technical set-up of the conversation. Fundamental misalignment issues are those which are inherent to conversations about complex topics, such as our Semantic User Models.

There are two main reasons why we have not distinguished between these two types in our results up to this point. The first is that to truly make this distinction well, multiple different studies with different types of agents should be done to find out which issues remain when the methodology changes. This is beyond the scope of this paper. The second reason is that we argue that when we look at alignment in a human-agent interaction context, the methodology of how the agent is built and interacts is fundamental to the interaction. For human-human interaction it is perhaps possible to study the fundamentals of misalignment only. However, an agent is always built in a certain way, with a purpose, technological foundations and tools, as chosen by the developers. We would argue that this methodology of how the agent is built, for what it is built and in what context it is used is an essential part of how the agent communicates and, therefore, in how misalignments can occur. In this way, methodological issues cannot be completely removed from a study into misalignment between human and agent during conversation.

Nevertheless, we can and should consider which of our findings are particular to the methodology of the agent of this specific study, and which we would expect to be more fundamental. We firstly speculate that the categories of *misunderstanding of the concepts* and *misunderstanding of the structure*, as well as their connection are fundamental to dialogues about the types of user models employed in this study. These

user models are characterized as having different types of concepts which are given meaning by how they relate and are structured, so understanding of these things is fundamental. Of course, user models which are structured in a different way might give rise to other types. Secondly, we speculate that for the categories of *not knowing how to answer* and *system misunderstands the user* we see a combination of fundamental and methodological issues. Both of these types of issues are strongly influenced by the type and constraints around the current conversational system (such as text-to-speech systems only recognizing a given set of responses), which lacks flexibility in many ways. In this case, our methodology has probably contributed to some of the misalignments we found. However, we do expect that some of the issues here are more fundamental, especially those being caused by the other two, more fundamental types. As an example, if in a conversation between two people someone misinterprets a question, it is also possible that this leads to the other misinterpreting the answer.

To conclude, we believe that for all of the types of misalignment in our model, they would remain relevant to some extent for other conversational agents given that they are talking with users about the same type of user models. We would expect that especially the first two types are the least methodology-dependent, but to fully establish that more different studies would be needed.

7.2. Avoiding misalignment

Our results highlight that just asking users about concepts in Semantic User Models is not always enough to make them understand the concepts and structures involved. In the literature about transparency, this term is often used to just refer to sharing information about the system’s workings. The system asking about concepts and information can be seen as a way of sharing information about what it wishes to know, so as transparency. Our work shows, however, that to avoid misalignment and achieve full user understanding, sometimes more is necessary. Sharing information without checking if the user has understood what the system intended will not necessarily lead to better results. If we use the concepts as laid out by Verhagen, Neerincx, and Tielman (2021), we see that in our case, transparency in what the system needed to know did not always lead to a more interpretable system due to misunderstandings about what was shared. This finding highlights the need for a feedback loop and an understanding of possible misalignment if we truly wish to achieve more transparent communication.

Moreover, through the changes we have made from the first to the second experiment we have already explored several design solutions for addressing misalignments in this paper. The overarching theme that can be identified in these design solutions is to *restrict* the underlying generic modelling language for capturing the user’s actions, values and capabilities to what is needed in the domain of application. This concerns the language used (Design solution 1), the structure of the model (Design solutions 2 and 4), and the concepts themselves (Design solution 3). This highlights the tension between expressivity of user models and transparency or understandability for the user. The generic model gives the user potentially more freedom to express themselves, but it makes the models much harder to grasp. We posit that generic frameworks for semantic user modelling are useful to investigate, as the theoretical properties can be studied and they can be applied in different domains (see also Section 7.3). However, when using them in a concrete application domain, the language, structure and concepts used need to be adapted to align with what is really needed in that context. For

Misalignments 5, 6, and 7 we have not explored design solutions in this paper. These challenges are connected with the use of conversational interfaces in general, and results from this area can be integrated in our context to address these issues. Moreover, as highlighted above, the investigation of grounding strategies and feedback loops will be an important component of addressing this problem.

Other possible avenues for how to avoid misunderstandings exist as well. For our user group, haptic interfaces could be further explored. Although not as easily accessible currently as voice interfaces, work is being done to explore haptic feedback for touchscreen devices (Palani, Fink, and Giudice (2020)). Moreover haptic interfaces have been shown to improve student learning as shown by Nam, Li, Yamaguchi, and Smith-Jackson (2012), indicating that might also be used to present the tree structures in our model, for instance. Although less applicable for the visually impaired user group, non-verbal cues such as gaze and gestures can help in a human understanding a system. But they could also allow a system to recognize when the user does not understand something (Campbell (2020); Esfandiari-Baiat, Hunyadi, and Esposito (2020)). Such modelling of the user’s understanding could also guide explanations to reduce these misunderstandings (Abdulrahman et al. (2019); Faulkner et al. (2018)). More inspiration could be drawn from literature on human teamwork, for instance by identifying how misunderstandings are avoided or fixed in such contexts, see for instance Menekse, Purzer, and Heo (2019). Work from the perspective of dialogue management could help in how to adapt the dialogue when misunderstandings occur as well (Yamaoka, Hara, and Abe (2015)). Ultimately, there is evidence that difficult verbal tasks are not inherently more difficult to perform with a computerized teammate (Palanica, Thommandram, and Fossat (2019)), indicating that these issues should be solvable.

7.3. Generalizability

Although this study was performed in the context of travel for visually impaired users, the type of user model and resulting misalignment model is relevant for other domains of supportive technology as well. In this work, the interaction took place via speech only, as any visual or non-verbal communication would not work for our user group (Azenkot and Lee (2013)). Although this limitation does not always hold, it does provide an opportunity to truly look at the modality of speech and the difficulties that can arise during verbal communication. Given the increasing prevalence of conversational agents as interfaces (Emarketer (2017)), our results are generally relevant. The use of a multi-modal user interface or different modalities for non-blind user groups could have impact on better understanding in higher levels of our misalignment model (answering and system understanding), but not so in two lower levels of understanding (concepts and structure). Therefore, we believe that our misalignment model can be generalized towards non-blind user groups interacting with conversational interfaces.

In our study, the conversational agent was the interface between a formal user model and the user itself. The goal of having such a direct and interactive interface between the user model and the user is to allow a system to personalize in a more transparent way, as well as to gain information which can only originate from the user. We propose that our misalignment model might also be relevant for other types of interfaces trying to achieve the same goal. Even in a graphical interface showing the structure of the user model visually, the user misunderstanding the structure might eventually result in the user entering the wrong type of data in the wrong place. Although more research is needed to study if exactly the same types of misalignments occur in practise, we

would hypothesise that the conceptual structure of our model is still relevant.

7.4. Limitations and future work

To fully appreciate the work presented in this paper, it is necessary to also consider the limitations. Firstly, the current studies were conducted only with participants who were not English but Czech native speakers. This choice introduces a bias in the understanding of the (in English communicating) system as a whole, specifically in our model of misalignment between user and the system. For future experiments a control group of native English speaking participants should be included. Secondly, we used speech to text technology in our second experiment which caused no small amount of error, which are clearly listed in the section 6.1.4. Perhaps choosing different technology could solve some of the problems. We did not make any advanced analysis of these technologies in advance, but we chose the technology that appeared most available at the moment. In the future more attention should be paid to the choice of these technologies. Following, this study was performed with a relatively small user group. As our purpose for this study was to qualitatively see how the elicitation of this kind of user model would work, these small numbers were sufficient for our current purposes. Future iterations of the system should, however, also be evaluated statistically with larger groups. Finally, one of the goals of eliciting a user model explicitly through conversation is to increase transparency in and understanding of the system. In this study, we did not explicitly measure transparency in any way, and instead performed a qualitative analysis on what misalignments could lead to a lack of mutual understanding. To evaluate how transparent our system was, we would have needed to compare our system to not having a conversation at all, which was beyond our scope. However, it would be very interesting and relevant to further investigate what exactly the difference in transparency is between a learned user model and one that is elicited through direct communication.

7.5. Conclusion

In this paper, we study the elicitation of a user model with a conversational agent for visually impaired travellers. Such a model representing activities and values could be used for personalisation of navigation support systems. Moreover, the elicitation of such a model via conversation immediately gives the user transparency into how the system works. Our goal was to study which types of misalignment could occur between user and the system during conversational elicitation of user models, and the relationships between them. Our results from two iterations of qualitative studies reveal four levels of misalignment, namely: misunderstanding the general structure of the user model and dialogue; misunderstanding the concepts used in the model; the user not understanding of how to talk to the agent and finally the system's misunderstanding of the user's utterances. Misunderstanding on these 4 levels can eventually lead to missing or incorrect information being stored in the user model, and a loss in transparency as the user does not truly understand what the system wants to know. Importantly, we found that these levels are highly interdependent, i.e. misunderstanding on one level leads to misunderstanding on the next level, following a predictable pattern. These results provide insights into how misalignments can be avoided, and specifically highlight the importance of ensuring that the user understands the conceptual structure and concepts in user models. If we wish to elicit knowledge for formal

user models via a conversational interface and create a system which is both more personalized and transparent, mutual understanding between system and user throughout the conversational process is key.

Funding

This work is part of the research programme CoreSAEP, with project number 639.022.416, which is financed by the Netherlands Organisation for Scientific Research (NWO).

This research has been supported by projects Navigation of handicapped people funded by grant no. SGS19/178/OHK3/3T/13 and Research Center for Informatics (reg. n. CZ.02.1.01/0.0/0.0/16_019/0000765).

Notes on contributor(s)

Jakub Berka implemented the conversational agent for the second study, performed the second study, analysed the data and played a large part in writing the initial version of the paper. Jan Balata implemented the conversational agent for the first study, performed the first study and reviewed the writing. Catholijn M. Jonker provided valuable input on the study design and execution, and reviewed the writing. Zdenek Mikovec provided input on the study design and execution, analyzed the data and contributed some sections of the writing. M. Birna van Riemsdijk provided valuable input on the study design and execution, and contributed some of the writing and positioning of the paper. Myrthe L. Tielman implemented the user model for both studies, analyzed the data and played a large part in writing and revising the paper.

Notes

Additional files containing participant's statements and data about the user models can be found in: Berka et al. (2020).

Acknowledgements

We would like to thank our participants for their time and valuable input, without whom we would not have been able to conduct this study. We thank Pei-Yu Chen, PhD student at TU Delft, for her insights on user modelling and personalization, which has helped us position our work.

Biographies

Jakub Berka is a UX Designer and Front-End developer in a private company in Prague, Czech Republic. He received his master degree in Human-Computer interaction, with focus on navigation of visually impaired people. He is an author of several research papers published on international conferences in HCI field.

Jan Balata is a UX Researcher in a private enterprise in Prague, Czech Republic. His main interest is human-computer interaction with special focus on non-visual user interfaces, accessibility, mobile user interfaces. He co-authored number of articles in academic journals and over ten papers on international conferences in HCI field.

Catholijn M. Jonker is professor at Interactive Intelligence at Delft University of technology and part-time professor at the Leiden Institute for advanced computer science, both in the Netherlands. Her research goal is to achieve long-term human-agent teamwork, aiming for value-sensitive co-evolution between humans and agents.

Zdenek Mikovec is Associate Professor at the Czech Technical University in Prague. His expertise is in the field of human-computer interaction, with a focus on interaction in special environments. His research interests are navigation of people with limited orientation capabilities, multi-modal user interfaces, mobile user interfaces, and accessibility of ICT.

M. Birna van Riemsdijk is associate professor Intimate Computing at University of Twente, the Netherlands. Her research investigates how to take into account human vulnerability in the design and run-time reasoning of intimate technologies. She was awarded a Vidi personal grant and the Dutch Prize for Research in ICT 2014.

Myrthe L. Tielman is assistant professor at the Interactive Intelligence group at Delft University of Technology, the Netherlands. Her research interest lies in how AI can better understand people with the goal of developing truly trustworthy AI. To this end, her focus lies on topics like trust, explainability and values.

References

- Abdulrahman, A., Richards, D., Ranjartabar, H., & Mascarenhas, S. (2019). Belief-based agent explanations to encourage behaviour change. In *Proceedings of the 19th acm international conference on intelligent virtual agents*.
- Ahmetovic, D., Guerreiro, J., Ohn-Bar, E., Kitani, K. M., & Asakawa, C. (2019). Impact of expertise on interaction preferences for navigation assistance of visually impaired individuals.
- Azenkot, S., & Lee, N. B. (2013). Exploring the use of speech input by blind people on mobile devices. In *Proceedings of the 15th international acm sigaccess conference on computers and accessibility* (pp. 11:1–11:8). New York, NY, USA: ACM.
- Azenkot, S., Prasain, S., Borning, A., Fortuna, E., Ladner, R. E., & Wobbrock, J. O. (2011). Enhancing independence and safety for blind and deaf-blind public transit riders. In *Chi 2011* (pp. 3247–3256).
- Balata, J., Mikovec, Z., & Slavik, P. (2018). Conversational agents for physical world navigation. In *Studies in conversational ux design* (pp. 61–83). Springer.
- Bangor, A., Kortum, P., & Miller, J. (2008). The system usability scale (sus): An empirical evaluation. *International Journal of Human-Computer Interaction*, 24(6), 574–594.
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114–123.
- Berka, J. J., Balata, J. J., Jonker, C. C., Mikovec, Z. Z., van Riemsdijk, M. B., & Tielman, M. (2020, Sep). *Additional files belonging to the paper: Misalignment in user model elicitation via conversational agents*. 4TU.ResearchData. Retrieved from <https://doi.org/10.4121/12901496.v1>
- Brachman, R., & Levesque, H. (2004). *Knowledge representation and reasoning*. Morgan Kaufmann.
- Brooke, J. (1996). Sus: A "quick and dirty" usability scale. In P. Jordan, B. Thomas, I. McClelland, & B. Weerdmeester (Eds.), (p. 189-194). Taylor & Francis.
- Bujacz, M., Baranski, P., Moranski, M., Strumillo, P., & Materka, A. (2008). Remote guidance

- for the blind – A proposed teleassistance system and navigation trials. In *Hsi 2008* (pp. 888–892).
- Campbell, N. (2020). The temporal structure of multimodal communication. intelligent systems reference library. In S. I. Hunyadi L. (Ed.), (p. vol 164). Springer.
- Converse, S. A., Cannon-Bowers, J. A., & Salas, E. (1991). Team member shared shared mental models: A theory and some methodological issues. *Proceedings of the Human Factors Society 35th Annual Meeting*, 1417-1421.
- Cranefield, S., Winikoff, M., Dignum, V., & Dignum, F. (2017). No pizza for you: Value-based plan selection in BDI agents. In *International joint conference on artificial intelligence*.
- Dignum, V. (2017). Responsible autonomy. In *Proceedings of the 26th international joint conference on artificial intelligence* (p. 4698–4704). AAAI Press.
- Emarketer. (2017, May). *Alexa, say what?! voice-enabled speaker usage to grow nearly 130% this year*. Retrieved from <https://www.emarketer.com/>
- Esfandiari-Baiat, G., Hunyadi, L., & Esposito, A. (2020). The temporal structure of multimodal communication. intelligent systems reference library. In S. I. Hunyadi L. (Ed.), (p. vol 164). Springer.
- Fan, X., & Yen, J. (2010). Modeling cognitive loads for evolving shared mental models in human-agent collaboration. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41, 354 - 367.
- Faria, J., Lopes, S., Fernandes, H., Martins, P., & Barroso, J. (2010). Electronic white cane for blind people navigation assistance. In *Wac 2010* (pp. 1–7).
- Faulkner, T. K., Niekum, S., & Thomaz, A. (2018). Asking for help effectively via modeling of human beliefs. In *Companion of the 2018 acm/ieee international conference on human-robot interaction*.
- Friedman, B., Jr., P. H. K., & Borning, A. (2006). Human-computer interaction and management information systems: Foundations advances in management information systems, volume 5 (advances in management information systems). In P. Zhang & D. Galletta (Eds.), (p. 348-372). M.E. Sharpe.
- Georgiou, T., & Demiris, Y. (2017). Adaptive user modelling in car racing games using behavioural and physiological data. *User Model User-Adapted Interaction*, 27, 267–311.
- Golledge, R. G. (1993). Geography and the disabled: a survey with special reference to vision impaired and blind populations. *Tran. of the Inst. of British Geographers*, 63–85.
- Golledge, R. G. (1999). *Wayfinding behavior: Cognitive mapping and other spatial processes*. JHU press.
- Guerreiro, J., Ohn-Bar, E., Ahmetovic, D., Kitani, K., & Asakawa, C. (2018). How context and user behavior affect indoor navigation assistance for blind people. In *Proceedings of the internet of accessible things* (p. 2).
- Haller, H., Nguyen, V.-B., Debizet, G., Laurillau, Y., Coutaz, J., & Calvary, G. (2017). Energy consumption in smarthome: Persuasive interaction respecting user’s values. In *The 9th ieee international conference on intelligent data acquisition and advanced computing systems: Technology and applications*.
- Harbers, M. (2011). *Explaining agent behavior in virtual training* (Unpublished doctoral dissertation).
- IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). *Ethically aligned design - a vision for prioritizing human well-being with autonomous and intelligent systems, version 2*. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Retrieved from <https://ethicsinaction.ieee.org/>
- Kayal, A., Brinkman, W.-P., Neerincx, M. A., & Riemsdijk, M. B. V. (2018, May). Automatic resolution of normative conflicts in supportive technology based on user values. *ACM Trans. Internet Technol.*, 18(4), 41:1–41:21.
- Kließ, M. S., Stoelinga, M., & van Riemsdijk, M. B. (2019). From good intentions to behaviour change: Probabilistic feature diagrams for behaviour support agents. In *Prima 2019: Principles and practice of multi-agent systems* (pp. 354–369). Springer International Publishing.

- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., ... Coiera, E. (2018). Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25, 1248-1258.
- Lüttich, K., Mossakowski, T., & Krieg-Brückner, B. (2004). Ontologies for the semantic web in Casl. In *Recent trends in algebraic development techniques, 17th international workshop (wadit'04)* (Vol. 3423, pp. 106–125). Springer.
- Menekse, M., Purzer, S., & Heo, D. (2019). An investigation of verbal episodes that relate to individual and team performance in engineering student teams. *International Journal of STEM Education* volume, 6.
- Nam, C. S., Li, Y., Yamaguchi, T., & Smith-Jackson, T. L. (2012). Haptic user interfaces for the visually impaired: Implications for haptically enhanced science learning systems. *International Journal of Human-Computer Interaction*, 28(12), 784-798. Retrieved from <https://doi.org/10.1080/10447318.2012.661357>
- Ohn-Bar, E., Guerreiro, J., Kitani, K., & Asakawa, C. (2018). Variability in reactions to instructional guidance during smartphone-based assisted navigation of blind users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3), 131.
- Palani, H. P., Fink, P. D. S., & Giudice, N. A. (2020). Design guidelines for schematizing and rendering haptically perceivable graphical elements on touchscreen devices. *International Journal of Human Computer Interaction*, 36(15), 1393-1414. Retrieved from <https://doi.org/10.1080/10447318.2020.1752464>
- Palanica, A., Thommandram, A., & Fossat, Y. (2019). Adult verbal comprehension performance is better from human speakers than social robots, but only for easy questions(2019). *International Journal of Social Robotics*, 11.
- Pasotti, P., Jonker, C. M., & van Riemsdijk, M. B. (2017). Action identification hierarchies for behaviour support agents. In *Workshop on cognitive knowledge acquisition and applications*.
- Pasotti, P., van Riemsdijk, M. B., & Jonker, C. M. (2016). Representing human habits: towards a habit support agent. In *European conference on artificial intelligence*.
- Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 640:1–640:12). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/3173574.3174214>
- Scheutz, M., DeLoach, S. A., & Adams, J. A. (2017). A framework for developing and using shared mental models in human-agent teams. *Journal of Cognitive Engineering and Decision Making*, 11, 203-224.
- Schwartz, S. (1992). Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, 25, 1-65.
- Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). "hey alexa, what's up?": A mixed-methods studies of in-home conversational agent usage. In *Proceedings of the 2018 designing interactive systems conference* (pp. 857–868). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/3196709.3196772>
- Serramia, M., Lopez-Sanchez, M., Rodriguez-Aguilar, J. A., Rodriguez, M., Wooldridge, M., Morales, J., & Ansotegui, C. (2018). Moral values in norm decision making. In *International conference on autonomous agents and multiagent systems*.
- Stephanidis, C. C., Salvendy, G., of the Group Margherita Antona, M., Chen, J. Y. C., Dong, J., Duffy, V. G., ... Zhou, J. (2019). Seven hci grand challenges. *International Journal of Human-Computer Interaction*, 35(14), 1229-1269. Retrieved from <https://doi.org/10.1080/10447318.2019.1619259>
- Tielman, M., Jonker, C., & van Riemsdijk, M. (2018). What should I do? Deriving norms from action, values and context. In *Mrc - tenth international workshop modelling and reasoning in context, held at faim*.
- Tuttle, D. W., & Tuttle, N. R. (2004). *Self-esteem and adjusting with blindness: The process of responding to life's demands*. Charles C Thomas Publisher.
- van de Poel, I. (2013). Translating values into design requirements. In D. P. Michelfelder,

- N. McCarthy, & D. E. Goldberg (Eds.), (chap. Philosophy and Engineering: Reflections on Practice, Principles and Process). Springer.
- van de Poel, I. (2015). Handbook of ethics, values and technological design. In J. van den Hoven et al. (Ed.), (p. 89-115). Springer.
- van Riemsdijk, M. B., Jonker, C. M., & Lesser, V. (2015). Creating socially adaptive electronic partners. In *International conference on autonomous agents and multiagent systems*.
- Verhagen, R., Neerincx, M., & Tielman, M. (2021). A two-dimensional explanation framework to classify ai as incomprehensible, interpretable, or understandable. In *Extraamas*.
- Völkel, T., & Weber, G. (2008). Routechecker: personalized multicriteria routing for mobility impaired pedestrians. In *Proceedings of the 10th international acm sigaccess conference on computers and accessibility* (pp. 185–192).
- Vtyurina, A. (2019). Towards non-visual web search. In *Proceedings of the 2019 conference on human information interaction and retrieval* (pp. 429–432). New York, NY, USA: ACM.
- Wald, M. (2021). Ai data-driven personalisation and disability inclusion. *Frontiers in Artificial Intelligence*, 3.
- White, R. W., & Grant, P. (2009). Designing a visible city for visually impaired users. *Proc. of the 2009 Int. Conf. on Inclusive Design*.
- WHO. (2009). *ICD update and revision platform: change the definition of blindness*. Retrieved from [http://www.who.int/blindness/Change the Definition of Blindness.pdf](http://www.who.int/blindness/Change%20the%20Definition%20of%20Blindness.pdf)
- Wycherley, R. J., & Nicklin, B. H. (1970). The heart rate of blind and sighted pedestrians on a town route. *Ergonomics*, 13(2), 181-192. Retrieved from <https://doi.org/10.1080/00140137008931131> (PMID: 5432363)
- Yamaoka, M., Hara, S., & Abe, M. (2015, Dec). A spoken dialog system with redundant response to prevent user misunderstanding. In *2015 asia-pacific signal and information processing association annual summit and conference (apsipa)* (p. 223-226).