

## Article

# I Let Go Now! Towards a Voice-User Interface for Handovers between Robots and Users with Full and Impaired Sight

Dorothea Langer <sup>1,\*</sup> , Franziska Legler <sup>1</sup>, Philipp Kotsch <sup>2</sup>, André Dettmann <sup>1</sup>  and Angelika C. Bullinger <sup>1</sup><sup>1</sup> Ergonomics and Innovation, Chemnitz University of Technology, 09107 Chemnitz, Germany<sup>2</sup> HFC Human-Factors-Consult GmbH, 12555 Berlin, Germany

\* Correspondence: dorothea.langer@mb.tu-chemnitz.de; Tel.: +49-371-531-32217

**Abstract:** Handing over objects is a collaborative task that requires participants to synchronize their actions in terms of space and time, as well as their adherence to social standards. If one participant is a social robot and the other a visually impaired human, actions should favorably be coordinated by voice. User requirements for such a Voice-User Interface (VUI), as well as its required structure and content, are unknown so far. In our study, we applied the user-centered design process to develop a VUI for visually impaired humans and humans with full sight. Iterative development was conducted with interviews, workshops, and user tests to derive VUI requirements, dialog structure, and content. A final VUI prototype was evaluated in a standardized experiment with 60 subjects who were visually impaired or fully sighted. Results show that the VUI enabled all subjects to successfully receive objects with an error rate of only 1.8%. Likeability and accuracy were evaluated best, while habitability and speed of interaction were shown to need improvement. Qualitative feedback supported and detailed results, e.g., how to shorten some dialogs. To conclude, we recommend that inclusive VUI design for social robots should give precise information for handover processes and pay attention to social manners.



**Citation:** Langer, D.; Legler, F.; Kotsch, P.; Dettmann, A.; Bullinger, A.C. I Let Go Now! Towards a Voice-User Interface for Handovers between Robots and Users with Full and Impaired Sight. *Robotics* **2022**, *11*, 112. <https://doi.org/10.3390/robotics11050112>

Academic Editor: Anna Esposito

Received: 31 August 2022

Accepted: 13 October 2022

Published: 15 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** verbal communication; inclusive design; user-centered design; robot object handover; voice-user interface; voice dialog

## 1. Introduction

Currently, extensive effort is invested in the development of functional and usable assistive robots, as these are hoped to solve serious current problems. One example is understaffing and the skills shortage in the health care sector, where people with all sorts of impairments could benefit from an assistive robot. Additionally, in the production industry, social robots are becoming team members of collaborative mixed human–robot teams [1,2]. “Social (or sociable) robots are designed to engage people in an interpersonal manner, often as partners, in order to achieve positive outcomes [3] (p. 1936)”. They work alongside human workers and facilitate an interpersonal relationship due to the coordination of behavior as well as verbal and nonverbal communication [3].

Handing over objects to a human is a necessary key capability for many different robots [4–6] and an example of a complex collaborative task. It requires temporal and spatial synchronization of actions and is strongly based on visual and haptic information. Therefore, it is necessary to ensure both partners are aware and certain of the attention of their cooperation partner. Humans use eye contact, head orientation, body posture, and hand and arm positions as cues of reciprocal attention during collaborative tasks [7] and, especially, to interpret an intention to handover [8]. Examining the current state of the art, active, coordinating gaze behavior of the robot is considered a precondition for human–robot interaction during collaborative tasks [9]. In short: if the robot cannot detect the human’s eyes, handover is not started [10,11]. However, coordinating gaze behavior between humans and robots cannot be ensured universally. Certain working

postures in the production sector cause restrictions in sight (e.g., [6]). Likewise, blind and visually impaired (BVI) people cannot perform eye contact with a robot. Thus, conditions of work and the individual prerequisites of the users of the robot themselves underline the potential of using other modalities to support a successful handover from robots to humans. One suitable modality is speech. In the context of interaction with smart devices, speech was shown to be a natural interaction modality that increases the overall user satisfaction for BVI people [12]. Ref. [13] points out that speech is the preferred primary interaction modality for people with disabilities. Still, it is unclear if this holds true for handover between humans and robots, as this task is normally based on visual and haptic information. With our study, we address this issue and take a step towards the design and development of a Voice-User Interface (VUI) for social robots that meets the requirements of BVI people. In this context, the VUI is understood as the structure and content of a dialog to verbally coordinate the process of robotic handover interaction. It is therefore independent of the applied speech recognition software and the deployed technology for the robotic handover task.

### *1.1. Research on Object Handover under Visual Restrictions*

Reviewing the literature, we were not able to identify research done to support handover between robots and BVI people. Although BVI users desire to use robots as assistants for finding and passing objects to support their everyday lives [14], most existing robots have been designed for users with full visual abilities. This is mirrored in a set of studies where robots for object handover have been designed and evaluated with sighted users, e.g., [4,15,16]. We found only one study in the area—Ref. [8] studied blinded subjects, to simulate visual impairment, performing human-to-human handovers to deduce knowledge for robot-to-human handovers. It was found that without visual feedback, handovers are performed much slower and more carefully. Still, the handover task was not supported by speech or sound.

Handover tasks are characterized by a high degree of passivity on the side of the receiving participant. Research on robot-to-human handovers with visual contact to the robot identified different process phases during handover: robot searching and grasping of the object, transporting it to the receiver (approaching), signaling readiness to handover, actual handover interaction with releasing the object, and ending of the interaction [4,8]. Ref. [5] defines object recognition, people detection, robot motion planning and detection of a user's grasp as the necessary phases during robotic handover. Users' active participation is only required during the actual handover of the object. During passive phases, users under visual restrictions should be acoustically supported. For BVI people, it is especially important to actively inform them about the current status of a system, or allow voice commands for retrieving this information [17]. Therefore, VUIs hold undisputed potential for BVI people, not only during active, but also passive phases of handover tasks.

It is known that during robotic handovers, humans prefer an orientation of the object that corresponds to human conventions [4]. For instance, an easy grasp (e.g., handing over a cup horizontally and with the handle first) by the receiver is called an 'etiquette factor' during robot-to-human handover [5].

Robotic handovers are especially complex when it comes to hazardous objects, as perceived risk affects perceived handover quality and safety. Handing over hazardous objects from robot to human resulted in lower perceived quality and safety [11]. Some object surfaces bear a risk of injury when handing over to a receiver, e.g., because they are pointed, sharp-edged, or hot. These need to be covered or handed averted from the receiver [5], which is especially important for handovers under visual restrictions.

In human-to-human object handovers, BVI users need more information compared to sighted users and are known to use specific strategies. Several strategies were identified: (1) active grasping of the receiver after the object is placed directly in front of the hand (instead of the giver 'pressing' the object into receiver's hand), (2) movement restricted to one participant at a time in midair handovers, (3) passive role of the impaired individ-

ual in midair handovers, (4) explicit verbal information about potential sources of harm from handover objects, and (5) additional verbal communication during the process and usage of acoustic signals [18]. Due to the preferred passive role of BVI people in midair handovers, the robot takes over the role of the sighted interaction partner and therefore has to coordinate the localization of the human and robotic hand.

As the scant literature illustrates, systematized knowledge about robot-to-human handovers under visual restrictions is not available. We proceed to illustrate the characteristics of Voice-User Interfaces (VUI) that allow speech and sound as a primary interaction modality in human–robot interactions in general.

### *1.2. Voice-User Interfaces in Human–Robot Interaction for Sighted and Visually Impaired Users*

BVI people are the power users and, consequently, experts of VUIs [19], with voice commands being their most frequently used interaction modality [20]. In a study, all included BVI subjects used TalkBack or VoiceOver for smartphone interaction in daily life [12]. Furthermore, a study in Germany showed that 88% of included BVI subjects regularly use common voice assistants such as Siri [19]. BVI users also use speech more frequently for input, generate longer messages, and are more satisfied with speech interfaces than sighted people [21]. Ref. [19] concludes that BVI people as power users of VUIs “may be able to correctly interpret and retrieve longer and more complex speech responses” (p. 2). Supporting daily activities, VUIs allowing for easy and quick entry [13] present a high potential to make impaired people more independent.

Applying verbal communication contributes to anthropomorphism of social robots, i.e., attributing human (intelligent) behavior to the machine [22]. Ref. [20] found that BVI people desire voice interaction for a more ‘natural’ interaction with technical systems. This also implies the usage of everyday language with regards to grammar and sentence structure, as well as the observance of dialect to enhance familiarity and user experience [20]. The same holds true for using characteristic features of intonation, pitch, and timbre to display different intentions and emotions of a robot [23].

Targeting natural speech interaction, the computational model underlying most collaborative robots today is Natural Language Processing (NLP). This “refers to the computational approach of analyzing, understanding and manipulating natural language text or speech” [15] (p. 3). In most NLP systems, a predefined dictionary is used to compare to input voice commands [24,25]. Therefore, non-matching words included in a speech command are simply ignored, which removes the necessity for a predefined simple speech grammar or predefined phrases/voice commands and reduces familiarization time prior to interaction [5]. Still, it has to be noted that failures in speech processing can easily lead to user frustration [20].

The aim of designing a VUI that is highly similar to human-to-human speech is in accordance with several guidelines cautioning designers against the development of ‘robotic conversations’ [13]. However, a balance between naturalness and a ‘too human-like’ conversation needs to be achieved. VUIs should not ‘trick’ users into a false belief that a robot is a human by applying strongly ‘human-like’ conversations [13], as this could lead to wrong estimations regarding the capabilities and intelligence of a robot. It also has to be taken into consideration that BVI people often use and are used to screen readers with synthesized robotic-sounding voices and a higher-than-natural output rate [13]. While this can be explained by a greater ability for serial memory found in BVI people [26], it also illustrates differing requirements for the interaction with VUIs dependent on the specific user group.

In most research on robot handover, only a few steps of a robotic handover are supported by speech interaction, as robots are designed for sighted people. For example, the work of [5] as well as [15] only support ordering, and the work of [4] supports the immediate timepoint of object handover. This is probably because most phases within a robotic handover interaction can usually be entirely supported by visual information. Nevertheless, even if visual information is available, speech interaction is useful. One study

showed that sighted subjects performing a handover task with a robot did not try to grasp the object until the robot verbally encouraged them to do so [4]. To conclude, supporting the whole process of handover with VUI interaction could enable differing groups of users to feel confident during robotic handovers.

Research on speech interaction with other assistive systems and human-to-human handovers offers some information that VUIs should give during robotic handovers under visual restrictions. These include the aforementioned need for status information and orientation of an object (see Section 1.1). Even if social robots implicitly apply human conventions regarding object orientation, it cannot be assumed that people under visual restriction trust in the robot's action and, accordingly, in their safety. Thus, explicit verbal information on object orientation should support a handover process. Additional insights from human-to-human handovers [18] show that BVI people want to be verbally informed about unusual properties of the object in advance, an intention to hand over an object right before, and want to give or receive a verbal reinsurance of safe grasp right at the moment of handover before letting it go. Furthermore, BVI people use and pay attention to acoustic sounds to locate a placed object and to perceive approaching or removing motions of their interaction partner.

Targeting a general and inclusive usage of robots, development of a voice dialog covering the complexity of a robot–human handover under visual restrictions is therefore highly desirable for all user groups.

### 1.3. Voice-User Interfaces Design for Diverse Users with User-Centered Design (UCD)

VUI design for object handover for diverse user groups is a complex task needing careful consideration of their differing demands. Impaired users are a heterogeneous group differing individually—e.g., in abilities, use cases, and preferences—and the combination of features causes individually unique challenges [27]. Moreover, requirements of non-impaired users might be differing or even opposing to those of impaired users. As described earlier, fully sighted people use coordinating gaze behavior as a cue of reciprocal attention [7,9] and to interpret an intention to handover [7]. A VUI that coordinates the handover process is purely additional for sighted people, whereas for BVI people, it is the only source of process information. This could lead to a differing tolerance concerning the amount of information conveyed by the VUI. As BVI people use speech more frequently and can possibly handle more complex speech information than sighted users, they might wish for more information per time.

However, despite differing requirements, BVI users prefer interfaces designed as closely as possible to customary products and feel uncomfortable if they are dependent on special solutions. Consequently, design activities should always target smooth interactions for all users while paying attention to inclusive approaches rather than developing special solutions for BVI users [12]. Hence, a procedure for the design of a suitable VUI is necessary that puts future, diverse users at the center.

User-Centered Design (UCD) [28] is a normed process developed for designing interactive systems. It provides an appropriate procedure, as it focuses on user requirements from the start with the overarching goal to ensure an ideal user experience. User-centered design is based on the six following fundamental principles:

- The design is based upon a precise understanding of users, their tasks, and their environments; deeply dive into their living environment instead of using vague ideas;
- Users are involved throughout the design and development process; users' perspectives affect the development and are not only requested for evaluation of the final product;
- User-centered evaluation drives and refines design; every prototype is evaluated by users and their feedback is used for further product development;
- This process is iterative; stages of the development process are not linear and are done more than once; user feedback creates the necessity to repeat stages multiple times; Design of products does not only address the ease of use, but the whole user experience; products create emotions, provide real solutions, and encourage repeated usage;

- The design team must include multidisciplinary skills and perspectives; designers, human factor experts, programmers, and engineers closely cooperate.

The UCD process consists of four general activities that are completed as subsequent stages at several times.

1. Analyzing: Information about relevant user groups is collected and summarized, to understand their tasks, goals, and why and in which context they will use the product;
2. Specifying: From the collected information, user requirements are derived to set design guidelines and goals;
3. Prototyping: Concepts and design solutions are created and implemented. Early prototypes are typically simple, with no or reduced functionality, becoming more and more realistic with every iteration cycle;
4. Evaluating: Users try out prototypes and give feedback to test the design against the user requirements.

Effects of user involvement are overall positive, especially regarding user satisfaction [29], but also regarding system success, as well as decreased time and cost of development [30]. User involvement for the development of assistive technology in general is considered paramount [31,32]. The more complex the developed system, the more user involvement has been found to be necessary [29]. Contrary to studies showing the benefits of user involvement in general, literature on the involvement of disabled users is inconclusive. Some authors postulate that their involvement should take place in every general activity of UCD, to empower them to solve their own problems [31]. Others argue that this is inefficient, due to the heterogeneity of impaired user groups and because it vastly increases the effort, diminishing the aforementioned advantage regarding time and cost of development [27,32]. Ref. [20] gives an overview of 25 papers describing interactive technologies for smart homes targeting BVI people. In 15 cases, voice or sound feedback is used as an interaction mode, confirming the frequent use of this modality for BVI people. However, just 12 of 25 papers report any user involvement, where only 7 involve users in the late evaluation stage. Moreover, when developing a smart home system with speech interaction, Ref. [33] tested voice commands without BVI users and only the final system evaluation included BVI users.

To derive user requirements, simulation of impairments is not sufficient in the case of BVI people [34], but is repeatedly used. Ref. [27], for instance, proposes the usage of acting studies to sensitize designers to requirements of impaired users instead of involving them in product development activities. Furthermore, it is common to simulate impairment by masking healthy sensory processing, e.g., by wearing dark painted glasses or blindfolds [8]. Ref. [34] has shown that usability experts and designers do not have sufficient insight into living environments of special groups such as BVI people. Especially for congenital blindness, simulation quickly reaches its limits, as eyesight-covered sighted experts still use and rely on their visual experience and knowledge within the design processes. It has to be noted, however, that BVI people apply strategies not available for sighted people that are only under temporary visual restrictions. It follows that a product or process that was designed by exclusively involving BVI users could therefore be inaccessible for sighted people, and vice versa. As a consequence, when designing an assistive robot for handover tasks under visual restrictions, BVI, blinded, and sighted people should be involved in the UCD, although not all user groups have to be involved in every development stage.

#### *1.4. Research Objectives and Paper Structure*

With our research, we target design, development, and evaluation of an inclusive VUI prototype to be installed in a handover robot. We follow the guidelines of the UCD process and involve BVI, blinded, and sighted users. Derived from research gaps, the following tasks and research questions are specified:



- (1) User requirements: What information conveyed by the VUI do users need to perform a handover interaction under visual restrictions? Which features of the speech interaction support the perception of the handover robot as a social partner?
- (2) Definition of VUI dialog: What dialog flow and commands are necessary to verbally coordinate the process of robotic handover interaction for diverse user groups?
- (3) Evaluating the VUI prototype: Does the VUI prototype enable a successful handover? How is the VUI evaluated by users regarding support ability and satisfaction with social characteristics of the social robot? Is the VUI inclusively applicable for BVI, blinded, and sighted users, despite not involving all user groups in every stage of the UCD process?

Summing up, the background is that handing over objects to a human as a necessary key capability of many different robots. The research problem is that a broadly applicable robot interface requires additional modalities aside from visual information to coordinate the complex process of robotic handover interaction. Under sight restriction, speech dialog is the preferred interaction modality. However, requirements for such a dialog and its necessary structure, as well as content for a successful and satisfying handover, are unknown so far. The UCD process provides a methodology for the design of a suitable VUI that puts future, diverse users at the center. Hence, we contribute the basics to successfully design and implement an inclusive handover function in any robot.

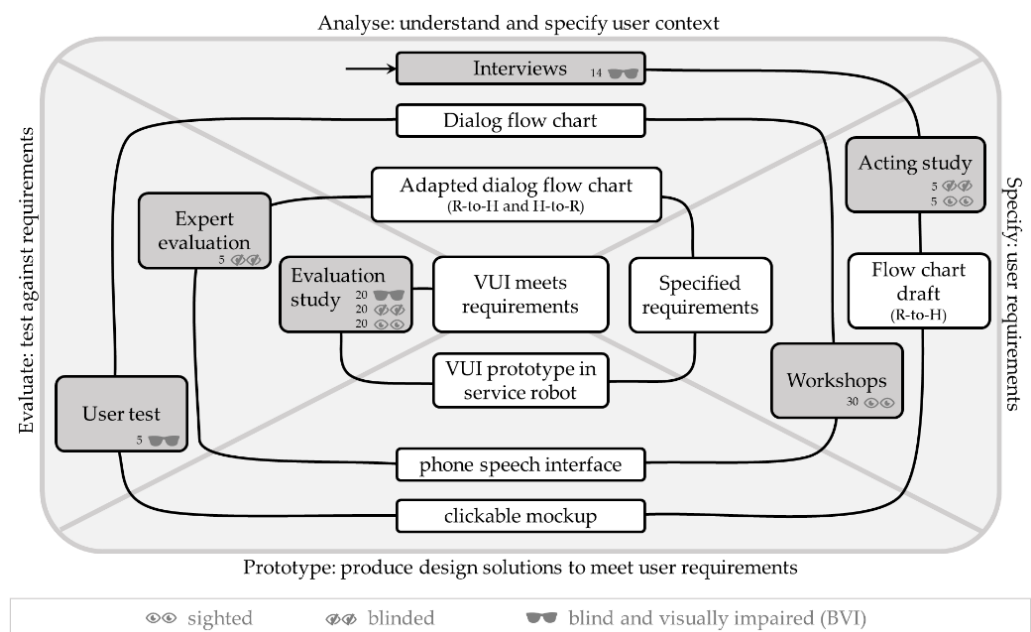
Subsequently, in Section 2, we present the methodology of the applied iterative UCD process including samples, materials, procedures, measures, and data analysis. Section 3 covers the results of this process with respect to user requirements for VUI design, required VUI dialog content and structure, and evaluation of the developed VUI prototype. In the subsequent Section 4, results are discussed regarding the identified research gaps. Furthermore, general as well as user-specific recommendations for inclusive VUI design supporting the complete process of handover tasks are given. Finally, Section 5 concludes with the contributions of the presented research to the advancement of handover functions in social robots.

## 2. Materials and Methods

This section describes the applied UCD process for the development of an inclusive VUI design for handover tasks with users with various visual abilities. Based on insights from a focus group regarding handover strategies of BVI people [18], six successive research steps with user involvement were conducted. Therefore, user groups were divided into BVI, blinded, and sighted people. As a result of each research step involving users, adaptations and specifications of the VUI for handover tasks were made. Three different prototypes with increasing functionality were applied during the process. Figure 1 visualizes the whole applied UCD process from the first interviews until a final evaluation study.

### 2.1. Samples

Overall, 123 subjects were involved during the UCD process. Within each step, user groups were involved selectively, with the aim of an economic research process (see Section 1.3). The group of BVI people is heterogeneous—the result of a visual test cannot perfectly be associated with remaining individual capabilities. Referring to legal definitions, the group of BVI people in this research either involved ‘blind people’ with less than 2% remaining sight or ‘strongly visually impaired people’ with up to 8% remaining sight in the less impaired eye. BVI people were acquired within local associations for the blind. Table 1 summarizes characteristics of samples within each UCD step.



**Figure 1.** Overview of the applied user-centered design (UCD) process for inclusive VUI design for handover tasks under visual restrictions. Iterative research steps with user involvement are marked grey. Numbers represent the number of involved users, dependent on user group. R-to-H = handover direction robot-to-human, H-to-R = handover direction human-to-robot.

**Table 1.** Sample characteristics within each user-centered design (UCD) step.

Step	Involved Groups and Sample Size	Gender Ratio (Female:Male)	Mean Age (in Years)	Group Comparison Regarding Age
Interviews	BVI (14)	6:8	~45	-
Acting Study	blinded (5) sighted (5)	5:5	~35	-
User Test	BVI (5)	2:3	~58	-
Workshops	sighted (30)	- 1	- 1	-
Expert Evaluation	blinded (5)	1:4	~35	-
Evaluation Study	BVI (20), blinded (20), sighted (20)	37:23	~42	$M_{BVI} = 40.3$ ( $SD_{BVI} = 15.3$ ) $M_{blinded} = 44.6$ ( $SD_{blinded} = 19.9$ ) $M_{sighted} = 42.2$ ( $SD_{sighted} = 19.8$ )

<sup>1</sup> Workshops were performed during networking meetings with voluntary participation. Demographic information was not collected. BVI = blind and visually impaired. Grey background for row discriminability.

## 2.2. Materials

**Clickable mockup.** To give subjects a first impression of the VUI supporting handover tasks and to test the appropriateness of the deduced dialog structure during early stages of the UCD, a clickable mockup was developed. Therefore, each dialog phase within the current available flowchart was specified by exemplary and concrete predefined sentences or questions that the robot may use to interact with users. This collection of predefined speech outputs was digitally recorded with two synthetic voices from the acapela group repertoire (<https://www.acapela-group.com/de/voices/repertoire/>, accessed on 19 July 2022). ‘Claudia’, a German female voice, and ‘Klaus’, a German male voice, were implemented. The resulting voice output files were implemented in a simple PowerPoint presentation, allowing experimenters to react to subjects’ utterances or questions by playing the according output file with the predefined and recorded output (Wizard-of-Oz method).

**Phone-based speech interface.** The implementation of a VUI into a robot requires the functional integration of robotic hardware and software components, as well as the respective user interfaces. To test the functionality and sufficiency of the VUI dialog content and structure as well as the performance of speech processing in advance, the VUI was

developed as a standalone prototype. Therefore, Voice over IP (VoIP) was used based on the freeware PhonerLite ([https://lite.phoner.de/index\\_de.htm](https://lite.phoner.de/index_de.htm), accessed on 19 July 2022). Experimenters were able to call a server where the VUI was running and subjects were able to speak to the VUI via a phone. The VUI only supported handovers in the direction robot-to-human. It consisted of (1) a welcoming and introduction with the name of the robot followed by a short manual, (2) the voice dialog supporting the handover tasks, and (3) several short voice commands such as “status” and “continue” that could be used during all times. If used, those commands were prioritized, leading to an interruption of the current dialog. ‘Claudia’ was used as speaker voice for the VUI. A tone was used to confirm a processed user input.

VUI prototype in robot assistant. For the final object handover under visual restriction, a collaborative robot was designed and built. Therefore, a robot type UR10 from Universal Robots [35] was installed on an automated guided vehicle platform. A flexible three-finger hand type BH8-280 of the brand BarrettHand™ (Barrett Technology, Newton, MA 02460 USA) was attached to the end effector of the robot. To enable a handover task under visual restrictions, the aim was to implement (1) a VUI as primary interaction modality, (2) an optical sensor-system for object identification including classification of hazardous surfaces, and (3) an intelligent body and hand recognition algorithm. All subcomponents of the robot were interconnected by an event-based distributed-state-machine. For further information, Appendix A describes the applied robot assistant using the taxonomy of [22]. The robot arm moved slowly and on an automated path. Trajectory was only restricted by objects in the test environment and a safety space at the subjects’ positions. The final VUI further used ‘Claudia’ as the speaker voice and the tone for confirmation of user input procession. It contained the same components as the phone-based speech interface and a comprehensive standard, as well as a shortened expert mode (see Section 3.2 and Appendix C).

### 2.3. Design, Procedures and Measures

With the exception of the evaluation study, all UCD steps were conducted including only one or two of the three user groups (BVI, blinded, sighted). The final evaluation study was conducted to assess successful handover coordination and user satisfaction with the VUI prototype. Table 2 summarizes the research aims, methodology, and procedure of each UCD step. Audio and video recording were used. For several UCD steps, three personas were characterized to design and develop the VUI prototypes by considering different capabilities and knowledge of users. The three personas were:

- Beginner: wants a detailed dialog; each activity of the robot should be commented on in detail by voice outputs; robot as assistant worker;
- Intermediate: wants articulated status messages, but no detailed explanations; robot as an intelligent machine;
- Expert user: wants short commands and short messages; robot as a tool.

**Table 2.** Research aims, methodology, and procedure of each applied UCD step.

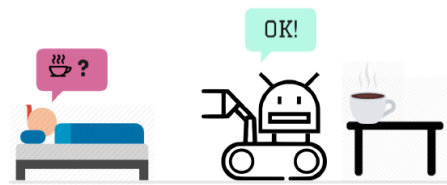
Step	Research Aims	Method/Design/Material <sup>1</sup>	Procedure
Interviews (BVI)	<ul style="list-style-type: none"> <li>• Basic needs of user groups regarding human and robotic handovers under visual restrictions</li> </ul>	M: structured interviews	Structure of interviews (1) Hazardousness of objects during midair handovers; (2) General hazardousness under visual restriction; (3) Daily strategies for handovers; (4) Attitudes towards and demands for communicating with the robot.



Table 2. Cont.

Step	Research Aims	Method/Design/Material <sup>1</sup>	Procedure
Acting Study (blinded, sighted)	<ul style="list-style-type: none"> <li>Observation of natural handover processes</li> <li>Abstraction of basic dialog structure during handovers</li> <li>Collection of literal utterances and questions to train speech recognition</li> </ul>	M: laboratory study D: 2 × 2 within-design: <i>Role of giving participant human assistant vs. assistant robot</i> <i>Modality</i> <i>midair handover vs. placing on table</i>	Role play <ul style="list-style-type: none"> <li>Scenario setting: BVI person in need of assistance would like to receive an object that is inaccessible to him/her.</li> <li>Blinded person 1 (receiver) with blindfold and sighted person 2 (giver) who played a human assistant or assistant robot dependent on scenario.</li> <li>Task: object handover in unstandardized interaction</li> <li>Experimenter specified the object within each handover trial.</li> <li>Open questioning at the end.</li> </ul>
User Test (BVI)	<ul style="list-style-type: none"> <li>Testing success of flow chart draft for handover dialog structure</li> <li>Collection of literal utterances and questions to train speech recognition</li> <li>Inserting collected literal utterances into flow chart draft, resulting in a dialog flow chart</li> </ul>	M: Wizard-of-Oz (WOZ) ser test (structured, unstandardized) D: 3 handover scenarios (within) <i>reach (midair interaction) vs. take (midair interaction) vs. place (indirect interaction)</i> MA: clickable mockup	<ol style="list-style-type: none"> <li>Forced selection between a female and male speaker voice of VUI;</li> <li>Three handover scenarios using the VUI mockup with selected speaker voice to hand over a cup of water.</li> </ol> <ul style="list-style-type: none"> <li>Experimenters played the robot and controlled the VUI (WOZ) by clicking an appropriate predefined speech output.</li> </ul>
Workshops (sighted)	<ul style="list-style-type: none"> <li>Collect literal variations (words, grammar) to train speech recognition</li> <li>Handover dialogs in case of errors to extent the dialog flow chart</li> <li>Specification of flow chart</li> </ul>	M: remote workshop D: 3 × 4 mixed-design: <i>Persona—between (beginner vs. intermediate vs. expert user)</i> <i>Scenario—within (ideal scenario vs. speech recognition failure vs. problems finding object vs. path to user is blocked)</i> MA: digital whiteboard (real-time, collaborative)	<ol style="list-style-type: none"> <li>Two persons (1 receiver, 1 giver) virtually handing over a “cup” (picture of a cup was moved from giver to receiver, see Figure 2), using verbal interaction to support handovers in both directions (robot-to-human vs. human-to-robot) by considering the assigned persona;</li> <li>Other persons of the group transcribed verbal utterances;</li> <li>Discussion on dialog process and utterances.</li> </ol>
Expert evaluation (blinded)	<ul style="list-style-type: none"> <li>Finding usability failures</li> <li>Iterative adaptation of the VUI after each expert</li> <li>Collection of literal variations</li> </ul>	M: Unstandardized expert evaluation MA: phone-based speech interface	<ol style="list-style-type: none"> <li>Interaction with VUI by Voice over IP;</li> <li>Experts were told to imagine handover situations with a robot;</li> <li>Choice between predefined objects, handover directions, and handover modalities;</li> <li>After several trials: interview about the overall interaction, focus on difficulties, remembering and appropriateness of short voice commands, missing process information, grammar, wording;</li> </ol> <ul style="list-style-type: none"> <li>Experimenters observed experts’ reactions during interaction with the VUI (mimic, verbal).</li> </ul>
Evaluation study (BVI, blinded, sighted)	<ul style="list-style-type: none"> <li>Evaluation of success of handover support by VUI, satisfaction with VUI support and with social characteristics dependent on user group</li> </ul>	M: standardized experiment D: 3 × 3 × 2 mixed design <i>User groups—between (BVI vs. blinded vs. sighted)</i> <i>Objects—within (knife vs. cup vs. spanner)</i> <i>Modalities—within (midair vs. placing)</i> MA: VUI prototype in assistant robot	<ol style="list-style-type: none"> <li>Pre-survey;</li> <li>Random assignment of sighted subjects to group ‘blinded’ or ‘sighted’;</li> <li>Six handover (3 objects x 2 modalities) interactions with the robotic prototype, each paused by short post-scenario questions;</li> <li>Post-experiment questionnaire and open questions.</li> </ol>

<sup>1</sup> M = Method, D = Design, MA = Material. Grey background for row discriminability.



**Figure 2.** Visualization of remote handover roleplay on the real-time collaborative whiteboard applied in workshops. Subjects were able to move objects by selecting and shifting on the board.

Evaluation study. Due to the complexity of the final evaluation study, the study is described in detail in the following section.

A quasi-experimental design was applied. Subjects' vision served as the between-group variable. The group 'blind and visually impaired (BVI)' consisted of blind or severely visually impaired persons (see Section 2.1). Sighted subjects were randomly assigned to the groups 'blinded' (eyes covered with a blindfold) and 'sighted'. Objective success of handover tasks, ordering time, and handover time, as well as subjective rating of satisfaction with VUI support and with social characteristics were used as dependent variables.

The final VUI was implemented in an assistant robot (see Section 2.2). The VUI, object delivery, and handover actions of the robots were fully functional and autonomous.

Before the experiment, subjects were informed about the procedure of the study through a written participant information sheet and signed a declaration of consent and privacy policy. Screen reader-capable documents were used. Subjects were welcomed in an anteroom. Due to pandemic reasons, experimenters controlled for compliance with the hygienic concept of the study. Subjects received financial compensation for effort and filled in a pre-survey for demographics. For standardization across subjects, all surveys in this study were read out loud, subjects verbally answered and experimenters digitally recorded answers. Within the experimental room, subjects were encouraged to become familiar with the robot by touching the robotic parts if wanted and were given a verbal explanation of functionalities. Here, the robot did not communicate with the subjects. After that, subjects sat down on a hospital-like couch and were given a headset loosely placed around the neck for speech input. The experimenters instructed the subjects to freely choose the standard or expert mode of the VUI (further description see Section 3.2) in each of the six following trials. The robot welcomed and introduced itself to the subjects by reading out a manual to explain its functions. The robot also asked subjects to speak loudly and vigorously, and explained the available short commands. Experimenters instructed the subjects which object they should order from the robot (cup, knife, spanner). The subjects were free to choose between handover modality 'midair handover' or 'placing object on the table' (Block 1). These objects were chosen from pre-study results where hazardousness was rated for several objects during imagined handovers, showing that knives are evaluated as hazardous, spanners as unhazardous, and cups as neutral. Later, the same objects were ordered and handed over again with the handover modality rejected before (Block 2). The sequence of handover objects was predetermined for each subject and varied randomly. Each trial consisted of ordering the object and receiving it from the robot without any further information from experimenters. In the end, subjects verbally completed a post-experimental survey covering the quantitative evaluation of the VUI. For the robotic subcomponents 'object identification' and 'object grasping', the Wizard of Oz methodology was applied: The second experimenter placed the according object within the robotic hand and manually sent the command to close the hand. The sighted subjects were not able to see the manual actions of the experimenter behind a partition wall. After all handover trials, a post-experimental interview was performed with the subjects.

In the pre-survey, demographic information regarding sex, age, and impairments of sight were captured. Error rate of handovers was used as an objective measurement of handover success. Additionally, ordering time, and handover time were measured and the frequency of using the VUI expert mode was analyzed. For the subjective evaluation of satisfaction with VUI support and with social characteristics, Subjective Assessment

of Speech System Interfaces (SASSI) [36] was used because the survey equally focuses on pragmatic qualities of the VUI and the resulting affect, emotions, and frustration [37]. Due to the relatively standardized interaction in an experimental setting, subscale Cognitive Demand was not applied, resulting in a total of 29 items, all rated on a seven-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree. In the end, open feedback from the subjects was requested and several specific questions regarding the VUI were asked: ‘Which adjective would you choose to describe the interaction with the robot?’, ‘How did you feel about the voice outputs?’, ‘Was there any missing or redundant information from the robot?’.

#### 2.4. Data Analysis

To analyze qualitative data during the UCD process, inductive and deductive category formation were applied. Therefore, qualitative data were audio and/or video recorded to enable a retrospective in-depth analysis. A relevant sense of the conversations was summarized directly while listening and later from audio data. In the case of interviews, answers to the questions of researcher were directly assigned to the pre-defined category of interest in the interview guideline (deductive). In the case of unstandardized research methods, e.g., the acting study, the process of handovers was observed and divided into sub-processes (inductive) and compared to the pre-drafted flow chart for handovers. Finally, to collect literal variations for the improvement of speech recognition, utterances were transcribed word-for-word, discussed by researchers, and assigned to the appropriate dialog phase within the dialog flow chart.

For data analysis of evaluation data, the statistics software R [38] was used. In case of skewed distributions, Median (Mdn) and Median Absolute Deviation (MAD) were calculated and nonparametric tests were applied. Error rate for the human–robot interaction was measured by summing all unsuccessful handover trials and dividing this number by the total number of all trials. If subjects did not receive the specific ordered object, the trial was defined as unsuccessful. Handover failures due to technical errors of the robot (e.g., problem of opening robotic hand) were not included in the measure. Overall, for 60 subjects performing six experimental trials each, 360 trials were planned during the experiment. Due to technical problems of the demonstrator, only 338 trials without technical limitations could be performed. Objective time data and use of the VUI expert mode were read out script-based from the robot logfiles. The whole human–robot interaction was divided into the phases ordering, search and delivery of the object, handover, and returning to home position. Only the duration of two phases was entirely dependent on the VUI performance and is reported here. Therefore, ordering time was defined as the time period between the robot asking for the subjects’ order and the robot indicating search for the requested object. Handover time was defined as the time period between the robot requesting handover readiness and the final opening of the robotic hand.

### 3. Results

This chapter first summarizes deduced requirements for an inclusive VUI design to support handover tasks under visual restriction (Section 3.1). Although some requirements were repeatedly found during the UCD process, each requirement is reported at the research step of first mentioning. After, the VUI prototype dialog content and structure resulting from iterative development is described (Section 3.2). Third, the results of the evaluation study conducted by involving all three user groups are reported regarding the success of handover support with VUI and satisfaction with VUI support and social characteristics.

#### 3.1. Requirements for VUI Design

Table 3 summarizes the deduced requirements and links all detailed requirements to the UCD steps of mentioning.

**Table 3.** Classified requirements for VUI design identified during the UCD process.

Design Requirement and General Insights	UCD Step <sup>1</sup>
<b>Dialog content</b>	
Placing on table: feedback on orientation of objects on table using clock direction or verbal front/side directions	U, (I)
Reassurance about safe and stable grasp before hand opening	I
Announcing if object is potentially hazardous	I
Communicating specific feature of objects (e.g., sharp, sticky, wet)	I
Requirement for position of object in relation to receiver hands	I
Standardization of robotic dialog to develop a routine	I
Introduction with guidance on desired speech input	EE
If equal robotic actions are performed during various phases of handover (e.g., ‘opening hand’ to take up object and release object), a one-to-one mapping of speech outputs with specific phase by using literal speech variations is necessary (e.g., “I am opening my hand” during object grasping vs. “I am letting go now” during handover)	EE
Continuous verbal communication about current robotic action	A
Selectable option: offering placing on the table instead of midair handover	I
Usage of communication to support both passive as well as active phases during robotic handovers (for details see Section 3.2)	I, A, U, W, EE, EV
Acoustic confirmation of command processing	U
Information regarding system failures always being speech-based	U
Information on orientation of objects in robot hand	I, (A), (U)
Confirmation loop for releasing the object during handover	I, (EE)
Usage of a tone to explicate input confirmation	U
Information on robot intent to handover	I
<b>Social manners and etiquette</b>	
Task-oriented speech outputs, no chattering	U
Short speech output	U
No overuse of “please” in robotic outputs	U
Kind and nice	U
Commanding tone for input confirmation	U
Formal speech (age-independent), no colloquial utterances	U
Possibility to thank the robot and robotic reaction to gratitude by users	U
<b>Aesthetics/anthropomorphism</b>	
Slight preference for female voice	U
Introduction of the robot by name	U
Reaction to calling the name of the robot	EE
Desire to use easy, but complete and grammatical correct speech inputs at first contact	U
Natural communication that feels ‘warm’	I
<b>Supporting passive user phases</b>	
Possibility to actively retrieve status information during passive times	I, (U)
If true: usage of artificial sounds made by the VUI to replace quiet mechanics of the robot	EV
Usage of sounds instead of voice outputs to reduce output length	U
<b>Short commands</b>	
Possibility to interrupt robot speech at all times (interrupt-events with prioritization)	U
Emergency call	I
<b>Individualization</b>	
Speaker voice	U
Amount of speech outputs (novice/expert)	U, (W), (EE)
Replacing of unneeded voice outputs during all phases by sounds for experienced users	W
Speed of speaker voice (BVI prefer faster output)	U
Individualizing name of robot	U
Blinded users: more communication to ensure a situation awareness	I
Acoustic guidance (by sounds) for user hand positioning during midair handovers	EV
<b>Speech Recognition and Speech Output</b>	
Short time latencies between two speech outputs to prevent unintended interruptions by users	EE
Acoustic signal of the VUI to show readiness for voice input	EV
Good natural language recognition	EV

<sup>1</sup> I = Interviews, A = Acting Study, U = User Test, W = Workshops, EE = Expert Evaluation, EV = Evaluation Study.  
 Grey background for category and row discriminability.

Within the interviews, BVI subjects mentioned that sharp objects such as a knife, hot objects such as filled cups, and breakable objects such as thin glasses or even broken pieces are always dangerous to hand over. This information was further used to select handover objects for the following UCD steps with practical handover interactions. BVI subjects had an overall belief that they could successfully communicate with robots, but were concerned that the robot's voice is "machine-like" and "artificial". Opinions about the acoustic sound of the robot arm were ambivalent, as it could enable localization of the robot, but it was also expected to be annoying. Users emphasized the importance of communication in general and mentioned mandatory information and feedback, e.g., the necessity for verbal reassurance about safe and stable grasp before opening the hand.

The acting study included blinded subjects for the first time. Results showed that midair handovers led to great uncertainty on the user's side, showing an increased demand for verbal communication. Framing the interaction as 'interacting with a robot' compared to framing it as 'interacting with a person' resulted in a change of communication from familiar conversation style to a command mode. Subjects mentioned necessity of a detailed description of the current robotic action to increase trust in the robot and to ensure proper situation awareness during interaction at all times. From study results, a flow chart structuring the pure technical substeps that should be supported by dialog was made (see Appendix B, Figure A2). These undefined dialog phases were integrated as placeholders for the later specified user inputs and robot outputs. Additionally, for each dialog phase, frequently used literal utterances and questions from the acting study were captured to design a clickable VUI mockup.

During the user test, subjects perceived the speaker voices as clear and friendly, so no further improvements on speaker voice was done within the next UCD steps. All five BVI subjects were able to successfully complete the handover process with predefined voice outputs triggered by experimenters (Wizard of Oz). They expressed an overall confidence that the voice dialog is sufficient to successfully perform the handover task. Tones were found to be a good extension of speech to replace robotic output and to confirm user inputs. From each interaction scenario, literal variations of speech input were collected from subjects and used for training of the speech recognition algorithm. Therefore, key words within speech input commands were analyzed. The findings of the user test resulted in an adaption of the flow chart draft towards a dialog flow chart for handover interaction.

Due to the usage of personas during the workshop, it was possible to develop an expert mode for the VUI that eliminated several robotic outputs or even entire dialog phases that seemed unnecessary for the successful handover, with the condition of prior knowledge of the VUI (for examples, see Appendix C). This further contributed to an individualization of VUI interaction. Additionally, subjects specified dialogs for erroneous interaction situations. Therefore, the dialog flow chart was extended once again. Finally, all used literal utterances and questions were analyzed and used for training of speech recognition.

Regarding the expert evaluation, results indicated that the dialog phases of the VUI described in Section 3.2 were sufficient to successfully finish the interaction, and comprehensibility and sentence structure of the VUI were judged to be satisfying. If the interaction stopped prior to 'returning to home station', general speech recognition due to unidentified speech input (e.g., specific wording), accent, or technical problems such as microphone calibration were the reasons. Therefore, the speech recognition was optimized and further parameterized. It was obvious that the experts continuously formulated requirements separately from novice first-time users and more experienced users. Experts expressed insecurities in situations of VUI silence. These situations were analyzed and used for an extension of explicit information given by the VUI for the first-time users of the intended robot assistant.

### 3.2. VUI Prototype Dialog Content and Structure

The VUI dialog covered the ordering of several daily objects with two handover modalities (midair or placing on a table). After a welcoming and the possibility to assess a



verbal manual, the VUI asks for orders. An order is finally processed if the VUI receives all three pieces of information—object, handover direction, and handover modality—from the user. An example of a complete voice dialog for a robot-to-human handover under visual restrictions can be found in Appendix C. Summing up results from all research steps with user involvement, the VUI prototype dialog for handover included nine successive parts of verbal communication. They are shown in Table 4.

**Table 4.** Successive communication parts of the VUI prototype dialog for object handover.

Part	Communication
1	Welcoming/start of interaction including introduction by name
2	Obtaining order information
3	Passive status information during searching and grasping of the object
4	Approaching towards the human, keeping a distance to the user
5	Confirmation loop for intention to hand over
6	Actual approach towards user's hand including an expression of required user actions
7	Information regarding orientation and features of the handover object and an expression of required user actions
8	Confirmation loop for safe grasp during handover
9	Final announcement of forthcoming hand opening

Two dialog versions, a novice and an expert mode, were available (see Appendix C). While every action of the robot is supported by speech output in the novice mode, speech output during passive times of users (e.g., while the robot is searching for and grasping an object) is reduced. Still, both modes support the short command ‘status’, which can be used also during expert mode to recall a speech output about the current action of the robot in case of uncertainty. Other available short commands were ‘go on’ (shortens down the current speech outputs and acts as confirmation phrase), ‘pause’ (pausing the interaction till continue), ‘call for help’ (connectivity to an internal or external emergency call center if required), ‘emergency stop’ (immediate stop of robot movements in case of danger or discomfort), and ‘back to base’ (direct ending of the interaction and returning to base station). All short commands were prioritized and interrupted the current robotic actions.

### 3.3. Evaluation of VUI Prototype

This section summarizes the results of the final evaluation study. Only six trials out of 338 were classified as unsuccessful, resulting in an overall error rate of 1.8%. In group ‘sighted’, no unsuccessful trials occurred. In comparison, the error rate was 1.8% in group ‘BVI’ and 3.4% in group ‘blinded’.

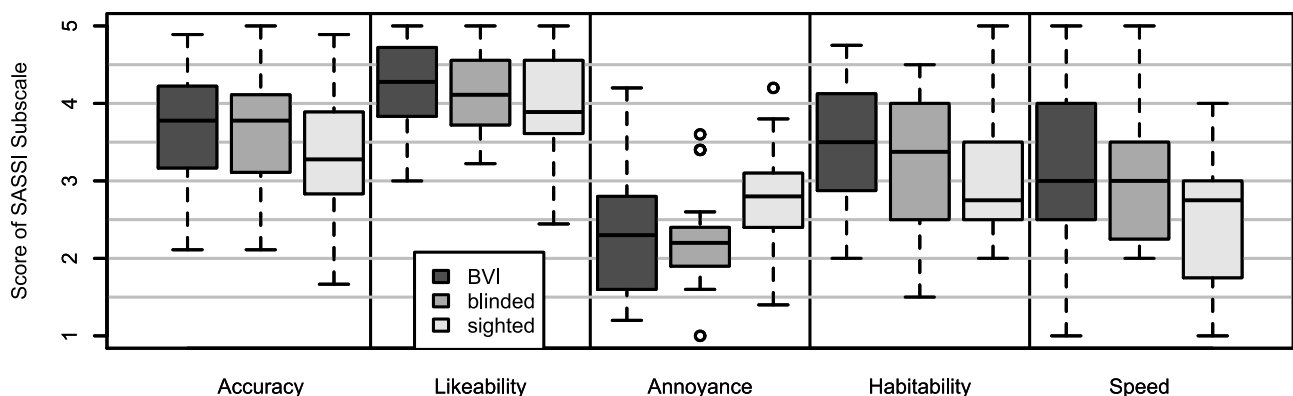
Data from robot logfiles showed a median duration for ordering of 54 s ( $MAD = 21.4$ ). The Kruskal–Wallis test showed no significant group difference in ordering time ( $X^2(2) = 2.32$ ,  $p = 0.314$ ). Median handover time was 75 s ( $MAD = 15.5$ ). The Kruskal–Wallis test showed no significant group difference in handover time ( $X^2(2) = 1.40$ ,  $p = 0.497$ ). The expert mode was chosen by subjects in 7% of all trials. These trials with expert mode were performed by 17% of the subjects, including two BVI, four blinded, and four sighted subjects. Analyzing the frequency of using the expert mode within subjects, more than half of all trials with expert mode were performed by sighted subjects.

The subscales of the SASSI questionnaire (System Response Accuracy, Likeability, Annoyance, Habitability, and Speed) were analyzed depending on group. Table 5 shows the results of the group comparisons and Figure 3 visualizes the distribution of data. Only the subscale Annoyance resulted in a significant difference; post-hoc test with Bonferroni correction ( $\alpha_{adj} = 0.016$ ) showed that the difference occurred between group ‘blinded’ ( $M = 2.25$ ) and ‘sighted’ ( $M = 2.82$ ). Therefore, sighted subjects evaluated the VUI as significant more annoying than blinded subjects ( $t = 2.79$ ,  $p = 0.008$ ,  $d = 0.712$ ). In Figure 3, for group ‘blinded’, a much lower variance in data is obvious, showing higher agreement about the Annoyance of the VUI. Likeability was evaluated best, with mean values all

clearly above the scale midpoint. In contrast, the subscale *Speed* showed mean values close to the scale midpoint. Overall, variances in data were high regarding ratings for System Response Accuracy, Habitability, and Speed.

**Table 5.** ANOVA for SASSI subscales dependent on group.

SASSI Subscale	<i>F</i>	<i>p</i> -Value	$\eta^2_p$
System Response Accuracy	2.14	0.149	0.036
Likeability	0.84	0.362	0.014
Annoyance	5.35	0.024	0.084
Habitability	2.34	0.132	0.039
Speed	3.86	0.054	0.062



**Figure 3.** Results of subjective evaluation of the VUI. BVI = blind and visually impaired.

The presentation of qualitative feedback is divided into the three user groups. Within all groups, subjects mentioned performance problems of the VUI because speech inputs of the subjects were not correctly understood and subjects had to repeat or correct the processed order information.

**BVI:** BVI subjects described using the speech output of the VUI for locating the robot within the room and would use this information for an estimation of the remaining time until handover. The spoken manual of the VUI was expressed to be important. Overall, subjects described VUI interaction as “good”, although also “cumbersome” because the robot had to finish a current speech output before it reacted to a new speech input of the subjects. The VUI interaction was characterized as “too slow”. The general mode of expression was described as “clear” and “comprehensible” and only a few subjects felt the speaker voice was “artificial”. All information given by the VUI was evaluated as equally important and highlighted the necessity of communicating the current task and location of the robot, as well as an explicit communication of expected subject behavior. BVI subjects asked for a more explicit cue for positioning the hand to support the handover task. In addition to that, the idea of an acoustic guidance of the subject’s hand was expressed. Furthermore, they requested more flexible natural language processing to allow even more natural dialog of the VUI, an acoustic signal to know that the VUI is ready for a voice input instead of only confirming inputs and an individualization of speaking rate of the VUI.

**Blinded:** Blinded subjects could imagine using the expert mode, but only after an appropriate familiarization time. Overall, the speech interaction was evaluated as “good”, which also included the speech recognition of the VUI. Again, some subjects perceived the speech interaction as “too slow” and they would have to “get used to it”. The general mode of expression was described as “clear” and “comprehensible”, but it was also mentioned that the robot talked too much. They highlighted the importance of the VUI’s explicit communication of expected subject behavior. Order confirmation was highlighted as important. Speech outputs during passive phases of subjects were evaluated as redundant.

A few subjects requested a clearer speaker voice and, similar to BVI subjects, an acoustic signal of the VUI to know that the VUI is ready for a voice input. Furthermore, they asked for artificial acoustic sounds during robot movements to be able to acoustically locate the robot. A further individualization of speech input, for example, by using codes for regular orders, was desired.

**Sighted:** Sighted subjects stated that they prefer the expert mode because it subjectively shortens the necessary time for the interaction. Some subjects characterized the speech interaction as “pleasant”, “appropriate for the intended purpose”, and “relatively clear”, while other characterizations were overshadowed by performance problems in the speech recognition of the VUI. The general mode of expression regarding social manners was described as “friendly”, “helpful”, and “formal”, but, on the contrary, the speaker voice was characterized as ‘monotonous’ and ‘jerky’. The confirmation of intended actions of the robot seemed important to the subjects, while speech outputs during passive phases of the subjects were evaluated as redundant. Subjects asked for more flexible natural language processing to allow even more natural dialog, including a larger vocabulary. An individualization of speaker voice and artificial acoustic sounds during robot movements was also requested, as the mechanics of the moving robot were too quiet.

#### 4. Discussion

With our research, we targeted design, development, and evaluation of an inclusive VUI prototype. This was installed in a handover robot by following the guidelines of the UCD process and involving BVI, blinded, and sighted users. The results of this process are user requirements, a VUI prototype designed and developed on the basis of these requirements, and evaluations by BVI, blinded, and sighted users. These results are condensed into a set of recommendations for inclusive VUI development and dialog design. Finally, the success of this VUI prototype is discussed.

##### *4.1. Recommendations for Inclusive VUI Design for Robot Handover under Visual Restriction*

Overall, the requirement analysis highlighted the importance of communicating with the robot in general and gave insights into mandatory information for the VUI to ensure a successful and satisfying object handover under visual restrictions. The results also showed that subjects adjusted their communication style from a familiar conversation to a command mode, showing that subjects expected the robot to understand less than a human partner. Additionally, including experts during the design process added value, as they formulated requirements separately from novice first-time users and experienced users. Furthermore, testing of the VUI as a standalone phone-based prototype by experts highlighted insecurities in situations of VUI silence. These expressed insecurities revealed dialog phases that caused insufficient situational awareness due to missing visual and environmental sound cues (e.g., sounds from a moving robot arm). To address this issue, these dialog phases were further improved by adding VUI outputs. Applying this method, requirements for a wide range of situations (e.g., loud ambient noises) and users could be collected. The requirements were not always free of contradictions, showing the necessity to make careful design decisions for VUI design dependent on user group and context. Still, to enable an inclusive design, the VUI has to be adapted towards the weakest user. However, the result of lower error rates of BVI people during object handover compared to blinded people shows that, in the examined use case, this is not necessarily the group of BVI people. A possible reason is that BVI people have a set of specific handover strategies [18], which blinded users probably do not have.

The collected requirements are condensed into general and user-group specific recommendations for inclusive VUI design considering (a) handover success and (b) user satisfaction with social characteristics.

(a) Necessary information for handover success: Users need certain information conveyed by the VUI to successfully perform a handover interaction under visual restrictions.

- Actively provide status and position information during passive user phases. This confirms the literature, as recalling status information from an assistive system was found to be key information and a goal of VUIs [17]. Depending on the complexity of information, this can be done by continuous verbal communication about current robotic action or unique sounds instead of unnecessary voice outputs. If the mechanics of the robot are noiseless, offer generated artificial sounds to facilitate position estimation in case of visual restrictions.
- Before handover, provide precise information to the user about the object, its orientation, and its position. Users required that orientation of an object in the robot hand should be communicated, which was already mentioned by [4], along with the recommendation to orient objects corresponding to human conventions. As [18] found before, unusual properties of the object should be announced, with users mentioning properties such as being hazardous, sharp, sticky, or wet. Additionally, this research found that the position of the object should be communicated by giving clockwise (e.g., “The object is at the 9 o’clock position in front of your hand.”) or front/side direction in relation to receiver’s hand (e.g., “The object is in front, slightly left from your hand.”).
- Announce and request active confirmation of users for their readiness to handover, reinsurance of safe grasp, and for releasing. Already, Ref. [18] postulated communicating intention to handover and reinsurance of safe grasp, which was found in this research as well, complemented by the requirement to request active confirmation for releasing.
- Use sounds to show readiness for voice input and to immediately confirm command or input processing. Users require immediate confirmation of command processing and input confirmation to avoid user insecurity as to whether they are heard.
- Allow users to interrupt robot speech, but avoid interrupting user speech. Prioritized short commands for interruption of robot speech should be possible at all times, e.g., for emergency call or correcting input. Shortening time latencies between two speech outputs can prevent unintended interruptions by users.
- Enable good natural language recognition by understanding easy, but grammatical correct inputs. This further contributes to anthropomorphism and personification of the social robot [22], as do the features of speech interaction described next.

(b) Considering social characteristics for user satisfaction: In line with literature, social manners and etiquette, aesthetics, and anthropomorphism, as well as individualization, need to be satisfied, supporting the perception of the handover robot as a social partner. This is because these robots engage users in an interpersonal manner to achieve positive outcomes and facilitate an interpersonal relationship through the coordination of behavior, as well as through verbal and nonverbal communication [3].

- When designing voice output, pay attention to social manners and etiquette. Output should be kind and nice, but stay task-oriented. There should be no chattering or colloquial language, and no excessive use of the word “please”. For voice input, allow a commanding style, but also kind communication, such as the possibility to thank the robot with an appropriate answer.
- Include anthropomorphic features in social robot communication. Users like the introduction of a robot by name and expect it to react when calling this name. They require warm and natural communication, with only slight preference for female voice. Hence, if possible, offering female and male voices is recommended, which users can choose according to individual preference.
- Allow individualization to address differing needs of diverse user groups. Expert and novice modes, the latter with more extensive communication, ensure a continuous understanding of the situation for new or BVI users. In line with [13], which showed that BVI people often use screen readers with synthesized robotic-sounding voices and a higher than natural speaking output rate, voice output tempo should be adjustable.

An acoustic guidance system could provide sounds at the robot gripper to facilitate localizing the object for BVI users.

#### 4.2. Recommendations on Dialog Flow and Commands

Second, in the results, a VUI prototype for handover tasks under visual restrictions was specified. Its structure can be condensed into recommendations on dialog flow and commands that verbally coordinate the complete process of handover tasks under visual restriction for diverse user groups.

The specified VUI prototype for robot handover under visual restriction included verbal communication for nine serialized parts. Mapping them to previously known process phases during handover [4,8], it becomes apparent that those previously known phases do not cover the complete process of handover tasks. The first two parts of the VUI prototype—‘welcoming and start of interaction’ and ‘obtaining order information’—are neglected in previous handover process phases. However, for BVI people, they are especially important, as they are verbal in nature and can offer important presets, facilitating the handover process later on. For example, verbal welcoming by name can help BVI users in locating the robot and perceiving the robot as a social partner. Furthermore, during ordering, offering midair handover versus placing the object on a table can make it easier for BVI people to locate and take the object. Hence, dialog for handover tasks under visual restrictions should start at this point. Ref. [4] start their handover process phases with the robot’s searching and grasping of the object. In line with the general recommendations listed in Section 4.1, VUI dialog should give status information about the robot during searching and grasping the object. The next handover process phase is ‘transporting the object to the receiver (approaching)’, which corresponds with the structure of the VUI prototype. Next, the handover process phase signaling readiness to handover is supplemented by a confirmation loop for the intention to hand over in the VUI dialog. The last handover process phase is actual handover interaction, with releasing the object and ending of the interaction. For sighted people, this phase is strongly based on visual information [7,8]. For users under visual restriction, temporal and spatial synchronization of actions needs to be supported acoustically. Thus, VUI dialog structure in this handover process phase should be more detailed, containing five parts: a confirmation loop for intention to hand over; actual approach to the user’s hand, including an expression of required user actions; information regarding orientation and features of handover object and an expression of required user action; confirmation loop for safe grasp during handover; and final announcement of forthcoming hand opening.

To offer shortcuts in VUI dialog or to address urgent user needs during a handover process, prioritized short commands should be included, interrupting the current robot action. The short commands ‘go on’ and ‘back to base’ offer shortcuts to speed up or cancel the handover process. The short commands ‘pause’ and ‘emergency stop’ enable users to react to unexpected situations during the handover process. Finally, the short command ‘call for help’ is necessary for BVI users, for example, in case of functional failures or not knowing where the robot is or what it does.

#### 4.3. Success of VUI Prototype

Third, in the results, the VUI prototype was evaluated with three main success criteria: enabling successful handover coordination; users’ satisfaction with task support and social characteristics; and inclusive applicability for all users, despite not involving all user groups in every stage of the UCD process.

Enabling successful handover coordination: The VUI prototype enabled a successful handover. The error rate was very low, with only 1,8% unsuccessful handovers.

Users’ satisfaction with task support: Nevertheless, users’ evaluation regarding the support ability of the social robot was mixed. On the one hand, the speech interaction was characterized by positive attributes such as “good”, “pleasant”, “appropriate for the intended purpose”, and “relatively clear”. All information given by the VUI were



considered important by BVI people. Blinded and sighted people mentioned confirmation of intended actions of the robot, the VUI's explicit communication of expected subjects' behavior, and order confirmation as especially important. Of this information, BVI people only mentioned communication of expected behavior, and additionally highlighted the necessity of communicating the current task and location of the robot. However, it is perfectly possible that this group took explicit confirmation regarding some information for granted and not worth mentioning. Furthermore, blinded and sighted people evaluated speech outputs during passive phases as redundant, which BVI people did not, possibly showing different ways of thinking between BVI people and people with overall good visual abilities. Additionally, within all groups, subjects mentioned performance problems of the VUI because speech inputs of the subjects were not correctly understood and subjects had to repeat or correct the processed order information. Although there were no significant group differences in ordering and handover time, the speed of the VUI was evaluated differently between user groups. Whereas some BVI subjects evaluated the VUI as being far too slow, this was not the case for blinded subjects. One possible explanation is that BVI people have been found to have greater serial memory, enabling them to correctly interpret and retrieve longer and more complex speech responses [19].

**Users' satisfaction with social characteristics:** In line with [3,13,22], we evaluated social characteristics and emotional value of the interaction with the VUI, as being regarded as a social partner is a crucial precondition of satisfaction with social robots.

**Users' evaluation regarding social characteristics of the social robot** was positive. In the SASSI questionnaire, likeability was evaluated best. In qualitative feedback, the general mode of expression regarding social manners was described as "friendly", "helpful", "formal", "clear", and "comprehensible". Only a few users felt the speaker voice was "artificial", "monotonous", and "jerkily", or that the robot talked too much.

**Inclusive applicability for all users:** The VUI was found to be inclusively applicable for BVI, blinded, and sighted users, despite not involving all user groups in every stage of the UCD process. Handover success, ordering and handover time, and most subjective measures in SASSI questionnaire found no difference between user groups. Only the subscale Annoyance resulted in a significant difference, with sighted subjects evaluating the VUI as significantly more annoying than blinded subjects. The reason for this result could be that the speech outputs during passive phases of subjects were perceived as redundant by sighted users. In accordance with this, more than half of all trials with expert mode were performed by sighted subjects, although as many blinded as sighted subjects tried the expert mode. However, only sighted subjects stated that they prefer the expert mode because it subjectively shortens the necessary time for interaction.

To conclude, it is worth applying UCD with diverse user groups; at the same time, involving users of the diverse groups selectively into the development process was successful.

#### *4.4. Limitations and Future Research*

The presented research also has some limitations. Only short-term interaction with the inclusive VUI for robot handover tasks under visual restriction was examined, so that requirements and recommendations for long-term use cannot be derived. Nevertheless, in the near future, the mentioned interactions without prior experience are the more probable use case of social handover robots. For example in hospitals or hotels, which are unfamiliar situations and environments for BVI people, a handover robot could provide reasonable help. For wide long-term power use of such systems at home, to date, social robots are too expensive.

Further, in the UCD process, user involvement was not varied systematically. The whole UCD process was only undergone once. Therefore, from this research, we cannot say for sure if involving all user groups during each UCD step would have led to different results. Additionally, external conditions such as the COVID-19 pandemic limited possibilities for user involvement, leading to unusual choices of methods such as online workshops instead of roleplay workshops. Nevertheless, in line with [27], successful applicability of

the VUI prototype implies that the designers of the prototypes were sufficiently sensitized to the requirements of BVI people and applied them sufficiently in stages without BVI user involvement.

The present study's findings should be replicated and complemented in other social robots. For example, the findings were only tested with a small assistant robot. Therefore, before applying them in a heavy-load robot in an industrial context, the transferability should be tested. Furthermore, it can be expected that social manners are culture-dependent. Therefore, in another culture, user requirements and some details of dialog content regarding social manners might differ and need to be adapted. Some interdependencies of dialog with the technical implementation of handover functionality should be tested, especially regarding BVI people. They might perceive handover speed differently than sighted people and require differing speed-dependent information. Furthermore, adding haptic feedback could be helpful for BVI people, e.g., an airstream directed at the user's hand when approaching—the human, to support object location.

## 5. Conclusions

The presented research contributes to the development of social robots in different aspects. Knowledge regarding requirements for robotic support during the complex process of handovers with BVI people was developed, to make sure that social robotic assistance becomes accessible for this group. More specific, a voice dialog covering the whole process of a robot-to-human handover with and without visual restrictions was developed, which can be included in existing or newly developed robots with handover tasks. This dialog was tested by users in an evaluation study. It enabled successful robot handover, and quantified satisfaction with the voice interaction was positive, both for sighted and BVI people. Finally, it was demonstrated that, at least for inclusive VUI development of robotic handover, user involvement of disabled users is not necessary during every development stage to accomplish successful applicability. Overall, a speech interface was demonstrated successfully, supporting a handover interaction strategy for potentially hazardous objects between a robot and a human under visual restriction. This contributes to the basics of successfully designing and implementing an inclusive handover function, independent of the applied robotic system. It makes robotic handover inclusively accessible to BVI people, as well as to working situations in the production sector that cause restrictions in sight. Moreover, it also contributes to improving robot handover for sighted people, as verbal communication increases anthropomorphism of a social robot and enables users to feel confident during robotic handovers. Therefore, this research made a valuable contribution to the development of accessible and cross-applicable social robots.

**Author Contributions:** Conceptualization, D.L. and F.L.; methodology, D.L., F.L. and P.K.; validation, D.L. and F.L.; formal analysis, D.L. and F.L.; investigation, D.L., F.L. and P.K.; data curation, F.L. and D.L.; writing—original draft preparation, F.L. and D.L.; writing—review and editing, P.K., A.D. and A.C.B.; visualization, D.L. and F.L.; supervision, D.L. and F.L.; project administration, D.L. and F.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by German Federal Ministry of Education and Research, grant number 16SV7969K. The article processing charges for open access publication were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 491193532 and the Chemnitz University of Technology.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to restricted consent provided by subjects on the use of personal data pursuant to the General Data Protection Regulation in the European Union and the European Economic Area (EU General Data Protection Regulation).

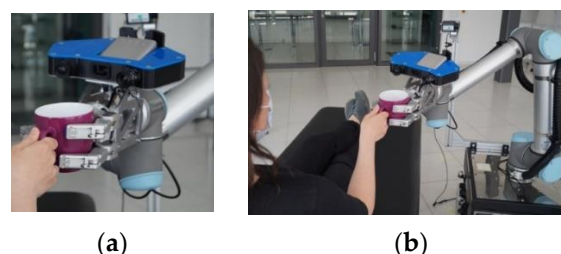
**Acknowledgments:** We thank Daniel Wimpff from Sikom Software GmbH for the technical implementation of the speech interface. Additionally, we thank the researchers of the Department of Cognitive Human–Machine-Systems at Fraunhofer IWU Chemnitz for the realization of the robotic demonstrator. Furthermore, we thank YOUSE GmbH for the cooperation in conceptualization, carrying out the separate UCD steps, and analyzing qualitative data. Many thanks to Theresa Langer for drawing the graphical abstract.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Appendix A. Description of the Robot for Handover under Visual Restrictions

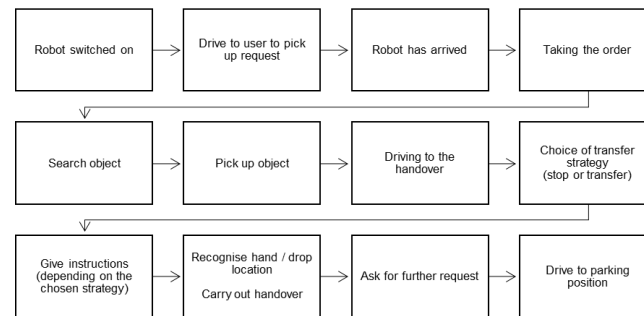
The robot was targeted to support people in unfamiliar situations where people are not able to use their everyday strategies during handover tasks. An exemplary application scenario is assisting BVI patients in a hospital or nursing home. Additionally, workers in industrial productions sites are strongly fluctuating and workers have to familiarize themselves to changing work stations quickly. There, an assistive handover robot could contribute to more ergonomic assembling by successively handing over different required tools or assembling components in the correct order, meaning that the worker doesn't need to turn around and remains focused on the assembling task.

According to the taxonomy of [22], the robot can be described as a social robot (field of application) that is targeted for a human collaborator (human role) in a one-to-one interaction (team composition). The robot has an embodied shape and the interaction is intended to take place in the field (exposure to robot/setting). Regarding the robot's task specification, handing over objects requires information exchange, precision for hand recognition, and transportation and manipulation of objects. The robot receives acoustic information due to speech recognition and optical sensor information (communication channel input). As handovers take place under visual restriction, only the acoustic output of the robot can be used by users (communication channel output). The robot physically approaches the user (physical proximity). Both interaction partners act synchronously in the same area (temporal proximity). The degree of robot autonomy can be described as overall high regarding information acquisition, information analysis, decision-making, and action implementation. Finally, regarding robot morphology, the robot appears both anthropomorphic (camera system seems like a head with eyes) and technical (UR-10 type robot). The same applies to communication due to natural grammar usage but technical, standardized dialog of the VUI and several speech input and outputs which have to be confirmed by the user. The movement is technical in nature and the robot is framed as a technical support system (context information). Figure A1 shows the robot to support handover tasks under visual restrictions.



**Figure A1.** Robot assistant for handover tasks under visual restrictions. (a) Detailed picture of the robot hand reaching for a cup and (b) entire robot with AGV platform, lightweight robot arm with Barret Hand, and sensor systems.

## Appendix B. Flowchart Draft



**Figure A2.** Flow chart—early draft of a rough sequence of a successful robotic handover interaction. Each step should be supported by VUI dialog.

## Appendix C. Example of the Final VUI Dialog

**Table A1.** Example of the final VUI dialog for (1) midair robot-to-human handover and (2) placing of a cup. Robot speech output in italics, user speech output written in bold.

	<p>1: User is awakening the robot by phone call.</p> <p>2: <i>Hello, my name is [name]. I can bring and hand over objects to you or pick up them up from you. Thereby I support two types of handover. Either directly into your hand or indirectly, by placing the object on a table. You can also interrupt me while I am speaking.</i></p> <p>3: Manual following on demand by saying 'Manual' and explaining that the robot asks for three information—object, handover direction, and handover modality—and introduction of available short commands.</p> <p>4: <i>How can I help you? [e]</i><sup>2</sup></p> <p>5: <b>Hello, [name]. Please bring me a cup of coffee. [e]</b> [2 out of 3 necessary information given]</p> <p>6: <i>I'm supposed to bring you a cup. Shall I hand you the cup directly or put it on the table? [e]</i></p>
<p><b>7. Directly into my hand. [e]</b></p> <p>8. <i>I'm supposed to bring you a cup of coffee. The transfer should take place directly from hand to hand. Did I understand the task correctly? [e]</i></p>	<p><b>7. On the table, please. [e]</b></p> <p>8. <i>I'm supposed to bring you a cup of coffee. The transfer should be done indirectly by placing it on the table. Did I understand the task correctly? [e]</i></p>
<p>9. <b>Yes [Yes, you did]. [e]</b></p> <p>10. <i>I am looking for the cup.</i></p> <p>11. <i>I have localized the cup.</i></p> <p>12. <i>I am moving to the cup and getting it for you.</i></p> <p>13. <i>I am opening my hand.</i></p> <p>14. <i>I am moving my arm to the object.</i></p> <p>15. <i>I am closing my hand.</i></p> <p>16. <i>I gripped the cup safely.</i></p> <p>17. <i>I am moving back a little.</i></p>	
<p>18. <i>I am approaching you for the handover, but still keeping a little distance. [e]</i></p> <p>19. <i>Confirmation required. I would now like to hand the cup directly into your hand. Say 'Go' as soon as you're ready to go. [e]</i></p> <p>20. <b>Go. [e]</b></p> <p>21. <i>Okay. Please hold out your hand to me so that I can navigate towards you. [e]</i></p> <p>22. <i>I hold the cup directly in front of your hand with the handle first. Follow the light air flow <sup>1</sup> with your hand and reach for it. [e]</i></p> <p>23. <i>Confirmation required. Say 'Go' if I may now open my hand and let go of the cup. [e]</i></p> <p>24. <b>Okay, go. [e]</b></p> <p>25. <i>I'm letting go now. [hand is opening after finishing the sentence] [e]</i></p>	<p>18. <i>I am going to you now to put the item on the table.</i></p> <p>[only robotic sounds perceptible]</p> <p>19. <b>Current status?</b></p> <p>20. <i>I am going to you now to put the item on the table.</i></p> <p>21. <i>I put the cup down on the table and let go now. [e]</i></p>

Table A1. Cont.

- 
26. **Thanks.**  
 27. *You are very welcome.*  
 28. *I am moving back a little.*  
 29. *Is there anything else I can do for you?*  
 30. **No, not at the moment.**  
 31. *I am driving back to my base station.*  
 31. *I have reached my base station. [e]*
- 

<sup>1</sup> The robot was equipped with a simple haptic feedback, implemented as a continuous air flow, located at the robotic hand to ease grasping by users. <sup>2</sup> Voice output and confirmation loops also occurring during expert mode are marked with [e].

## References

- Richert, A.; Shehadeh, M.; Müller, S.; Schröder, S.; Jeschke, S. Robotic Workmates: Hybrid Human-Robot-Teams in the Industry 4.0. In Proceedings of the International Conference on E-Learning, Kuala Lumpur, Malaysia, 2–3 June 2016.
- Müller, S.L.; Schröder, S.; Jeschke, S.; Richert, A. Design of a Robotic Workmate. In *Digital Human Modeling. In Applications in Health, Safety, Ergonomics, and Risk Management: Ergonomics and Design*, 1st ed.; Duffy, V., Ed.; Springer: Cham, Switzerland, 2017; Volume 1, pp. 447–456. [\[CrossRef\]](#)
- Breazeal, C.; Dautenhahn, K.; Kanda, T. Social Robotics. In *Springer Handbook of Robotics*, 2nd ed.; Siciliano, B., Khatib, O., Eds.; Springer: Cham, Switzerland, 2016; Volume 1, pp. 1935–1972. [\[CrossRef\]](#)
- Cakmak, M.; Srinivasa, S.S.; Lee, M.K.; Kiesler, S.; Forlizzi, J. Using spatial and temporal contrast for fluent robot-human handovers. In Proceedings of the 6th International Conference on Human-Robot Interaction, Lausanne, Switzerland, 6–9 March 2011.
- Aleotti, J.; Micelli, V.; Caselli, S. An Affordance Sensitive System for Robot to Human Object Handover. *Int. J. Soc. Robot.* **2014**, *6*, 653–666. [\[CrossRef\]](#)
- Koene, A.; Remazeilles, A.; Prada, M.; Garzo, A.; Puerto, M.; Endo, S.; Wing, A. Relative importance of spatial and temporal precision for user satisfaction in Human-Robot object handover Interactions. In Proceedings of the 50th Annual Convention of the AISB, London, UK, 18–21 April 2014.
- Langton, S.; Watt, R.; Bruce, V. Do the eyes have it? Cues to the direction of social attention. *TiCS* **2000**, *4*, 50–59. [\[CrossRef\]](#)
- Käppler, M.; Deml, B.; Stein, T.; Nagl, J.; Steingrebe, H. The Importance of Feedback for Object Hand-Over Between Human and Robot. In *Human Interaction, Emerging Technologies and Future Applications III. IHET 2020. Advances in Intelligent Systems and Computing*; Ahram, T., Taiar, R., Langlois, K., Choplin, A., Eds.; Springer: Cham, Switzerland, 2021; Volume 1253, pp. 29–35. [\[CrossRef\]](#)
- Cochet, H.; Guidetti, M. Contribution of Developmental Psychology to the Study of Social Interactions: Some Factors in Play, Joint Attention and Joint Action and Implications for Robotics. *Front. Psychol.* **2018**, *9*, 1992. [\[CrossRef\]](#)
- Strabala, K.; Lee, M.K.; Dragan, A.; Forlizzi, J.; Srinivasa, S.S.; Cakmak, M.; Micelli, V. Toward seamless human-robot handovers. *JHRI* **2013**, *2*, 112–132. [\[CrossRef\]](#)
- Bdiwi, M. Integrated sensors system for human safety during cooperating with industrial robots for handing-over and assembling tasks. *Procedia CIRP* **2014**, *23*, 65–70. [\[CrossRef\]](#)
- Costa, D.; Duarte, C. Alternative modalities for visually impaired users to control smart TVs. *Multimed. Tools Appl.* **2020**, *79*, 31931–31955. [\[CrossRef\]](#)
- Branham, S.M.; Roy, A.R.M. Reading Between the Guidelines: How Commercial Voice Assistant Guidelines Hinder Accessibility for Blind Users. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19), Pittsburgh, MD, USA, 29 October–1 November 2019.
- Bonani, M.; Oliveira, R.; Correia, F.; Rodrigues, A.; Guerreiro, T.; Paiva, A. What My Eyes Can't See, A Robot Can Show Me: Exploring the Collaboration Between Blind People and Robots. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18), Galway, Ireland, 22–24 October 2018.
- Angleraud, A.; Sefat, A.M.; Netzev, M.; Pieters, R. Coordinating Shared Tasks in Human-Robot Collaboration by Commands. *Front. Robot. AI* **2021**, *8*, 734548. [\[CrossRef\]](#) [\[PubMed\]](#)
- Choi, Y.S.; Chen, T.; Jain, A.; Anderson, C.; Glass, J.D.; Kemp, C.C. Hand it over or set it down: A user study of object delivery with an assistive mobile manipulator. In Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan, 27 September–2 October 2009. [\[CrossRef\]](#)
- Leporini, B.; Buzzi, M. Home automation for an independent living: Investigating the needs of visually impaired people. In Proceedings of the 15th International Web for All Conference (W4A '18), New York, NY, USA, 23–25 April 2018.
- Walde, P.; Langer, D.; Legler, F.; Goy, A.; Dittrich, F.; Bullinger, A.C. Interaction Strategies for Handing Over Objects to Blind People. In Proceedings of the Human Factors and Ergonomics Society (HFES) Europe Chapter Annual Meeting, Nantes, France, 2–4 October 2019.



19. Oumard, C.; Kreimeier, J.; Götzelmann, T. Pardon? An Overview of the Current State and Requirements of Voice User Interfaces for Blind and Visually Impaired Users. In Proceedings of the Computers Helping People with Special Needs: 18th International Conference (ICCHP), Lecco, Italy, 11–15 July 2022.
20. Oliveira, O.F.; Freire, A.P.; Winckler De Bettio, R. Interactive smart home technologies for users with visual disabilities: A systematic mapping of the literature. *Int. J. Comput. Appl.* **2021**, *67*, 324–339. [\[CrossRef\]](#)
21. Azenkot, S.; Lee, N.B. 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, New York, NY, USA, 21–23 October 2013.
22. Onnasch, L.; Roesler, E. A Taxonomy to Structure and Analyze Human–Robot Interaction. *Int. J. Soc. Robot.* **2021**, *13*, 833–849. [\[CrossRef\]](#)
23. Jee, E.S.; Jeong, Y.J.; Kim, C.H.; Kobayashi, H. Sound design for emotion and intention expression of socially interactive robots. *Intel. Serv. Robotics* **2010**, *3*, 199–206. [\[CrossRef\]](#)
24. Majumdar, I.; Banerjee, B.; Preeth, M.T.; Hota, M.K. Design of weather monitoring system and smart home automation. In Proceedings of the IEEE International Conference on System, Computation, Automation, and Networking (ICSCA), Pondicherry, India, 6–7 July 2018.
25. Ramlee, R.A.; Aziz, K.A.A.; Leong, M.H.; Othman, M.A.; Sulaiman, H.A. Wireless controlled methods via voice and internet (e-mail) for home automation system. *IJET* **2013**, *5*, 3580–3587.
26. Raz, N.; Striem, E.; Pundak, G.; Orlov, T.; Zohary, E. Superior serial memory in the blind: A case of cognitive compensatory adjustment. *Curr. Biol.* **2007**, *17*, 1129–1133. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Newell, A.F.; Gregor, P.; Morgan, M.; Pullin, G.; Macaulay, C. User-sensitive inclusive design. *Univ. Access Inf. Soc.* **2011**, *10*, 235–243. [\[CrossRef\]](#)
28. Deutsches Institut für Normung. *DIN EN ISO 9241-210 Ergonomie der Mensch-System-Interaktion—Teil 210: Menschzentrierte Gestaltung Interaktiver Systeme*; Beuth Verlag GmbH: Berlin, Germany, 2019.
29. Harris, M.A.; Weistroffer, H.R. A new look at the relationship between user involvement in systems development and system success. *CAIS* **2009**, *24*, 739–756. [\[CrossRef\]](#)
30. Kujala, S. User involvement: A review of the benefits and challenges. *Behav. Inf. Technol.* **2003**, *22*, 1–16. [\[CrossRef\]](#)
31. Ladner, R.E. Design for user empowerment. *Interactions* **2015**, *22*, 24–29. [\[CrossRef\]](#)
32. Philips, G.R.; Clark, C.; Wallace, J.; Coopmans, C.; Pantic, Z.; Bodine, C. User-centred design, evaluation, and refinement of a wireless power wheelchair charging system. *Disabil. Rehabil. Assist. Technol.* **2020**, *15*, 1–13.
33. Vacher, M.; Lecouteux, B.; Chahuara, P.; Portet, F.; Meillon, B.; Bonnefond, N. The Sweet-Home speech and multimodal corpus for home automation interaction. In Proceedings of the 9th Edition of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 26–31 May 2014; European Language Resources Association: Paris, France, 2014; pp. 4499–4506.
34. Miao, M.; Pham, H.A.; Friebe, J.; Weber, G. Contrasting usability evaluation methods with blind users. *UIAIS* **2016**, *15*, 63–76. [\[CrossRef\]](#)
35. Universal Robots (2016) Die Zukunft ist Kollaborierend. Available online: <https://www.universal-robots.com/de/download-center/#/cb-series/ur10> (accessed on 8 December 2021).
36. Hone, K.S.; Graham, R. Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). *Nat. Lang. Eng.* **2000**, *6*, 287–303. [\[CrossRef\]](#)
37. Kocaballi, A.B.; Laranjo, L.; Coiera, E. Understanding and Measuring User Experience in Conversational Interfaces. *Interact. Comput.* **2019**, *2*, 192–207. [\[CrossRef\]](#)
38. R Core Team. R: A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*; Chapman and Hall/CRC: Vienna, Austria, 2019. Available online: <https://www.R-project.org/> (accessed on 16 December 2019).