

AUTOMATIC TASK ASSISTANCE FOR
PERSONS WITH COGNITIVE DISABILITIES
IN BASIC ACTIVITIES OF DAILY LIVING

BY

CHRISTIAN PETERS

Doctoral thesis

submitted to

Faculty of Technology at Bielefeld University

in partial fulfillment of the requirements for the degree of

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

May 2013

Christian Peters: *Automatic Task Assistance for Persons with Cognitive Disabilities in Basic Activities of Daily Living*, © May 2013

ABSTRACT

Persons with cognitive disabilities such as Autistic Spectrum Disorders (ASD) and intellectual disabilities tend to have problems in sequencing and coordinating steps in the execution of basic Activities of Daily Living (ADLs) due to limited capabilities in cognitive functioning. In order to successfully perform basic ADLs, these persons are highly reliant on the assistance of a human caregiver. This leads to a decrease or even a loss of independence for care recipients and imposes a high burden on caregivers. Assistive Technology for Cognition (ATC) aims to compensate for decreased cognitive functions. ATC systems provide automatic assistance in the execution of ADLs by delivering appropriate prompts which enable the user to perform basic ADLs without any assistance of a human caregiver. This leads to an increase of the user's independence and to a relief of caregiver's burden.

In this thesis, we describe the design, development and evaluation of a novel ATC system. The TEBRA (TEeth BRushing Assistance) system supports persons with moderate cognitive disabilities in the execution of brushing teeth by providing audio-visual prompts to the user.

In order to reveal the characteristics of the task and the involved users, we conduct Interaction Unit (IU) analysis, a structured method of task analysis. We iteratively refine the initial design decisions based on the results of IU analysis in intermediate evaluations where we follow a user-centered design: in a Wizard of Oz study, we evaluate the reaction behaviors of persons with cognitive disabilities to system prompts. In an interview study, we ask professional caregivers about appropriate modalities and content of prompts.

We incorporate the design decisions into the implementation of the TEBRA system. A main requirement for the acceptance of an ATC system is context awareness: an explicit feedback from the user is not necessary in order to provide appropriate assistance. We allow for context awareness by implementing a user behavior recognition component which deals with the variations in the execution of behaviors such as different movement characteristics and different velocities: we infer user behaviors based on states of objects involved in the task which we apply in a Bayesian Network classification scheme. A dynamic timing model allows for different velocities of users and adapts to a user's velocity during a trial.

We evaluate a fully functioning prototype of the TEBRA system in a study with persons with cognitive disabilities. The main aim of the study is to analyze the technical performance of the TEBRA system and the user's behavior in the interaction with the system with regard to the main hypothesis: Is the TEBRA system able to increase the independence of users in the execution of brushing teeth?

ACKNOWLEDGMENTS

I am deeply obliged to Cognitive Interaction Technology Center of Excellence (CITEC) for providing me a scholarship.

I would like to thank my supervisors PD Dr. Sven Wachsmuth and Dr. Thomas Hermann for giving me the possibility to work on this exciting and demanding project: Sven and Thomas, I very much appreciated your advice and support at any time during my PhD studies. I also thank my colleagues from the Applied Informatics, the Ambient Intelligence group and the CITEC Central Lab Facilities for providing an enjoyable work climate.

I would like to particularly thank Prof. Jesse Hoey whom I got to know while I was working on my Diploma thesis and who inspired me to conduct research in the field of Assistive Technology for Cognition: Jesse, thank you very much for your advice and practical hints and most of all for your help in organizing and mentoring my research stay in Waterloo, Ontario which was an unforgettable experience.

Furthermore, I would like to thank Prof. Andrew Monk for his support concerning the Interaction Unit analysis, Simon Schulz for his help and support in all issues with regards to the TEBRA system hardware as well as for the integration of the sensor module into the electric toothbrush. Thanks to Jan Moringen and Johannes Wienke for all issues concerning RSB. I would like to thank JProf. Dr. Friederike Eyssel and Dr. Dieta Kuchenbrandt for their help in preparing and implementing the interview study. I would also like to thank Janusz Leidgens, Benjamin Dosch and Anja Durigo for their active support as student assistants in parts of the project and especially in the studies.

I would like to thank our cooperation partner *v. Bodelschwingsche Stiftungen Bethel* and, in particular, the inhabitants of the residential home *Haus Bersaba* who made this project possible by participating highly motivated in the studies. Furthermore, I would like to thank Prof. Dr. Michael Seidel, Dr. Michelina Bocchicchio, Gabriele Hüttemann-Völker, Matthias Volmer, Gudrun Möller, Gisela Knopf and Sabrina Bebermeier for enabling the cooperation and for helping me with organizing and conducting the studies. This thesis would not have been possible without the encouragement of my family and friends. I would like to thank my parents and my brother Michael for their absolute support throughout my whole private and professional life. Thanks to my friends Ilse and Sven who have always listened to my worries regarding the project and beyond.

Finally, I would like to thank my wife Rebekka and my daughter Marla. Rebekka, thanks for loving me as I am, for your advice, your patience and your trust in me and in us. Marla, you are the sunshine of my life. Seeing you happy makes me happy.

Contents

1	INTRODUCTION	1
2	ASSISTIVE TECHNOLOGY FOR COGNITION	5
2.1	What is Assistive Technology for Cognition?	5
2.2	Review of existing ATC	10
2.2.1	PEAT	11
2.2.2	Autominder	11
2.2.3	Archipel	12
2.2.4	COACH	13
2.2.5	Summary	14
2.3	Scenario selection	14
3	METHODS	17
3.1	Environmental perception	17
3.1.1	Single-layer approaches	18
3.1.2	Hierarchical approaches	21
3.2	Planning and decision making	21
3.3	Actuation and prompting	25
3.3.1	Delivery of prompts	26
3.3.2	Content of prompts	26
3.4	Summary	27
4	DESIGN OF THE TOOTH BRUSHING SYSTEM	29
4.1	In-situ observations and Interaction Unit analysis	29
4.2	Setup	35
4.2.1	Environmental sensors	36
4.2.2	Sensor-enhanced toothbrush	37
4.3	Wizard of Oz study	38
4.4	Interviews with caregivers	41
4.5	Data recording with complete sensors	43
4.6	Summary	45
5	TEBRA SYSTEM	47
5.1	Overall system architecture	47
5.2	User behavior recognition	50
5.2.1	Feature extraction and discretization	51
5.2.2	Bayesian network classification	55
5.2.3	Quantitative results	59
5.2.4	Enhancements in discretization	61
5.3	Planning and decision making	62
5.3.1	Tracking the user's overall progress	63
5.3.2	Dynamic timing model	65

5.3.3	Prompt selection	70
5.4	Summary	74
6	STUDIES	77
6.1	Study with regular users	77
6.1.1	System performance	78
6.1.2	Appropriate prompting	80
6.1.3	Technical enhancements	81
6.2	Study with persons with cognitive disabilities	83
6.2.1	Study design	84
6.2.2	Results and discussion	85
7	CONCLUSION AND OUTLOOK	105
A	INTERVIEW MANUAL FOR THE STUDY WITH CAREGIVERS	109
B	DOCUMENTS FOR THE STUDY WITH TARGET GROUP USERS	113
B.1	Sheet of information	113
B.2	Declaration of consent	115
B.3	Questionnaire	117
	BIBLIOGRAPHY	119

List of Figures

Figure 1	Taxonomy of activity recognition.	18
Figure 2	Milestones in the design and development of the TEBRA system.	30
Figure 3	Example image of the in-situ observation.	30
Figure 4	Washstand setup equipped with sensor technology.	35
Figure 5	View perspectives of the frontal camera (left image) and the overhead camera (right image).	37
Figure 6	Electric toothbrush and x-imu sensor.	38
Figure 7	Graphical user interface (GUI) used in the Wizard of Oz study.	39
Figure 8	Selection of prompts which were presented to the caregivers in the interviews.	42
Figure 9	Pictogram prompts and example of a video prompt used in the TEBRA system.	44
Figure 10	Components and data flow in the TEBRA system.	48
Figure 11	Functional overview of the TEBRA system.	49
Figure 12	Overview of the user behavior recognition component.	51
Figure 13	Feature extraction and discretization in the TEBRA system.	52
Figure 14	Bounding box hypotheses and predefined image regions.	53
Figure 15	Bayesian networks with IU-based, Naive Bayes and Holistic structure.	56
Figure 16	Distinction between <i>yes_sink</i> and <i>yes_face</i> for variable <i>brush_movement</i> .	63
Figure 17	Initial and final state of the state space including example transition.	65
Figure 18	Subdivision of user behaviors in the phases validation, pre-effect and post-effect.	65
Figure 19	Finite State Machine modeling the timing characteristics of user behaviors.	66
Figure 20	Cumulative distribution function (CDF) of behavior <i>rinse_mouth_wet</i> for velocity <i>fast</i> showing the different timing parameters and according phases of the FSM.	69
Figure 21	Ordering constraint graph depicting partial orderings of user behaviors in the brushing task.	73

Figure 22	Number of independent steps per trial for individual users.	87
Figure 23	Spatial variance in the execution of <i>clean_brush</i> .	89
Figure 24	Spatial variance in the execution of <i>use_towel</i> .	90
Figure 25	Two users leaning over heavily in <i>brush_teeth</i> .	90
Figure 26	Distribution of classification rates of <i>brush_teeth</i> over trials for individual users.	91
Figure 27	Internal states of the TEBRA system in a section of a trial of user 5.	94
Figure 28	Internal states of the TEBRA system in a section of a trial of user 5.	95
Figure 29	Internal states of the TEBRA system in a section of another trial of user 2.	96
Figure 30	Average number of prompts in the SYS scenario and reaction behavior for individual users.	97
Figure 31	Ratio of semantically correct system prompts and reaction behavior for individual users.	98
Figure 32	Ratio of correct reactions to semantically correct pictogram and video prompts for individual users.	100
Figure 33	Results of the questionnaire: helpfulness and acceptance of the TEBRA system according to users and caregivers.	102

List of Tables

Table 1	Classification of higher-level cognitive functions according to the ICF definition.	8
Table 2	Overview of state-of-the-art ATC systems reviewed in this thesis.	16
Table 3	Results of the IU analysis for brushing teeth.	32
Table 4	<i>Behavior</i> and <i>progress</i> variables extracted from the environmental configuration.	33
Table 5	Preconditions and effects of user behaviors.	34
Table 6	User's reactions in the Wizard of Oz study.	40
Table 7	Durations of user behaviors in the dataset with regular users.	45
Table 8	User behaviors and relevant <i>behavior</i> variables according to IU analysis.	57

Table 9	Classification rates of the Bayesian networks with different structures. 59
Table 10	Confusion matrix for BN with <i>IU-based</i> structure in the Bayesian filtering approach. 60
Table 11	Confusion matrix for BN with <i>Naive Bayes</i> structure in the Bayesian filtering approach. 61
Table 12	Parameters of the dynamic timing model in seconds for user behaviors in the different velocities. 70
Table 13	Set of ordering constraints and according causal links calculated by the partial-order planner. 72
Table 14	Classification rates of user behaviors in %. 78
Table 15	Values indicating the performance of the system in the <i>free</i> and <i>collaborative</i> scenario. 79
Table 16	Plots showing the number of independent steps in the CG and SYS scenario. 86
Table 17	Average number of independent steps in the CG and SYS scenario for individual users. 88
Table 18	Classification rates of user behaviors in %. 89
Table 19	Classification rates of user behaviors of eleven trials that have a classification rate of <i>brush_teeth</i> above 75%. 93
Table 20	Minimum, maximum and average duration of user behaviors. 93

INTRODUCTION

The human life expectancy will increase in the following decades: in Germany, the average life expectancy at birth will rise from 73.4 years in 1980 to about 81.4 years in 2020¹. The main reason for this increase in industrialized countries “is a reduction in death rates among the elderly” [115, p. 1]. Main factors are improved healthcare services ranging from medical provision to successful treatment of various illnesses. Consequently, the life expectancy of persons with cognitive disabilities increases likewise [117] resulting in longer healthcare provision. The number of persons with age-dependent cognitive disabilities such as dementia in general and Alzheimer’s disease in particular increases as well: in Western Europe, the proportionate increase of persons with dementia will be 43% from 2001 to 2020 [29]. Such effects lead to higher economic costs for healthcare provision. Due to the demographic shift to an aging population, the burden of covering the expenses and providing high-quality healthcare is distributed amongst a decreasing number of people in the working age population.

Persons with cognitive disabilities form a primary group of receiving healthcare due to their limited capabilities in cognitive functioning such as perception, reasoning and remembering [32]. Problems related to this functioning appear in a human’s daily routine where the successful execution of Activities of Daily Living (ADLs) is an integral part of an autonomous and self-determined life. ADLs refer to everyday tasks which can be distinguished into two categories: *basic* and *instrumental* ADLs. Basic ADLs involve aspects of fundamental functioning and include tasks such as eating, dressing or personal hygiene. Instrumental ADLs are related to an independent life as part of a community including financial management, housework and transportation.

A major problem for persons with cognitive disabilities in the execution of ADLs is task sequencing. Task sequencing refers to the ability to decompose tasks into sub steps. For a successful execution of the overall task, the sub steps need to be combined in an appropriate order. For most tasks such as hand washing, tooth brushing and dressing, the sub steps can be combined in a flexible way which allows for different ways of task execution. In a dressing task, for example, a user might put on the left sock first and the right sock afterwards, or vice versa.

Flexibility in task execution imposes a high risk of erroneous behavior for persons with cognitive disabilities: users forget steps or get stuck in

Activities of Daily Living (ADL)

Basic ADL

Instrumental ADL

Task sequencing

¹ United Nations Statistics Division, [http://data.un.org/Data.aspx?q=life+expectancy&d=PopDiv&f=variableID:68;crID:276;timeID:107,108,109,110,111,112,113,114&c=2,4,6,7&s=_crEngNameOrderBy:asc,_timeEngNameOrderBy:desc,_varEngNameOrderBy:asc&v=1]; accessed at 05-April-2013]

task execution due to limited capabilities in cognitive functioning. In such cases, an external intervention of a human caregiver is necessary for a proper task execution. Hence, an inability of executing ADLs leads to a decrease or even a loss of independence and makes persons with cognitive disabilities highly dependent on a human caregiver. According to Blair, a high dependence on a caregiver “is associated with mental health problems such as low self-esteem [...], a negative state of well-being [...], and problems with mood” [14, p. 2]. Furthermore, an inability of performing ADLs might impose security risks for the well-being of persons: for example, a person with Alzheimer’s disease aims to prepare tea, but forgets to pour water into the kettle. Turning on a kettle without any water inside is a potential fire hazard. Professional caregivers as well as informal caregivers such as family members worry about the well-being of the care recipients. This leads to an emotional burden of caring persons which might result in chronic stress and consequential diseases [59].

*Assistive Technology
for Cognition (ATC)*

Assistive Technology for Cognition (ATC) refers to technical interventions which compensate for decreased or missing cognitive capabilities by providing prompts which assist the user in the execution of ADLs. The application of ATC aims at increasing the independence of persons with cognitive disabilities from a human caregiver. This leads to an increase in self-esteem and self-determination in a care recipient’s life and, furthermore, to a relief of caregiver burden due to the prolonged independence of the care recipient [97].

The development of ATC systems is not entirely new. For example, the COACH system assists persons with dementia in the task of hand washing [39]. The Archipel system supports persons with intellectual disabilities in meal preparation [7]. The main goal of such systems is to foster the independence of the user by providing appropriate prompts when necessary for a successful task execution. A prompt is necessary in three situations: firstly, a person might forget a step in the overall task which leads to inappropriate follow-up behaviors. For example, a user rinses hands in a hand washing task without having taken soap first. Secondly, a person might not be able to terminate a sub step of the task due to obsessive behavior. Thirdly, a person is not able to focus on the task and loses track of the overall progress due to environmental distractions such as noise. In these situations, a prompt is necessary to assist the user in task execution.

Context awareness

Context awareness enables a system to detect such situations without explicit feedback of the user about completed steps: a context aware system infers a user’s current behavior as well as the overall progress in the task based on sensory information obtained in the environment. Existing systems differ in the amount of context awareness: for example, context awareness in the COACH system is limited since the behaviors of users are inferred based on a simple heuristic involving the user’s hands. The Archipel system provides context awareness for single behaviors, but is

not able to track the user's overall progress in the task. The implementation of a context-aware behavior is difficult since an ATC system needs to deal with the huge spatial and temporal variance which persons with cognitive disabilities show in task execution. In this thesis, we refer to spatial and temporal variance in the context of a user's behavior in ATC: spatial variance refers to differences in the execution of behaviors due to different motor abilities which result in different movement characteristics amongst individual user. Temporal variance denotes differences in the velocities of task execution which may vary greatly between individual users. For example, one user might perform sub steps of a task very slow, but another user might be very quick in execution.

Spatial and temporal variance

In this thesis, we describe the design, development and evaluation of a novel context-aware ATC system which is robust with regard to spatial and temporal variance of users: the *TEBRA* system (TEeth BRushing Assistance system) assists persons with cognitive disabilities in the execution of *brushing teeth*. Brushing teeth is an important basic ADL since (1) disregarding oral hygiene can lead to severe medical problems and (2) persons with cognitive disabilities usually have problems with brushing teeth due to the flexibility and complexity of the task: brushing teeth involves several objects such as paste and brush which are used in different sub steps during the task. The sub steps can be combined in a flexible way for successful task execution.

TEBRA system

The target group of users are persons with moderate cognitive disabilities such as behavioral disorder, intellectual disabilities and Autistic Spectrum Disorder. We cooperate with the residential home *Haus Bersaba* belonging to *v. Bodelschwingsche Stiftungen Bethel*², a clerical foundation in Bielefeld, Germany. 35 persons with mild to moderate cognitive disabilities live in *Haus Bersaba* and receive permanent care by professional caregivers including assistance in brushing teeth: a caregiver stands beside the person and assists during the brushing task by providing verbal and visual prompts. In a study with target group users, we evaluate the technical system performance including the recognition and tracking of a user's behaviors in the overall task as well as the appropriateness of prompts. Furthermore, we analyze the user's reactions to prompts and discuss aspects of usability and acceptance of the *TEBRA* system.

The contributions of this thesis are the following:

- We design and implement the *TEBRA* system, a novel ATC system which assists persons with moderate cognitive disabilities in the execution of brushing teeth.
- A main challenge in the development of an overall system is to combine and to coordinate components in a way that the system assists users in their individual way of performing the task. We deal with

² [www.bethel.de]

the huge spatial variance of persons with cognitive disabilities in a behavior recognition component where we infer a user's behaviors based on the states of objects manipulated during the behaviors.

- We also deal with the huge temporal variance in task execution by using a dynamic timing model which allows for different velocities of users by explicitly modeling a user's velocity.
- We evaluate a fully functioning prototype of the TEBRA system in a study with target group users. We analyze the rich corpus of interaction data obtained in the study with regard to the main hypothesis that the TEBRA system increases the independence of users in the brushing task by providing appropriate assistance in the overall task.

This thesis is structured as follows: chapter 2 gives an overview of *Assistive Technology for Cognition* including a brief history of the field and state-of-the-art systems. Chapter 3 provides a methodological overview of the two main components of ATC which are *environmental perception* and *planning and decision making*. Chapter 4 highlights the user-centered design process of the TEBRA system. We describe in-situ observations with the target group, the construction of a washstand setup in which the TEBRA system is integrated, a preliminary Wizard of Oz study and interviews with caregivers about prompting modalities. Chapter 5 describes the implementation of the main system components and the interplay in the overall system. In chapter 6, we describe and evaluate the main studies with the TEBRA system: a pre-study with regular users conducted at Bielefeld University and the main study with persons with cognitive disabilities conducted at *Haus Bersaba*. Chapter 7 concludes the thesis and provides an overview of future work.

2.1 WHAT IS ASSISTIVE TECHNOLOGY FOR COGNITION?

In the United States' Assistive Technology Act 2004¹, *Assistive Technology* (AT) is defined as

Assistive Technology

any item, piece of equipment, or product system, whether acquired commercially, modified, or customized, that is used to increase, maintain, or improve functional capabilities of individuals with disabilities.

The general term *disability* refers to both physical impairments and cognitive disabilities: physical impairment of a person's body describes "any impairment which limits the physical function of limbs or fine or gross motor ability."²

In this thesis, we aim to develop AT for persons with cognitive disabilities. Cognitive disability is a broad term including many different disorders related to cognitive functioning including perception, reasoning and remembering [32]. *Assistive Technology for Cognition* refers to technological interventions in the area of cognitive disabilities by compensating for lost or decreased cognitive functioning.

*Assistive Technology
for Cognition (ATC)*

The term *Assistive Technology for Cognition* (ATC) was first coined by Lo-Presti et al. as an umbrella term subsuming *cognition orthosis* and *cognitive prosthetics* [57]: *cognition orthosis* and *cognitive prosthetics* were used in the context of clinical rehabilitation. Kirsch et al. define a *cognition orthosis* as a compensatory strategy for functional deficits of patients with acquired brain injury [43]. In comparison to previous strategies in cognitive rehabilitation, the cognitive orthosis was meant to be "active, temporally proximate, designed for specific tasks, and interactive" [43, p. 2].

Cognitive prosthetic was defined by Cole and Lynch in [24] and [58], respectively. Cole and Lynch focus specifically on computer-based strategies. According to Cole, a cognitive prosthetic "uses computer technology, [...] is designed specifically for rehabilitation purposes, [...] directly assists the individual in performing some of their everyday activities and [...] is highly customizable to the specific needs of the individual" [24, p. 3]. Lynch provides a more general definition of the term in a review of computer-assisted cognitive retraining: a cognitive prosthetic is "any computer-based system that has been designed for a specific individual to accomplish one or more designated tasks related to activities of daily

¹ Assistive Technology Act (PL 108-364); [www.gpo.gov/fdsys/pkg/STATUTE-118/pdf/STATUTE-118-Pg1707.pdf; accessed 27-November-2012]

² Wikipedia: Physical Disability [http://en.wikipedia.org/wiki/Physical_disability; accessed 13-December-2012]

living (ADL), including work” [58, p. 5].

LoPresti et al. extend the application area of ATC from cognitive rehabilitation in clinical settings to a user’s individual home and community settings. Here, ATC interventions provide assistance as part of a user’s everyday life [57] where performing ADLs is a key requirement for independent living [16]. Inability to perform ADLs is mostly related to the loss or decrease of cognitive functions [79]. According to Scherer et al, ATC is

a special subclass of interventions that is designed to increase, maintain, or improve functional capabilities for individuals whose cognitive changes limit their effective participation in daily activities. [97, p. 3]

Hence, ATC aims to provide compensation for missing or decreased cognitive functions in order to “improve performance of functional activities that are critical components of independent community life, that contribute substantially to quality of life, or that significantly reduce caregiver burden” [97, p. 3].

Scherer et al. refer to cognitive changes who limit the effective participation in daily activities. Cognitive changes and the resulting cognitive disabilities have a variety of different characteristics. In the following list, we give a brief overview of the most common cognitive disabilities. The list is far from being complete since the term *cognitive disability* covers an extremely broad range of disabilities. An accurate assessment of these disabilities is beyond the scope of the thesis.

Cognitive disability

INTELLECTUAL DISABILITY The American Association on Intellectual and Developmental Disabilities (AAIDD) defines an intellectual disability as “a disability characterized by significant limitations both in intellectual functioning (reasoning, learning, problem solving) and in adaptive behavior, which covers a range of everyday social and practical skills. This disability originates before the age of 18”³. Intellectual disability is also referred to as mental retardation. Due to the negative connotation of mental retardation, intellectual disability is the preferred term to use.

LEARNING DISABILITY The British Department of Health defined learning disability by “the presence of a significantly reduced ability to understand new or complex information, to learn new skills (impaired intelligence) with a reduced ability to cope independently (impaired social functioning) which started before adulthood, with a lasting effect on development”.⁴ Common forms of learning disabilities include disorders in language (aphasia), usage of tools (apraxia), reading (dyslexia), writing (dysgraphia) and math/calculation (dyscalculia).

³ American Association on Intellectual and Developmental Disabilities (AAIDD); [www.aaidd.org/content_104.cfm]; accessed 28-November-2012]

⁴ British Department of Health; [www.northyorks.gov.uk/CHttpHandler.ashx?id=508&p=0]; accessed 28-November-2012]

NEURODEGENERATIVE DISEASES Neurodegenerative disease (ND) is an umbrella term for disorders “in which the nervous system progressively and irreversibly deteriorates.”⁵ ND includes all forms of dementia, also Alzheimer’s disease which is the most common form of dementia among elderly people.

AUTISTIC SPECTRUM DISORDER Autistic Spectrum Disorder (ASD) is a form of pervasive developmental disorders which is characterized by (1) qualitative impairment in reciprocal social interaction, (2) impairment in communication in imaginative activity, and (3) restricted repertoire of activities and interests. [3].

ACQUIRED BRAIN INJURY Acquired brain injury including traumatic brain injuries and cerebrovascular accidents (stroke) can cause severe limitations in cognitive and psychosocial functions. Acquired brain injuries can “result in impairments in one or more areas such as cognition, language, memory, attention, reasoning, abstract thinking, judgment, problem solving, sensory/perceptual/motor abilities, psychosocial behavior, physical function and information processing”. [56, p. 2]

Persons with various cognitive disabilities tend to show problems in the execution of ADLs. The term ADL covers both basic and instrumental activities. According to Spector et al., basic ADLs refer to activities of fundamental functioning such as dressing, eating and personal hygiene amongst others [104]. Instrumental ADLs refer to activities that enable a user to actively participate in society. Example of instrumental ADLs include shopping, transportation and house-keeping amongst others [104]. ADLs involve one or more cognitive functions which are necessary to perform the ADL successfully.

In recent years, there were several attempts to classify ATC interventions according to cognitive functions involved in the execution of ADLs [57, 56, 32]: LoPresti et al. distinguish between ATC compensating for executive function impairments and information processing impairments [57]. Executive function impairments “are typically associated with effective adaptation and accommodation to changing environmental demands through the appropriate and efficient integration of more basic cognitive skills [... such as] planning, task sequencing and prioritization, task switching, self-monitoring, problem solving and self-initiation and adaptability.” [57, p.10] [50] Information processing impairments include difficulties with sensory processing as well as social and behavioral issues.

Gillespie et al. provide a more systematic classification according to the International Classification of Functioning, Disability and Health (ICF) [32]. ICF is a classification of health and health-related domains given by the

⁵ Medical dictionary: Neurodegenerative disease [<http://medical-dictionary.thefreedictionary.com/Neurodegenerative+disease>; accessed 13-December-2012]

World Health Organization (WHO)⁶. Gillespie et al. identified applications of ATC in the following areas of cognitive functions referring to the ICF classification: attention, calculation, emotional functions, experience of self and time, and higher-level cognitive functions.

This thesis describes a contribution by developing ATC compensating for higher-level cognitive functions.

Hence, we focus the review of related work on this area. For an overview of ATC in the other areas, see [32].

Table 1 gives an overview of higher-level cognitive functions according to the ICF definition [32]. The ICF distinguishes abstraction, organization

Table 1: Classification of higher-level cognitive functions according to the ICF definition.

Cognitive function	ICF definition
Abstraction	Creating general ideas, qualities or characteristics out of, and distinct from, concrete realities, specific objects or actual instances.
Organization and planning	Coordinating parts into a whole, of systematizing; the mental function involved in developing a method of proceeding or acting.
Time management	Ordering events in chronological sequence, allocating amounts of time to events and activities.
Cognitive flexibility	Changing strategies, or shifting mental sets, especially as involved in problem-solving.
Insight	Awareness and understanding of oneself and one's behavior.
Judgment	Mental functions involved in discriminating between and evaluating different options, such as those involved in forming an opinion.
Problem-solving	Identifying, analyzing and integrating incongruent or conflicting information into a solution.

and planning, time management, cognitive flexibility, insight, judgment and problem-solving. Most ATC identified by Gillespie et al. assist in one of the two areas: organization and planning, and time management. Time management refers to scheduling a user's daily routine dealing with temporal constraints between different tasks. Organization and planning relates to the execution of single tasks in which the subtasks need to be

⁶ World Health Organization: ICF [<http://www.who.int/classifications/icf/en/>; accessed 13-December-2012]

structured and performed in a temporal order for a successful task execution. The application of ATC in these areas is feasible since ATC can serve as a structuring aid on a macro-level for several tasks in the daily routine (time management) and on a micro-level for single tasks (organization and planning).

The aim of ATC as a structuring aid is to provide assistance to the user by giving appropriate prompts, e.g. reminding the person to take medication, to keep a doctor's appointment or to perform a sub step in the execution of a task. In order to increase the independence of users, ATC aims to deliver prompts to the user only when necessary. *Context awareness* is a key concept of an ATC to decide whether a prompt is necessary for a user. Context awareness describes the ability of a system to adapt its behavior based on contextual information [98]. In the domain of ATC, a context-aware system needs no or very little feedback from the user about performed activities. Instead, the system perceives feedback via environmental sensors and infers a user's behavior based on sensory information. For example, consider an ATC system assisting in preparing tea. A context-aware ATC would perceive whether the water in the kettle boiled using environmental sensors. No explicit feedback of the user is necessary. Mihailidis et al. refer to such ATC systems as *Zero-effort technology* (ZET). Zero effort means "that the people who use them [ATCs] do not have to change how they go about their daily lives and the ZETs will support them appropriately" [62, p. 3]. Users need not to change or modify their behavior for the ATC system to work properly.

Context awareness

In order to implement a context-aware behavior, an ATC system needs to be *robust* with regard to an individual user's behavior. ATC needs to deal with huge variances in spatial and temporal execution of tasks. For example, one user may take much longer to perform a desired task than another. Furthermore, the performances of a single task may vary greatly in spatial execution due to different motor abilities amongst users.

Robustness

In both time management and organization and planning, complex tasks which consist of several sub steps may be performed in a number of ways: different sequences of sub steps lead to a successful execution. ATC needs to allow for flexible sequencing of sub steps by checking whether a user's current behavior is consistent with the overall plan to achieve a desired goal. In the case of an inconsistent behavior, ATC provides a prompt to the user which needs to be appropriate in timing and modality. Appropriate timing of prompts is mandatory for the acceptance of the ATC [109]. Furthermore, the modality of prompts needs to be adjusted to the users' abilities. For example, one user might be able to react correctly to a video prompt, but another user might be highly distracted by the video.

Appropriate prompting behavior

In the following section, we will exemplarily review existing ATC systems with regard to the key concepts of *context awareness*, *robustness* and *appropriate prompting behavior* in the time management and organization and planning domain.

2.2 REVIEW OF EXISTING ATC

In recent years, various surveys of ATC were published with a focus on studies [32] and technical implementations [111, 62, 90]. Gillespie et al. reviewed studies conducted with ATC in a clinical population [32]. The reported studies were categorized based on cognitive functions which the ATC intervention compensates for.

Tsui et al. reviewed ATC applied as memory aids for task sequencing [111]. They mainly describe ATC along the technical level ranging from no-tech devices such as picture books to high-tech devices such as personal electronics using artificial intelligence techniques. According to Tsui et al., future ATC systems need to incorporate human factors such as assessment of the user's mood, encouragement and reward in order to provide appropriate prompting.

Mihailidis et al. provided an exhaustive survey of ATC [62] incorporating aspects such as design paradigms, technological approaches and study results. They conclude that a main factor in the development of ATC systems will be the adaptability of the system to the user's needs and capabilities. The survey of Rashidi et al. [90] focuses on ambient assisted living for older adults. The paper covers the whole processing chain of ATC from acquisition of sensor data to the generation of prompts and reviews techniques and algorithms applied in each step. Rashidi et al. see future perspectives in the development of ATC in the areas of sensor technology and applied algorithms: for an application in real-world settings, unobtrusive sensor technology need to be constructed and integrated into a user's environment. Algorithms used in ATC system need to be improved with regard to reliability and accuracy. Furthermore, "some simplifying assumptions should be relaxed, such as the assumptions regarding single resident homes and availability of labeled data". [90, p. 8]

In the remainder of the section, we will be oriented towards the surveys of Rashidi and Mihailidis [90, 62]. We will focus our review on existing ATC systems which implement a fully functioning prompting system. Several works aim towards a fully functioning system, but implement only sub-components of such systems: for example, PROACT developed by Philipose et al. infers Activities of Daily Living (ADLs) from object usage [85]. Activity sequences are represented as probabilistic models which are generated from plain English descriptions of activities. PROACT implements a probabilistic engine inferring activities based on RFID sensors. However, no prompting component is implemented, yet, which generates prompts to a user about erroneous steps during an activity.

The PUCK system aims to assist users in diverse activities in a smart home such as watering plants, writing a birthday card and preparing meal [26]. PUCK uses a data-mining approach for prompt generation: the focus of the project is on learning the timing of prompts given data of a variety of sensors applied at furniture and objects in a smart home. The sensor data

needs to be manually labeled with the user's activity by a human annotator. An evaluation of the sensor patterns and a generation of prompts is not possible at run-time at the moment. Hence, PUCK can't be deployed as a real-time prompting system, yet.

The systems described in the following subsections implement fully functioning systems which are state-of-the-art in the field of ATC.

2.2.1 *PEAT*

PEAT (Planning and Execution Assistant Trainer) is a scheduling aid for persons with brain injury [52]. PEAT structures a user's daily routine by providing visual and audible cues using a mobile phone. A user/caregiver defines scripts for sequential daily routines including temporal information such as start/end time and expected duration. Scripts may contain choice points where a user has to decide between pre-specified alternatives. For example, in a dinner task, PEAT would make a suggestion of restaurants from which the user can choose. A planning component based on PROPEL (PROgramming Planning and Execution Language) [51] simulates variations of these scripts including different user choices and picks the variation which maximizes goal achievement.

The PEAT system is not context-aware since PEAT has no sensory information to perceive whether a task was performed. At the described choice points of a script, the user has to provide explicit feedback to the system. PEAT is robust by allowing for flexible execution of daily tasks: additionally to the initial scripts, caregivers/users can enter new tasks during the day. Conflicting tasks, e.g. due to overlapping temporal constraints between tasks or due to a task exceeding the planned time interval, are recognized. PEAT resolves these conflicts by replanning and providing an updated plan. A hand-held device is used to deliver audio-visual prompts generated by the system. Prompts include the starting and stopping time of a task. The user has the opportunity to skip or reschedule prompts using explicit feedback via the hand-held device.

2.2.2 *Autominder*

Similar to the PEAT system, Autominder schedules daily activities of persons with mild to moderate memory impairments [87]. Autominder models a user's daily plan, tracks a user's execution of the plan and decides whether to provide a prompt to the user. The Autominder system contains of three components implementing the above mentioned functions: the *Plan Manager* stores a user's daily plan. Tasks are modeled using disjunctive temporal problems (DTPs) allowing "for both quantitative (metric) and qualitative (ordering) constraints, as well as conjunctive and disjunctive combinations of them" [87, p. 3]. Tasks can be defined to occur at specific times or intervals, e.g. dinner at 5pm and toileting between 11am and 11.15am, respectively. Similar to the PEAT system, plans can be up-

dated during a day. The plan manager can handle modifications to plans such as inserting/deleting activities, successful executions and activities exceeding the desired time interval. The plan manager resolves possible conflicts and replans to meet the temporal constraints. The *Client Modeler* monitors the execution of the plan. In comparison to the PEAT system, the client modeler incorporates observations (the user's location) for activity recognition. The client modeler infers the user's progress in the plan using a Quantitative Temporal Bayesian network (QTBN) [23]. A QTBN contains a Dynamic Bayesian network (DBN) and a standard Bayesian network modeling causal and temporal relationships, respectively. Interface functions distribute information between the networks. The *Personal Cognitive Orthotic* (PCO) decides when to deliver a reminder to the user. PCO generates reminders using a Planning by rewriting (PbR) approach [2]. PbR uses a set of rules to filter different candidate plans. The rules are chosen in a way that four criteria are met [60]: (1) ensure that the user is aware of activities, (2) achieve a high-level of user and caregiver satisfaction, (3) avoid ineffective plans, (4) avoid making the user dependent on the system by providing too many prompts.

The Autominder system uses a simple heuristic to implement context awareness: a user's behavior is inferred on the location of the user. For example, if toileting is scheduled for a specific time and the user enters the bathroom, Autominder will conclude that the user has performed the scheduled activity. Autominder doesn't check whether the behavior was actually performed.

Similar to the PEAT system, Autominder allows for flexible execution of tasks by replanning when the daily plan is modified. Autominder also recognizes conflicting tasks and allows for temporal variance in the execution of tasks by specifying intervals in which the task can be completed.

2.2.3 Archipel

The Archipel system assists users with intellectual disabilities in meal preparation using a hierarchical planning approach [33, 7]. The task is modeled using a hierarchical representation with a tree-like structure: each node of the tree is an atomic action which has a set of constraints added such as a time slot when to execute the behavior. Archipel uses a set of sensors installed in an apartment which is part of the DOMUS lab of University of Sherbrooke, Canada. Sensors such as electro-magnetic contacts at doors, RFID tags, UWB tags, a flow meter and power line communication devices are installed in the environment and tools. Based on the sensory information, the system observes whether the constraints of the action are met and the current action was performed correctly. Hence, Archipel implements a context-aware behavior.

Archipel is designed to assist a user when necessary in order to foster the user's independence. However, a user needs to demand assistance explic-

itly by using a touchscreen device. For example, the user might ask for the next action or the location of an object in the environment. The prompts given by the system are designed based on study results where four main deficits in the execution of ADLs were identified [7]: initiation, planning, attention and memory deficits. In order to compensate for attention and memory deficits, an object's location will be highlighted using LEDs attached to the object. For planning deficits, an image or video of the desired action will be shown on the touchscreen device. The prompting modalities were found to be appropriate with regard to the user's abilities [86]. The hierarchical structure of the meal preparation task is predefined. Hence, Archipel doesn't allow for flexible task execution with regard to an individual ordering of actions.

2.2.4 COACH

The COACH (Cognitive Orthosis for Assistive aCtivities in the Home) is an automatic prompting system for persons with dementia which assists in the task of hand washing [39]. COACH uses computer vision techniques for environmental perception and a Partially Observable Markov Decision Process (POMDP) for planning and decision making. COACH tracks the user's hands and the towel using a flock-based color tracker [37]. The positions are discretized with regard to static regions which are pre-specified. The discretized positions are passed to a belief monitor which maintains a belief of the user's progress in the overall task. The belief monitor simulates a belief update based on observations. If the simulated belief differs significantly from the current belief over a period of time, a real belief update will be triggered and the POMDP generates a system action based on an offline-learned policy. A system action is either to do nothing, deliver a prompt or to call for a human caregiver. COACH delivers prompts using different levels of details: audio, personalized audio prompt using a person's name, and a video prompt. COACH incorporates estimates about the mental state of the user including responsiveness and awareness. COACH is context-aware because no explicit feedback from the user about completed sub steps is necessary. Sub steps are modeled implicitly using pairs of pre/post-actions: for example, if the hands enter (pre action) and leave (post action) the soap region, COACH infers that the user has taken the soap with a pre-specified probability. Hence, COACH allows for different temporal executions of sub steps, but the spatial variance is limited because objects such as soap dispenser are fixed at a certain location. For a user having an individual way of performing a sub step or a motor impairment, using a fixed object might not be feasible during a task. COACH allows for a flexible execution of sub steps, e.g. in the beginning of hand washing, a user can either take the soap first and then open the tap, or vice versa. For flexible task execution, each transition in the belief state of the POMDP needs to be specified with a certain probability. The manual specification of the probabilities is very hard and time-consuming. For

more complex tasks involving more objects, more sub steps and flexible execution, the specification process might become intractable. Hoey et al. use a *syndetic assistance process* (SNAP) to specify the probabilities of the POMDP using a knowledge-driven approach with a relational database which limits the number of probabilities to be specified manually [38, 35]. The COACH system was evaluated in a user study with 6 participants having moderate-to-severe dementia [61]. The participants' performance was tested in two alternating conditions: (1) baseline without COACH system, and (2) intervention with COACH system. The average rate of hand washing steps completed independently was increased by 11% in the intervention compared to the baseline scenario. Furthermore, intervention of a caregiver was decreased by 60% when using the COACH system.

2.2.5 Summary

Table 2 summarizes the ATC systems described in the previous subsections with regard to the three main concepts of context awareness, robustness and appropriate prompting behavior. Context awareness refers to the ability of a system to perceive a user's behavior based on sensory information. Despite spatial and temporal variance in the execution of the task, an ATC system needs to be able to robustly recognize a user's behavior. Based on the current behavior and a user's progress in the task, an ATC system needs to provide prompts which are appropriate in time and suit a user's capabilities.

The ATC systems reviewed in the previous sections implement context awareness, robustness to an individual user's behavior and appropriate prompting behavior in various degrees. PEAT allows for flexible behavior, but implements no context awareness. Archipel and COACH are context-aware, but show limitations in the robustness in behavior recognition.

To the best of our knowledge, no ATC system so far provides context awareness, robustness to an individual user's behavior and appropriate prompting behavior in a single system. This thesis describes our approach towards the development of such a system which assists a heterogeneous group of persons with cognitive disabilities in the complex scenario of brushing teeth. In the following section, we will analyze the characteristics of the scenario and describe the scenario selection process.

2.3 SCENARIO SELECTION

The ADLs for which ATC systems provide assistance, differ significantly in the degree of complexity and flexibility: a task will be denoted as complex, if it consists of a sufficient number of steps which have to be combined for successful performance. Furthermore, the different steps might include different objects to be manipulated during task execution. With increasing number of involved objects, the environmental perception gets more complex since the different objects need to be observed using sen-

sory information. Flexibility describes that users have to coordinate steps while different ways of execution are possible.

In this thesis, we chose the task of *tooth brushing*, a basic ADL which is both complex and flexible. In order to investigate the characteristics of the task, we analyzed tooth brushing using a task analysis technique called Interaction Unit (IU) analysis which is described in section 4.1. According to IU analysis, brushing teeth consists of eight sub steps: *fill_mug*, *wet_mouth*, *paste_on_brush*, *brush_teeth*, *clean_mug*, *clean_brush*, *rinse_mouth* and *use_towel*. The eight sub steps can be combined in various ways for a successful task execution. The sub steps include the manipulation of four objects during tooth brushing: mug, towel, toothpaste and toothbrush. Different cognitive functions are involved in the brushing task which makes the execution error-prone for persons with cognitive disabilities. Firstly, users need to recognize involved objects and realize the affordance of objects during the task, e.g. the recognition of the mug including the mug's ability to store water. Secondly, users need to remember that a sub step has been completed without having an environmental cue which indicates substep completion. For example, the user has rinsed the mouth with water. In comparison to sub steps such as *paste_on_brush* where the user has a visual cue (paste is on the brush), there is no visual cue in *rinse_mouth* which indicates the completed sub step.

In comparison to Archipel and COACH, two state-of-the-art systems in task sequencing, the task of brushing teeth is more complex than preparing meal and washing hands, respectively: for example, washing hands consists of five sub steps (wet hands, take soap, water on, water off, dry hands) in which two objects (soap, towel) are involved. Preparing meal is similar in complexity compared to the TEBRA system. However, Archipel doesn't allow for flexible task execution since the order of sub steps is fixed. Hence, flexibility of the task is not modeled in the Archipel system. Beside the scientific relevance due to the complexity and flexibility of the task, tooth brushing is important from a medical perspective: tooth brushing is an important activity in the daily routine since disregarding oral hygiene usually leads to severe medical problems.

Tooth brushing

Table 2: Overview of state-of-the-art ATC systems reviewed in this thesis.

System	Application area	Task	Context awareness	Robustness	Prompting behavior
PEAT	Time management	Scheduling daily routines	No, a user's explicit feedback is necessary	Allows for flexible behavior by replanning	Audio-visual reminders delivered via hand-held device
Autominder	Time management	Scheduling daily routines	Limited: behavior inferred based on a user's location only	Allows for flexible behavior by replanning	audio-visual prompts provided when necessary
Archipel	Organization and Planning	Meal preparation	Yes, environmental perception using various sensors	Limited: predefined sequence of the task; no flexible execution possible	Prompts provided when necessary as requested by the user; different types of prompts for different errors
COACH	Organization and Planning	Hand washing	Environmental perception uses simple heuristic of pre-/post-actions to recognize behaviors	Yes, but flexibility in spatial execution is limited due to fixed position of the soap dispenser	Prompting hierarchy; modality of prompts chosen according to the user's performance

METHODS

In the previous chapter, we identified the main properties of ATC systems which enable appropriate assistance with regard to a user's individual behavior and abilities: context awareness, robustness and appropriate prompting behavior. Context awareness and robustness are important aspects for a user's acceptance of an ATC system. No or very little feedback from the user is necessary about completion of steps. Furthermore, the ATC system is able to deal with the huge spatial and temporal variance in the execution of tasks robustly. Environmental perception aims to implement such properties. The key aspect of environmental perception is the recognition of human activities with a focus on the particular question: *what is the user's current activity?*

In order to give appropriate prompts, a *planning and decision making* component needs to track the user's progress in task execution based on the environmental perception. The component provides prompts to the user when necessary. The content and modality of prompts is chosen with regard to the overall progress and the user's individual behavior.

In this chapter, we will review methods for environmental perception, planning and decision making and actuation and prompting.

3.1 ENVIRONMENTAL PERCEPTION

Environmental perception in the area of ATC systems is closely linked to the recognition of human activity since the ATC systems aim to support a user in performing desired activities. Human activity recognition is a long-studied field of research with applications in sports analysis [28, 75], surveillance [55, 47], ADL monitoring [21, 123, 83] and many more. Recognition approaches in the different fields are based on a variety of sensors including RFID [20, 76, 85], pressure sensors [54], inertial sensors [122, 6, 12, 84], cameras [39, 15, 46] and combinations of such sensors in mobile or wearable devices [121, 22]: Pham et al. recognize activities in a food preparation task using objects equipped with accelerometers [84]. Bieber et al. use accelerometers in a mobile phone to recognize activities such as walking, jogging, jumping and cycling [12]. RFID tags installed at everyday objects are used by Buettner et al. [20]. Based on the traces of object use, they infer activities such as reading a book, watching TV and preparing coffee.

Most work in human activity recognition is based on analyzing camera images using computer vision techniques. In recent years, several surveys were published inferring human activities based on visual information [63, 112, 1]. Here, we orient towards the survey of Aggarwal and Ryoo

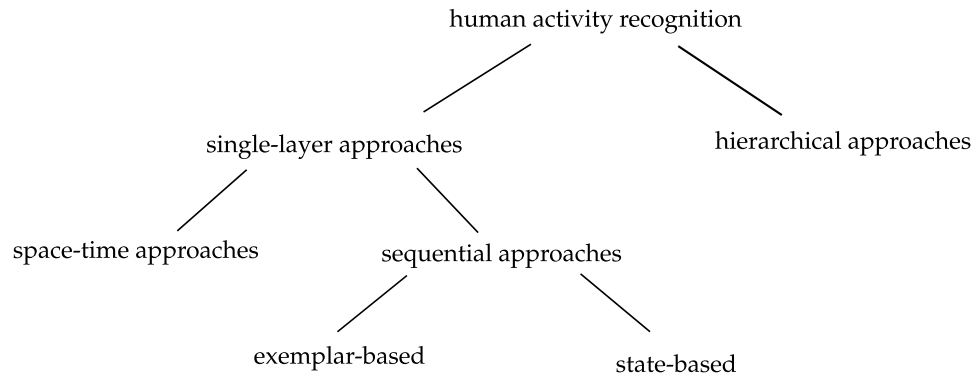


Figure 1: Taxonomy of activity recognition taken from the survey paper of Aggarwal and Ryoo [1].

[1] who provide an approach-based taxonomy of activity recognition as shown figure 1. Aggarwal and Ryoo distinguish activity recognition into hierarchical and single-layered approaches which are described in the following subsections.

3.1.1 *Single-layer approaches*

Single-layer approaches work on sequences of images. Activities are represented by sequential patterns extracted from the image sequences. Activities are classified by comparing patterns of an unknown image sequence to pre-trained patterns representing activities. Aggarwal and Ryoo distinguish between different ways of identifying sequential patterns explained in the following subsections.

3.1.1.1 *Space-time approaches*

In space-time approaches, patterns are extracted from 3-dimensional space-time volumes which are spanned by consecutive images over time. Laptev detects spatio-temporal interest points in the space-time volumes which are stable over different spatial and temporal scales [45]. The scales on which the interest points were detected, form an event descriptor. Laptev uses the descriptor to recognize activities with periodically occurring movements such as walking where the movements of the legs are periodical.

Ryoo et Aggarwal use spatio-temporal relationship match, a kernel measuring the similarity of two feature sets representing activities [95]. The features are interest points in the space-time volume which represent changes in the appearance of a video patch. Activities represented by sets of interest points are classified by calculating the similarity between an unknown feature set and a pre-trained prototype of an activity.

3.1.1.2 Sequential approaches

In comparison to space-time approaches, sequential approaches extract sequential patterns from “a sequence of observations (i.e. feature vectors), and deduce that an activity has occurred in the video if they are able to observe a particular sequence characterizing the activity” [1, p. 16]. Aggarwal and Ryoo subdivide sequential approaches into exemplar-based and state-based approaches.

EXEMPLAR-BASED APPROACHES In exemplar-based approaches, activities are represented using template sequences. An unknown sequence is classified by comparing it to the template sequence of each activity. A classical technique for matching sequences is the Dynamic Time Warping (DTW) algorithm [100, 11]. DTW searches for an optimal match of two sequences by aligning the sequences maximizing a similarity measure.

*Dynamic Time
Warping (DTW)*

Veeraraghavan et al. use DTW for recognizing simple activities such as kicking, waving and turning around [113]. They address the problem of temporal variance in execution by explicitly modeling intra- and inter-personal variabilities: Veeraraghavan et al. learn an average trajectory called a nominal activity trajectory which represents an activity.

Blackburn and Ribeiro apply an adapted form of the DTW algorithm matching silhouette patterns representing activities such as walking and jumping [13].

Exemplar-based approaches can deal with variations in temporal executions of activities by providing a set of template sequences for each activity. The template sequences are usually learned based on training data. Exemplar-based approaches can extract template sequences from small sets of training data which is a mandatory property for an ATC system where training data is hard to acquire.

STATE-BASED APPROACHES State-based approaches represent high-level activities using probabilistic models composed of a number of hidden states. The states are interconnected and generate observations (feature vector). Both the transitions between states and the generation of observations are probabilistic. The probabilities are learned based on training data. A common state-based model is a Hidden Markov Model (HMM) which was originally used in speech recognition [89]. An HMM is a tuple $\lambda = (S, O, T, E, \pi)$ with S and O denoting a finite number of states and observations, respectively. T is a matrix of transition probabilities and the entry T_{ij} denotes the probability of a transition from state i to state j . E is a matrix of emission probabilities: E_{kj} describes the probability of making observation k in state j . π is a probability distribution denoting the initial probabilities of states. An HMM usually represents a single activity as a sequence of hidden states which emit an observation in each time step (e.g. frames in a video). The transition and emission probabilities are learned based on sample data using the Baum-Welch algorithm [8]. An unknown sequence of observations is classified as an activity using the Viterbi algo-

*Hidden Markov
Model (HMM)*

rithm [31]. The algorithm finds the sequence of hidden states which has most likely generated the given sequence of observations.

Yamato et al. were the first to use HMMs in activity recognition for sports scenes [119]. Image frames showing tennis actions such as forehand stroke, backhand stroke are binarized into foreground and background parts. On each image frame, feature vectors are calculated counting the number of background pixels in each mesh. Vector quantization is used to generate prototypes of feature vectors. Each prototype corresponds to an observation generated by a hidden state in the HMM.

Various extensions of classical HMMs have been applied to activity recognition: Coupled HMMs (CHMMs) are used in the recognition of martial arts activities [19] and interaction activities of two persons in a surveillance scenario [71]. CHMMs model two interacting processes by coupling the states of two HMMs. In a CHMM, two different states at a time are possible in comparison to a regular HMM where the process can only be in a single state at a time. The probabilities between states of different HMMs model the influence of the processes on each other. Timing variability in terms of duration of states is implicitly modeled in HMMs and CHMMs. Hidden semi-Markov models (HSMM) explicitly model the duration of states. Natarajan and Nevatia use a combination of CHMM and HSMM to recognize sign language [66].

Dynamic Bayesian network (DBN)

HMMs are the simplest type of Dynamic Bayesian networks (DBNs) which consist of a single hidden variable and a single observation variable. DBNs model scenarios with a number of hidden and observable variables. A DBN is an extension of Bayesian network (BN) to the temporal domain. A BN is a probabilistic graphical model representing a joint probability distribution of random variables. Conditional independence relations between variables are modeled using a directed acyclic graph. A DBN is formed by duplicating BNs over time. Interconnections between variables in subsequent time-slices model the dependence of variables over time. Wu et al. use DBNs to recognize household activities based on video data in combination with RFID data of manipulated objects [118].

Bayesian network (BN)

The advantage of state-based approaches is their capability of modeling complex high-level activities. However, complex models require a huge amount of sample data in the training of such models which is usually very hard to acquire in an ATC scenario.

In ATC systems, the capability of dealing with spatial variance is important. In this thesis, the term spatial variance describes the different ways of executing activities since persons with cognitive disabilities tend to show huge intra- and inter-personal variation amongst users. In both exemplar-based and state-based approaches, the capability of modeling spatial variance is limited since such approaches don't generalize well on the different movement characteristics in the execution of activities: different movement characteristics need to be represented by different prototypes (in exemplar-based approaches) and state models (in state-based approaches). Hence, the sample data from which the representations are constructed, need to

incorporate a vast number of possible movement characteristics which is often infeasible in ATC scenarios.

3.1.2 Hierarchical approaches

In comparison to single-layered approaches, hierarchical approaches “represent high-level human activities by describing them in terms of other simpler activities, which they generally call sub-events” [1, p. 2]. An example of hierarchical approaches is a hierarchical Hidden Markov models (HHMMs): atomic actions are modeled by lower-level HMMs. Complex activities are represented using sequences of atomic actions where an atomic action is modeled as a single state in a higher-level HMM.

Nguyen et al. use HHMMs in the recognition of indoor activities based on movement trajectories of users [69]. A variant of the HHMMs is a multi-layer Hidden Markov Model (ML-HMM). The layers of the ML-HMMs model activities in different temporal granularities in order to cope with temporal variance in activity execution. Oliver et al. apply ML-HMMs to recognize activities in meeting room situations such as face-to-face communication between users or a presentation of a single user [70]. Their recognition is based on heterogeneous sensor data including video, audio and keyboard/mouse interaction.

Hierarchical models are suitable in domains where the activities can be decomposed into smaller action units. However, hierarchical models have the disadvantage that the capability of modeling flexible behavior is limited. For example, a user prematurely terminates a behavior due to problems in sequencing the task which is a common situation in an ATC scenario with persons with cognitive disabilities. In a hierarchical HMM, the model of the low-level behavior will be activated if the user starts the behavior. A premature termination of the low-level behavior might lead to a state in the HHMM which is locked in the model of the low-level behavior. This might result in an erroneous perception of subsequent behaviors.

3.2 PLANNING AND DECISION MAKING

Automated planning deals with the creation of plans in a planning domain. A plan is a sequence of actions executed by an agent which transfers the state of the world from a given initial state to a given goal state. Planning domains can be classified along different properties. These properties determine appropriate techniques to solve the underlying planning problem: the state of the world is modeled using random variables which are either *discrete* or *continuous*. The agent can either *fully* or *partially observe* the state of the world. In a fully observable world the agent knows the state of the world at any time. In a partially observable world, the agent can't observe the state of the world directly, but reasons about the state using sensory information. Based on the state of the world, the agent chooses actions to execute. The effect of actions on the

world state are either *deterministic* or *probabilistic*. Effects of deterministic actions are static regardless of the time and the state of the world. For probabilistic actions, multiple effects are possible which are modeled by using probability distributions over the outcomes of actions.

Classical planning

Several planning strategies have emerged over the last decades to deal with different properties of planning domains. We focus on planning domains in which only a single agent is involved in the planning process. For an overview of multi-agent planning problems, see chapter 11 in [94]. Classical planning deals with problems in which the world is fully observable and actions are deterministic. Additionally, there are no exogenous actions which means that only the agent's actions can change the state of the world. Good introductions to classical planning are provided by the books of Russel and Norvig [94] or Poole and Mackworth [88].

One of the simplest strategies in classical planning is forward planning. A forward planner searches a state space graph from an initial state to a goal state. In a state space graph, nodes represent states and arcs correspond to actions between states. Search techniques include both uninformed and informed methods. Uninformed methods such as depth-first/breadth-first search or iterative deepening do not take into account any cost or distance from a current node to a goal node. Informed methods such as A* search include cost information via a heuristic function. Disadvantage of planning using search strategies is the huge consumption of memory and computation time. In complex domains, the state space grows rapidly suffering from the combinatorial explosion. Hence, the number of computation steps and the amount of memory to search the state space increases as well.

In order to overcome this disadvantage, a planning problem can be formulated as a constraint satisfaction problem (CSP). A CSP uses a factored representation of states where each variable has a value from a domain D. The variables underlie constraints which describe allowable combinations of values. CSPs can usually be solved much faster since the state space can be pruned with regard to the constraints. For an overview of CSP applications, we refer to the survey paper of Brailsford et al. [18]. In both forward planning and constraint satisfaction problems, the resulting plan follows a total order of actions. In many domains, there is no reason to enforce a total order. For example, consider a dressing scenario: the goal is to put on socks and shoes. It is irrelevant for goal achievement whether the right sock or the left sock are put on first.

Partial-order planning (POP)

Partial-order planning is a planning technique which allows for a partial ordering of actions. Partial-order planning only commits to an ordering of actions when necessary which is also called strategy of *least commitment*. A partial-order planner (POP) consists of an initial state I, a goal state G and a set of STRIPS actions A. STRIPS is an action-centered representation of the dynamics in the planning domain [30]. An action is modeled describing the preconditions and effects of actions on the state of the

world. For an action to occur, the preconditions need to be fulfilled. The effects describe which state variables are affected when an action occurs. Initial state I and goal state G are configurations of the variables which represent the state of the world. A POP creates a set of ordering constraints O (action a before b), a set of causal links C (action a provides condition x for b) and a set of variable bindings B (variable $v = c$ where c is a constant). Any total order of actions which is consistent with the ordering constraints is a valid plan. An extension to POP is UCPOP which supports conditional effects and universal quantification [78].

The advantage of a POP is its ability to allow for different plans without modeling all plans explicitly which lead to a successful task execution. This is especially helpful in the domain of ATC. Tasks such as brushing teeth are very flexible in a way that different ways of task execution are possible. Furthermore, persons with cognitive disabilities show a huge variety in task execution, also following unconventional ways of execution due to cognitive problems. Hence, modeling all possible plans of execution might lead to an overly complex representation of the problem.

In classical planning domains, plans can be generated offline because the state of the world is known at every time in the future due to deterministic actions and full observability of the world. In real-world domains such as ATC scenarios, plans need to be generated online since the state of the world is only partially observable using error-prone sensory information. For example, in the tooth brushing task, we have no sensor which accurately measures the water filling level of the mug. Actions in an ATC scenario are highly probabilistic: an ATC system provides a prompt to the user, but it's unclear whether the user will react to the prompt correctly, incorrectly or at all. *Decision-theoretic planning*, often denoted as *planning under uncertainty*, deals with relaxed assumptions in planning domains such as partial observability and probabilistic actions.

In decision-theoretic planning, problems can be separated into two categories: (1) episodic and (2) sequential decision making problems. Episodic problems consist of a set of episodes which are independent of each other. In each episode, the agent decides for a single action based on the history during the current episode (single-step plan). For example, consider an agent that needs to sort out defective parts in an assembly line¹. Based on the current status of the part, the agent decides either to sort out the defective part or keep it on the line.

Sequential decision making deals with the creation of multi-step plans. For example in an ATC scenario for brushing teeth, multiple prompts are usually necessary to assist a person in the execution of the task. In this section, we deal with sequential decision making problems since the majority of ATC scenarios such as hand washing and brushing teeth require sequential decisions. Furthermore, the number of actions to take (in an

¹ example taken from p. 43 of chapter 2 in the book by Russell and Norvig [94]

ATC scenario, the number of prompts to give) is not known in advance which is called an infinite horizon problem.

*Markov Decision
Process (MDP)*

A common technique for sequential decision making in a fully observable, probabilistic world with an infinite horizon is a Markov Decision Process (MDP) [10, 40]. An MDP is a tuple (S, A, P, R) with S being a finite set of states and A being a finite set of actions. P is a transition model with $P(s'|s, a)$ denoting the probability of transitioning to a state s' at time $t + 1$ when the state at time t is s and the agent takes action a . R specifies the reward function with $R(s, a, s')$ describing the expected immediate reward of getting from state s to state s' when the agent takes action a . Sequences of rewards are combined into a single value V . Commonly used is the discounted reward $V = \sum_{i=1}^{\infty} \gamma^{i-1} \cdot r_i$ where γ is a factor between 0 and 1 discounting future rewards. The solution of an MDP is a policy $\pi : S \rightarrow A$ which provides an action for any possible state. An optimal policy yields the highest expected rewards amongst all policies. Solution techniques finding an optimal policy such as value iteration, policy iteration or variances of these are discussed in book of Sutton and Barto [106].

*Partially observable
Markov Decision
Process (POMDP)*

In most real-world applications such as ATC scenarios, the world's state is only partially observable via environmental sensors. Partially observable Markov Decision Processes (POMDPs) are an extension of MDPs to domains where the state of the world is only partially observable [41]. A state generates an observation which can be perceived via sensory information. Formally, a POMDP is a tuple (S, A, P, R, O, Ω) . S , A , P and R are defined according to the MDP definition. O denotes a set of observations and Ω is an observation model with $P(o|s)$ is the probability of observing o given the state s . Since the state of the world is only partially observable, the agent doesn't know exactly which state it is in. Instead, it maintains a belief state B , a probability distribution over possible states. The solution to a POMDP is to find a policy which chooses an action based on the agent's belief state. Hence, a policy in a POMDP is a function $\pi : B \rightarrow A$ mapping a belief state to an action. Finding a policy for a POMDP is much more complicated than for an MDP since the belief state B is continuous and encodes an infinite number of possible states. Techniques for solving POMDPs are discussed in [64].

The application areas of POMDPs are diverse: Ross et al. use a POMDP for a robot navigation task [93]. Zhang et Sridharan apply a POMDP in vision-based scene analysis [120]. Thomson and Young enhance a spoken dialog system by modeling the internal system state with a POMDP [110]. Business applications include the generation of stock investment policies [5] and decisions in road maintenance [124]. Hoey et al. apply a POMDP model in an ATC scenario: the COACH system as described in section 2.2.4 assists persons with dementia in the task of hand washing [39].

An advantage of POMDPs is the applicability to real-world problems by dealing with partial observability and probabilistic actions. However, such real-world problems are usually complex due to a large number of states and actions involved in the domain. Hence, the specification process of a

POMDP is complex since the probabilities in the transition model T and the observation model Ω as well as the reward function R have to be specified manually based on expert knowledge. Especially in ATC domains, expert knowledge is hard to gain since persons with cognitive disabilities show a huge variance in the execution of tasks with often unpredictable behavior.

In ATC scenarios such as task assistance, the planner performs actions in sequential order. When a prompt is issued, the system evaluates upcoming sensor information before providing a further action. Hence, actions are non-interleaved and non-concurrent. Interleaved and concurrent actions are subject of *Temporal Planning* which is discussed in chapter 14 in [67]. However, timing is an important factor in an ATC system: the above-mentioned planning techniques operate on a discrete time scale without modeling the duration of states and actions explicitly in the planner. In contrast, tasks such as hand washing or brushing teeth are continuous and the durations of a user's individual behaviors vary extremely, especially for persons with cognitive disabilities. Hence, the dynamics of the planning component needs to be aligned with the task dynamics. In the COACH system, Hoey et al. use a simple heuristic: whenever the belief state of the POMDP changes significantly, a belief update is simulated. When the simulated belief is persistent over a certain time, a real belief update is triggered and an action is chosen.

In the TEBRA system, we maintain an explicit timing model including durations of behaviors. Modeling the task dynamics explicitly has the advantage of decreasing the model complexity of the planner. For example, including task dynamics variables in the POMDP model would lead to computational burden due to an increased state space size.

In order to provide proper assistance to the user, the content of prompts is as important as the time at which prompts are delivered. A review on prompt contents is given in the following section.

3.3 ACTUATION AND PROMPTING

In the field of cognitive rehabilitation, instructional strategies used by human caregivers are widely explored [103, 17]. However, strategies in automatic prompting systems such as ATC are not very well studied. Seelye et al. discuss strategies and problems which arise by transferring instructional strategies from a human caregiver to the prompting behavior of an ATC system [99]. They focus on two aspects of automatic prompting: *delivery* and *content* of prompts which are described in the following subsections.

3.3.1 *Delivery of prompts*

Delivery of prompts refers to the technological methods used in the generation of prompts starting from low-level and mid-level prompts to more sophisticated high-level prompts. Low-level prompts are time-based reminders using an alarm clock or a mobile phone. Wilson et al. describe the NeuroPage system which provides reminders for daily activities for persons with memory impairment [116]. Szymkowiak et al. present a PDA device to prompt for daily activities which is remotely accessible by a caregiver [108].

Mid-level prompts are more sophisticated and incorporate time and the user's location to deliver appropriate prompts. For example, the Memo-Clip is a wearable device reminding a user of a task which is associated with a specific location [9]. When the user enters the location, an audio signal and a text message are triggered. The mapping between tasks and locations is prespecified by the user.

In comparison to mid-level prompts, high-level prompts are increasingly context-aware with regard to a user's activities. Hence, they provide more sophisticated assistance to the user in terms of appropriate timing of prompts. For example, a user is at a specific location, but is involved in a meeting situation. A high-level prompt would consider the inappropriateness of a prompt in the current situation and postpone the prompt until the meeting is finished. High-level prompts enable a system to provide feedback when necessary. Hence, they foster the user's independence in the execution of tasks which is desirable in ATC systems. Artificial Intelligence and Machine Learning are common techniques which are involved in the generation of high-level prompts. The reviewed ATC systems described in section 2.2 provide high-level prompts by including various planning approaches.

3.3.2 *Content of prompts*

Modality Seelye et al. distinguish the contents of prompts into two main properties [99]: (1) modality and (2) level of information. Modality refers to the sensory channel affected by the prompt. Common modalities in ATC systems are verbal instructions, visual prompts or multi-modal combinations of these types.

The GUIDE system provides step-by-step assistance for persons with cognitive impairments using verbal prompts [72]. GUIDE aims to emulate a conversational situation between a caregiver and a user by asking questions about the completion of sub steps. The user can answer with "yes" or "no". The latter indicates that the user needs further assistance. GUIDE was applied in assisting in morning routines [73] as well as donning a limb [74].

The *Activity Compass* is an automatic transportation routing system for persons with mild cognitive disabilities [77]. The possible destinations

and transportation options such as bus or taxi are prompted via visual cues such as images and pictograms on a cellphone.

Various systems provide multimodal prompts combining verbal and visual cues, e.g. the PEAT system [52] and the COACH system [39] described in the previous chapter.

Level of information refers to the amount of information provided in a single prompt. Seelye et al. differentiate between indirect and direct prompts: a direct prompt such as *“Fill the mug with water.”* addresses a problem by mentioning the desired behavior. An indirect prompt such as *“There is a problem in the execution of the task.”* gives no hint of what behavior is expected by the user. Direct prompts are either minimal or specific: a minimal prompt provides only basic information (e.g. *“Fill the mug with water.”*). A specific prompt provides detailed information of what to do (e.g. *“Paul, take the red mug from the counter and fill it with water using the tap.”*). Users respond differently to the levels of prompts due to individual cognitive and functional abilities. According to Demchak et al., an ATC system needs to suite an individual user’s abilities by providing a graded prompt hierarchy [27]. In each level of the hierarchy, different amount of information is provided to the user. A common strategy is *least-to-most* prompting originally used in applied behavior analysis in order to teach persons with cognitive disabilities to execute tasks in daily living [36, 65]. For example, an indirect prompt might be given to the user first. If the user doesn’t react correctly, the prompt hierarchy is escalated and a prompt with an increased amount of information is delivered. A least-to-most prompting fosters the independence of the user as far as possible since the minimum amount of necessary information is delivered to make the user progress in the task.

Level of information

In the COACH system, Hoey et al. use a prompt hierarchy consisting of three prompts. A minimally specific verbal prompt (e.g. *“Use the soap.”*), a maximally specific verbal prompt (e.g. *“John, use the soap on your left in the pink bottle.”*) and a video prompt demonstrating the desired behavior [39].

3.4 SUMMARY

In this chapter, we gave an overview of the main components of an ATC system which are environmental perception, planning and decision making, and actuation and prompting. We reviewed different approaches used in the implementation of such components with regard to their applicability in ATC scenarios.

In chapter 2, we identified context awareness, robustness to spatial and temporal variance in task execution and appropriate prompting behavior as the main functional requirements in order to develop an ATC system with high usability and acceptance of the user.

In this thesis, we aim to design and develop the main components of the TEBRA system with regard to the functional requirements and integrate

the single components in an overall system. Environmental perception is realized in a user behavior recognition component which aims to implement context awareness and robustness to spatial variance in task execution. We use a Bayesian network approach which we adapt according to the requirements of context awareness and robustness to spatial variance: we model a user's behaviors based on states of objects which are manipulated during the behavior. In order to recognize a user's behaviors, we use a Bayesian network in a Bayesian filtering approach.

The planning and decision making component needs to implement appropriate prompting behavior and deal with temporal variance: we use a deterministic planner which is based on a partial-order planner. We apply a dynamic timing model which allows for different velocities of users by modeling dynamics in the user's execution of the task explicitly. We implement a prompting hierarchy similar to the one used in the COACH system. We use pictogram and real-life video snippets in an escalation hierarchy to prompt users appropriately with regard to their abilities. We will describe the implementation of the components in more detail in chapter 5.

Besides specifying the functional requirements of an ATC system, analyzing the characteristics of the task and the involved users is equally important. In the following chapter, we will describe the design process of the TEBRA system incorporating a user-centered design approach in which we take into account the characteristics of the task and the involved users.

DESIGN OF THE TOOTH BRUSHING SYSTEM

The overall goal of ATC systems is the assistance of persons with cognitive disabilities to facilitate their everyday life. The demands and abilities of users play an important role for the acceptance of ATC in a user's everyday life. According to Scherer et al., psychosocial factors such as disregarding the user's requirements during the design process is a common reason why ATC application is abandoned quickly after deployment [97]. *User-centered* design is a methodology¹ which incorporates the user's demands and abilities early into the design process [34].

*User-centered
design*

Design decisions need to take into account the characteristics of the task. Task analysis is an important technique to reveal task characteristics and provide initial design decisions. In the design of the TEBRA system, we use a task analysis method called Interaction Unit analysis to reveal characteristics of the brushing task using a structured methodology. Interaction Unit analysis is described in section 4.1. The initial design decisions are iteratively refined based on user-centered evaluations: we conducted intermediate evaluation studies during the design process towards a fully functioning prototype in order to test the system design and the components at different levels of development. The intermediate studies are described in sections 4.3 and 4.4, respectively. Throughout the whole design process, we take into account expert knowledge of professional caregivers. For example, in a questionnaire study with the caregivers of *Haus Bersaba*, we asked the caregivers about appropriate content and modalities of prompts. Milestones in the design process of the TEBRA system, from a paper mockup to a fully functioning prototype used in a study with target group users, are depicted in figure 2.

4.1 IN-SITU OBSERVATIONS AND INTERACTION UNIT ANALYSIS

Designing an ATC system based on common-sense knowledge about brushing teeth is not sufficient. Users of the ATC are persons with cognitive disabilities who usually show special characteristics in task execution: firstly, due to decreased motor abilities which often coincide with cognitive disabilities [44], target group users might show uncommon usage of objects. We aim to take into account such differences to common behavior as far as possible in the design of the TEBRA system. Secondly, persons living in a residential home commonly rely on the assistance of a caregiver while brushing their teeth. Caregivers aim to impart a routine in the execution of the brushing task which suits the user's abilities. We analyze the caregiver's way of task assistance and consider important aspects in

¹ according to ISO standard *Human-centered design for interactive systems* (ISO 9241-210, 2010)

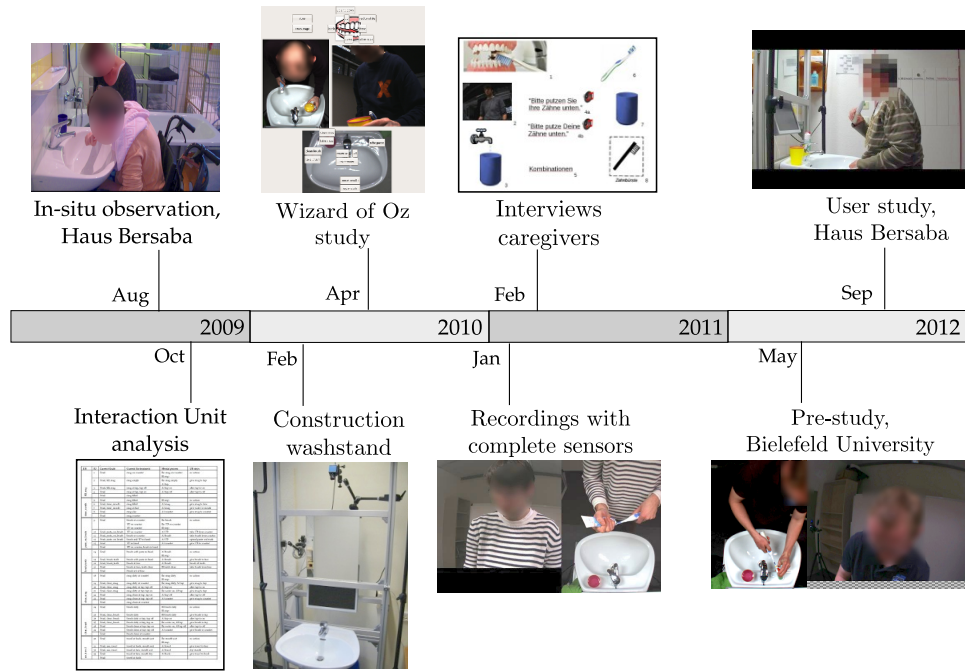


Figure 2: Milestones in the design and development of the TEBRA system.

the design phase. We conduct a qualitative data analysis on in-situ observations made at the residential home *Haus Bersaba* where persons with moderate cognitive disabilities permanently live. In-situ observations are a common way to study a user’s behavior in a natural environment [48]. Each observation is a video which shows a user brushing teeth while being observed and supported by a caregiver. Figure 3 depicts an example image. We recorded 23 trials performed by eight users at three different days



Figure 3: Example image of the in-situ observation.

where seven users conducted three trials each and one user conducted two trials. The users are supported by two caregivers assisting in 10 and 13 trials, respectively.

Interaction Unit analysis

We use Interaction Unit (IU) analysis proposed by Ryu and Monk as a method of task analysis [96]. IU analysis models user-machine interaction with cycles of interaction called interaction units. A user executes actions

in order to achieve a desired goal. Actions are triggered using both visible cues of the environment and mental processes of a user. IU analysis describes actions, goals, environmental states and mental process in a single model and allows to “describe the intimate connection between goal, action, and the environment in user-machine interaction” [96, p. 1].

Hoey et al. use an adapted form of IU analysis to facilitate the specification process of an automatic prompting system using a POMDP [38]. We use a similar form of IU analysis to extract task-relevant information which we incorporate in the design of the TEBRA system. The results of IU analysis, given in table 3, were obtained by iteratively analyzing the recorded videos.

We decompose the brushing task into seven subtasks given in column *UB*. We will refer to the subtasks as *user behaviors* in the following. User behaviors are *paste_on_brush*, *fill_mug*, *rinse_mouth*, *brush_teeth*, *clean_mug*, *clean_brush* and *use_towel*. Column *Current goals* describes a user’s goal stack where *Final* means the overall goal of getting the tooth brushed properly. Whenever a user behavior is initiated, the behavior is added to the goal stack as the user’s current goal. When the user behavior is completed, the goal is removed from the stack and *Final* is the current goal again. Each user behavior is further subdivided into single steps described in column *UB steps*. For example, performing *rinse_mouth* consists of a sequence of three steps: mug is moved to the face, the user rinses his/her mouth and the user moves the mug away from the face. Column *Mental processes* describes the mental processes involved to initiate user behavior steps. Ryu and Monk distinguish between three mental processes: *recognition*, *recall* and *affordance*.

RECOGNITION(RN) Recognition means that the user can directly perceive an object’s state in the environment, e.g. mug is empty in IU 2 in table 3.

RECALL(RL) The user needs to remember a certain state of the environment which is not directly observable. For example, the user has to recall that the mug is dirty in IU 18 because it was used in a previous step.

AFFORDANCE(AF) Affordance describes the recognition of the meaning of an object and the way to use it, e.g. the tap can be altered to on which makes the water flow in IU 20.

Column *Current environment* describes the environmental configuration as preconditions of single user behavior steps. Performing the step changes the environmental configuration, for example in the first step of *paste_on_brush*: the toothpaste tube is on the counter and taking the tube changes the toothpaste location to ‘in hand’.

We utilize the environmental configuration given in column *Current environment* to extract environmental states in terms of discrete variables as depicted in table 4. We distinguish between *behavior* and *progress* variables:

Table 3: Results of the IU analysis for brushing teeth. TT = toothpaste tube, Rn = Recognition, Rl = Recall, Af = Affordance. See text for a detailed description of the table.

UB	IU	Current goals	Current environment	Mental processes	UB steps
fill_mug	1	Final	mug on counter	Rn mug on counter Rl step	no action
	2	Final, fill_mug	mug empty	Rn mug empty Af tap	give mug to tap
	3	Final, fill_mug	mug at tap, tap off	Af tap on	alter tap to on
	4	Final	mug at tap, tap on	Af tap off	alter tap to off
rinse_mouth		Final	mug filled		
	5	Final	mug filled	Rl step	no action
	6	Final, rinse_mouth	mug filled	Af mug	give mug to face
	7	Final, rinse_mouth	mug at face	Af mug	give water to mouth
	8	Final	mug else	Af counter	give mug to counter
paste_on_brush		Final	mug counter		
	9	Final	brush on counter TT on counter TT on counter	Rn brush Rn TT on counter Rl step	no action
	10	Final, paste_on_brush	TT on counter	Af TT	take TT from counter
	11	Final, paste_on_brush	brush on counter	Af brush	take brush from counter
	12	Final, paste_on_brush	brush and TT in hand	Af TT	spread paste on brush
	13	Final	TT in hand	Af counter	give TT to counter
brush_teeth		Final	TT on counter, brush in hand		
	14	Final	brush with paste in hand	Af brush Rl step	no action
	15	Final, brush_teeth	brush with paste in hand	Af brush	give brush to face
	16	Final, brush_teeth	brush at face	Af brush	brush all teeth
	17	Final	brush at face, teeth clean	Rl teeth clean	take brush from face
clean_mug		Final	brush not at face		
	18	Final	mug dirty at counter	Rl mug dirty Rl step	no action
	19	Final, clean_mug	mug dirty at counter	Rn mug dirty, Af tap	give mug to tap
	20	Final, clean_mug	mug dirty at tap, tap off	Af tap on	alter tap to on
	21	Final, clean_mug	mug dirty at tap, tap on	Rn water on, Af tap	give mug to tap
	22	Final	mug clean at tap, tap on	Af tap off	alter tap to off
clean_brush	23	Final	mug clean at tap, tap off	Af counter	give mug to counter
		Final	mug clean at counter		
	24	Final	brush dirty	Rn brush dirty Rl step	no action
	25	Final, clean_brush	brush dirty	Rl brush dirty	give brush to tap
	26	Final, clean_brush	brush dirty at tap, tap off	Af tap on	alter tap to on
	27	Final, clean_brush	brush dirty at tap, tap on	Rn water on, Af tap	give brush to tap
use_towel	28	Final	brush clean at tap, tap on	Rn water on, Af tap off	alter tap to off
	29	Final	brush clean at tap, tap off	Af counter	give brush to counter
		Final	brush clean at counter		
use_towel	30	Final	towel at hook, mouth wet	Rn mouth wet Rl step	no action
	31	Final, use_towel	towel at hook, mouth wet	Af towel	give towel to face
	32	Final, use_towel	towel at face, mouth wet	Af towel	dry mouth
	33	Final	towel at face, mouth dry	Af hook	give towel to hook
	Final	towel at hook			

Table 4: *Behavior* and *progress* variables extracted from the environmental configuration in table 3.

State variable	Values
behavior	
mug_position	counter, tap, face, else, no_hyp
towel_position	hook, face, else, no_hyp
paste_movement	no, yes
brush_movement	no, yes_sink, yes_face
tap_condition	off, on
progress	
mug_content	empty, water
mug_condition	dirty, clean
mouth_condition	dry, wet, foam
brush_content	no_paste, paste
brush_condition	dirty, clean
teeth_condition	dirty, clean

we apply *behavior* variables to recognize user behaviors in a recognition component which we will describe in section 5.2. The *progress* variables are hard to observe using sensory information due to reasons of robustness: for example, it is very error-prone to visually detect whether the *brush_condition* is *dirty* or *clean*. A specialized sensor at the brushing head is not desirable due to hygienic reasons. However, the *progress* variables are important since they are part of the environmental state during the task. We utilize *progress* variables to monitor the user's progress in brushing teeth which will be described in section 5.3.1.

We abstract from the recognition of single behavior steps as given in column *UB steps* in table 3. Instead, we infer the user's behavior based on the *behavior* variables which express states of objects manipulated during a behavior. From column *Current environment*, we extract five *behavior* variables describing important object states: *mug_position*, *towel_position*, *paste_movement*, *brush_movement* and *tap_condition*. The upper part of table 4 shows the five variables and their according discrete values. For *brush_movement*, we have the values *no*, *yes_sink* and *yes_face*. The latter ones are important to discriminate between the user behaviors *paste_on_brush* and *brush_teeth* based on the movement of the brush. The values of the variables *mug_position* and *towel_position* are the different regions identified in column *Current environment* where the mug and towel appear during task execution. *No_hyp* is used if no hypothesis about the mug/towel position is available.

The lower part of table 4 shows *progress* variables and their according discrete values which we use to monitor the user's progress in the task. At

*Behavior and
progress variables*

Progress state space

each time in task execution, the user's progress is modeled by the set of six *progress* variables which we will denote *progress state space* in the following. The occurrence of a user behavior during the execution of the task leads to an update of the progress state space: we define necessary preconditions and effects of user behaviors in terms of *progress* variables. When a user behavior occurs, we check whether the preconditions are met and, if so, update the progress state space with the effects of the current behavior. Table 5 shows the preconditions and effects for user behaviors in terms of *progress* variables extracted during IU analysis. We distinguish

Table 5: Preconditions and effects of user behaviors extracted from the environmental configuration in table 3.

User behavior	Preconditions	Effects
paste_on_brush	brush_content=no_paste	brush_content=paste
	teeth_condition=dirty	brush_condition=dirty
fill_mug	mug_content=empty	mug_content=water
clean_mug	mug_content=empty	mug_condition=clean
	mug_condition=dirty	
	teeth_condition=clean	
rinse_mouth_clean	mug_content=water	mug_condition=dirty
	mouth_condition=foam	mouth_condition=wet
	teeth_condition=clean	mug_content=empty
rinse_mouth_wet	mug_content=water	mug_condition=dirty
	mouth_condition=dry	mouth_condition=wet
brush_teeth	brush_content=paste	teeth_condition=clean
	teeth_condition=dirty	brush_content=no_paste
	mouth_condition=wet	mouth_condition=foam
		brush_condition=dirty
clean_brush	brush_condition=dirty	brush_condition=clean
	teeth_condition=clean	brush_content=no_paste
use_towel	mouth_condition=wet	mouth_condition=dry
	teeth_condition=clean	

between *rinse_mouth_wet* and *rinse_mouth_clean*: the behaviors are equal with regard to object usage, but differ in the semantics based on the time at which the behaviors are executed within the overall task. Video analysis showed that wetting the mouth with water using the mug (before brushing the teeth) is a common step as part of the user's regular daily routine. If a user forgets this steps, the caregiver will intervene and prompt the



Figure 4: Washstand setup equipped with sensor technology.

user to do so. This step is described as *rinse_mouth_wet* whereas cleaning the mouth after the brushing step is *rinse_mouth_clean*. The preconditions and effects of *rinse_mouth_wet* and *rinse_mouth_clean* differ. Hence, we differentiate between these behaviors in tracking a user's overall progress in the task.

The main findings of the IU analysis are three-fold: firstly, we decomposed the brushing task into eight user behaviors given in table 5. Secondly, we identified variables as given in table 4 which describe important objects and according discrete states that are relevant during task execution. Thirdly, we determined preconditions and effects of user behaviors shown in table 5 in order to track a user's progress in the task.

In the following section, we describe the construction of the washstand setup and the equipment of the setup with sensor technology in order to recognize behaviors identified in the IU analysis.

4.2 SETUP

We built a washstand setup as depicted in figure 4. The frame of the washstand is constructed of aluminum profiles². We installed a customary washbowl with a single-lever mixer tap and a mirror. All installations

² manufactured by Bosch Rexroth [<http://www.boschrexroth.com>; accessed 23-November-2012]

comply with the DIN 18024-2 norm for sanitary areas which are accessible for people with impairments: the top edge of the washbowl needs to be at most 80 cm in height. The washbowl has to be equipped with a single-lever tap. The setup has to provide legroom with a depth of 30 cm and a height of 67 cm below the washbowl for wheelchair access.

We equipped the washstand with a TFT display including speakers as a device to prompt the user during task execution. As shown in figure 4, the TFT display is installed between the mirror and the sink. We integrated the prompting device into the setup in a central position because we don't want to shift the user's attention away from the washstand during prompting.

In order to give appropriate prompts to the user, the system needs to implement a context-aware behavior and needs to recognize the user behaviors identified in the IU analysis as given in table 3. Hence, the washstand is equipped with a set of sensors for environmental perception. We use a combination of sensors which we installed in the environment and tools as described in the following subsections.

4.2.1 *Environmental sensors*

The equipment of the washstand setup with environmental sensors is sensitive with regard to privacy concerns. Privacy issues arise due to the retrieval and storage of sensitive personal data in a user's bathroom. In the design and development process of the TEBRA system, storing a user's data is necessary to evaluate and enhance system performance. We obtained the user's declaration of consent before collecting sensitive data throughout the studies described in this thesis.

The washstand is equipped with a set of unobtrusive sensors. Unobtrusive means that the sensors are smoothly integrated into the environment without attaching sensors to the user's body directly. We avoid such wearable sensors because we don't want to disturb the user in the execution of the task.

The sensors chosen for the TEBRA system need to provide a sufficient amount of data to detect the states of important objects identified during IU analysis. We equipped the washstand with two cameras to visually capture the important areas involved in tooth brushing: one camera (Imaging-Source DBK 21BF04 Firewire) observes the environment from an overhead perspective and captures the counter and the sink region. A second camera (PointGrey Flea2 13S2C Firewire) with a frontal perspective observes the upper body part of the user including the face. Figure 5 shows example images.

The overhead and the frontal camera grab images of resolution 480x640 and 800x600, respectively. The grabbing frequency of 15 Hz is chosen with respect to real-time constraints. 15 Hz is a trade-off between processing a huge amount of data on the one hand and providing a suitable percep-

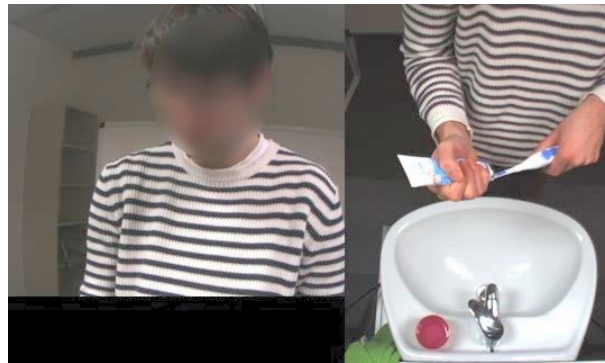


Figure 5: View perspectives of the frontal camera (left image) and the overhead camera (right image).

tion of a user's behavior on the other hand: the higher the frequency, the more computation time is needed to process the data. This might lead to a delayed processing of images. However, if the frequency is too low, the visual recognition components might miss important parts of the users' movements. In both cases, an inappropriate grabbing rate might lead to a faulty perception and, consequently, to an erroneous system performance. According to table 4, the state of the tap (*tap_condition*) and the toothbrush (*brush_movement*) are important for the recognition of user behaviors in tooth brushing. In order to determine the *tap_condition*, we installed a flow sensor (Gentech FCS-03) at the water supply to the tap. The flow sensor measures the water flow providing a binary on/off signal.

In order to distinguish between the three states of the *brush_movement* variable, we installed a sensor module with nine degrees of freedom into the toothbrush which is described in the following subsection.

4.2.2 Sensor-enhanced toothbrush

The toothbrush used in the TEBRA system is a commercially available, electric toothbrush as shown in figure 6. The brush is equipped with an x-imu sensor module manufactured by x-io technologies³ as shown in the bottom right of figure 6. The sensor module has nine degrees of freedom: a gyroscope measuring change in orientation, an accelerometer providing gravitational acceleration and a magnetometer measuring the earth's magnetic field in x,y and z-axis each. The x-imu unit is equipped with a Bluetooth module which is used for wireless data transfer. The module is very compact with dimensions 33 x 42 x 10 mm. However, the module didn't fit into the original plastic housing of the brush. In order to have an ordinary looking toothbrush without extensions which is comfortable to handle, we integrated the module into the brush⁴: the gripping surface of the brush was removed and replaced by a handle in which the sensor module was

³ <http://www.x-io.co.uk/>

⁴ This work was done by Simon Schulz from the Central Lab Facilities (CLF) of the Cognitive Interaction Technology Center of Excellence (CITEC) at Bielefeld University.

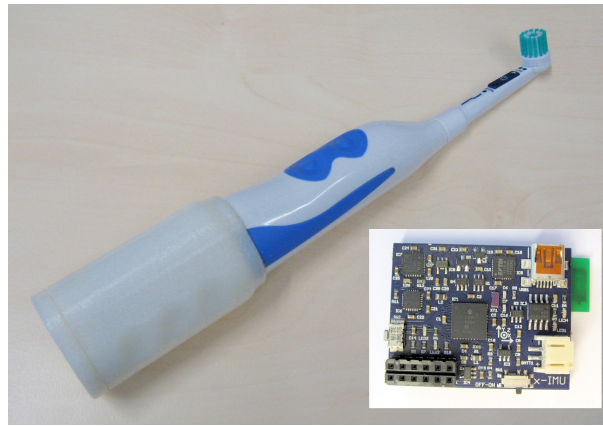


Figure 6: Electric toothbrush used in the TEBRA system. The x-imu sensor shown in the bottom right of the image is installed into the handle of the brush.

integrated. The handle was manufactured from acryl-based photopolymer using a CAD-based design. Acryl-based photopolymer is ideally suited for use in a brushing scenario since it is water-proof and non-toxic according to the DIN norm EN 71-3. The sensor module and the engine in the brush are powered using a lithium-ion polymer battery which is also integrated into the constructed handle. The setup was applied in a first study which is described in the following section.

4.3 WIZARD OF OZ STUDY

A key principle of user-centered design is evaluating and refining components very early in the design and implementation phase. We conducted a study with inhabitants of *Haus Bersaba* where we aimed to evaluate the users' reactions to system prompts in comparison to direct caregiver prompts [80]. At the time of the study, the washstand was equipped with two cameras and a microphone for environmental perception. Both the flow sensor and the sensor module in the brush had not been installed, yet. Hence, the recognition of user behaviors and the planning component were not implemented, too.

We followed the *Wizard of Oz* (WOz) methodology [42] to emulate full functionality of the system: the user brushes teeth at the washstand. The user thinks that he/she is faced with a fully functioning system. Instead, the caregiver - the wizard in our scenario - operates the system via a graphical user interface (GUI) as depicted in figure 7. The caregiver sits behind a room-divider and is not able to observe the washstand directly. Instead, the GUI provides live streamed images and audio from the sensors installed at the washstand. The caregiver assists the user in the brushing task by generating prompts via the GUI: pressing a button on the GUI triggers a prompt which is delivered instantaneously to the user via the TFT display. The GUI is divided into two sections: the lower part of the GUI depicts buttons for sequencing the task into different behaviors,

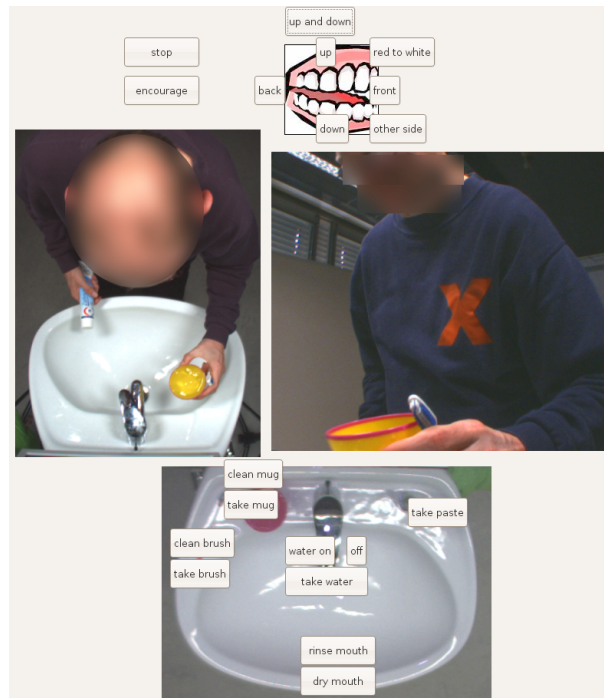


Figure 7: Graphical user interface (GUI) used in the Wizard of Oz study.

e.g. *clean_brush* or *clean_mug*. Each button triggers an audio-visual prompt of the desired behavior. We decided for an audio-visual modality of prompts based on qualitative analysis of the in-situ observations: we identified verbal commands paired with either haptic or visual feedback as the main prompting modality used by the caregivers. Visual feedback is provided with both deictic gestures (e.g. caregiver points to an object of interest) and iconic gestures (e.g. caregiver demonstrates a specific movement which can be adapted by the user). Haptic feedback includes touching the user to attract attention. Since we don't want an automatic prompting system to directly actuate in the user's environment, we avoid haptic feedback. Hence, we decided for real-life video snippets analogous to a caregiver's visual feedback paired with an audio command. The snippets were prerecorded by the author of the thesis. The upper part of the GUI depicts buttons for assisting in the main brushing phase, e.g. brush in the *back* or on the *other side*. Since the distinction between slight differences, e.g. brush *back* and brush *up*, is hard to obtain on a video snippet, we use pure verbal commands for assistance during the brushing phase. In the further development of the TEBRA system, we focus on sequencing the task of brushing teeth into behaviors. We regard the length of the brushing phase, but disregard the exact quality of brushing the different parts of the mouth which is behind the scope of this thesis.

The aim of the WOz study was to evaluate whether and how persons with cognitive disabilities react to system prompts instead of direct caregiver prompts. We conducted six WOz trials with three users performing

Table 6: User’s reactions to audio-visual (AV) and audio (A) prompts in the Wizard of Oz (WIZ) trials and the trials with caregiver assistance (CG).

User	Reactions WIZ(CG)			Reactions AV(A)		
	correct	false	no	correct	false	no
1	85 (44)	5 (44)	10 (11)	100 (77)	0 (8)	0 (15)
2	26 (43)	32 (29)	42 (29)	10 (44)	60 (0)	30 (56)
3	43 (67)	14 (33)	43 (0)	33 (50)	33 (50)	33 (0)

two trials each. We compare the results to trials where the caregiver directly prompts the user. Each of the three users conducted a single trial with caregiver assistance. We will refer to the two scenarios with WIZ (for system prompting generated by the wizard) and CG (for direct caregiver prompting), respectively.

The number of samples in the study is low. Hence, taking into account the average reaction behavior in the two scenarios is not meaningful. Table 6 gives an overview of the user’s individual reactions. We categorize the reactions to the prompts into three classes: *no* reaction, *false* and *correct* reaction to a prompt. *Correct* reaction means that the user adapts his/her behavior according to the prompt. In a *false* reaction, the user switches to a behavior which is different to the prompt.

We identified two main behaviors: user 1 shows a shift from false reactions in the CG scenario to correct reactions in the WIZ scenario with similar no reaction. Correct reactions are highly increased in WIZ (85%) compared to CG (44%). False reactions in CG (44%) are highly decreased in WIZ (5%). We found user 1 having very good abilities to understand system prompts. Users 2 and 3 show a different behavior: user 2 for example has a similar false rate in both scenarios, but no reaction is highly increased in the WIZ scenario (42%) compared to CG (29%). User 3 always reacted to the prompts in CG, either correctly or falsely. However, we see a shift from correct/false reactions to no reaction in WIZ. Users 2 and 3 seem to be distracted from task execution by system prompting.

For a more detailed description of the user’s individual reaction behavior in the WIZ scenario, we take into account the reaction to audio prompts on the one hand and combined audio-visual prompts on the other hand shown in table 6: we observe differences in the reaction to audio prompts compared to audio-visual prompts: user 2 has a moderate rate of correct reactions to audio prompts of 44% and 60% false reactions to audio-visual prompts. User 2 has severe problems following both task execution and audio-visual prompting. We believe that the video snippets distract user 2 in task execution instead of assisting him. User 1, however, shows 100% correct reactions on audio-visual prompts. The video snippets combined with audio seem to be suitable to prompt user 1 in task execution.

The results of the study suggest that an ATC system needs to provide

prompts of different modalities and level of information in order to provide appropriate assistance for users with different cognitive abilities. Furthermore, the responsivity of persons with cognitive disabilities typically varies from day to day. Hence, we aim to implement prompts with different modalities in a graded escalation hierarchy. With increasing level in the escalation hierarchy, the prompts contain an increasing amount of information in order to foster the independence of the users by providing as less assistance as necessary.

In order to investigate the appropriateness of prompts with regard to different modalities and contents, we take into account expert knowledge: we conduct an interview study with caregivers of *Haus Bersaba* described in the following section.

4.4 INTERVIEWS WITH CAREGIVERS

A successful prompt in an ATC system needs to be suitable in modality and level of information in a way that the user can understand and react correctly to the prompt [99]. We aimed to find out about appropriate modalities and levels of information of prompts by conducting interviews with three caregivers of *Haus Bersaba*. The interview manual can be found in appendix A.

Caregivers are experts in prompting since they provide professional nursing care in the task of brushing teeth as part of their daily routine. We interviewed the caregivers independently of each other and recorded the interviews in order to evaluate the caregiver's answers. In each interview, we presented prompts of three modalities: audio prompts, visual prompts and audio-visual combinations.

AUDIO We chose an audio modality due to two reasons: firstly, users are familiar with audio prompts since caregivers mainly use verbal instructions. Secondly, O'Neill and Gillespie argue that "prompting in the verbal medium rather than the visual medium provides a more direct augmentation of executive function" due to a close relationship between language and executive function in the human brain [73, p. 9]. We used audio prompts in terms of verbal commands which were prerecorded by the author of the thesis. We presented commands with different levels of detail ranging from short, specific commands (e.g. "Clean mug.") to more sophisticated instructions (e.g. "Please, clean the mug in front of you."). We asked the caregivers about different properties of the commands: (1) Is a male or a female voice more appropriate for prompting? (2) Is an unknown or a known voice more suitable?

VISUAL Visual prompts are cognitively more demanding than audio prompts since they might shift the user's attention away from the task. However, visual prompts can be very effective since a wide range of visualizations from simple cues such as images to dense

information presentations such as videos are possible. We presented two types of task-related visualizations to the caregivers including different levels of information: images of objects aim to trigger the user's memory and activate a user's routine of task execution by giving appropriate hints. We presented pictograms showing a behavior, cartoon-like images and images of real-life objects. A video comprises much more information in a single prompt than an image: we recorded videos which show the author of the thesis performing a behavior. Hence, the user can directly follow the behavior shown in the video which constitutes a more direct way of assistance. Figure 8 depicts a selection of visual prompts which were presented to the caregivers in the interviews.

AUDIO-VISUAL Audio-visual prompts are combinations of the above-mentioned audio and visual prompts, e.g. a cartoon-like image paired with a verbal command. As a special type of audio-visual prompts, we augmented an audio command with embodiments of prompts such as a virtual agent or a cartoon-like character. For example, figure 8 (d) depicts the virtual agent MAX developed at Bielefeld University [49].

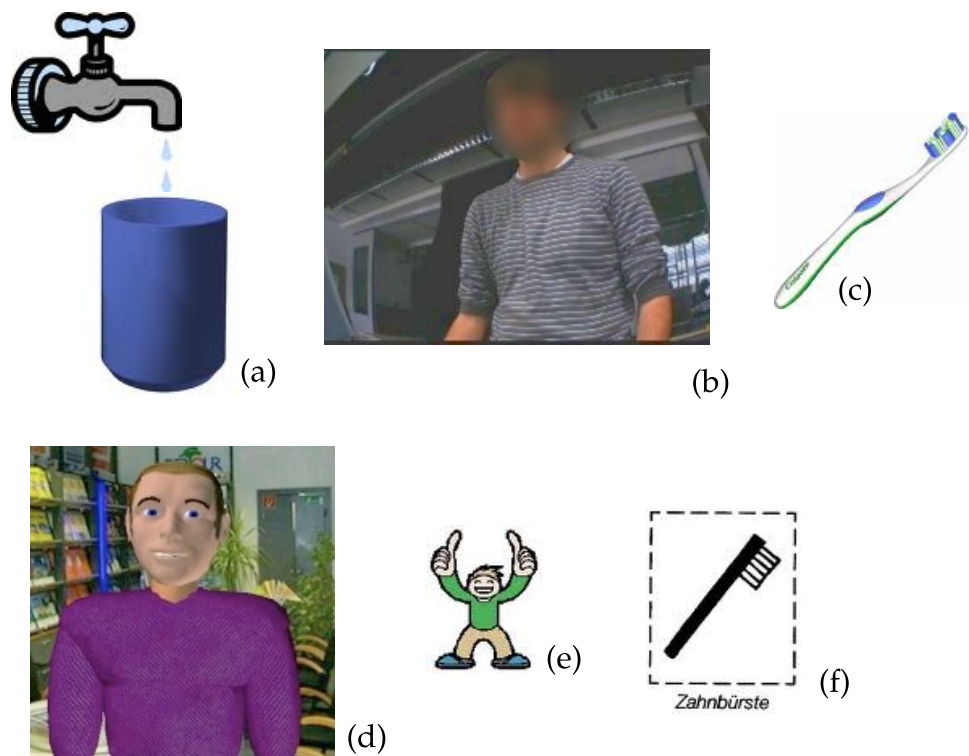


Figure 8: Selection of prompts which were presented to the caregivers in the interviews. (a) cartoon-like image, (b) picture from a real-life video, (c) real-life image, (d) virtual agent MAX, (e) cartoon-like character, (f) pictogram.

The qualitative analysis of the recorded interviews revealed that an audio command is necessary to attract the user's attention. A visual cue only is likely to fail since the user might miss the visual cue. All three caregivers favored short commands in which the textual information is reduced to a minimum. For example, "Clean mug." is preferred to "Please, clean the mug in front of you." since the shorter command is less cognitive demanding than a longer one. Furthermore, male voice is preferred to female voice according to the caregivers. It is negligible whether the voice is known or unknown since the effect of an unknown voice will be obsolete after a few trials with the system.

The caregivers argued that a verbal command should be accompanied by a visual cue. Two types of prompts were favored: *pictogram* prompts and *real-life videos* showing the desired behavior. Pictograms are most likely to suit most of the user's abilities since users are already familiar with pictogram prompts: such prompts are already part of a user's daily routine in *Haus Bersaba*. However, some users might not be able to understand pictogram prompts, but need a more sophisticated visualization: real-life videos showing the desired behavior are appropriate for such users according to the caregivers.

*Pictogram and
real-life video
prompts*

Two of three caregivers perceived an embodiment of audio commands such as a virtual agent and a cartoon-like character as inappropriate since the characters attract the attention of the users, but do not provide a visual cue of the desired behavior. Additionally, users might feel infantilized by a cartoon-like character.

We incorporated the results of the interviews in the development of a two-level prompting hierarchy. Graded prompting hierarchies are a common way to foster a user's independence in task execution by increasing the specificity of prompts during assistance [27]. On the first level, we present pictogram prompts paired with an audio command. If the user doesn't react to a prompt, the TEBRA system will escalate in the prompting hierarchy. On the second level, we present a real-life video of the desired behavior paired with an audio command. Figure 9 shows the exact wording of the audio commands in German and the according translation in English. Furthermore, the pictogram prompts used in the TEBRA system are shown and, exemplarily, the real-life video of *clean_brush* using a film strip visualization. For a single behavior, the same audio command is used in the pictogram prompt and the real-life video prompt.

4.5 DATA RECORDING WITH COMPLETE SENSORS

In the Wizard of Oz (WOz) study described in section 4.3, the sensory equipment of the washstand setup comprised the two cameras only. The flow sensor and the sensor module in the toothbrush were not installed at that time. In parallel to the development of the prompting modalities, we completed the equipment of the washstand setup and installed the flow sensor and the sensor module in the toothbrush.

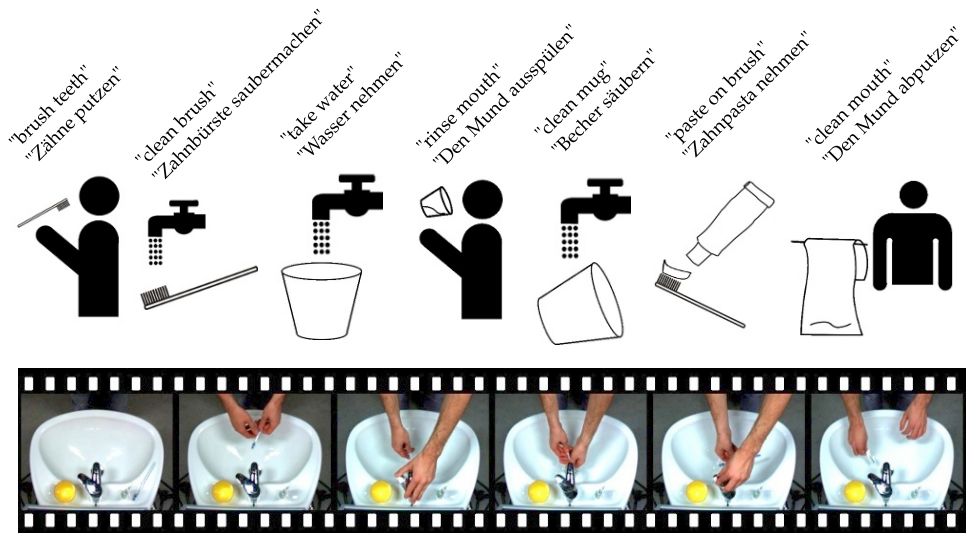


Figure 9: Pictogram prompts and exact wording of audio commands in English and German. The filmstrip visualization at the bottom exemplarily shows the real-life video of *clean_brush* in single images.

We recorded a dataset with regular users brushing their teeth. We aim to develop the user behavior recognition component and the planning and decision making component based on the dataset and the results of IU analysis given in section 4.1. We use data of regular users due to two reasons, here: firstly, a sufficient amount of data with persons with cognitive disabilities is hard to acquire. However, we consider the peculiarities of the target group because IU analysis is conducted on videos with persons with cognitive disabilities. Secondly, regular users show similar characteristics in terms of spatial and temporal execution compared to persons with cognitive disabilities due to individual preferences in task execution. We recorded 18 trials conducted by nine users: two users conducted four trials, three users performed two trials and four users conducted a single trial each. We manually annotated the dataset with the user behaviors given in table 7. The analysis of the annotated data revealed two characteristics of the dataset as given in table 7. Firstly, the total durations of behaviors vary extremely: for example, the total durations of all *brush_teeth* and *clean_brush* behaviors are 1188 s and 114 s, respectively. Hence, the amount of training data which can be used in a supervised classification mechanism, differs between behaviors. A recognition component needs to cope with such imbalance in the training dataset. Secondly, the average durations of behaviors differ significantly: *brush_teeth* and *use_towel* have an average length of 66 s and 4.9 s, respectively. Hence, the behavior recognition component needs to be robust with regard to the duration of user behaviors. Furthermore, the planning and decision making component needs to keep track of the user's progress despite varying durations of behaviors due to an individual user's performance.

Table 7: Total, average, minimum and maximum duration of user behaviors in the dataset with regular users. Column #UB denotes the number of occurrences of each behavior in the dataset.

User behavior	#UB	length in s	
		total	avg (min/max)
paste_on_brush	16	183	11.4 (7.9/15.6)
rinse_mug_fill	14	43	3.1 (1.6/4.4)
rinse_mug_clean	6	17.5	2.9 (1.5/4.2)
rinse_mouth_clean	18	34	1.9 (1.2/2.6)
rinse_mouth_wet	11	22	2 (0.9/3.1)
brush_teeth	18	1188	66 (23.7/174.4)
clean_brush	24	114	4.8 (1.8/26.3)
use_towel	13	64	4.9 (2.2/10.1)

4.6 SUMMARY

In this chapter, we described the design process of the TEBRA system. We followed a user-centered design approach by revealing the characteristics of the task and the involved users: we applied IU analysis as a method of task analysis based on in-situ observations of the users in brushing teeth. The results of the IU analysis are threefold: firstly, we decomposed the task into eight behaviors given in table 5. Secondly, we identified preconditions and effects of these behaviors as shown in table 5. Thirdly, we observed important states of objects which are manipulated during the execution of behaviors. These states are encoded in *behavior* and *progress* variables given in table 4.

We conducted two intermediate evaluation studies during the design process: in a Wizard of Oz study, we evaluated a users' reactions to system prompts. We found out that users show different reaction behaviors to prompts of audio and visual modality due to individual cognitive abilities. In order to provide appropriate prompting for users with different cognitive abilities, we decided to implement an escalation hierarchy incorporating prompts with different modalities and level of information. In order to investigate appropriate modalities of promptings for the escalation hierarchy, we involved professional caregivers as experts in assisting a user in task execution. We conducted an interview study about appropriate prompting modalities with caregivers of *Haus Bersaba*. Caregivers found pictogram prompts and real-life videos paired with an audio command as appropriate for prompting users with moderate cognitive disabilities.

The results of IU analysis, the Wizard of Oz study and the interview study with caregivers revealed that designing an ATC system for persons with cognitive disabilities based on common-sense knowledge is not feasible. An ATC system which disregards the special characteristics of the users,

is most likely to lack in usability and acceptance. For example, users who don't understand the prompts of an ATC system won't most likely feel well assisted by the system.

In the following chapter, we will give a technical description of the TEBRA system. We will describe the main components *user behavior recognition* and *planning and decision making* in detail and show how we incorporate the design decisions obtained in the chapter into the implementation of the TEBRA system.

Real-time capability is one of the major technical requirements for systems which are deployed in real-life situations. Real-time systems can be classified into two categories [105]: *hard* real-time systems have strong constraints with regard to the response behavior of the system. Extending a response deadline is regarded as a system failure. In comparison to hard real-time systems, constraints with regard to response behavior are weakened in *soft* real-time systems: according to Shin et al., soft real-time systems are capable of reacting to situations without perceivable delay for the user [101].

*Soft real-time
capability*

In this thesis, the term real-time capability refers to soft real-time capability. For an ATC system, real-time capability refers to the property of delivering contemporary prompts which integrate into a user's execution of the task at appropriate times and support a user in successful task execution. We avoid measuring response times of the TEBRA system, but qualitatively assess the appropriateness of prompts with regard to timing in the overall task.

In the following section, we give an overview of the overall architecture of the TEBRA system and highlight problems arising from real-time capability.

5.1 OVERALL SYSTEM ARCHITECTURE

One of the key challenges in the implementation of a real-time capable system is the synchronization of data flow between different components of the system. Sensory information need to be processed in a timely manner to provide a coherent system behavior in terms of contemporary prompts. Figure 10 depicts the data flow and processing components in the TEBRA system from sensory information to prompt delivery. We utilize the Robotics Service Bus (RSB) [114] to manage the data flow in the TEBRA system. RSB is an event-driven middleware which provides functionality for data synchronization, data transportation between system components and logging capabilities. RSB uses the *Spread* toolkit for data transportation. Spread "is a group communication system [...] and provides a range of reliability, ordering and stability guarantees for message delivery" [4, p. 1]. The TEBRA system runs on a single machine (DELL Optiplex 990, Quad-Core i5-2500@3.3GHz, 8GB RAM) and uses RSB to communicate data between components of the system.

We use GStreamer¹ to grab data from the Firewire cameras. GStreamer is an open source multimedia framework where different components

¹ <http://gststreamer.freedesktop.org>

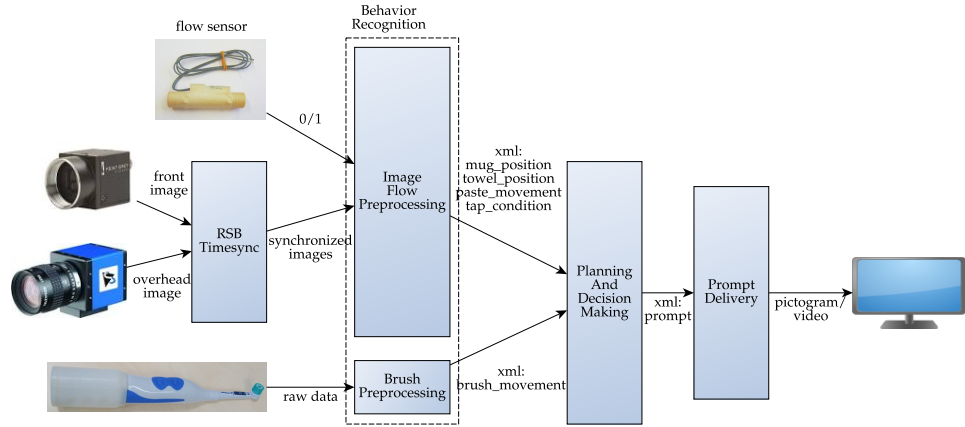


Figure 10: Components and data flow in the TEBRA system. See text for a detailed description.

(plug-ins) are combined to tree-like processing chains. The image data is preprocessed and published to other components using an RSB plug-in for GStreamer. The images are fed into an *RSB Timesync* component which synchronizes the images based on the creation time stamp and provides synchronized data as pairs of images. The synchronized images are passed to the component *ImageFlowProcessing* which queries the flow sensor whenever new image pairs are available. *ImageFlowProcessing* is part of the *Behavior Recognition* and computes a subset of the *behavior* variables given in table 4: *towel_position*, *mug_position*, *paste_movement* and *tap_condition*. The values of the variables are computed and published using a string representation in XML format. The computation of *brush_movement* was transferred to an external component *BrushPreprocessing*. The brush sensor is queried much more frequently than the other sensors with a frequency of about 50Hz. Changes in the orientation of the brush occur with high frequencies. We aim to detect slight changes in orientation by adapting the query frequency to 50Hz in order to detect *brush_movement* robustly.

The *behavior* variables are passed to the *Planning and Decision Making* component which synchronizes *brush_movement* to the residual *behavior* variables: we queue recent observations of *brush_movement*. We synchronize *brush_movement* to the residual *behavior* variables using the creation time stamp of the data.

If the *Planning and Decision Making* component decides for prompting the user, an XML message containing a string representation of the prompt is passed to the *Prompt Delivery* component. *Prompt Delivery* chooses the according pictogram/video prompt and uses GStreamer to show the prompt on the display installed at the washstand.

During system development, logging and storage of system data is important to analyze and improve the system's performance. RSB provides a set of libraries and tools for logging, storage and introspection

of data in the *RSBag* project which we use in the TEBRA system. We store the synchronized image containers, the raw data from the brush and flow sensor, the XML-based configuration of the *behavior* variables and the prompts in order to reproduce the system's behavior. The data is stored in the tide (time-indexed data entries) format which allows for introspection and playback of timely correlated data of different sources.

A single trial of the brushing task generates approximately three to seven GB of data, depending on the length of the trial. The image data makes up the major proportion of the data (~95%). Due to the huge amount of data, the recording component needs to store the data at runtime instead of buffering the data in memory and writing the data to a hard disc after the trial. Buffering might result in a memory overflow. In order to store the huge amount of data in a timely manner, we write the data to a solid state disc which is favorable to a common, magnetic hard disc in terms of writing speed.

The requirement of real-time capability demands the proper interaction of functional components which operate on different temporal levels. Figure 11 depicts the TEBRA system from a functional perspective. Sensor

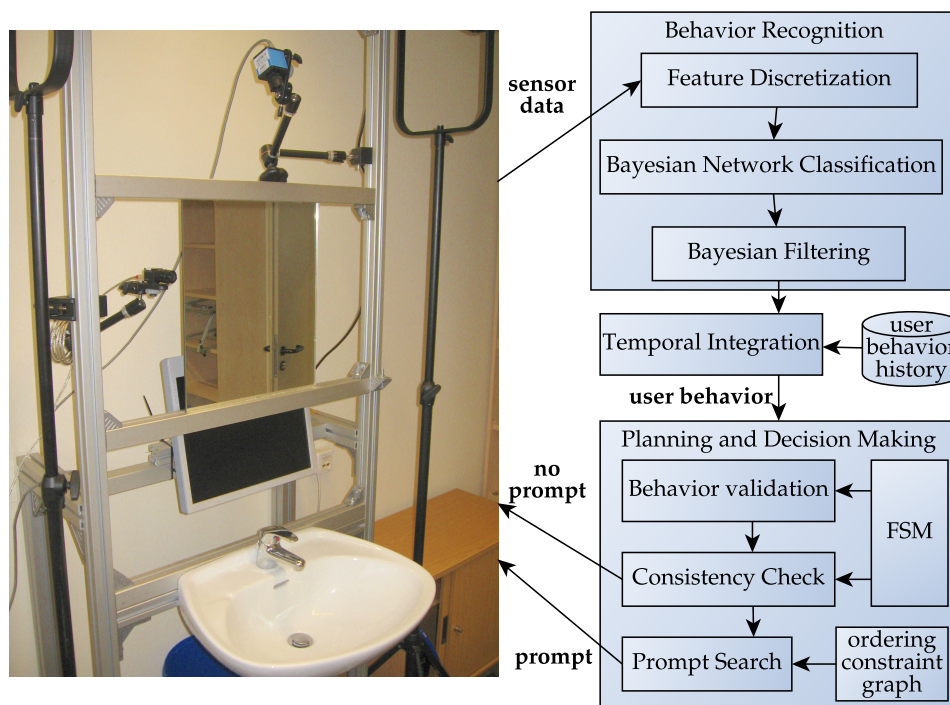


Figure 11: Functional overview of the TEBRA system.

data is provided with a rate of 15 Hz. Each time new sensory information is available, the behavior recognition component computes the current user behavior. A temporal integration mechanism accumulates the user behaviors over time and provides the duration of the behavior in seconds since the planning and decision making component operates on a coarser time scale: the user's progress is tracked in the range of seconds.

We maintain a local history of user behaviors to which the current behavior s is added. We calculate the occurrence rate o_s of s in the history. As long as $o_s \geq k$, the occurrence rate o_s is above a threshold k , we will measure the duration of s in seconds. If $o_s < k$, we reset the current behavior time to 0. k models the sensitivity of the behavior recognition component to perception errors: for example, $k = 1$ doesn't allow for any perception errors during the recognition of a behavior. We tested different values of k between $k = 0.6$ and $k = 0.9$ in test trials. We found $k = 0.8$ to be appropriate in order to allow for misperceptions in the behavior recognition component without resetting the current behavior duration in case of single sensor errors. The planning and decision making component checks the consistency of the current behavior with regard to the user's progress in the task. If the behavior is inconsistent, a prompt will be selected and delivered to the user via the display at the washstand. If the behavior is consistent, the TEBRA system will not prompt, but instead start a new iteration cycle receiving sensor data.

The behavior recognition component and the planning and decision making component will be described in detail in the following sections.

5.2 USER BEHAVIOR RECOGNITION

In an ATC system, the recognition of a user's behaviors is important in order to track the user's overall progress in the task and give appropriate prompts. User behavior recognition is challenging due to the huge spatial variance in the execution of the task: firstly, a user shows an individual way of performing single behaviors. For example, one user may take the paste with the left hand while spreading the paste on the brush. Another user might use the right hand which results in completely different movement characteristics. Additionally, recognizing behaviors of persons with cognitive disabilities is challenging since cognitive disabilities might coincide with motor impairments leading to an even more individualized execution of behaviors [44]. For example, we recorded a person in the in-situ observations described in section 4.1 who was not able to use the left hand due to left-sided hemiparesis. Hemiparesis is a partial paralysis of one side of the body. For performing *paste_on_brush*, the person clenched the brush between the teeth and spread the paste on the brush with the healthy hand.

We abstract from recognizing specific movements by tracking objects or the user's hands due to the huge variance in execution. Instead, we infer a user's behavior based on the environmental configuration which is expressed by states of objects manipulated during a behavior [81]. Interaction Unit (IU) analysis provides both the user behaviors to recognize and the according environmental configuration: we decomposed the tooth brushing task into single user behaviors as described in table 3. Each user behavior is further subdivided into single steps. IU analysis combines semantic information about the behaviors with environmental con-

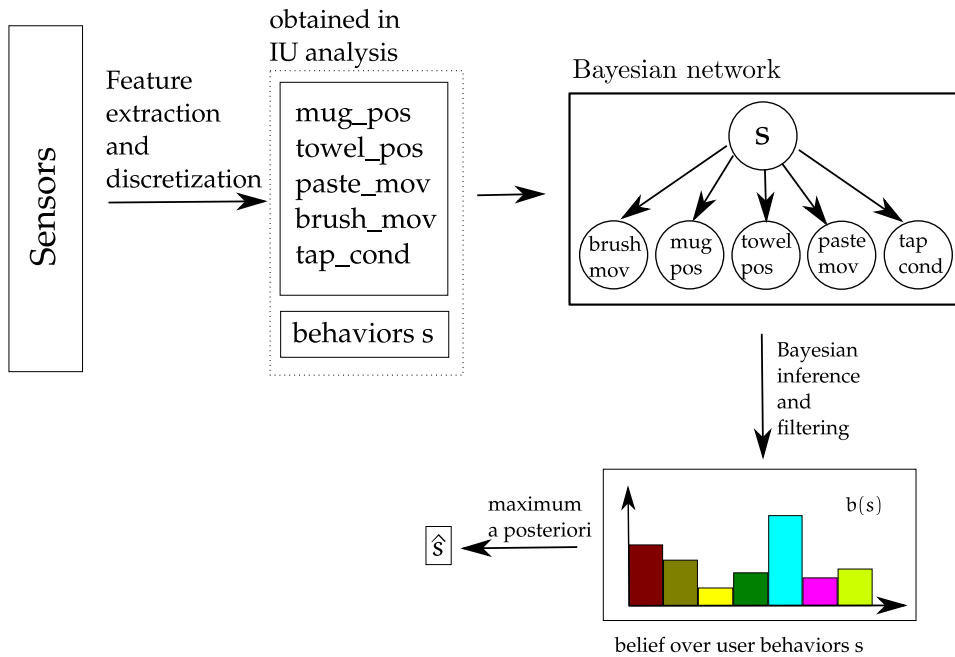


Figure 12: Overview of the user behavior recognition component. See text for a detailed description.

figurations. From the environmental configuration, we extracted the *behavior* variables *mug_position*, *towel_position*, *tap_condition*, *brush_movement* and *paste_movement* as given in table 4 which we use as an intermediate representation in our recognition component.

Figure 12 gives an overview of the recognition component. We extract features from the sensory information which we discretize into the *behavior* variables denoted *mug_position*, *towel_position*, *paste_movement*, *brush_movement* and *tap_condition*. The *behavior* variables were obtained from the environmental configuration in the IU analysis. We recognize a user behavior s based on the *behavior* variables using a Bayesian network classification scheme. A Bayesian inference and filtering approach provides a belief $b(s)$, a discrete probability distribution over user behaviors s . $b(s) = P(S = s | O = \mathbf{o})$ is the probability that the user behavior is s while \mathbf{o} is being observed where $\mathbf{o} = o_1, \dots, o_5$ are the values of the *behavior* variables. We choose the maximum a posteriori hypothesis $\hat{s} = \arg \max_s (b(s))$ which is the behavior with the highest probability in $b(s)$. In the following section, we will describe the recognition component in more detail starting with the feature extraction and discretization.

5.2.1 Feature extraction and discretization

Figure 13 depicts the extraction and discretization of features in the behavior recognition component of the TEBRA system. We use computer vision techniques to detect the mug and towel position as well as the movement of the paste based on camera images.

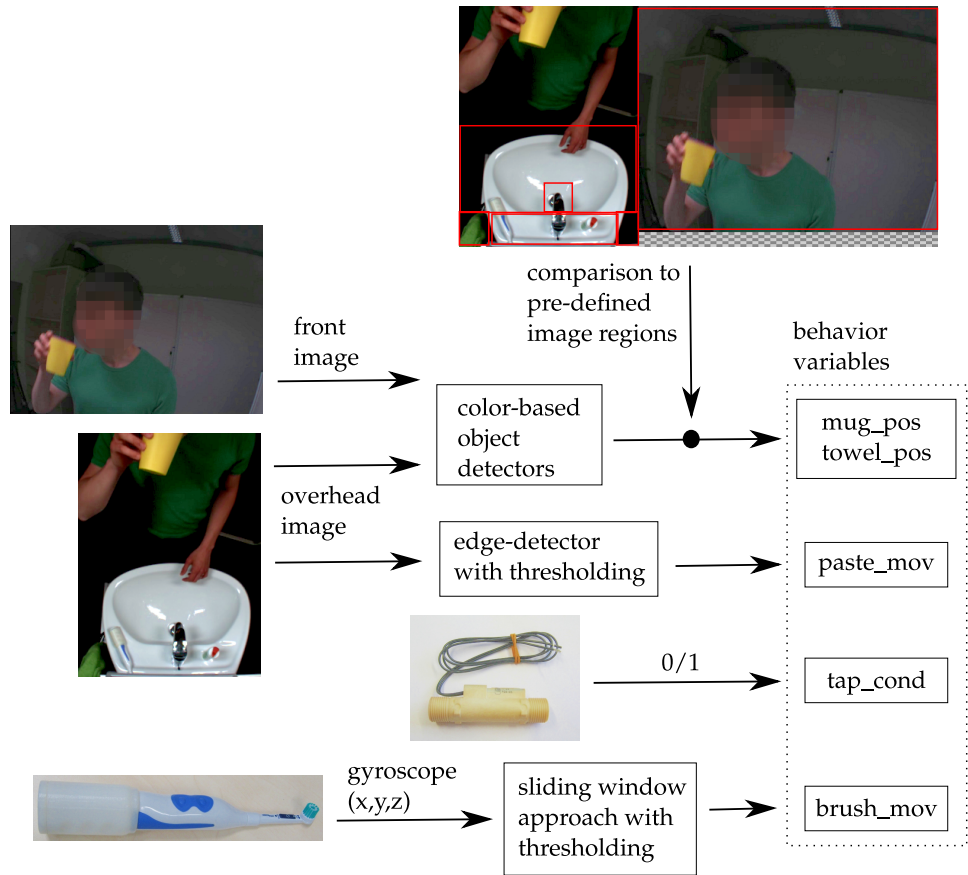


Figure 13: Feature extraction and discretization in the TEBRA system.

Color-based object detector

For the position of the mug and the towel, we apply a color-based object detector as proposed in [102]: the result of the detector is a bounding box which describes the location of an object in an image based on a color distribution model of the object. For each object, a single color distribution model is computed: the color model uses the rg chromaticity space with $r = \frac{R}{R+G+B}$ and $g = \frac{G}{R+G+B}$. The rg space is represented using visual words where each visual word covers an area in the rg space. Visual words are uniformly distributed and cover the whole rg space. Each object is represented as a histogram of visual words.

A known object is detected on a new image using a sliding window approach applied on different image scales. For each window, the color distribution histogram is computed and compared to the learned color model using histogram intersection [107]. The resulting similarity value is entered into a similarity map which is computed over the different scales. The similarity value is the number of pixels in the bounding box agreeing to the learned color model. The similarity map is thresholded to find hypotheses of the object's location. Hypotheses with a size smaller than a defined minimum size are disregarded. The minimum sizes are chosen with regard to minimum sizes of objects when they appear in the sink region.

We utilize an instance of the color detector on both the overhead and front image for the mug and towel position each. On the overhead image, the

detector generates hypotheses for objects on the lower half of the image only. The upper part of the image, which is an overlapping area with the frontal image, is disregarded by the detector in order to avoid confusion of hypotheses when the object is detected on both images as shown in figure 14. We combine the hypotheses of the two detectors by choosing the hypothesis with highest similarity value and, hence, with highest confidence. We chose the color distribution detector due to several reasons: firstly, the detector is robust to partial occlusions. This is important in the tooth brushing scenario where objects are often occluded by a user's hands or body during the execution of the task. Secondly, the detection needs low computational effort which is desirable in a real-time capable system. Thirdly, the detector is partly scale- and rotation-invariant as long as the color distribution of the object is similar from different perspectives. This is given in our scenario since both the mug and the towel are unicolored.

Figure 14 (a) depicts detector results for the mug and towel location in terms of bounding box hypotheses. We compare the center position (x, y)

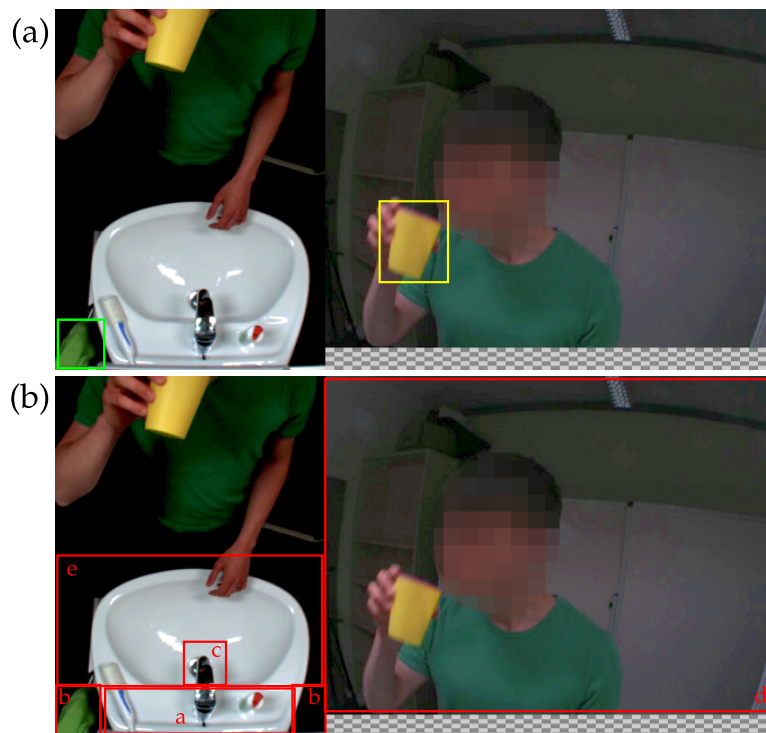


Figure 14: (a) Bounding box hypotheses for mug and towel. (b) Predefined, static image regions used in the discretization of features. a - counter, b - hook, c - tap, d - face, e - else.

of the best hypothesis of an object to a set of predefined, static regions depicted figure 14 (b). Important regions in the brushing scenario are extracted from the IU analysis results. We identified the *counter*, *hook*, *tap*, *face*, and *else* region denoted with a-e in figure 14 (b). In the following, we describe the discretization of each *behavior* variable in detail.

MUG_POSITION For the detection of the mug, important image regions and, hence, important values of the *mug_position* are *counter*, *tap*, *face* and *else*. In our approach, the whole frontal image is the *face* region since we don't explicitly detect a user's face due to two reasons: firstly, the user's face might be occluded by objects or the user's hands for a certain amount of time during the brushing task which makes the face recognition error-prone. Secondly, some users lean forward during the brushing task. Hence, their faces disappear completely from the frontal image which makes a face recognition unreasonable. If the center point of the bounding box hypothesis is in one of the regions, the variable *mug_position* will be set to the according value. For example in figure 14, *mug_position* is set to *face*. If the detector doesn't provide a hypothesis on any image, the variable is set to *no_hyp*.

TOWEL_POSITION The position of the towel is determined similarly to the position of the mug. The static image regions used for towel detection are the *face* and *else* region as described in the previous section. Additionally, two areas left and right of the counter region are treated as a common region *hook* where the towel is usually hang up. Similar to *mug_position*, we set *towel_position* to *no_hyp* if the detector does not produce a valid hypothesis for one of the images.

PASTE_MOVEMENT Movement of the paste is discretized into values *yes* and *no*. We assume that the paste is placed on the counter unless the user applies the paste. The exact region where the paste is located is not important since the paste is only involved in a single behavior *paste_on_brush*. Hence, a binary decision of the movement of the paste is sufficient. If the paste is in the *counter* region, the number of edge pixels is increased compared to the case when the paste is outside the region due to a manipulation by the user. Since the size of the *counter* regions is static, the number of edge pixels in the *counter* region indicates whether the paste is used or not. *Paste_movement* is detected by simply thresholding this number: if the number of edge pixels is below a threshold t_p , the variable is set to *yes*, otherwise to *no*. t_p was determined based on test trials. t_p is very sensitive with regard to lighting conditions: varying illumination of the washstand might lead to erroneous discretizations of *paste_movement*. Hence, we use two spotlights which provide stable lighting conditions.

BRUSH_MOVEMENT We detect whether the brush is moving using the gyroscope data of the sensor module installed in the brush. The gyroscope provides angular velocity in x , y and z direction. We calculate $\sigma = \sqrt{\frac{1}{3} \cdot \sum_{i \in \{x,y,z\}} (\bar{g} - g_i)^2}$, the standard deviation of the gyroscope data in x , y and z direction where $\bar{g} = \frac{1}{3} \sum_{i \in \{x,y,z\}} g_i$ is the mean value. We calculate $\mu_\sigma = \frac{1}{3} \sum_{j=1}^3 \sigma_j$ where μ_σ is the average of σ in three consecutive time steps. If μ_σ is below a threshold

t_b , we set *brush_movement* to *no*. Otherwise, we infer that the brush is moving due to the change in the standard deviation of the angular velocity. Hence, t_b denotes a maximum standard deviation with regard to the angular velocity. In order to distinguish between the behaviors *paste_on_brush* and *brush_teeth* (the brush is moving in both behaviors), we use a Support Vector Machine (SVM) for a more fine-grained classification into the two classes *yes_sink* and *yes_face*: The input features are the 9-dimensional features from the brush module: acceleration, gyroscope and magnetometer data in x, y and z direction each. The SVM with an RBF kernel is trained with a leave-one-out cross-validation scheme on manually labeled data: here, we use the sample data described in section 4.5 which contains a total of 18 trials performed by nine users. The γ and C parameters of the RBF kernel are calculated using an extensive grid search in the parameter space.

TAP_CONDITION The discretization of the *tap_condition* is trivial: if the flow sensor returns 0, *tap_condition* is set to *off*, otherwise to *on*.

In the following section, we describe the recognition of behaviors using Bayesian networks in which we incorporate the values of the *behavior* variables.

5.2.2 Bayesian network classification

IU analysis decomposes the brushing task into user behaviors as given in column *UB* of table 3. We subsume the user behaviors *fill_mug* and *clean_mug* to a common user behavior *rinse_mug* in the recognition component because the *behavior* variables involved as well as the according object states are nearly identical for both user behaviors: the mug is given to the tap and the water is turned on. The distinction between filling and cleaning the mug is not observable with the computer-vision techniques used in the TEBRA system. However, we need to distinguish between *fill_mug* and *clean_mug* in the planning and decision making component in order to properly track the user's progress in the task.

In a regular trial of brushing teeth, user behaviors don't follow exactly on each other, but mostly alternate with transition behaviors: for example, a user's hands approach or leave a manipulated object. We consider these transition behaviors by adding a user behavior *nothing* which we treat as any other behavior in our recognition model.

We apply a Bayesian network (BN) to classify user behaviors denoted with variable *S* based on the *behavior* variables which are referred to as O_1, \dots, O_5 in the following. A BN is a probabilistic graphical model representing a joint probability distribution of random variables. The joint distribution can be visualized using directed acyclic graph (DAG). Each node in the DAG is a random variable. A missing edge between two nodes de-

*Bayesian network
(BN)*

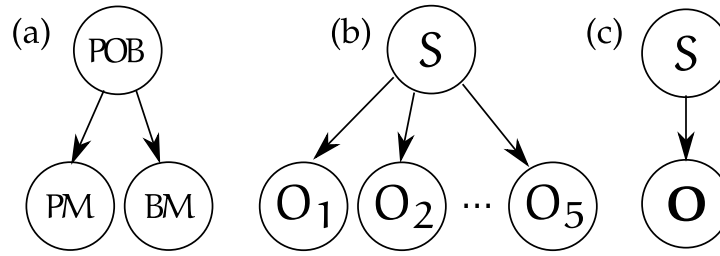


Figure 15: Bayesian networks with three different structures: (a) IU-based structure, exemplarily shown for *paste_on_brush*. *BM* and *PM* are brush and paste movement, respectively. (b) Naive Bayes (c) Holistic.

describes a conditional independence between the random variables. In a BN, the joint distribution can be factorized as a product of density functions of each variable conditional on its parents in the DAG. The factorization leads to a much more compact representation of the joint probability distribution using a product of conditional probabilities.

Probabilistic inference or reasoning in BNs is done by querying the joint probability distribution with regard to instantiations of variables. The complexity of probabilistic inference is highly dependent on the structure of the BN: exact inference based on the elimination of variables is possible in singly connected networks where at most a single path exists between any two nodes of the network. Exact inference algorithms such as variable elimination or clique tree propagation become intractable for multiply connected networks in which multiple paths between nodes exist. In such cases, approximate inference algorithms such as loopy belief propagation make a trade-off between computational complexity and accuracy of the query. For an overview of inference algorithms, we refer to the books of Neapolitan [68] and Darwiche [25].

A BN is ideally suited to model the structural relations between user behaviors S and *behavior* variables O_i . The inclusion of a *behavior* variable O_i in the BN of user behavior S describes that O_i is relevant for classifying S . We extract the relevance relations from the results of IU analysis: for example, the *behavior* variables *paste_movement* and *brush_movement* are involved in *paste_on_brush*. Hence, we consider them as relevant for recognizing *paste_on_brush*. All other *behavior* variables are not regarded in the structure of the BN for *paste_on_brush* which is depicted in figure 15 (a). For each user behavior, we maintain a BN with a structure according to table 8 where we list relevant variables for each user behavior. IU analysis does not provide a structural relationship for behavior *nothing* because IU analysis doesn't model transition behaviors explicitly as user behaviors. Since the transition behaviors can occur between all behaviors, we declare all observation variables relevant for classifying *nothing*. We denote the approach using relevance relations in the BN structure as *IU-based*. We compare the *IU-based* approach with a *Naive Bayes* approach where all *behavior* variables are included in the BN of any user behavior as shown in figure

Table 8: User behaviors and relevant *behavior* variables according to IU analysis.

User behavior	Relevant behavior variables
paste_on_brush	paste_movement, brush_movement
rinse_mug	mug_position, tap_condition
rinse_mouth	mug_position
brush_teeth	brush_movement
clean_brush	brush_movement, tap_condition
use_towel	towel_position

15 (b). Each *behavior* variable O_i is treated as conditionally independent given the user behavior:

$$P(o_1, \dots, o_5, s) = \prod_{i=1}^5 P(o_i|s) \cdot P(s) \quad (1)$$

We are interested in the posterior probability of user behavior s given observation \mathbf{o} which is

$$b(s) = P(s|\mathbf{o}) = \frac{P(\mathbf{o}|s) \cdot P(s)}{\sum_s P(\mathbf{o}|s) \cdot P(s)} \quad (2)$$

The BN with *Naive Bayes* structure has the ability to deal with small training sets since the probability of each o_i depends only on the user behavior s . This is important in our work, because some user behaviors like *clean_brush* are rare compared to other behaviors. Hence, the amount of available training data is limited.

We compare the BNs with *IU-based* and *Naive Bayes* structure to a *Holistic* BN in which the model structure is simplified: we subsume the observations $O_1 \dots O_5$ in a vector \mathbf{O} and treat \mathbf{O} as a single observation as shown in figure 15 (c). A disadvantage of *Naive Bayes* and *Holistic* is the inclusion of variables in the network structure which are not manipulated during the behavior according to IU analysis. For example, the position of the mug doesn't change during the behavior *use_towel*. Hence, the position of the mug doesn't contribute to the recognition of the isolated behavior *use_towel* because the position of the mug is not important for the recognition of the beginning and the end of *use_towel*.

The *Holistic* and *Naive Bayes* BNs are prone to faulty observations which happen occasionally in the discretization of features into *behavior* variables. In the *Holistic* BN, faulty observations are more crucial due to subsuming the observations $O_1 \dots O_5$ to a single observation \mathbf{O} . The *Holistic* BN requires a huge amount of training data in order to represent the conditional probabilities in the BN appropriately despite faulty observations. In *IU-based*, *Naive Bayes* and *Holistic* BN, faulty observations might lead to rapid changes in the belief b from one time step to the next. This is not desirable in our scenario, because transitions between user behaviors

are rather smooth due to the nature of the task: for example, the behavior *brush_teeth* doesn't follow directly on *paste_on_brush* since the toothpaste is usually put away to the counter. Hence, we apply a transition model which takes into account the belief of the preceding time step. This results in a Bayesian filtering approach similar to the forward algorithm in a Hidden Markov Model as the simplest type of a Dynamic Bayesian network. The belief \mathbf{b} is updated to a consecutive belief \mathbf{b}' for each user behavior s' as shown in equation 3. In the following, we will use lower case letter s and o_i denoting instantiations of the random variables S and O_i .

$$\mathbf{b}'(s') = \frac{Z(s', \mathbf{o}) \cdot \sum_{s \in S} T(s', s) \cdot \mathbf{b}(s)}{C} \quad (3)$$

with the normalization term $C = \sum_{s' \in S} Z(s', \mathbf{o}) \cdot \sum_{s \in S} T(s', s) \cdot \mathbf{b}(s)$. $Z(s', \mathbf{o})$ is the probability of making observation \mathbf{o} and the user behavior is s' .

For the *IU-based* approach, $Z(s', \mathbf{o}) = P(o_{\overline{i_{s'}}}) \cdot P(o_{i_{s'}}|s') = \prod_{i_{s'}} P(o_{i_{s'}}|s') \cdot P(s')$ with $i_{s'}$ are the variables i which are relevant for user behavior s' and $\overline{i_{s'}}$ are the variables i which are irrelevant for user behavior s' according to IU analysis. Since $P(o_{\overline{i_{s'}}})$ is independent of behavior s' , the term can be placed outside the sum in the denominator of equation 3 and canceled out.

For *Naive Bayes* and *Holistic*, $Z(s', \mathbf{o}) = \prod_{i=1}^5 P(o_i|s') \cdot P(s')$ and $Z(s', \mathbf{o}) = P(\mathbf{o}|s') \cdot P(s')$, respectively. $T(s', s) = P(s'|s)$ is the probability of a state transition from user behavior s to user behavior s' .

The observation model $Z(s', \mathbf{o})$ is learned on manually annotated training data using Maximum Likelihood (ML) estimation.

$$P_{ML}(O_i = o_j | S = s_k) = \frac{n_{ijk}}{N_{ik}} \quad (4)$$

where n_{ijk} is the number of observations of variable i with value j in behavior k and N_{ik} is the number of observations of variable i when user behavior is k . We apply a leave-one-trial-out cross validation scheme on data of 18 brushing trials to estimate the parameters: the test set consists of data of a single trial and the residual data forms the training set.

Learning the transition model T from data is similar to learning the observation model because a behavior s and the subsequent behavior s' are annotated in the training set:

$$P_{ML}(S = s' | S = s) = \frac{n_{ss'}}{N_s} \quad (5)$$

where $n_{ss'}$ is the number of transitions from user behavior s to s' and N_s is the total number of transitions from s . The ML estimation results in a very peaked state transition distribution: the probability of self transitions is very high. Transitions from one state to another have a very low probability because the number of occurrences of different user behaviors is small compared to the length. Hence, the transition model in equation 3 leads to smooth state transitions between user behaviors because single

Table 9: Comparison of classification rates for the different approaches: OT - Bayesian filtering, O - Bayesian network classifier without Bayesian filtering. IU - BN structure obtained in IU analysis, NB - Naive Bayes, HO - Holistic denote the different BN structure approaches, RMg - rinse_mug, UT - use_towel, PB - paste_on_brush, RMth - rinse_mouth, BT - brush_teeth, CB - clean_brush, N - nothing.

approach	RMg	UT	PB	RMth	BT	CB	N	Avg	
O	IU	73.1	88.1	75.2	81.4	79	48.1	25.9	67.3
	NB	91.6	89.5	90.4	91	71.3	87.6	47.4	81.3
	HO	74.2	89.1	77.9	75.7	72	84.1	49.5	74.7
OT	IU	70.4	91.3	72.1	80.3	89.9	38	28.6	67.2
	NB	94.1	95.1	85.4	95.7	81.7	88.8	50.4	84.5
	HO	79	94.4	75.3	79.3	81.9	85.5	55	78.6

faulty observations can't rapidly change the entire belief from one time step to the next.

In the following section, we will describe the evaluation of the BNs with *IU-based*, *Holistic* and *Naive Bayes* structure in the recognition of user behaviors.

5.2.3 Quantitative results

We show results for the classification of user behaviors based on BNs with different structures on a dataset of 18 trials described in section 4.5. Each trial shows a single brushing task performed by a regular user. Since we abstract from the recognition of specific movements by tracking objects or the user's hands, data of regular users can be used for evaluating our component in a first development cycle because regular users show similar characteristics in terms of a flexible and highly user-dependent execution of the task. We will evaluate the performance of the recognition component with users with cognitive disabilities in section 6.2.2.1.

The 18 trials were performed by nine users where two users conducted four trials each, three users conducted two trials each and four users conducted a single trial each. Table 9 shows the classification rates for individual user behaviors as well as average rates. *IU*, *NB* and *HO* refers to the *IU-based*, *Naive Bayes* and *Holistic* network structure, respectively. We compare the approaches with two different variants of our recognition component: *OT* describes the Bayesian filtering variant where the belief b is updated according to equation 3 using the observation model Z and the transition model T . *O* describes the variant using the Bayesian network without a transition model between behaviors.

The average classification rates for *NB* and *HO* are slightly increased in *OT* compared to *O* except for *IU* where rates are similar. Obviously, the

transition model can deal with faulty observations by suppressing rapid belief changes from one time step to the next which increases the classification rates. The transition model favors smooth belief changes which is desirable for our system due to the nature of the underlying task. In the following, we concentrate on the analysis of OT to compare the approaches in more detail.

The *NB* method leads to the highest classification rates with an average of 84.5%. The BN with *Holistic* structure has a rate of 78.6% which is better compared to the BN with *IU*-based structure with 67.2%, but worse compared to *NB* with 84.5%. For *IU*, *NB* and *HO*, the classification rates for *nothing* are significantly decreased in comparison to other user behaviors. *Nothing* serves as a transition behavior between any two behaviors during the task and models the characteristics of any states of objects manipulated in a transition between two behaviors. Since the various transitions between two behaviors are subsumed in *nothing*, the classification rate of *nothing* is poor in comparison to other behaviors.

The classification rates for single user behaviors are also decreased in *IU* compared to *NB* and *HO*. Furthermore, user behaviors that have an equal or similar set of relevant observation variables are mixed up in the *IU* method as shown in the confusion matrix in table 10. *Clean_brush*

Table 10: Confusion matrix for BN with *IU*-based structure in the Bayesian filtering approach. See table 9 for abbreviations.

	RMg	UT	PB	RMth	BT	CB	N
RMg	70.4	1.4	6.3	3.5	2.4	16.1	0
UT	0	91.3	5.9	0	1.1	0	1.6
PB	0	1.4	72.1	0	21.6	0	4.8
RMth	0	0	6.4	80.3	13.3	0	0
BT	0	0	10.1	0	89.9	0	0
CB	0	3.6	24.1	0	34.2	38.1	0
N	1.7	7.1	24.9	12.7	23.5	1.6	28.6

(for which *brush_movement* and *tap_condition* are relevant) is misclassified as *brush_teeth* (*brush_movement*) with 34.3% and *paste_on_brush* (*brush_movement* and *paste_movement*) with 24.1%. Obviously, the variables which are treated as irrelevant according to *IU* analysis provide important information for distinguishing user behaviors where the same set of objects are manipulated. Hence, irrelevance according to *IU* analysis does not imply conditional independence in the structure of the BN. Obviously, all *behavior* variables are relevant in a way that they contribute information with regard to the recognition of behaviors in the overall context of the task. Although the relevance relations between user behaviors and *behavior* variables are not taken into account in the *NB* and *HO* approaches, the majority of results obtained in *IU* analysis are incorporated in the *NB*

and HO, e.g. the decomposition of the task into behaviors and the environmental configuration of important objects in terms of *behavior* variables. The classification rates of *NB* given in the confusion matrix in table 11 underline the assumption that all *behavior* variables are relevant in the recognition of any behavior: *clean_brush* is misclassified as *brush_teeth* with only 1.3% and *paste_on_brush* with 3.5%. Both *NB* and *HO* make use of the full

Table 11: Confusion matrix for BN with *Naive Bayes* structure in the Bayesian filtering approach. See table 9 for abbreviations.

	RMg	UT	PB	RMth	BT	CB	N
RMg	94.1	0	0.1	1.2	0	2	2.5
UT	0.1	95.2	1	0	0.1	0	3.7
PB	0	0.6	85.4	0.1	6.8	0	7.1
RMth	1.1	0	0.2	95.7	1.2	0	1.8
BT	0.2	0	12.2	0.5	81.7	0	5.4
CB	0.1	0	3.5	6.3	1.3	88.8	0.1
N	1.9	5	14.6	8.9	15.7	3.8	50.4

information available by incorporating all *behavior* variables for each user behavior. However, *NB* has a higher average rate with 84.5% compared to *HO* with 78.6% as shown in table 9. Apparently, *NB* is more suited to deal with small amounts of training data for certain user behaviors in our scenario: due to the conditional independence assumption in the *NB* approach, the probabilities for each *behavior* variable given the user behavior can be calculated independently of the other variables. This leads to a more accurate prediction of the underlying probabilities and a higher classification rate compared to the *HO* method where the *behavior* variables are subsumed in a single observation. As shown in table 9, the *NB* produces good classification results for single user behaviors despite a huge difference in length according to table 7 in section 4.5. The classification rates of *rinse_mouth* and *use_towel* are excellent with 95.7% and 95.2%, respectively. However, the rates of behaviors where *paste_movement* and *brush_movement* are involved according to IU analysis such as *brush_teeth* and *paste_on_brush* are lower with 81.7% and 85.4%. In order to improve the recognition rates of *brush_teeth* and *paste_on_brush*, we aim to enhance the discretization processes for *paste_movement* and *brush_movement* since these variables contribute most information with regard to recognizing the beginning and end of *brush_teeth* and *paste_on_brush*.

5.2.4 Enhancements in discretization

For the movement of the paste, we revised the discretization process based on edge pixels since the thresholding method turned out to be inaccurate when lighting conditions changed while users were leaning

towards the sink. We use an instance of the color distribution detector which we already applied for detecting the mug and towel position. In order to detect the paste robustly, we replaced the tube of toothpaste with a dispenser because a dispenser provides a bigger area to detect compared to a tube of toothpaste which is narrow from an overhead perspective. We ensured that using a paste dispenser is appropriate for our target group users: the caregivers of *Haus Bersaba* didn't see a negative impact in task execution since all persons are capable of using a dispenser.

We apply a single detector instance to recognize *paste_movement*. We aim to distinguish whether the paste is on the counter or in any other region of the image. If the paste is not on the counter, we will infer that the paste is manipulated by the user. Hence, paste detection on the overhead image is sufficient. If the center point of the best hypothesis for the paste is located in the *counter* region, *paste_movement* will be set to *no*, and otherwise, to *yes*.

We revised the calculation of *brush_movement*. The discretization based on the gyroscope data and the subsequent SVM classification of gyroscope, accelerometer and magnetometer data was error-prone for distinguishing *paste_on_brush* and *brush_teeth*.

We distinguish between *yes_sink* and *yes_face* based on the orientation of the brush in order to lower the confusion between *paste_on_brush* and *brush_teeth*. The sensor module in the brush provides *Euler angles* which describe the relative orientation of the brush. We exploit the differences in relative orientation of the brush in *paste_on_brush* and *brush_teeth*. *Yes_sink* refers to the case when the brush is oriented towards the mirror of the washstand as is usually done in *paste_on_brush*. For *yes_face*, the brush is oriented towards the user which is characteristic for *brush_teeth*. We will set *brush_movement* to *yes_sink* if the orientation of the z component of the brush is $g_z \geq -90$ and $g_z \leq 80$ as illustrated in figure 16. Otherwise, we set *brush_movement* to *yes_face*. We determined the threshold values based on test trials where we evaluated different parameter values.

We use a calibration routine prior to every trial which sets the zero orientation according to a fixed initial orientation of the brush. Hence, we ensure that the zero point of the orientation is persistent over all trials and our orientation-based approach is suitable.

In the studies with the first prototype of the TEBRA system described in chapter 6, we apply the enhanced discretization approaches for *paste_movement* and *brush_movement* in Bayesian networks with *Naive Bayes* structure in combination with the Bayesian filtering approach (*OT*) which provided the best results in the behavior recognition.

5.3 PLANNING AND DECISION MAKING

The planning and decision making component aims to track the user's progress in the overall task. We utilize the *progress* variables obtained in

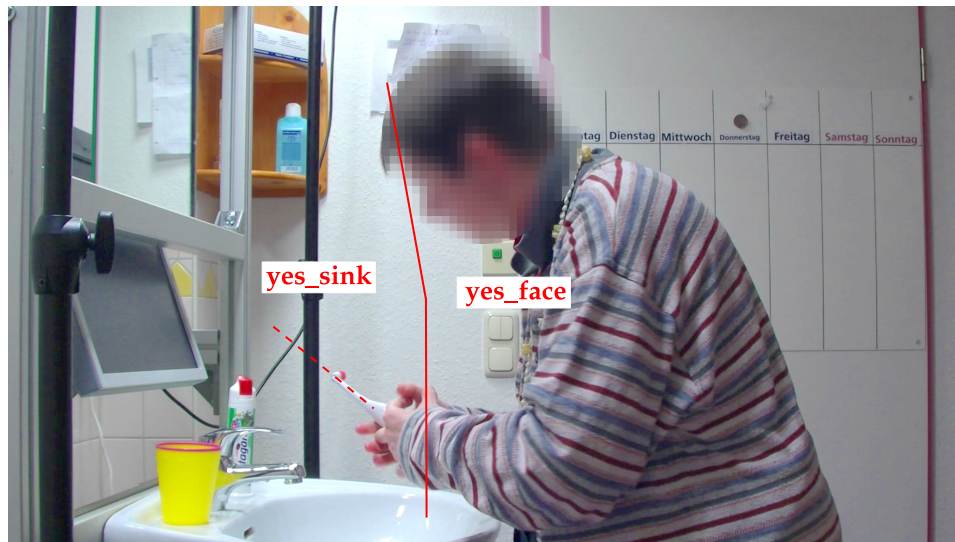


Figure 16: Distinction between *yes_sink* and *yes_face* for variable *brush_movement*. In this example, the orientation of the brush (dashed line) points into the sink region and *brush_movement* is set to *yes_sink*.

the IU analysis as described in section 4.1 into tracking a user's progress which will be described in section 5.3.1. The main challenge in the planning component is the huge temporal variance in the execution of the task. For example, one user may perform *paste_on_brush* much slower compared to another user due to individual abilities. Furthermore, the time a user takes to perform a behavior, might vary from day to day due to the daily mood or the influence of medication. In order to deal with the intra- and inter-personal variance, we use a generalized timing model which we will describe in section 5.3.2. Based on the user's progress in the task and a user's current behavior, the planning and decision making component generates prompts to assist the user in task execution. The main paradigm in the TEBRA system is that a prompt should only be given when necessary in order to foster the independence of the user. A prompt is necessary when the user (1) gets stuck in task execution, (2) performs a behavior at an inappropriate time during task progress or (3) can't terminate a behavior. We will describe the approach for prompt selection in section 5.3.3.

5.3.1 Tracking the user's overall progress

In the behavior recognition component, we can't distinguish between *rinse_mouth_clean* and *rinse_mouth_wet* because the *behavior* variables are nearly identical for both behaviors. Hence, we subsumed the behaviors *rinse_mouth_clean* and *rinse_mouth_wet* to a common behavior *rinse_mouth*. In order to track a user's progress in the overall task properly, we need to distinguish between *rinse_mouth_clean* and *rinse_mouth_wet* since the behaviors have different semantics in the course of the task: *rinse_mouth_wet* describes taking water using the mug before brushing

teeth. *rinse_mouth_clean* denotes removing the foam after brushing by rinsing the mouth with water. Furthermore, the behaviors are different in terms of preconditions and effects as given in table 5: *rinse_mouth_wet* has the preconditions *mug_content=water* and *mouth_condition=dry*. In comparison to the preconditions of *rinse_mouth_wet*, *rinse_mouth_clean* has the preconditions *mug_content=water*, *mouth_condition=foam* and an additional precondition *teeth_condition=clean*. The preconditions *mouth_condition=foam* and *teeth_condition=clean* can only be provided by the behavior *brush_teeth*. Hence, *brush_teeth* serves as a logical border between the behaviors *rinse_mouth_wet* and *rinse_mouth_clean* during task execution. We use this fact in a heuristic in order to distinguish between these behaviors: when *rinse_mouth* is classified by the recognition component, it will be set to *rinse_mouth_wet* if *brush_teeth* has already been recognized during the trial. Otherwise, *rinse_mouth* will be set to *rinse_mouth_clean*.

We apply the same heuristic in order to distinguish between *rinse_mug_fill* and *rinse_mug_clean*. These behaviors are also subsumed to a common behavior *rinse_mug* in the recognition component due to similarities in the involved *behavior* variables: *brush_teeth* is the only behavior which provides precondition *teeth_condition=clean* for *rinse_mug_clean*. When *rinse_mug* is classified by the recognition component, it will be set to *rinse_mug_clean* if *brush_teeth* has already been recognized during the trial. Otherwise, *rinse_mug* will be set to *rinse_mug_fill*.

We use the *progress* variables described in table 4 to track the user's overall progress in the brushing task. The six *progress* variables *mug_condition*, *mug_content*, *mouth_condition*, *brush_content*, *brush_condition* and *teeth_condition* obtained in the IU analysis form a discrete state space. The values of the variables describe the states of objects manipulated during the user behaviors. For example, *mug_content* denotes whether the mug contains *water* or is *empty*.

A successful execution of the tooth brushing task is a transition from an initial state to a final state in the state space. Transitions between states are triggered based on the occurrence of user behaviors: we are not able to robustly recognize whether the effects of user behaviors have occurred due to the limited capabilities of the sensor technology. For example, it is nearly impossible to detect whether a user has spread paste on the brush based on computer vision techniques. Furthermore, an additional sensor for this purpose is not desirable due to hygienic reasons. Hence, we infer the occurrence of behavior effects based on the duration of behaviors. When a behavior is recognized for a certain period of time, we infer that the user has successfully performed the behavior and update the progress state space with the effects of the behavior.

Preconditions and effects of behaviors are extracted from the IU analysis and provided in table 5. Figure 17 depicts the initial and the final state as well as an update of the progress state space for behavior

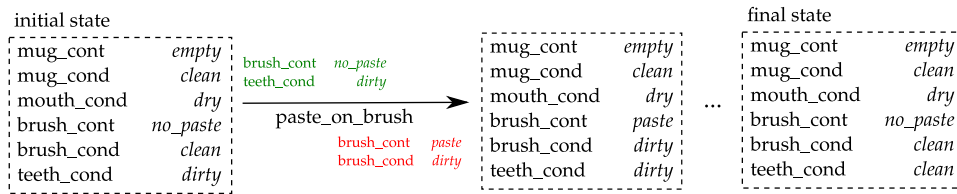


Figure 17: Initial and final state of the state space. Example transition from the initial state to a subsequent state due to behavior *paste_on_brush*.

paste_on_brush. If *paste_on_brush* is recognized, the variables *brush_content* and *brush_condition* of the state space are set to *paste* and *dirty*, respectively. All other variables remain unchanged. The update procedure of the state space is deterministic: the effects of a user behavior on the state space are identical for each point in time during the trial.

In the TEBRA system, we explicitly model the timing characteristics of user behaviors in a dynamic timing model [82] described in the following subsection.

5.3.2 Dynamic timing model

In order to track a user's overall progress in the task, we need to appropriately update the state space based on the occurrence of user behaviors. An appropriate update is challenging with regard to the huge temporal variance in the execution of behaviors due to (1) different durations of behaviors and (2) different velocities of users in task performance. We explicitly model the timing characteristics of user behaviors in a *dynamic timing model* to track a user's progress in the task properly with regard to the following principle: we aim to prevent a user from performing an erroneous behavior by checking the consistency of the behavior as early as possible. If the consistency check is too late, the behavior effects might have been erroneously occurred already. This might lead to an inconsistent state space and erroneous prompts during the remainder of the task.

Dynamic timing model

We subdivide user behaviors into three phases depicted in figure 18: validation, pre-effect and post-effect. Transitions between two phases denote

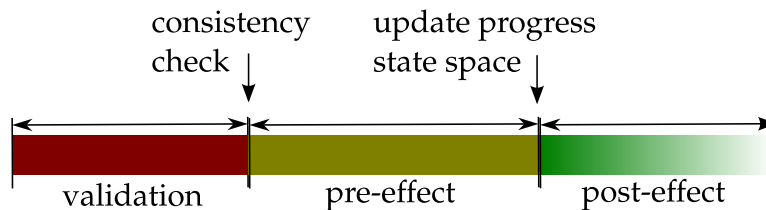


Figure 18: Subdivision of user behaviors in three different phases: validation, pre-effect and post-effect. Transitions between phases denote important events in the planning and decision making component. See text for detail.

important events in the planning and decision making component. At the transition from the validation phase to the pre-effect phase, we check the consistency of the current user behavior with regard to the progress state space. The duration of the validation phase ensures that a user's current behavior is persistent over a period of time. Hence, we avoid delivering erroneous prompts due to temporary errors in the recognition component. At the transition from the pre-effect to the post-effect phase, we update the progress state space with the effects of the current behavior.

We model timing characteristics of user behaviors using a Finite State Machine (FSM). A FSM models states and state transitions of a system. The system can only be in a single state at each point in time. A FSM is fully specified by a finite set of states and a set of conditions at each state which trigger transition between states.

Here, we apply a FSM with five states which is depicted in figure 19. A FSM is suitable because it allows for modeling the different phases

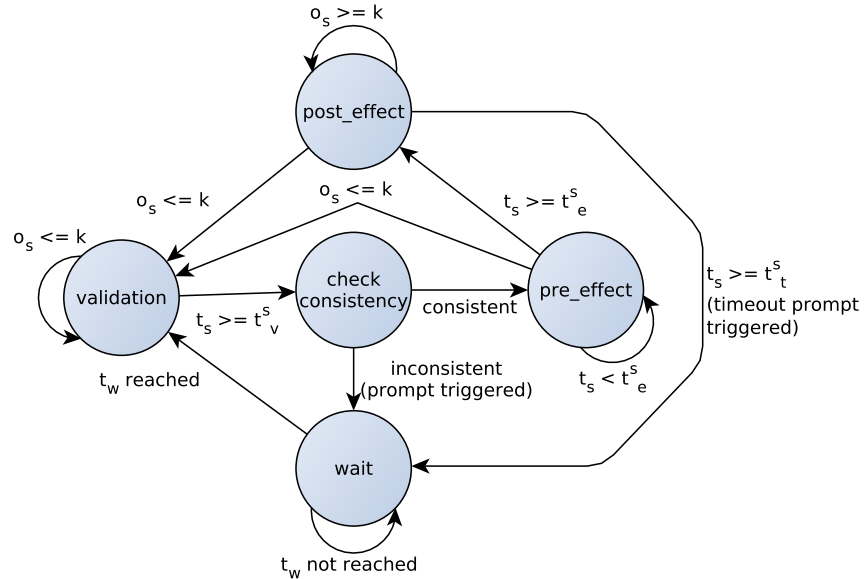


Figure 19: Finite State Machine modeling the timing characteristics of user behaviors. t_s is the duration of behavior s ; t_v^s , t_t^s and t_e^s - validation, timeout and effect time of behavior s , respectively; o_s is the occurrence rate of behavior s in the behavior history; $k = 0.8$ models the sensitivity of the behavior recognition component to perception errors; $t_w = 5$ s is a fixed wait time after a prompt.

during a behavior as well as the transitions between the phases: at system initiation and in the beginning of each user behavior, the FSM is in the *validation* state. The end of the validation phase of the current behavior s is reached after a *validation time* t_v^s by comparing the duration t_s of s to t_v^s : if $t_s \geq t_v^s$, the FSM transits to state *consistency_check*: we perform a consistency check of the current behavior s by comparing the preconditions of s with the progress state space. If the preconditions are not fulfilled, s is inconsistent which means that s is not an appropriate

behavior at that time. For example, consider the initial state given in figure 17. The user performs *brush_teeth* which has three preconditions: *brush_content=paste*, *teeth_condition=dirty* and *mouth_condition=wet*. The preconditions *brush_content* and *mouth_condition* are not fulfilled since the variables are set to *no_paste* and *dry* in the state space. Hence, the behavior is inconsistent at that time. In case of an inconsistent behavior, a prompt is delivered to the user. We describe the selection process of a prompt in detail in subsection 5.3.3.

If the preconditions of s are fulfilled, s is consistent and the FSM transits to state *pre_effect*. The effects of a user behavior occur after a minimum duration of the behavior which we call *effect time* t_e^s . State *pre_effect* denotes that the duration of s is too short for the effects to occur because t_e^s is not reached. If $t_e^s \geq t_s$, we update the progress state space by applying the effects of s . The FSM transits to state *post_effect*. For any user behavior, a *timeout* t_t^s may occur in the *post_effect* state. A timeout t_t^s denotes that the user might not be able to terminate the behavior, e.g. due a user's obsessiveness in task execution. If the duration $t_s \geq t_t^s$, a timeout prompt will be selected and delivered to the user. After a prompt, the FSM transits to a state *wait* for a fixed time $t_w = 5s$ in order to wait for the user to receive the prompt and react properly.

In order to cope with the huge variance in the duration of individual behaviors, we maintain a timing model $\mathbf{t}^s = (t_v^s, t_e^s, t_t^s)$ for each user behavior s . For example, the duration of *use_towel* is usually much shorter compared to *brush_teeth*. Hence, the effect time t_e^s and timeout t_t^s of the behaviors are completely different. The validation time t_v^s can be set higher for longer behaviors to avoid a misdetection of the behavior due to perception errors.

In addition to different durations of user behaviors, users show different velocities in the execution of behaviors due to individual abilities. In the TEBRA system, we allow for different user velocities by maintaining timing models \mathbf{t}_i^s for three different user velocities where $i = \{f, m, s\}$ corresponding to *fast*, *medium* and *slow* execution velocity. The three velocity categories were chosen manually by the author based on the in-situ observations described in section 4.1.

The timing parameters are estimated using an unsupervised learning approach based on durations of user behaviors: the sample data was extracted from 49 trials of brushing teeth which were performed by both regular users (33 trials) and persons with cognitive disabilities (16 trials). We include data of regular users in the estimation of timing parameters since they show similar characteristics in terms of different velocities amongst user compared to persons with cognitive disabilities.

We apply a k-means algorithm which clusters the durations of each user behavior s into $k = 3$ classes corresponding to *fast*, *medium* and *slow* execution velocity. The analysis of the k-means clustering with different values for k underlines that the choice of $k = 3$ clusters based on in-situ observa-

tions is appropriate here: the reconstruction error is $e = \sum_C \sum_l d(x_C^l, \mu_C)$ where $d(x_C^l, \mu_C)$ is the squared Euclidian distance between the data sample x_C^l of cluster C and the according cluster prototype μ_C . For $k = 3$ clusters per user behavior, $e = 2.2 \cdot 10^{10}$. For $k = 2$ clusters, the reconstruction error $e = 4.8 \cdot 10^{10}$ is much higher. For $k = 4$ clusters, $e = 1.3 \cdot 10^{10}$ is lower compared to the reconstruction error for $k = 3$ clusters, but four of the clusters (two for behavior *rinse_mouth_clean*, one for *use_towel* and one for *rinse_mug_clean*) do not contain any data point due to the distribution of durations for these user behaviors. Since we can't utilize empty clusters in the calculation of the timing parameters, we choose $k = 3$ clusters per user behavior here.

We fit a Gaussian distribution $\mathcal{N}_i^s(\mu, \sigma^2)$ over the members of each cluster. We calculate the timing parameters $\mathbf{t}_i^s = (t_{v_i}^s, t_{e_i}^s, t_{t_i}^s)$ using the parameters of the Gaussian distributions. We calculate the validation time $t_{v_i}^s$ of behavior s in velocity model i with

$$t_{v_i}^s = g \cdot \mu_i^s \quad (6)$$

where μ_i^s is the mean duration of the user behavior s in velocity model i and $g = 0.3$. $t_{v_i}^s$ denotes that we validate s in velocity model i for a duration of 0.3 times the mean duration of s in i . $g = 0.3$ was found to be appropriate based on test trials.

The effect time and timeout are set with respect to the inverse cumulative distribution function (invCDF) of $\mathcal{N}_i^s(\mu, \sigma^2)$. For a given probability p , invCDF returns the duration at which the cumulative distribution function (CDF) is p . Exemplarily, we depict the CDF of *rinse_mouth_wet* for velocity *fast* in figure 20: The effect time and timeout are calculated using

$$t_{e_i}^s = \text{invCDF}(p_e) \quad (7)$$

$$t_{t_i}^s = h \cdot \text{invCDF}(p_t) \quad (8)$$

with $p_e = 0.3$, $p_t = 0.9$, $h = 2.5$ and $\text{invCDF}(x) = \mu_i^s + \sigma_i^s \cdot (-1) \cdot \sqrt{2} \cdot \text{erfcInv}(2x)$ and erfcInv is the inverse complementary error function. $t_{e_i}^s$ denotes that the effects of behavior s in velocity model i occur after a duration of $\text{invCDF}(p_e)$. We chose a small parameter $p_e = 0.3$ in the calculation of the effect time of a behavior due to the following reason: missing a successful behavior and an update of the progress state space due to a large effect time is not desirable because it leads to an incorrect progress state space. Setting the effects of a behavior earlier than necessary is less critical since we already validated the user behavior and we assume that the user doesn't terminate the behavior abruptly prior to the effect time. The parameters $p_t = 0.9$ and $h = 2.5$ are chosen sufficiently big in order to prevent the TEBRA system from delivering a timeout prompt too early during a behavior. Choosing the parameters p_t and h too small might result in a prompt which interrupts and confuses a user in regular

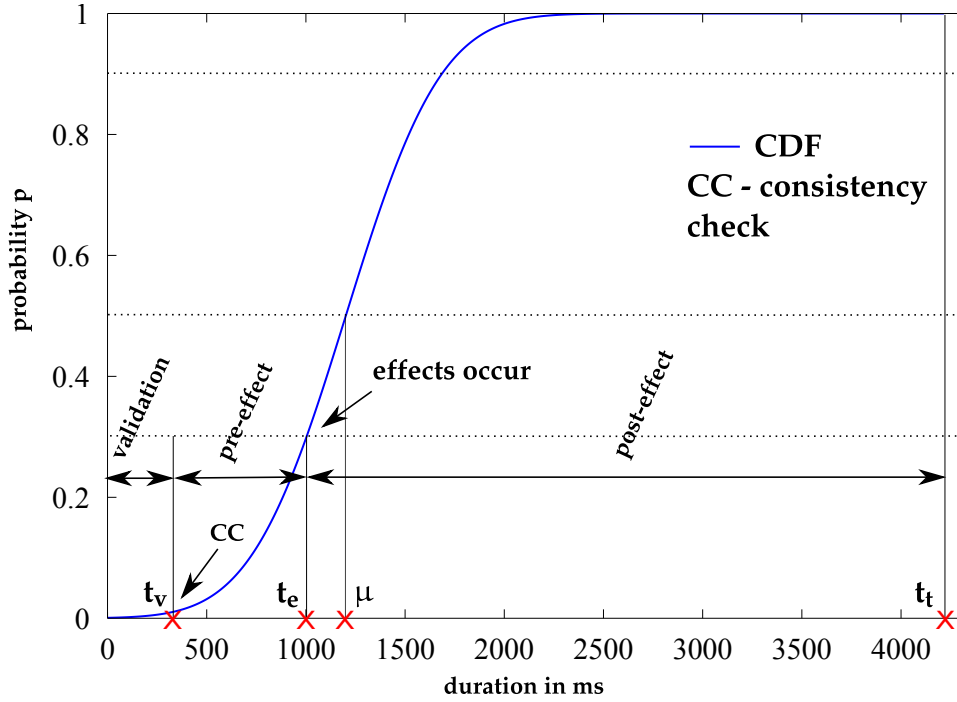


Figure 20: Cumulative distribution function (CDF) of behavior *rinse_mouth_wet* for velocity *fast* showing the different timing parameters and according phases of the FSM.

execution of a behavior.

We have a total number of $N = 72$ timing parameters: for each of the eight user behaviors given in table 5, we have three velocities with three timing parameters each. We calculate the parameters based on four meta-parameters manually set to $g = 0.3$, $p_e = 0.3$, $p_t = 0.9$ and $h = 2.5$ used in equations 6, 7 and 8. Table 12 gives an overview of the timing parameters in seconds. We manually adjusted the timing parameters in two ways: firstly, we set a minimum time for behavior *brush_teeth* proposed by the caregivers in order to ensure that the teeth are sufficiently cleaned. Hence, we set the effect time $t_{e_i}^s = 60s$ for behavior $s = brush_teeth$ in each velocity model i . Secondly, we check the consistency of a user behavior after a maximum behavior duration of 5s in order to prevent a user from performing an inconsistent behavior over a long period of time. Hence, we set the validation time, after which a consistency check is triggered, to $t_{v_i}^s = \max(t_{v_i}^s, 5)$ for each behavior s in each velocity model i . The adjustments of the validation time affected behavior *paste_on_brush* in velocity *slow* and *brush_teeth* in velocities *medium* and *slow*.

We apply the learned timing parameters in a dynamic timing model which chooses the timing parameters of the FSM according to the user's current velocity in a trial. When the user terminates a behavior s , we determine the duration t_s . We categorize the duration into one of the velocity

Table 12: Parameters of the dynamic timing model in seconds for user behaviors in the different velocities. t_v , t_t and t_e - validation, timeout and effect time. The meta-parameters used in the calculation are set to $g = 0.3$, $p_e = 0.3$, $p_t = 0.9$ and $h = 2.5$.

User behavior	fast			medium			slow		
	t_v	t_e	t_t	t_v	t_e	t_t	t_v	t_e	t_t
paste_on_brush	1.4	3.7	17.5	3.4	10.3	35.8	5.0	24.0	60.5
rinse_mug_fill	0.5	1.6	6.4	1.1	3.3	11.0	2.3	7.3	21.9
rinse_mug_clean	0.6	1.9	6.8	1.2	3.5	12.0	2.4	7.3	22.9
rinse_mouth_wet	0.4	1.4	4.4	0.7	2.0	6.2	1.0	3.1	9.2
rinse_mouth_clean	0.5	1.6	5.6	0.9	2.6	8.7	2.2	6.1	25.0
brush_teeth	3.1	60.0	55.7	5.0	60.0	194.7	5.0	60.0	426.5
clean_brush	0.5	1.4	6.6	1.7	5.0	18.6	4.5	11.8	56.0
use_towel	0.8	2.3	10.3	1.7	5.1	17.7	3.1	9.7	30.0

classes *fast*, *medium* and *slow*: we use the probability density functions of the Gaussian distributions of behavior s to calculate the probability

$$p_i^s(t_s) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t_s-\mu)^2}{2\sigma^2}} \quad (9)$$

for each velocity class i . $p_i^s(t_s)$ is the likelihood that the Gaussian distribution modeling behavior s in velocity i generated the behavior with duration t_s . The velocity class of behavior s is $c = \arg \max_i p_i^s(t_s)$ - the velocity class that has most likely produced the behavior with the current duration.

During a trial, we count the number of occurrences of behaviors of each velocity class i . We set a user's current velocity by applying a winner-takes-all method on the velocity counts which chooses the velocity occurring most frequently during the trial so far. In the beginning of a trial, we don't use prior knowledge about a user's velocity in former trials. Hence, we allow for differences in a user's velocity between trials which might arise due to daily mood or effects of temporary medication.

5.3.3 Prompt selection

The main paradigm in the prompting behavior of the TEBRA system is to provide prompts to the user when necessary in order to foster a user's independence. Hence, prompts are triggered in two situations during a trial: firstly, a user can't terminate a behavior and $t_s > t_s^t$, the duration t_s of behavior s exceeds the timeout t_s^t . Secondly, a user's current behavior is inconsistent with the progress state space at a consistency check. Inconsistent means that there are open preconditions of the current user behavior which are not fulfilled in the progress state space. In both situations, an appropriate prompt to the user regards the following aspects: *consistency* and *adaptability*.

CONSISTENCY The prompted user behavior needs to be consistent with the current progress state space in a way that the preconditions of the prompted behavior need to be fulfilled. Furthermore, the prompted behavior should assist the user to make progress towards the overall goal of the task.

ADAPTABILITY The prompted behavior provides at least a single open precondition of the current behavior as an effect since we aim to assist the user in his/her individual way of performing the task. We select a prompt which supplies the open precondition and allows a user to re-perform the desired behavior after correctly performing the prompted behavior. For example, the user has successfully performed *brush_teeth* and performs *use_towel* afterwards. *use_towel* is inconsistent because precondition *mouth_cond=wet* is not fulfilled (*mouth_cond=foam*). In this situation, two prompts are appropriate: *clean_brush* and *rinse_mouth_clean*. The prompt selection mechanism would then decide for *rinse_mouth_clean* because it provides the open precondition *mouth_cond=wet* as an effect. If the user performs *rinse_mouth_clean*, he/she can go on performing *use_towel* as desired which would not be the case with *clean_brush*.

The prompt selection approach in the TEBRA system uses a search strategy on a graph-based representation to find appropriate prompts with regard to the above-mentioned principles. We maintain an *ordering constraint graph (OCG)* which models a set of ordering constraints between user behaviors in the overall brushing task. An ordering constraint is a temporal relation $a \prec b$ where a and b are actions and \prec denotes that a precedes b . We calculate the OCG for the tooth brushing task in a semi-automatic procedure based on a *partial-order planner*.

Ordering constraint graph (OCG)

Partial-order planning is a planning technique which allows for a partial ordering of actions. Given an initial state I , a goal state G and a set of STRIPS-like actions A with preconditions and effects, the partial order planner calculates a set of ordering constraints O and a set of causal links C . A causal link describes that action a provides condition x for b . Any total ordering of actions which is consistent with the temporal ordering represented in the ordering constraints, is a valid plan transferring from the initial to the goal state.

Partial-order planner (POP)

In the TEBRA system, we use the results obtained in the IU analysis to specify the planning domain for the tooth brushing task. The user behaviors and according preconditions and effects as given in table 5 form the set of actions A . The initial state I and the goal state G are extracted from the IU analysis in table 3 and depicted in figure 17. The goal state is $G = [mug_content=empty, mug_condition=clean, mouth_condition=dry, brush_content=no_paste, brush_condition=clean, teeth_condition=clean]$. The initial state I differs only in the variable *teeth_condition* which is set to *dirty* in the beginning of the task.

The set of ordering constraints and causal links for the tooth brushing task are described in table 13. The ordering constraints and causal links

Table 13: Set of ordering constraints and according causal links calculated by the partial-order planner.

Ordering constraint	Causal link
init \prec rinse_mug_fill	mug_content = empty
init \prec paste_on_brush	brush_content = no_paste
rinse_mug_fill \prec rinse_mouth_wet	mug_content = water
paste_on_brush \prec brush_teeth	brush_content = paste
rinse_mouth_clean \prec brush_teeth	mouth_condition = wet
rinse_mouth_wet \prec brush_teeth	mouth_condition = wet
brush_teeth \prec use_towel	teeth_condition = clean
brush_teeth \prec clean_brush	brush_condition = dirty
rinse_mouth_wet \prec rinse_mug_clean	mug_condition = dirty
rinse_mug_clean \prec final	mug_condition = clean
clean_brush \prec final	brush_content = no_paste
use_towel \prec final	mouth_condition = dry

are partly redundant for behaviors which have similar preconditions and effects. For example, both *rinse_mouth_clean* and *rinse_mouth_wet* provide *mouth_condition=wet* for *brush_teeth*. However, only *rinse_mouth_wet* is appropriate as a predecessor of *brush_teeth* according to the results of the IU analysis. Hence, we revised the set of ordering constraints in consultation with the caregivers of *Haus Bersaba* in order to ensure that the ordering constraints are consistent with the caregiver’s daily routine of assistance.

We manually constructed the OCG as depicted in figure 21. An arrow in the OCG describes that the source behavior provides necessary preconditions for the target behavior. For example, *rinse_mug_fill* provides the effect *mug_content=water* which is a precondition of *rinse_mouth_wet*. The OCG depicts no strict execution plan of the task which the user has to follow, but models the ordering between behaviors in the overall task: for example, the behavior sequence *rinse_mug_fill*, *paste_on_brush*, *rinse_mug_fill* is consistent with respect to the partial ordering given in figure 21. Modeling the partial ordering is desirable in the TEBRA system since it allows a user to perform the brushing task in an individual way as long as the overall constraints represented in the OCG are met during task execution. Furthermore, the OCG representation is much more compact with regard to memory consumption in comparison to modeling any allowed transition from the initial state to the goal state explicitly.

We search for an appropriate prompt in the OCG as described in algorithm 1. We determine the open preconditions of the inconsistent user behavior *s*. We process each open precondition as described in algorithm

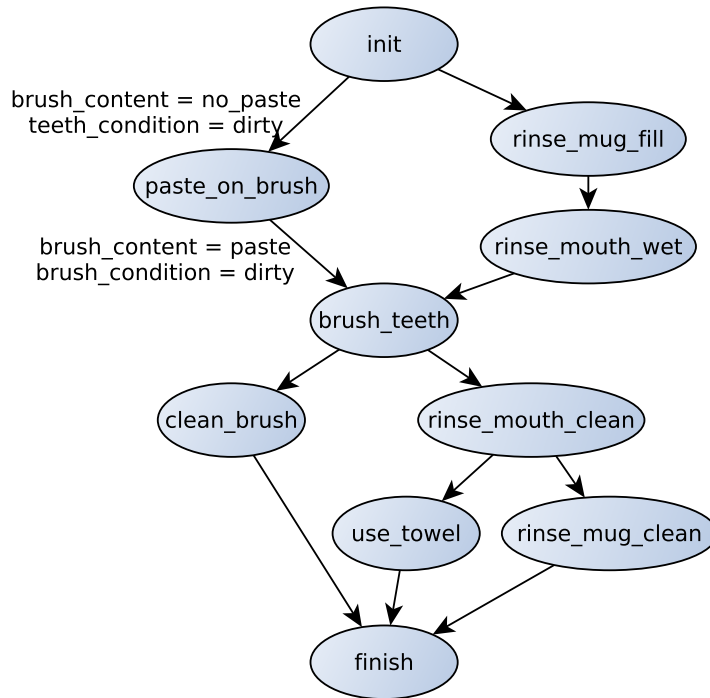


Figure 21: Ordering constraint graph depicting partial orderings of user behaviors in the brushing task. We depict the preconditions and effects of *paste_on_brush*, exemplarily.

2: we search for a user behavior s' which is a predecessor of s in the OCG and provides the open precondition. If s' exists, we check the consistency with regard to the progress state space. When s' is consistent, s' is an appropriate prompt. If more than a single behavior s' provides an open precondition, we will choose the behavior which is closest to the *finish* node in the OCG which represents the goal state. If s' is also inconsistent due to open preconditions, we recursively call *selectPrompt* with s' in order to find a behavior resolving the open preconditions of s' . By recursively calling *selectPrompt*, we resolve chains of open preconditions over several user behaviors by iterating backwards through the OCG. For example, the user already performed *paste_on_brush* and subsequently performs *brush_teeth* which has *mouth_condition=wet* as an open precondition. Function *ProcessPrecondition* finds *rinse_mouth_wet* for supplying the open precondition. However, *rinse_mouth_wet* has *mug_content=water* as an open precondition. In the recursive call to *SelectPrompt*, *rinse_mug_fill* is found to be consistent and provides the open precondition of *rinse_mouth_wet*. Hence, the user is prompted to perform *rinse_mug_fill*.

If no predecessor of s is found providing the open precondition, we search for a consistent behavior by iterating backwards through the OCG starting at the *finish* node. By starting at the finish node, we aim to find a consistent behavior which is most closely to the desired goal state. Furthermore, we avoid prompting for a behavior which the user has already performed or

Algorithm 1 Select appropriate prompt

```

1: function SELECTPROMPT(s)
2:   op  $\leftarrow$  GETOPENPRECONDITIONS(s)
3:   for all op do
4:     prompt  $\leftarrow$  PROCESSPRECONDITION(s,op[i])
5:     add prompt to valid prompts
6:   end for
7:   if |valid|  $\geq$  2 then
8:     prompt  $\leftarrow$  GETCLOSESTTOGOAL(valid)
9:     return prompt
10:  else
11:    return valid[0]
12:  end if
13: end function

```

Algorithm 2 Process precondition

```

function PROCESSPRECONDITION(s,op)
2:  s'  $\leftarrow$  FINDSUPPLYUB(s,op)
   if s' is empty then
4:    prompt  $\leftarrow$  FINDCONSISTENTPREDECESSOR(goal)
   return prompt
6:  else
   CHECKCONSISTENCY(s')
8:  if s' is consistent then
   return s'
10: else
   return SELECTPROMPT(s')
12: end if
   end if
14: end function

```

which doesn't yield progress in the overall task.

In case of a timeout, a user's current behavior is consistent without open preconditions since the behavior has already passed the consistency check during performance. Hence, the prompt selection mechanism directly searches for a consistent follow-up behavior starting at the *finish* node.

5.4 SUMMARY

In this chapter, we gave a technical description of the TEBRA system. One of the major requirements for the TEBRA system is real-time capability. In this thesis, real-time capability describes that the system operates without perceivable delay in a way that prompts are delivered appropriately to the user. In order to provide appropriate prompts, we track the user's progress in the overall task using a deterministic planner based on a partial-order planner and Finite State Machine. A dynamic timing model allows for different velocities of users during task execution which is characteristic for persons with cognitive disabilities. Furthermore, we take into account

different spatial variance in the execution of the task: we infer a user's behavior based on states of objects manipulated during behaviors which we utilize in a Bayesian network classification scheme and a Bayesian filtering approach.

The TEBRA system is robust with regard to occasional errors in the classification of behaviors. Such errors occur due to perception errors in the discretization of sensor data into *behavior* variables. Discretization errors are hard to avoid in our scenario due to the special characteristics of task execution reflected in occlusion or accidental misuse of objects during the task. We maintain a history of user behaviors and a parameter which describes the tolerable ratio of misclassifications during the recognition of a behavior over time.

Furthermore, the TEBRA system is able to deal with unexpected behaviors which are common for persons with cognitive disabilities. For example, a user inspects the brush instead of performing a goal-directed behavior with regard to tooth brushing. The unexpected behavior will be classified as *nothing*. If the duration of *nothing* exceeds 20 seconds, a timeout will be triggered and the TEBRA system provides a prompt which helps the user to make progress towards the overall goal of brushing teeth.

In the following chapter, we will describe the application of the TEBRA system in a pre-study with regular users and the main study with persons with cognitive disabilities.

STUDIES

In the preceding chapters, we described the design and development of the TEBRA system from a paper mock-up to a fully functioning prototype. We apply the first prototype of the system in a pre-study with regular users described in section 6.1. The pre-study is a first evaluation of the overall system and the interplay between the major components *user behavior recognition* and *planning and decision making*. We conduct a pre-study with regular users due to two reasons: firstly, we follow an iterative design and development process in which we refine the TEBRA system regularly based on evaluations with users. We chose regular users who are easier to acquire compared to persons with cognitive disabilities. Secondly, regular users show individual ways in the execution of the task which may not coincide with the system's framework of action. For example, a user rinses his/her mouth using the hands instead of the mug. In this case, the TEBRA system won't recognize *rinse_mug_fill* and issue a prompt in the course of the task. The user has to adapt to the prompts to successfully execute the task from a system's point of view. Hence, we provoke similar phenomena in terms of prompting and reaction behavior in a study with regular users compared to a study with persons with cognitive disabilities.

Based on the results of the pre-study with regular users, we technically enhanced the TEBRA system and conducted the main study of this thesis: we applied the fully functioning prototype of the TEBRA system in a residential home for persons with cognitive disabilities. The design and results of the study are described in detail in section 6.2.

6.1 STUDY WITH REGULAR USERS

The goal of the pre-study with regular users is the evaluation of the technical performance of the TEBRA system. We analyze the classification rates in the user behavior recognition component. We evaluate whether the planning and decision making component provides appropriate prompts to the user. Furthermore, we are interested in the users' reactions to prompts given by the system. The reaction of regular users to system prompts is a measure whether the prompts are semantically reasonable to a minimum degree: if prompts confuse regular users in the task, the prompts might most likely be inappropriate to assist persons with cognitive disabilities since the prompts need to be very precise and clear to suit the cognitive abilities of target group users.

We conducted 26 trials of the brushing task performed by 13 users. Each user performed a single trial in each of two different scenarios: in the *free*

scenario, users received the instruction to brush their teeth as they would regularly do. The system generates prompts if necessary according to the user’s task execution. The users were not explicitly told to follow the system’s prompts, but were free to choose whether they acted according to the prompts. In the *collaborative* scenario, the user was instructed to perform the brushing task in collaboration with the system: we encouraged the users to act according to system prompts whenever they found the prompts to be appropriate throughout task execution.

6.1.1 System performance

In order to calculate the classification rates of the behavior recognition component, we compare the recognition results with ground truth data which was manually annotated by the author of the thesis and a student assistant. Table 14 shows the classification rates of the user behaviors in the 26 trials. The classification rates of *rinse_mug_fill*, *paste_on_brush*,

Table 14: Classification rates of user behaviors in %. RMgC - *rinse_mug_clean*, RMgF - *rinse_mug_fill*, UT - *use_towel*, PB - *paste_on_brush*, RMC - *rinse_mouth_clean*, RMW - *rinse_mouth_wet*, BT - *brush_teeth*, CB - *clean_brush*, N - nothing.

	RMW	RMC	RMgF	RMgC	BT	PB	CB	UT	N
RMW	79.5	0.0	0.0	0.0	0.0	0.6	0.0	0.0	19.9
RMC	32.5	42.0	2.4	1.6	0.0	0.1	9.9	0.0	11.5
RMgF	1.9	0.8	86.0	4.4	0.0	0.0	3.9	0.0	3.0
RMgC	0.0	6.2	0.0	78.8	0.0	0.0	0.5	8.0	6.6
BT	0.6	0.0	0.1	0.0	70.0	25.6	0.3	1.0	2.4
PB	0.0	0.0	0.0	0.0	0.5	99.2	0.0	0.0	0.3
CB	0.8	0.0	0.2	1.8	1.9	6.6	85.4	0.1	3.2
UT	0.0	0.0	0.0	0.2	0.1	0.6	0.0	87.1	12.1
N	3.9	2.4	1.3	1.4	2.8	22.8	2.0	9.8	53.7

clean_brush and *use_towel* are very good with 86%, 99.2%, 85.4% and 87.1%, respectively. However, the rate of *rinse_mouth_clean* is very low with 42%. *Rinse_mouth_clean* is confused with *rinse_mouth_wet* in 32.5% of the cases. As described in section 5.3.1, we use a systematic heuristic to discriminate between *rinse_mouth_clean* and *rinse_mouth_wet*: when *rinse_mouth* is classified by the recognition component, it will be set to *rinse_mouth_wet* if *brush_teeth* was already performed during the trial. Otherwise, *rinse_mouth* will be set to *rinse_mouth_clean*. The heuristic is highly dependent on the recognition of *brush_teeth* which has a classification rate of 70%. The recognition of *brush_teeth* is challenging: the sensor module in the brush measures the change in orientation. The changes are integrated over time to

obtain Euler angles on which the *behavior* variable *brush_movement* is set. In the integration process, small errors are accumulated over time. For a behavior such as *brush_teeth* which usually has a long duration compared to other behaviors, the accumulation of errors leads to misclassifications: *brush_teeth* was mixed up with *paste_on_brush* in 25.6% of the cases.

The average classification rate over all user behaviors is 75.7% where *nothing* has the second lowest classification rate with 53.7%. *Nothing* is the only behavior which is confused with any other behavior in the recognition as shown in table 14. *Nothing* serves as a transition behavior between user behaviors in the brushing task. For example, a user's hands approach or leave an object which initializes or finalizes a behavior. Such phases are included in the training process of the Bayesian network (BN) responsible for *nothing* since we don't explicitly model the initialization and finalization phases of the user behaviors in the according BNs. Hence, the recognition rate of *nothing* is low with 53.7% since the training data consists of initialization and finalization phases of various behaviors. If we drop the rate of *nothing* in the average calculation, we get a classification rate of 78.5% which is a good result with regard to the huge variance in task execution. An important measure of the system's performance is the number of trials in which the user reaches the final state according to the system's framework of action for performing the tooth brushing task. The performance of the TEBRA system in the *collaborative* scenario is excellent as depicted in column *FSR(%)* in table 15. Each of the 13 users reached the final state in

Table 15: Values indicating the performance of the TEBRA system in the *free* and *collaborative* scenario; coll - collaborative scenario, #P - number of prompts, avg P - average no. of prompts, SC - semantically correct prompts, C - correct reaction to a prompt, CSC - correct reaction to a semantically correct prompt, dur - minimum (maximum) duration, FSR - final state reached.

	#P	avg P	SC(%)	C(%)	CSC(%)	Duration	FSR(%)
free	87	6.7	59	10	10	63 (184)	8
coll	117	9	66	75	85	142 (292)	100

this scenario where users ought to follow the prompts when appropriate. In the *free* scenario, only a single user reached the final state: regular users have an individual way of executing the task which may not coincide with the system's framework of action. For example, several users rinsed their mouth by taking water with their hands instead of using the mug. In order to reach the final state, users have to adapt to the prompts given by the system. In the *free* scenario, users were not explicitly encouraged to react to prompts, but were instructed to brush their teeth as they would regularly do. Since all users were capable of brushing their teeth independently of the system, all users except one didn't follow system prompts which leads to the rate of 8%.

The excellent results in the *collaborative* scenario show that the system is able to assist a user in trials which differ significantly in duration: the minimum (maximum) duration in seconds are 63 (184) seconds in the *free* and 142 (292) in the *collaborative* scenario. The trials not only differ in the overall durations, but also in the durations of single user behaviors. We cope with the different behavior durations using the dynamic timing model as described in section 5.3.2. Exemplarily, we show the advantage of the dynamic timing model in two situations. We will give a more detailed analysis of the dynamic timing model in section 6.2.2.1 where we will describe the study with persons with cognitive disabilities.

In the first situation in a trial of user 2, the timing model switched from user velocity *slow* to *fast*. A subsequent user behavior *rinse_mouth_wet* was correctly recognized using the timing parameters of *fast*. However, with the timing parameters of *slow*, *rinse_mouth_wet* would not have been recognized because the duration would have been too short for the effects to occur. The recognition of *rinse_mouth_wet* would have been missed by the system.

In a second situation in a trial of user 6, the timing model switched from user velocity *fast* to *slow*. In the subsequent user behavior *brush_teeth*, a perception error occurred and *brush_teeth* was erroneously recognized as *paste_on_brush* for a duration t . With the timing parameters of velocity *fast* or *medium*, the system would have erroneously recognized *paste_on_brush* as a new behavior due to the decreased validation time. With velocity *slow*, the duration t was too short for *paste_on_brush* to be recognized. The timing model avoided a perception error which would have led to erroneous follow-up prompts in the remainder of the task.

6.1.2 Appropriate prompting

An important measure of the system's prompting behavior is the number of prompts which are semantically correct. We will refer to a prompt as semantically correct if the prompted behavior is appropriate with regard to the user's progress in the task so far. For example, the user has successfully performed *brush_teeth*, but is not able to proceed in the task. A prompt for behavior *clean_brush* is triggered due to a timeout. The prompt is semantically correct because *clean_brush* is appropriate as a subsequent behavior of *brush_teeth*. We determine the semantic correctness by validating each prompt with regard to the ground truth annotation of the behaviors in the task.

In the *collaborative* scenario, 66% of the total number of 117 prompts were semantically correct as depicted in table 15. The 34% of semantically incorrect prompts contain follow-up prompts arising from perception errors: when a user behavior was successfully performed, but not recognized correctly, the system delivers follow-up prompts which are semantically incorrect. If we only regard the first prompts of the system in each trial, the number of semantically correct prompts increases to 92% which is a very

good rate based on the complexity of the task.

The users' reactions to prompts indicate whether the prompts are meaningful to the user. The reaction of the user is correct when he/she updates the behavior to what was prompted by the system. We found 75% of correct reactions in the *collaborative* scenario including semantically correct and incorrect prompts. Taking into account only the semantically correct prompts, correct reactions are much higher with 85%. The residual 15% in which the users' reactions were not correct, are spread over different users who already showed correct reactions to prompts. Hence, we conclude that the users are able to understand the prompts, but were not willing to react to the prompts in these situations.

The rate of correct reactions is an indirect measure of the appropriateness of prompts. Additionally, we directly asked the users to judge the prompts in terms of technical presentation and understandability in a questionnaire. Technical presentation refers to the duration and size of the prompts when displayed on the screen at the washstand. We utilize a seven point Likert scale [53] to assess the user's answers. A value of 1 denotes insufficient and 7 denotes perfectly good.

In the TEBRA system, we use two kinds of audio-visual prompts: pictograms and real-life videos, both paired with an audio command. Users judged the technical presentation of pictogram and real-life video prompts with 6.1 and 6.0, respectively. The understandability of prompts was evaluated with a score 5.9 for pictogram prompts and 5.5 for real-life video prompts. We conclude that the prompts are technically and semantically appropriate at least to regular persons. We will evaluate the appropriateness of prompts for persons with cognitive disabilities in section 6.2.2.2 where we describe the results of the study with target group users.

The results of the pre-study with regular users show that the TEBRA system is able to assist users through the entire task of brushing teeth by giving semantically and technically appropriate prompts.

6.1.3 Technical enhancements

Despite the promising results, the evaluation of the pre-study revealed potential to improve the performance in the major areas of the TEBRA system.

RECOGNITION We aimed to increase the robustness of the recognition component with regard to the spatial and temporal variance of task execution. We learned the parameters of the Bayesian networks on an extended training set where we included the data of the study with regular users. The inclusion of data of regular persons is feasible since regular users show similar characteristics in terms of spatial and temporal variance. We expect that the calculation of the conditional probabilities of the BNs on an extended data set leads to a more robust classification of behaviors.

The successful recognition of *brush_teeth* is very important for a good performance of the system. The classification rates of *rinse_mug_fill*, *rinse_mug_clean*, *rinse_mouth_wet* and *rinse_mouth_clean* are highly dependent on the recognition of *brush_teeth*. If *brush_teeth* is recognized correctly, *rinse_mug* (*rinse_mouth*) will be set to *rinse_mug_clean* (*rinse_mouth_clean*) and otherwise to *rinse_mug_fill* (*rinse_mouth_wet*). We identified problems in the recognition of *brush_teeth* when the user leaned forward towards the sink during the brushing phase. We aimed to increase the recognition rate by adjusting the parameters in the discretization of the data obtained from the sensor module in the brush: we set *brush_movement* to *yes_sink* if the orientation of the z component is $g_z \geq -90$ and $g_z \leq 60$. Otherwise, we set *brush_movement* to *yes_face*. By lowering the second parameter from 80 to 60, we diminish the region where *brush_movement* is set to *yes_sink*.

PLANNING A large proportion of semantically incorrect prompts turned out to be follow-up prompts due to perception errors. We intended to reduce the number of semantically incorrect follow-up prompts using a heuristic: if a user doesn't react to a pictogram prompt, we will escalate in the prompting hierarchy and deliver a video prompting for the same behavior. If a user doesn't react to the video prompts either, we will assume that the behavior was successfully performed by the user, but the recognition component has missed the behavior. Hence, we update the progress state space by applying the effects of the prompted behavior after presenting the third subsequent prompt of the same behavior. According to the caregivers, the heuristic is feasible since the probability of reacting to the third prompt (the second video prompt) of a behavior is very low when the user didn't react correctly to the previous pictogram and video prompt of the same behavior.

PROMPTING In the questionnaire, we included an open-format question asking for suggestions to improve the prompting behavior. Several users mentioned that the audio commands of the prompts are hard to understand when the electrical toothbrush is turned on. Hence, we increased the volume of the audio commands. Furthermore, we conducted a noise reduction in the audio channel to enhance the quality of the signal.

The caregivers suggested adding a further prompt: some of the users forget to turn off the water, but let the water flow for an unnecessary long time. Hence, we will issue a prompt *water off*, if the water is flowing for 30 s. We developed both a pictogram and a video prompt for *water off*.

SYNCHRONIZATION In the evaluation of the pre-study trials, we had synchronization problems between sensor data and subsequent computation results such as the discretized *behavior* variables. We manually

synchronized the data identifying time stamps which were closest in time. In order to facilitate the evaluation of trials in further studies, we use the time stamp of the synchronized image data as a reference time stamp. All time stamps of subsequent calculations which are based on the image data, are set to the reference time stamp. Hence, we synchronize the sensor data and the intermediate results throughout the whole processing chain from sensor data acquisition to prompt delivery.

6.2 STUDY WITH PERSONS WITH COGNITIVE DISABILITIES

The study with persons with cognitive disabilities described in this section is the first study where we deploy a prototype of the TEBRA system to target group users. We cooperate with the residential home *Haus Bersaba* belonging to the *v. Bodelschwingsche Stiftungen Bethel*, a clerical foundation in Bielefeld. 35 persons with cognitive disabilities live permanently in *Haus Bersaba* and receive professional nursing care in their everyday life.

We conducted our study at *Haus Bersaba* due to two reasons: firstly, we aimed to minimize the number of external factors which influence the results of the study. One of these factors is the location of the study. A user might behave differently in an external environment such as a laboratory in the university since the environment is unfamiliar. Hence, we chose the user's familiar environment as the study location. Secondly, we aimed to decrease the effort in transportation for inhabitants and caregivers participating in the study.

The recruiting of participants - called users in the following - was based on inclusion and exclusion criteria which we assessed in conjunction with the caregivers of *Haus Bersaba*. We included users who (1) are motivated to participate in the study, (2) are reliant on a caregiver for successful execution of the tooth brushing task, (3) show appropriate perception and responsiveness to react to verbal and visual assistance, (4) are aged between 18 and 75 and have an IQ greater than 35. Exclusion criteria were severe physical disabilities which prevent the user from fulfilling the task. For example, a user needs to have the motor skills to hold and use the toothbrush. Furthermore, a decreased ability in visual perception which prevents a user from perceiving prompts on the screen, as well as serious medical conditions such as heart deficiency and cancer are exclusion criteria.

We installed the TEBRA system in a vacant room of *Haus Bersaba*. The data recorded during the study is sensitive with regard to privacy concerns: we record data with different sensors including cameras which show users in tooth brushing which is a private activity in a user's bathroom. All participants in the study (caregivers and users/legal guardians) signed a declaration of consent and a sheet of information where we de-

scribed the study procedure as well as the privacy policy. The privacy policy includes that we (1) treat the acquired data strictly confidential, (2) restrict the data access to the investigators of the study, and (3) anonymize the data prior to evaluation. Furthermore, a user is able to terminate the participation in the study at any time without giving any reasons. The sheet of information and the declaration of consent can be found in appendix B.1 and B.2, respectively.

In order to ensure the appropriateness of the study with regard to privacy issues as well as ethical and nursing aspects, we applied for ethical approval at the ethics committee of *Westfälische Wilhelms-Universität Münster*. The ethics committee approved the application without limitation.

The aim of the study is the technical evaluation of the TEBRA system with regard to a user's execution of the task and the investigation of the acceptance and the user's behavior in the interaction with the system.

6.2.1 Study design

The target group in our study consists of seven users (3 male / 4 female) aged between 41 and 56 years. The target group is heterogeneous since the users have different types of moderate cognitive disabilities including behavioral disorders, intellectual disabilities and Autistic Spectrum Disorders. Due to the heterogeneous user group and the small sample size of seven users, general hypotheses in terms of diagnostic assessment and therapeutic treatment of users with specific disabilities is not feasible. Instead, we evaluate the influence of the TEBRA system on a user's individual behavior in brushing teeth.

Single-subject design

We follow a *single-subject design* approach widely used in behavioral science [91, 92]. In single-subject design studies, a user's individual behavior is evaluated under varying, but controlled conditions. For example, a patient with cancer suffers from various symptoms which are observed over a period of time (baseline phase). Afterwards, the patient gets a medication (intervention phase). If the symptoms differ in both phases, a functional relation based on the treatment with medication will be inferred.

In this study, we evaluated the user's behavior in a classical AB design where A and B correspond to the baseline and intervention phase, respectively. The treatment variable here is the entity that provides a user's assistance which is either the caregiver or the TEBRA system. In the caregiver (CG) scenario (baseline phase), users brush their teeth at the washstand. The TEBRA system is working in a way that sensor data is recorded and the user's overall progress in the task is tracked, but the delivery of prompts is suppressed. Instead, a caregiver standing beneath the washstand, assists the user in the brushing task. The CG scenario is the regular way of task assistance in *Haus Bersaba* since all users in our study are reliant on the assistance of a caregiver in brushing teeth during

their daily routine.

In the system (SYS) scenario (intervention phase), users are assisted by the TEBRA system which provides audio-visual prompts via the display installed at the washstand. A caregiver, who is hidden behind a room divider, is present in each SYS trial in order to intervene and take over the assistance in case of fatal system errors.

The seven users conducted trials on nine different days. Each user performed only a single trial in the recording session of a day. We ensured that the trials smoothly integrate into a user's daily routine in order to evaluate the user's behaviors in regular situations as far as possible. Hence, we aimed to align the study times with the regular tooth brushing times of the users as far as possible by conducting the trials in the evenings. We avoided to conduct the study in the morning since the morning routine is stressful for both caregivers and users: most of the users work in sheltered workshops and need to leave to work early. Hence, the caregivers need to provide assistance in the morning routine for working users in very short time. We don't want to increase the burden of caregivers by conducting the study in the morning additionally to the caregiver's stressful routine. We recorded a total of 55 trials: 20 in the CG scenario and 35 in the SYS scenario. One user skipped six trials (1 CG, 4 SYS) due to motivational reasons and participated only in two CG and SYS trials, each. Two trials of user 2 and a single trial of user 3 were terminated due to a system crash. The caregiver assisted the users in the remainder of the task. In CG, a single caregiver assisted in each of the 20 trials.

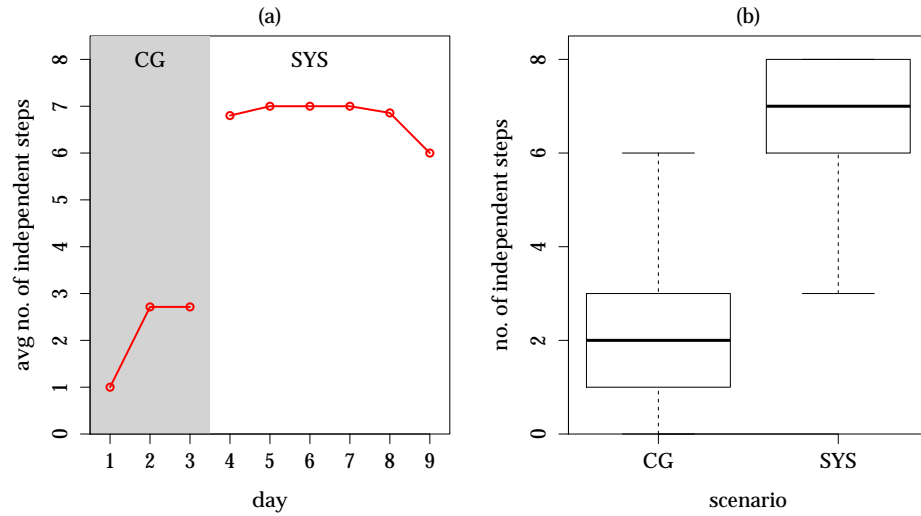
We present and discuss the results of the study in the following section.

6.2.2 Results and discussion

The TEBRA system aims to increase the independence of users and improve their self-confidence by providing appropriate assistance in task execution. An important measure of the influence of the TEBRA system is the number of independent steps - the number of steps which a user is able to perform without the help of a caregiver. For example, a user adapting his/her behavior due to a system prompt is an independent step of the user since no caregiver is involved in prompting.

In the SYS scenario, the number of independent steps is highly increased compared to the CG scenario. Figure 16 (a) shows the average number of independent steps on the nine days of the study where the CG scenario comprised three and the SYS scenario six study days. The average number of independent steps in the CG scenario is stable on days 2 and 3 with around 2.7 steps. On day 1 of the CG scenario, the average number is very low with 1.0 independent steps only. The users brushed their teeth at the unfamiliar washstand for the first time. According to the caregiver, users were highly excited due to the start of the study and the recording of their performance. Hence, the users were unconcentrated which resulted in a

Table 16: (a) Average number of independent steps per trial day. (b) Boxplot of number of independent steps in the CG and SYS scenario. The different steps of the brushing task are listed in table 5.



poor performance in terms of the low number of independent steps. The average number in the SYS scenario is stable over five days with around 7. The average result on the last day of the SYS trials is decreased due to a single user's performance: user 6, who completely skipped four SYS trials due to motivational issues, quit the trial after three steps and left the room due to unknown reasons. Up to the time where user 6 left the room, the performance of the system was not overly erroneous. We conclude that the user left due to personal reasons and not due to inappropriate assistance by the TEBRA system. The decreased number of independent steps in this trial decreased the average rate shown in figure 16 (a). In the following, we will drop the results of user 6 due to the limited amount of data available (only two CG trials and a single SYS trial).

The visual inspection of the average number of independent steps reveals a great difference between the CG and the SYS scenario. The statistical significance of the difference is tested using a non-parametric test. We apply a non-parametric test since the average number of independent steps is skewed and, hence, not normally distributed according to figure 16 (b).

We apply a Mann-Whitney U-Test in order to deal with data which is not normally distributed. The Mann-Whitney U-Test is a rank-based test which evaluates whether the data of two sample sets are drawn from the same distribution. Here, the null hypothesis states that the number of independent steps in the two scenarios are drawn from the same distribution. The test statistic is $U = \sum_{i=1}^m \sum_{j=1}^n S(X_i, Y_j)$ where $S(X_i, Y_j) = 1$ if $Y_j < X_i$, otherwise 0. We reject the null hypothesis if the

value of U is very low or very high: based on the empirical data, the test provides a significant result with $U = 16$ and $p = 3.5 \cdot 10^{-9}$. We reject the null hypothesis since the value of $p \ll 0.05$. We infer that the application of the TEBRA system has an effect in terms of an increased average number of independent steps of users.

The average results over all users hide variations between individual users. Figure 22 shows the number of independent steps for individual users over trials. A red cross denotes a trial in which the system crashed

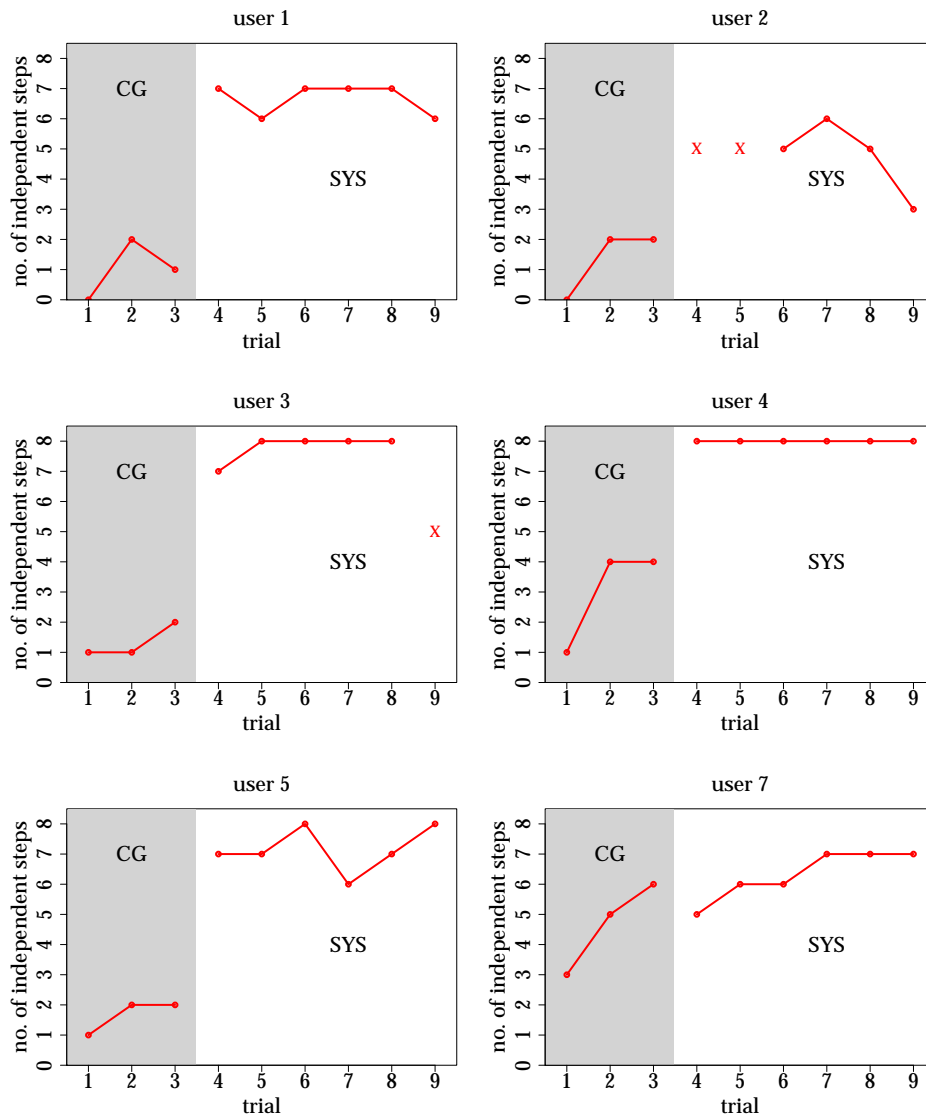


Figure 22: Number of independent steps per trial for individual users.

due to technical problems with the Bluetooth connection of the brush which occurred in three SYS trials.

Users 3 and 4 show excellent results using the TEBRA system: all trials of user 4 were perfect trials in a way that all eight sub steps of the task

were performed independently of a caregiver. User 3 has similar results with an average number of 7.8 independent steps per trial. In comparison to users 3 and 4, user 2, for example, has a lower number of independent steps with 5.5 per trial. In the last trial of user 2, the number of independent steps drops from about five or six independent steps in the previous SYS trials to three: in this trial, user 2 wore a yellow shirt which was very similar in color appearance compared to the yellow mug used in the trials. Parts of the yellow shirt were erroneously recognized as the mug on the frontal image. Hence, the discretization of the mug detector hypothesis into the position of the mug was error-prone throughout the whole trial. This resulted in errors in the classification of user behaviors and, hence, to an increased number of false prompts during the course of the trial. The false prompts confused user 2 in task execution which led to the decreased number of three independent steps in this trial. All users show an increase in the number of independent steps from the CG to the SYS scenario. The amount of increase varies between individual users as shown in table 17. User 7 shows the best performance in the CG scenario amongst all users

Table 17: Average number of independent steps in the CG and SYS scenario for individual users and increase in the number of independent steps from CG to SYS.

		user					
		1	2	3	4	5	7
avg no. of ind. steps	CG	1	1.3	1.3	3	1.7	4.7
	SYS	6.7	4.8	7.8	8	7.2	6.3
	CG to SYS	+5.7	+3.5	+6.5	+5	+5.5	+1.6

with 4.7 independent steps on average. However, the increase of independent steps from the CG to the SYS scenario is low with 1.6. The benefit of the TEBRA system is lower for user 7 compared to other users. For example, user 3 showed an increase of 6.5 independent steps from the CG to the SYS scenario which is an outstanding increase. Hence, the benefit of using the TEBRA system is great for user 3 who showed a low average of 1.3 independent steps in the CG scenario.

In the following subsections, we further analyze the overall performance by evaluating the recognition component and the TEBRA system's ability to deal with spatial and temporal variance in task execution.

6.2.2.1 Technical evaluation

A key challenge for providing appropriate prompting is the recognition of user behaviors. Table 18 shows the results of the user behavior recognition component for the trials in the SYS scenario. The average recognition result over all user behaviors is 69.3%. The classification rates of single behaviors vary extremely between 97.8% for *paste_on_brush* and 41% for

Table 18: Classification rates of user behaviors in the SYS scenario in %. RMgC - rinse_mug_clean, RMgF - rinse_mug_fill, UT - use_towel, PB - paste_on_brush, RMC - rinse_mouth_clean, RMW - rinse_mouth_wet, BT - brush_teeth, CB - clean_brush, N - nothing.

	RMW	RMC	RMgF	RMgC	BT	PB	CB	UT	N
RMW	82.5	0.0	4.8	0.0	0.0	1.2	0.6	0.0	10.9
RMC	31.6	41.0	1.1	5.3	0.0	0.9	0.1	2.6	17.4
RMgF	1.3	1.9	54.9	24.2	0.0	0.0	0.5	1.9	15.1
RMgC	0.0	0.0	1.2	80.4	0.0	2.3	8.1	3.0	5.0
BT	7.0	0.8	1.3	0.4	50.1	25.1	2.7	0.3	12.3
PB	0.3	0.0	0.0	0.0	1.5	97.8	0.0	0.0	0.4
CB	0.0	0.3	0.0	4.3	0.0	3.2	79.9	10.6	1.7
UT	0.0	2.2	0.0	3.3	0.0	2.2	2.5	75.5	14.3
N	4.6	3.0	1.6	2.4	4.4	13.2	4.1	4.7	62.0



Figure 23: Spatial variance in the execution of *clean_brush*.

rinse_mouth_clean.

The results of the behaviors *rinse_mouth_wet*, *rinse_mug_clean*, *clean_brush* and *use_towel* range from 75.5% to 82.5%. These results are good with regard to the huge spatial variance in task execution: spatial variance describes the different movement characteristics of individual users during behavior execution as depicted in figure 23. In the execution of *clean_brush*, one user was holding the tap while cleaning the brush. Another user cleaned the whole brush under the tap. Furthermore, the user's hands are partly or fully occluded as shown for *use_towel* in figure 24. A recognition using a hand or an object tracker would not be feasible due to occlusions. We abstract from the recognition of movement trajectories of objects or the user's hands, but instead infer user behaviors based on states of objects involved in the behaviors. Hence, our recognition component is able to deal with huge spatial variations in behavior execution. For example, for recognizing *use_towel*, the *towel_position* is set to *face* when the towel occurs in the frontal image. Different *use_towel* behaviors can be recognized despite partial occlusions of objects due to body parts of the user.



Figure 24: Spatial variance in the execution of *use_towel*.

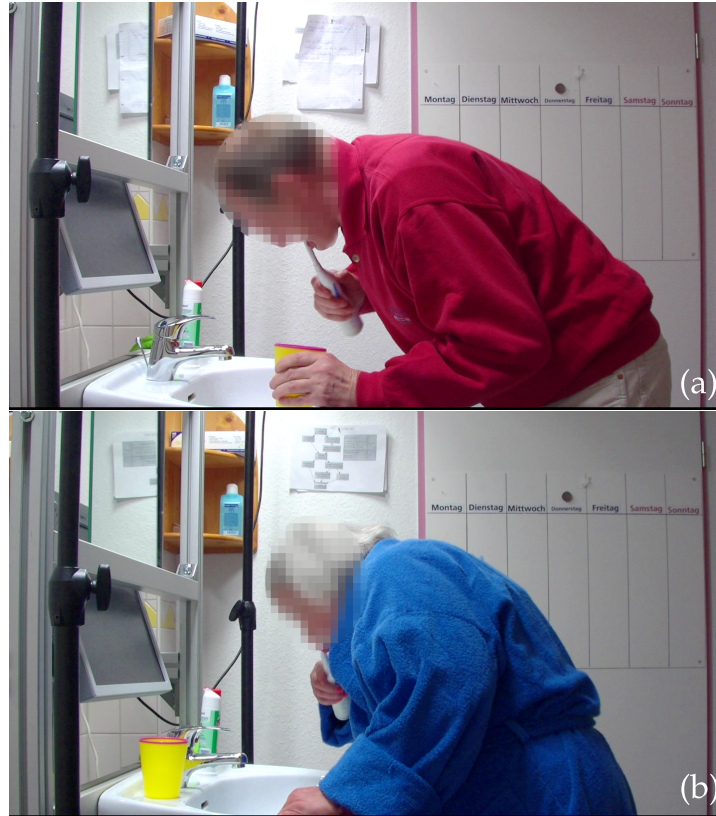


Figure 25: Two users leaning over heavily in the execution of *brush_teeth* which results in a misclassification of *brush_teeth* as *paste_on_brush*.

Furthermore, some users handled multiple objects during the execution of single behaviors. For example, one user was holding the mug while cleaning the brush. The recognition component is able to deal with such situations due to the Naive Bayes structure of the Bayesian network (BN): by including all *behavior* variables in the inference process of any behavior, the BN filters out the effects of the misleading variable.

However, the rates of *rinse_mouth_clean*, *rinse_mug_fill* and *brush_teeth* are poor with 41%, 54.9% and 50.1%, respectively. *Brush_teeth* is mixed up with *paste_on_brush* in 25.1% of the cases. Obviously, the classification based on the orientation of the brush described in section 5.2.4 is error-prone.

Figure 25 exemplarily depicts images from behavior *brush_teeth* for two users. In figure 25 (a), the user leans forward heavily. Hence, the brush is

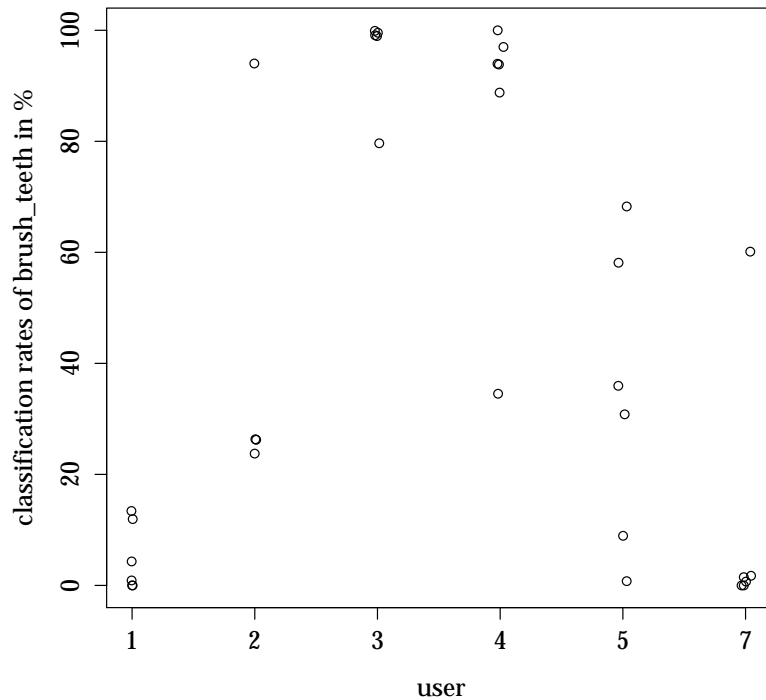


Figure 26: Distribution of classification rates of *brush_teeth* over trials for individual users. The points are randomly jittered on the x-axis for better visualization of overlapping points.

oriented in a way that the discretization of the *brush_movement* is set to *yes_sink* instead of *yes_face* which leads to a misclassification of *brush_teeth* as *paste_on_brush*. The user shown in figure 25 (b) shows a similar body pose during brushing. Furthermore, the user tends to turn the head to the right which changes the orientation of the brush even further towards the sink and reinforces the misclassification. The trials from which the images were taken, have a very poor classification rate of *brush_teeth* with 9% and 1%, respectively. However, the classification rates for some users are excellent as depicted in figure 26 which shows the distribution of classification rates of *brush_teeth* over trials of individual users. A point in the diagram denotes the classification rate in a single trial. The points are randomly jittered on the x-axis for better visualization of overlapping points.

For user 3, the rates are around 100% with a single outlier at 80%. For trials of user 1, the classification rates are below 20%. Hence, the results of specific users decrease the overall recognition rates of *brush_teeth* to 50.4%.

The recognition rate of *brush_teeth* influences the recognition rates of *rinse_mouth_clean* and *rinse_mug_fill* which are 41% and 54.9%, respectively. For the recognition of both behaviors, we use a heuristic

as described in section 5.3.1: when *rinse_mouth* (*rinse_mug*) is classified by the recognition component, it will be set to *rinse_mouth_clean* (*rinse_mug_clean*) if *brush_teeth* was already performed during the trial. Otherwise, *rinse_mouth* will be set to *rinse_mouth_wet* (*rinse_mug_fill*).

Rinse_mouth_clean was misclassified as *rinse_mouth_wet* with 31.6%. The misclassification concentrates on trials in which the recognition rate of *brush_teeth* is poor: *rinse_mouth_clean*, which is performed after *brush_teeth*, is classified as *rinse_mouth_wet* because *brush_teeth* was not recognized properly.

According to the problems with *rinse_mouth*, we expected a similar problem in the classification of *rinse_mug* in a way that the behavior after *brush_teeth* (*rinse_mug_clean*) is misclassified as the behavior prior to *brush_teeth* (*rinse_mug_fill*). However, we observed the contrary behavior: *rinse_mug_fill* is misclassified as *rinse_mug_clean* with 24.2%.

A closer look at the trial data revealed the following explanation: the misclassification mainly concentrates on trials in which *brush_teeth* was properly recognized. Users tended to wet their mouth prior to *brush_teeth* until no water was left in the mug. When they aimed to perform *rinse_mouth_clean* after a successful execution of *brush_teeth*, they started to fill the mug with water again. Hence, a regular *rinse_mug_fill* behavior was misclassified as *rinse_mug_clean* since the heuristic doesn't model these situations.

We justify the explanations of the misclassifications described above by evaluating eleven trials in which the classification rate of *brush_teeth* is above 75% or better. We manually chose the threshold since a classification rate of 75% is a good result to a minimum degree with regard to the challenging data. Here, the average rate of *brush_teeth* is 94.2% which is excellent. Table 19 shows the confusion matrix. The misclassification of *rinse_mouth_clean* as *rinse_mouth_wet* has dropped to 0.0%. This underlines the explanation that the misclassification focuses on trials where the recognition rate of *brush_teeth* is poor. Accordingly, the misclassification of *rinse_mug_fill* as *rinse_mug_clean* has increased to 32.7% since regular executions of *rinse_mug_fill* after *brush_teeth* were erroneously classified as *rinse_mug_clean*.

The good average rate of 78.6% shows that the recognition component used in the TEBRA system is able to deal with huge variances in spatial task execution when two aspects can be improved: firstly, the recognition rates of behaviors are highly dependent on the rate of *brush_teeth*. Hence, the improvement of recognizing *brush_teeth* is very important for a successful user behavior recognition in the overall task. Secondly, we need to improve the heuristic which discriminates between *rinse_mouth_wet* (*rinse_mug_fill*) and *rinse_mouth_clean* (*rinse_mug_clean*) in order to avoid misclassifications due to modeling errors. We will discuss possible improvements of these aspects in the outlook chapter.

Table 19: Classification rates of user behaviors in % for eleven trials in the SYS scenario that have a classification rate of *brush_teeth* above 75%. RMgC - *rinse_mug_clean*, RMgF - *rinse_mug_fill*, UT - *use_towel*, PB - *paste_on_brush*, RMC - *rinse_mouth_clean*, RMW - *rinse_mouth_wet*, BT - *brush_teeth*, CB - *clean_brush*, N - *nothing*.

	RMW	RMC	RMgF	RMgC	BT	PB	CB	UT	N
RMW	86.3	0.0	0.4	0.0	0.0	0.1	0.0	0.0	13.2
RMC	0.0	69.7	0.0	1.8	0.0	1.4	0.0	0.0	27.1
RMgF	0.6	5.0	51.5	32.7	0.1	0.0	0.0	0.0	10.1
RMgC	0.0	0.0	0.0	84.9	0.0	0.0	11.4	0.0	3.7
BT	0.1	0.0	0.0	0.0	94.2	3.6	0.7	0.0	1.4
PB	0.0	0.0	0.0	0.0	3.4	95.8	0.0	0.0	0.8
CB	0.0	0.9	0.0	0.2	0.0	0.5	95.1	0.0	3.3
UT	0.0	0.2	0.0	5.8	0.0	1.0	3.6	80.3	9.1
N	3.7	5.2	0.8	2.3	10.1	17.9	4.1	6.2	49.7

Additionally to the spatial variance, temporal variance is expressed in both inter-behavior and intra-behavior timing differences: inter-behavior differences are variations in the duration of behaviors amongst each other. Table 20 gives an overview of average durations of behaviors for all SYS trials. The average duration of individual behaviors in the brushing task

Table 20: Minimum, maximum and average duration of user behaviors.

User behavior	Durations in sec.		
	avg	min	max
<i>paste_on_brush</i>	9.8	2.8	28.4
<i>rinse_mug_fill</i>	2.5	0.5	9.5
<i>rinse_mug_clean</i>	3.2	0.8	7.9
<i>rinse_mouth_wet</i>	2.5	0.9	9.5
<i>rinse_mouth_clean</i>	2.4	0.6	8.6
<i>brush_teeth</i>	67.9	19.0	143.0
<i>clean_brush</i>	5.2	0.7	16.1
<i>use_towel</i>	12.0	1.8	73.3

ranges from 2.4s for *rinse_mouth_clean* to 67.9s for *brush_teeth*. As shown with the classification rates of table 18, the recognition component is able to deal with behaviors varying extremely in duration: for example, the average durations of *paste_on_brush* and *rinse_mouth_wet* are 9.8s and 2.5s, respectively. The classification rate for *rinse_mouth_wet* is very good with

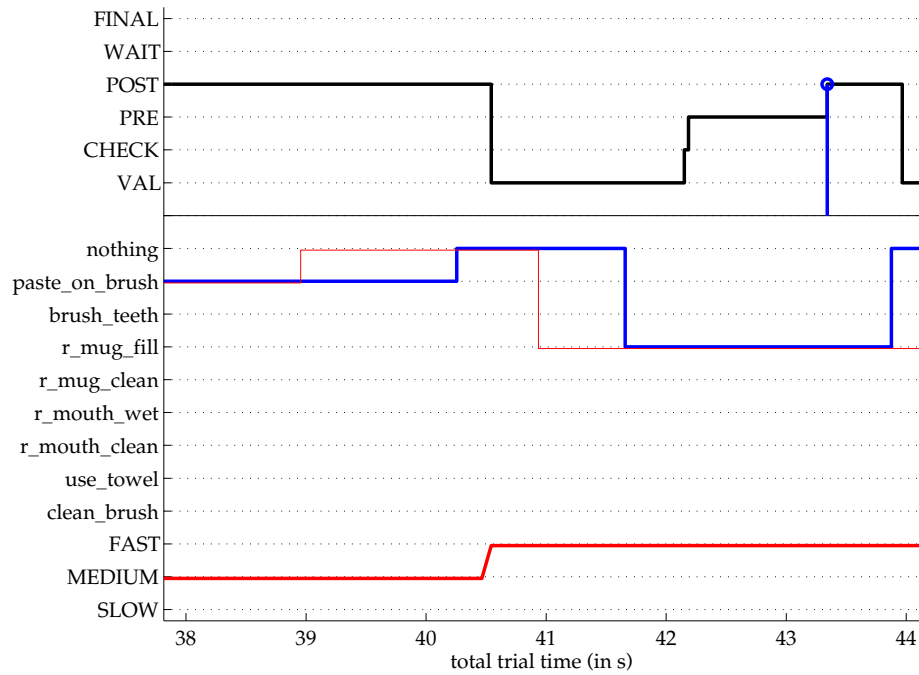


Figure 27: Section of a trial of user 5 showing the state of the Finite State Machine (black line), the estimate of the user’s behavior according to the recognition component (blue line), the estimate of the user’s velocity (thick red line), and the ground truth annotation of behaviors (thin red line). The vertical blue line denotes the update of the state space by applying the effects of the current user behavior which is *rinse_mug_fill* here.

82.5% and excellent for *paste_on_brush* with 97.8%.

The durations vary not only between different behaviors, but also in different executions of a single behavior which is called intra-behavior difference in the following. Intra-behavior difference arises from different velocities in task execution due to a user’s individual abilities. For example, the durations of single executions of *paste_on_brush* range from 2.8s to 28.4s. We apply a dynamic timing model to deal with intra-behavior variations and different velocities of users. The dynamic timing model changes the timing parameters of the Finite State Machine described in section 5.3.2. The validation, effect and timeout parameters for individual user behaviors are set based on a user’s execution velocity during the trial so far. We maintain a timing model for *fast*, *medium* and *slow* execution velocity. During a trial, we determine the velocity of the current user behavior and count the frequencies of behaviors which fall into the velocity classes *fast*, *medium* and *slow*. We apply a winner-takes-all method and apply the parameters of the timing model which has the highest count of behaviors during the trial so far. We will exemplarily describe the benefit of the dynamic timing model in three situations.

Figure 27 visualizes the state of the FSM (black line), the estimate of the user’s behavior according to the recognition component (blue line),

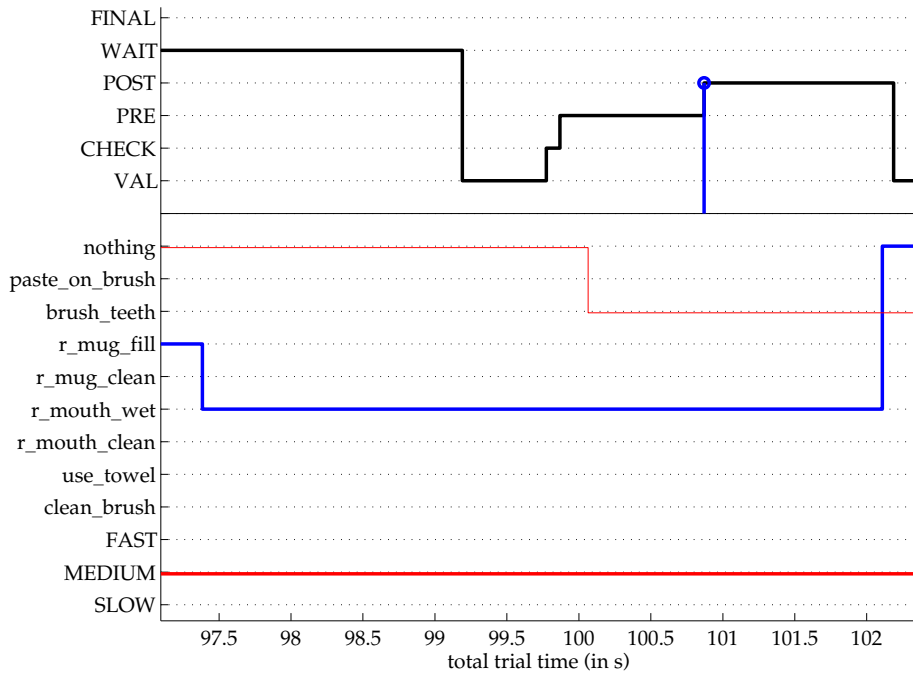


Figure 28: Section of a trial of user 2. For a description of the lines, see figure 27.

the estimate of the user's velocity (thick red line), and the ground truth annotation of behaviors (thin red line). The visualization covers an interval of about six seconds in a trial of user 5. User 5 finishes *paste_on_brush* at about 40.3s. Due to the duration of *paste_on_brush* and the velocities of the preceding behaviors, the velocity model is updated from *medium* to *fast* at 40.5s. At 41.8s, the user starts *rinse_mug_fill* which is performed for 2.2s. Due to velocity model *fast*, the effects of the behavior occur after 1.6s which is depicted by the vertical blue line at 43.4s. With the model for *medium* velocity, *rinse_mug_fill* would not have been recognized correctly since the effect time of 3.3s would not have been reached. Hence, the effects of *rinse_mug_fill* would not have been applied to the progress state space leading to erroneous prompts in the remainder of task execution.

A second situation is depicted in figure 28 showing a section of a trial of user 2. The velocity model of the user is *medium*. *Nothing* and *brush_teeth* are erroneously classified as *rinse_mouth_wet* for 4.7s. Using the velocity model *fast*, which is the initial model of a trial, a timeout would have been reached after 4.4s according to table 12. Hence, an erroneous prompt would have been issued based on the perception error. By adapting the velocity model to *medium* during the course of the trial, the dynamic timing model avoided the delivery of an erroneous prompt.

Figure 29 shows a third situation from a trial of user 2. The user performs *rinse_mug_fill* which is successfully recognized by the TEBRA system. The progress state space is updated with the effects of *rinse_mug_fill* after about 25.6s which is depicted by the vertical blue line. The user forgets to perform *rinse_mouth_wet* and *paste_on_brush*, and erroneously starts *brush_teeth* at 32s. Due to the inconsistency of *brush_teeth*, a pictogram

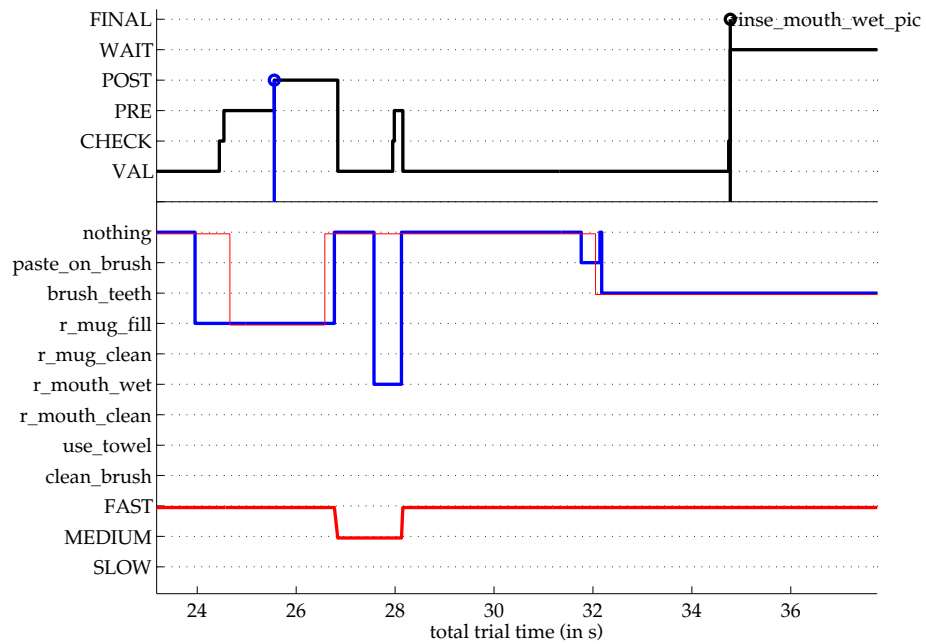


Figure 29: Section of a trial of user 2. For a description of the lines, see figure 27. The vertical black line denotes that a prompt was triggered by the TEBRA system at that time. Here, a pictogram prompt for behavior *rinse_mouth_wet* was delivered.

prompt for *rinse_mug_fill* is delivered at about 35s which is shown by the vertical black line. The dynamic timing model with velocity *fast* delivers a prompt which is appropriate in time in a way that the user is assisted in the erroneous performance of the task as soon as possible. With a *medium* or *slow* velocity, the prompt would have been delayed and the user would have performed the erroneous behavior for a longer period of time.

A disadvantage of the dynamic timing model is the inclusion of durations of erroneously classified behaviors in determining a user's velocity in a trial. For example, *brush_teeth* is misclassified as *paste_on_brush* for a duration of 3s. The duration of 3s is classified into velocity *fast*. Hence, the dynamic timing model erroneously increases the frequency counter of velocity model *fast* which leads to a skewed distribution of counts over the velocity classes. This might result in a wrong application of timing parameters and the delivery of false prompts in the remainder of the trial. However, as shown in the previous examples, the TEBRA system can deal with intra-behavior variances in temporal execution of behaviors by adapting to the user's velocity during task execution.

In the following subsection, we will analyze the prompting behavior of the system and the user's reaction behavior in detail.

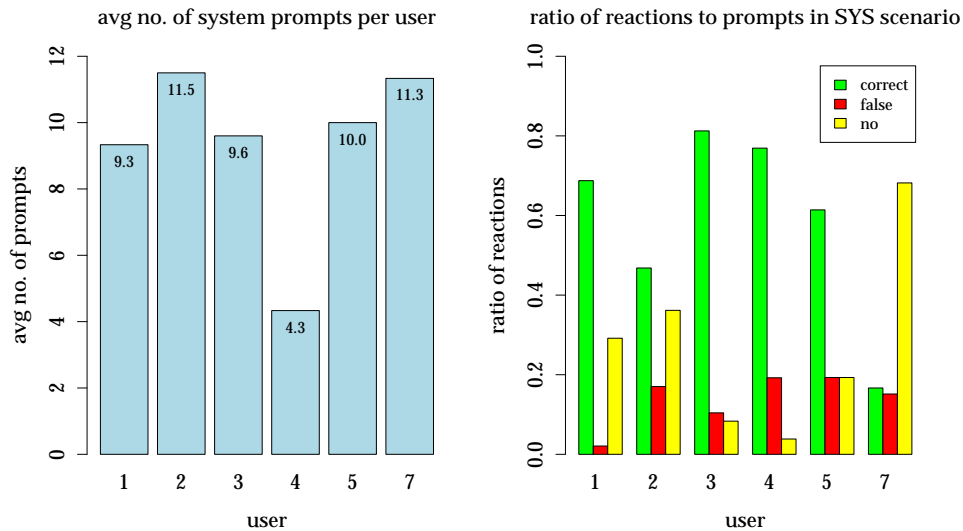


Figure 30: Left plot: average number of prompts in the SYS scenario for individual users; Right plot: ratio of reactions to prompts in the SYS scenario for individual users.

6.2.2.2 Prompting behavior and users' reactions

A user's individual capabilities in performing the task not only become apparent in the huge variance in task execution, but also in the amount of assistance needed by a user. The left plot in figure 30 shows the average number of prompts given in the SYS scenario for individual users. User 4 received a very low average number of prompts per trial with 4.3 compared to the other users. However, user 4 showed six perfect trials with the assistance of the TEBRA system as depicted in figure 22 in the last subsection. For the other users, the values are much higher, but on a similar level: The average values range from 9.3 prompts for user 1 to 11.5 for user 2. However, although the number of prompts is similar for users, the overall performance in the brushing task is different. For example, user 2 and user 7 were prompted 11.5 and 11.3 times on average, respectively. The average number of independent steps according to table 16 differs for user 2 with 4.8 and user 7 with 6.3. Obviously, individual users are able to deal with prompts differently which mainly depends on two factors: firstly, the appropriateness of prompts in timing and content, and, secondly, the responsiveness of individual users to the given prompts. In the following, we will evaluate these factors in more detail.

An important measure for a user's responsiveness to system prompts is the reaction behavior. We classify the reactions of users into three categories: *correct*, *false* and *no* reaction. A user's reaction to a prompt is correct when the user adapts his/her behavior according to the prompt by performing the behavior he/she was prompted for. If the user reacts to the prompt, but does not perform the desired behavior, the reaction will be classified as a false reaction according to the prompt. If the user does

not show a reaction at all, we will refer to it as no reaction.

The right plot in figure 30 shows the ratio of the three types of reactions to prompts in the SYS scenario. The ratio of false reactions to prompts is low for all users ranging from 1% for user 1 to 20% for users 4 and 5. Users 1 to 5 react correctly to the majority of prompts: users 3 and 4 show very good ratios of correct reactions to prompts with around 80%, each. We conclude from a correct reaction that the prompt is appropriate in both timing and content and that the user is capable of understanding and responding to the prompt. User 7 shows only 17% of correct reactions and a high ratio of no reactions with about 70%. No reaction to a prompt stems from three possible explanations: firstly, the user is not able to understand the prompt on a semantic level due to decreased cognitive abilities. Secondly, the prompt is appropriate, but the user is not willing to react to the prompt. Thirdly, the user doesn't react to the prompt because it is not appropriate for the user in timing or content.

In order to further evaluate the appropriateness of prompts, we take into account the number of semantically correct prompts as a measure of appropriateness. Semantically correct means that the type of prompt is appropriate with regard to the user's progress in the task so far. For example, a user has successfully filled the mug with water and gets stuck in task execution. An appropriate prompt in this situation would be either *rinse_mouth_wet* or *paste_on_brush*. We determine the semantic correctness by using a ground truth annotation of the behaviors in the task which was done by the author of the thesis and a student assistant.

*Semantically correct
prompt*

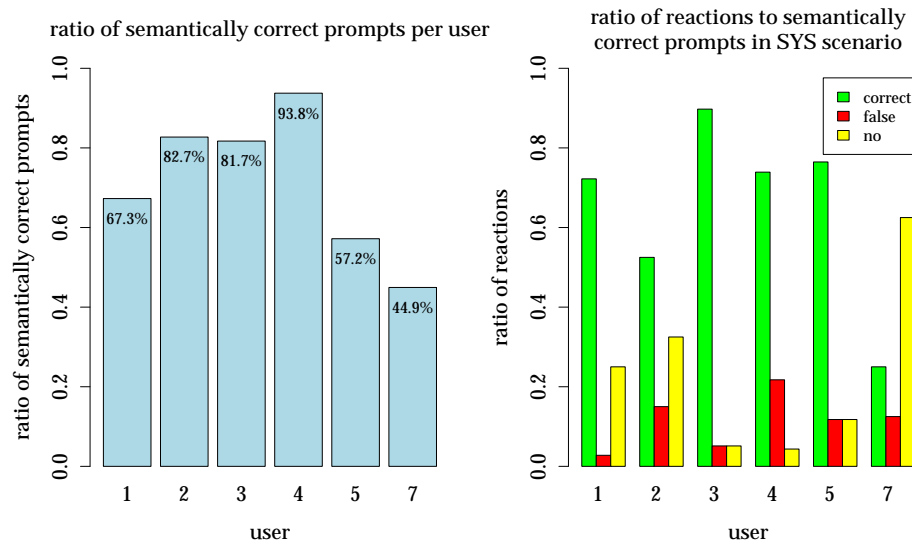


Figure 31: Left plot: ratio of semantically correct prompts in the SYS scenario for individual users; Right plot: ratio of reactions to semantically correct prompts in the SYS scenario for individual users.

The left plot in figure 31 shows the ratio of semantically correct prompts in the SYS scenario for individual users. The ratio of user 4 is excellent since 93.8% of the prompts are semantically correct. For users 2 and 3, the ratios of semantically correct prompts are good with 82.7% and 81.7%, respectively. However, the percentage for users 5 and 7 are decreased with 57.2% for user 5 and 44.9% for user 7. The low ratios of semantically correct prompts stem from erroneous follow-up prompts due to perception errors in the recognition component: for example, a user performs *rinse_mug_fill*, but the TEBRA system misses to recognize the behavior. The user performs *rinse_mouth_wet* subsequently which is a correct behavior according to the course of the trial. However, the system prompts the user to perform *rinse_mug_fill* which is semantically incorrect at that time. If the user does not react to the prompt, the system is likely to issue follow-up prompts for *rinse_mug_fill* which are semantically incorrect, too. The TEBRA system is able to limit the number of erroneous follow-up prompts using the following heuristic: after three consecutive prompts of the same behavior (one pictogram and two video prompts according to the escalation hierarchy), the system infers that it has made a perception error and applies the effects of the prompted behavior to the state space. Due to the heuristic, the system is able to recover from perception errors in which a user's behavior was missed during the execution of a trial.

In order to assess a user's responsiveness to prompts, we focus on reactions to semantically correct prompts because the appropriateness of semantically correct prompts is ensured. The right plot in figure 31 shows a user's reactions to semantically correct prompts. We expected a significant increase in the ratio of correct reactions to semantically correct prompts. However, the user's reactions to semantically correct prompts differ only slightly compared to the reactions to all prompts: the correct reactions increased slightly for any user except for users 3 and 4 who show a decrease from 90% to 82% for user 3, and from 78% to 75% for user 4. User 7 who showed no reaction to 70% of all prompts, has only a slightly decreased ratio of no reactions to semantically correct prompts with 60%. The correct reactions increased only marginally.

User 2 showed a similar characteristic: the ratio of correct reactions to semantically correct prompts is slightly increased from 45% to 51%. Two explanations are possible for the reaction behaviors of users 2 and 7: firstly, they might not be willing to react to the prompts given by the TEBRA system although the prompts are semantically correct. Secondly, they might not be able to understand and react correctly to the majority of system prompts since the presentation of prompts is inappropriate.

In the TEBRA system, we use pictogram and real-life videos to prompt the users. We analyze whether pictogram or video prompts are inappropriate for an individual user: figure 32 shows the ratio of correct reactions to semantically correct prompts for pictogram and video prompts. During

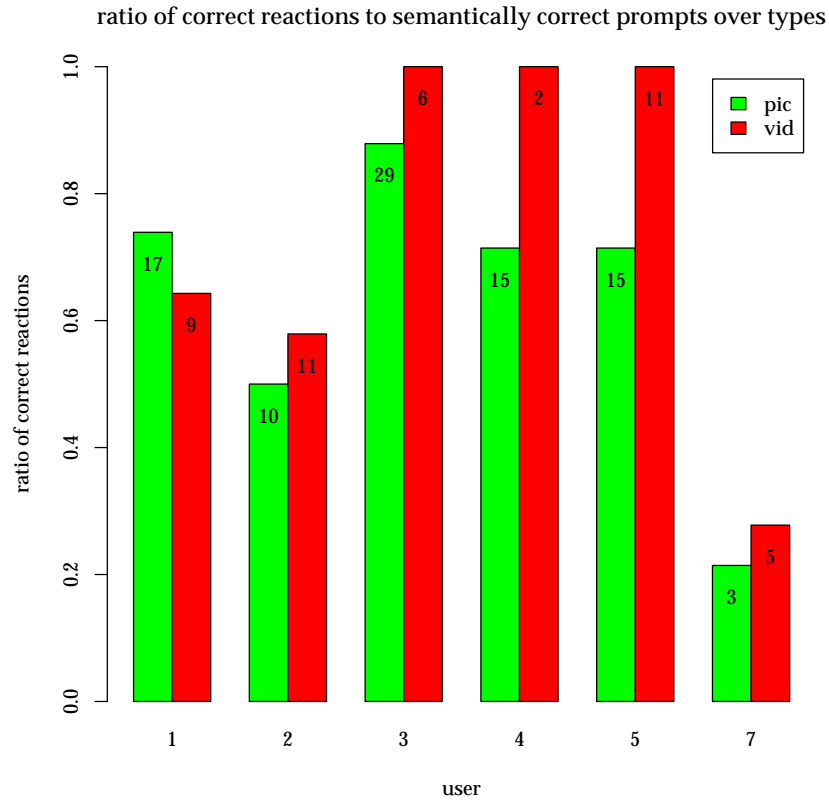


Figure 32: Ratio of correct reactions to semantically correct pictogram and video prompts for individual users. pic - pictogram prompts, vid - video prompts.

the analysis of trials with persons with cognitive disabilities, we observed that the TEBRA system provides prompts which are consistent with regard to a user's overall progress, but which are not necessary for the user because they were triggered due to perception errors for behaviors with a long duration: for example, a user initiated *brush_teeth* which is consistent at that time. The TEBRA system misclassifies *brush_teeth* as *paste_on_brush* prior to the effect time of 60s for *brush_teeth*. Since *paste_on_brush* is inconsistent, a *brush_teeth* prompt is triggered which is consistent with regard to the user's overall progress in the task. Although the prompt occurred due to a perception error and the prompt might not have been necessary for the user since he/she already performs *brush_teeth*, it is semantically correct with regard to the progress state space: the effect time of the behavior has not been reached and the progress state space has not been updated with the effects of the behavior, yet. We refer to such prompts as *random* semantically correct prompts. Such prompts are in contrast to *adequate* semantically correct prompts: adequate prompts help the user to initiate a next step when the user gets stuck in the task or interrupt an erroneous performance of the user during the task. Figure 32 contains both adequate and random semantically correct prompts. The figure contains a

total number of 89 pictogram and 44 video prompts.

User 3 shows a ratio of 89% correct reactions to pictogram prompts and 100% correct reactions to video prompts. Both kinds of prompts seem to be appropriate for user 3 with regard to the level of information provided in the prompts. Users 4 and 5 also show 100% correct reactions to video prompts, but only about 70% correct reactions to pictogram prompts. Pictogram prompts seem to be appropriate for users 4 and 5 in most situations. However, users 4 and 5 reacted incorrectly or not at all in 30% of the pictogram prompts. Video prompts seem to be more appropriate in such situations. Both user 4 and 5 reacted correctly in 100% of the cases.

User 2 shows a different reaction behavior: the ratio of correct reactions to pictogram prompts is 50%. For video prompts, the ratio is increased with 60% correct reactions. Video prompts seem to be more appropriate compared to pictogram prompts. We found three possible explanations: firstly, video prompts are better suited to grab the attention of users than pictogram prompts because the movement in the videos is more salient than the static pictogram prompts. Some users might miss the static pictogram prompts. Secondly, users might be able to react to a video prompt due to priming effects: a user might already be primed by a pictogram prompt of the same behavior which timely precedes a video prompt in any case. Thirdly, a video prompt provides a higher level of information about the behavior. Hence, video prompts might be more suited to a user's cognitive abilities. We are not able to reveal the reasons from the results of the study. We might investigate the reasons in more detail in future studies.

The reaction behavior of user 7 is poor to both pictogram and video prompts with 20% and 25%, respectively. We found two possible explanations for the user's behavior: firstly, user 7 might not be able to react to prompts at all: both pictogram and video prompts seem to be inappropriate for user 7. Secondly, the user might not be willing to follow the prompts given by the TEBRA system.

According to the caregivers, user 7 sticks to a strict routine in tooth brushing in which the user usually doesn't like distractions. This might indicate that user 7 is not willing to react to prompts. However, the exact reasons for the behavior of the user remain unclear.

The results show that the responsiveness to system prompts varies amongst individual users: some users react correctly to pictogram prompts, but other users need video prompts for proper assistance. The TEBRA system is able to deal with differences in the responsiveness of users by providing an escalation hierarchy which presents prompts with increasing level of information until the prompts provide appropriate assistance to a user.

6.2.2.3 Usability aspects

The application of an Assistive Technology for Cognition (ATC) highly depends on the usability of such a system. Usability in the context of ATC refers to the ease of use with regard to the overall goal of proper task as-

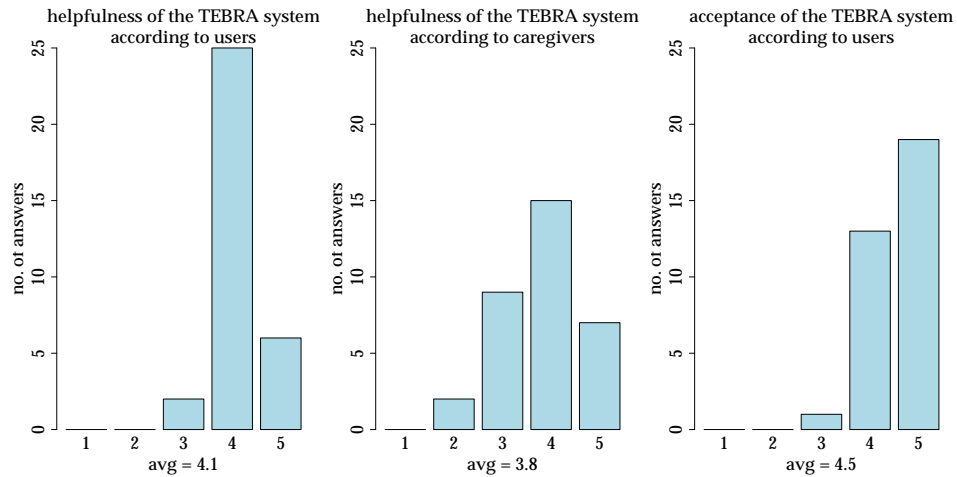


Figure 33: Results of the questionnaire: helpfulness of the TEBRA system according to users (left plot) and caregivers (middle), acceptance of the TEBRA system according to users (right). Answers on a 5-point Likert scale with 1 being *not at all* and 5 denotes *very good*.

sistance. The users' opinions are important in order to judge the usability of the TEBRA system. After each SYS trial, we asked the user whether the system was helpful in task execution as part of a questionnaire shown in appendix B.3. The question was asked by a caregiver who rated the answer on a 5-point Likert scale with 1 being no assistance at all and 5 denoting very good assistance.

The average value of the TEBRA system's helpfulness is 4.1. The left plot in figure 33 shows the distribution of answers on the 5-point Likert scale. Hence, the TEBRA system is helpful in task execution from a user's subjective point of view despite a number of semantically incorrect prompts due to perception errors. We distinguish between two types of errors which lead to an erroneous system behavior: false-positives and false-negatives. False-positive errors (also called false alarms) happen when the system delivers a prompt, but the prompt is not necessary at that time. False-negative errors occur in situations where the system misses a prompt although a prompt would have been appropriate.

Most of the erroneous prompts given by the system were prompts due to false-positive errors. We conclude that users accept false-positive errors when the system assists them properly throughout the remainder of the task by avoiding false-negative prompts. A trivial policy of avoiding false-negative prompts is providing prompts throughout the whole execution of the task. However, such a prompting behavior is not acceptable since the aim of an ATC system is increasing the independence of users by prompting when necessary. Hence, an appropriate prompting behavior requires a trade-off between minimizing false-negative prompts by providing steady prompting and increasing the independence of users by prompting when necessary. The results of the study and the questionnaire show that the TEBRA system implements an appropriate prompting behavior with re-

gard to this trade-off.

Additionally to the users, we asked the caregivers to judge whether the system helped the user in the brushing task. The distribution of answers is shown in the middle plot of figure 33. The average value of 3.8 is lower compared to the user's opinion with 4.1. However, 3.8 is a good result which shows that the assistance of the TEBRA system is appropriate from an expert's point of view.

A further aspect of usability is the user's acceptance of the system. We asked the users how much they liked to use the system as part of their daily routine. The right plot in figure 33 depicts the distribution of answers. An average value of 4.5 over all users underlines the good acceptance of the TEBRA system. Most of the users showed reactions such as smiling and laughing when they perceived system prompts. Furthermore, we observed that some of the users experienced the system as a kind of interaction partner: they talked to the system when a prompt was given or they reacted verbally to prompts by saying 'ok' or 'I will'. In two trials, we observed that the users were waiting with the execution of behaviors until the system prompted them what to do (due to a timeout). For these users, the interaction with the TEBRA system was a game-like situation where the users provoked a reaction of the system as an interaction partner. We encountered similar behavior in the CG trials where users tended to talk to the caregiver standing besides them. Users who talked frequently to the caregivers, were distracted more often and didn't focus on the proper execution of the task. According to the caregiver's comments, distraction due to verbal communication with the caregiver is one of the main sources for insufficient task execution. Since the TEBRA system is not able to respond to a user, the distraction due to verbal communication is minimized when using the TEBRA system. However, the communication between the caregiver and the user is an important social interaction for the user. Understanding the lack of such social interactions due to system use is an important issue in research of ATC systems, but is not taken into consideration in this thesis.

We like to finish the analysis of the study results by describing a special situation in the beginning of a SYS trial: a user was unsure whether to participate in the trial. The user didn't want to leave the room, but was not motivated to start the brushing task either. The caregiver asked the user to start the task several times, but the user didn't react to the requests. At a hint from the caregiver, we activated the TEBRA system which delivered a timeout prompt after a period of time. After the second timeout prompt, the user started with the task execution motivated by the prompts of the TEBRA system. According to the caregiver, the user wouldn't have started the task due to requests of a caregiver. We infer that the consistent prompts of the system offer a familiar and predictable environment which might be very important for persons with specific cognitive disabilities in the execution of tasks. Furthermore, three of six users performed much longer brushing phases in the overall task when using the TEBRA system. Ac-

According to the caregivers, these users have never performed tooth brushing as accurately and sustainedly with assistance of a human caregiver. Hence, users might benefit from the TEBRA system since they feel comfortable due to the highly predictable behavior of the system which leads to an increase in the quality of task execution.

6.2.2.4 Summary of results

We evaluated the TEBRA system in a study with persons with cognitive disabilities with regard to the following hypothesis: the TEBRA system is able to increase the independence of users by providing appropriate prompts which enable the user to perform the brushing task more independently. We were able to confirm the hypothesis since the number of independent steps is significantly increased from the CG to the SYS scenario. A closer look at the technical performance of the system revealed that the user behavior recognition can deal with spatial and temporal variance in task execution for some behaviors such as *paste_on_brush* and *rinse_mouth_wet*. However, the recognition of other behaviors such as *rinse_mug_fill* and *brush_teeth* need to be improved by more sophisticated detection of the states of objects involved in the different behaviors. Furthermore, the heuristic distinguishing between *rinse_mug_fill* (*rinse_mouth_wet*) and *rinse_mug_clean* (*rinse_mouth_clean*) needs to be improved to make the TEBRA system's behavior more robust.

We showed that the TEBRA system is able to deal with temporal variance in task execution. The system robustly recognizes behaviors with different durations and tracks a user's progress in the task appropriately: a dynamic timing model takes into account the different velocities of users in task execution.

We investigated the interaction behavior of users with the TEBRA system: the amount of assistance needed by individual users, which is expressed in the number of prompts given by the system, differs due to cognitive abilities. We distinguished between semantically correct and incorrect prompts. In order to evaluate a user's responsiveness to prompts, we focused on semantically correct prompts which are appropriate with regard to a user's progress in the task. We observed differences in the responsiveness to prompts amongst individual users: some users reacted correctly to pictogram prompts, but others needed video prompts which provide more detailed information of the desired behavior. However, all users were able to significantly increase the number of independent steps from the CG to the SYS trials. The TEBRA system is able to deal with variations in the responsiveness of users by implementing an escalation hierarchy of prompts which provides prompts with different levels of information.

Besides the technical evaluation, we asked the users and caregivers to assess the performance of the TEBRA system using a questionnaire. Both users and caregivers find the system helpful in task execution. Furthermore, users like to be assisted by the system which underlines the acceptance of the TEBRA system from a user's point of view.

CONCLUSION AND OUTLOOK

This thesis has described the design, implementation and evaluation of the TEBRA (TEeth BRushing Assistance) system. TEBRA is a novel Assistive Technology for Cognition (ATC) for persons with moderate cognitive disabilities. The TEBRA system provides assistance in the execution of *brushing teeth* by providing audio-visual prompts to users who are reliant on assistance in brushing teeth by a caregiver. Brushing teeth is an important basic Activity of Daily Living (ADL) since disregarding oral hygiene might result in severe medical problems. Brushing teeth is a complex and flexible task in a way that multiple steps and objects are involved which can be coordinated in different ways for a successful task execution.

State-of-the-art ATC systems such as the COACH system [39] assist in less complex and less flexible tasks such as washing hands and focus on a particular user group, for example persons with dementia. In this thesis, the target group of users is heterogeneous and includes persons with different cognitive disabilities such as behavioral disorder, intellectual disabilities and Autistic Spectrum Disorder (ASD).

In the design of the TEBRA system, we took into account the characteristics of the task and the involved users by applying Interaction Unit (IU) analysis which is a task analysis technique based on in-situ observations of the task. We subdivided the task into eight behaviors and identified preconditions, effects and environmental configurations of objects involved in the behaviors. We included the results of IU analysis in the user-centered design process of the TEBRA system. We evaluated and refined system components early in the design and implementation phase by conducting a Wizard of Oz (WOZ) study where we evaluated a user's reaction behaviors to system prompts. Furthermore, we conducted a questionnaire study with caregivers about appropriate prompting of the TEBRA system.

A key requirement for high usability and acceptance of the TEBRA system is context awareness. In a context-aware system, no explicit feedback from the user about the completion of sub steps is necessary. State-of-the-art systems such as COACH, Archipel, PEAT or Autominder [39, 7, 52, 87] have only limited or even no capabilities of context awareness. The TEBRA system provides context awareness by implementing a sophisticated behavior recognition component which is combined with a planning and decision making component to provide appropriate assistance. The main challenge in providing context-aware behavior is the

robust recognition of the users' behaviors. This is difficult due to the huge spatial and temporal variance in task execution which is characteristic for persons with various cognitive disabilities. Spatial variance refers to different movement characteristics of users in the execution of a functional unit in a task. Temporal variance denotes different durations of behaviors and different velocities of users in the task. In the TEBRA system, we deal with spatial variance by inferring behaviors based on states of objects involved in the behaviors instead of tracking a user's hands or objects. In order to recognize the user's behaviors, we use a Bayesian network in a Bayesian filtering approach. We also deal with temporal variance in terms of different durations of behaviors: a dynamic timing model allows for different velocities of users in task execution by modeling the velocities of users explicitly and adapting to the user's velocity during a task.

The main aim of the TEBRA system is to increase the independence of users from a human caregiver in the execution of brushing teeth. In order to evaluate its utility in this regards, we have conducted a study with seven persons of the target group being assisted by a fully functioning prototype of the TEBRA system. The study data comprises 20 trials with a caregiver's assistance and 35 trials with the TEBRA system's assistance which is a huge interaction corpus in the field of ATC. The results of the study showed that the TEBRA system is able to increase the independence of users in the tooth brushing task: *All of the users were able to perform significantly more steps of the task independently, when they had been assisted by the TEBRA system instead of a human caregiver.* The benefit of the system differs amongst users: one user showed only a slight increase of independent steps while another user was able to perform the brushing task completely independent in all trials with the system. The results of the study demonstrate the potential of the TEBRA system in assisting persons with cognitive disabilities in task execution.

However, the findings are restricted to the technical evaluation of the system and the user's reaction behavior to system prompts. We didn't aim to provide any diagnostic assessment or therapeutic treatment of users. Due to the limited number of participants in the study, we didn't focus on relations between a user's specific cognitive disabilities and his/her performance in the task. This might be subject of further research with the TEBRA system.

Research in the field of ATC has the overarching goal to develop systems which can be deployed permanently in the real life of users and provide user-tailored assistance whenever it is necessary. Ideally, such systems provide the following properties: an ATC system (1) is robust with regard to a user's individual performance of the task, (2) provides pervasive assistance in a way that the system is able to assist in multiple tasks, (3) adapts to a user's capabilities in task performance on runtime, (4) automatically initiates and terminates the assistance avoiding explicit

activation and shutdown of the system by a caregiver, and (5) integrates unobtrusive sensor technology.

In order to improve the TEBRA system towards an ideal ATC system, different perspectives for future development of the system might be taken into account:

ROBUSTNESS The usage of additional sensor data by applying further sensors allows for a more fine-grained representation of behaviors in terms of *behavior* variables. This might lead to an increase in robustness of the user behavior recognition with regard to special characteristics in task execution which can be observed for persons with cognitive disabilities.

PLANNING In a future version of the TEBRA system, the planning capabilities might be improved in a way that the planning component is able to reliably distinguish between behaviors which are similar in terms of object usage, but semantically different in the overall progress of the task: examples include *rinse_mug_fill / rinse_mug_clean* and *rinse_mouth_wet / rinse_mouth_clean*.

ADAPTABILITY/PERSONALIZATION The results of the study at *Haus Bersaba* revealed that an ATC system needs to suit the capabilities of individual users: we might enhance the adaptable behavior of the TEBRA system in the recognition and planning components by taking into account a user's mental state and an assessment of the user's cognitive abilities. Furthermore, learning techniques might be integrated into the TEBRA system which allow the system to adapt its behavior based on previous trials of individual users. For example, the TEBRA system might optimize the prompting behavior of the system by automatically choosing the modality of prompts based on a user's reactions to prompts in previous trials.

PERVASIVE ASSISTANCE Besides the improvements of the TEBRA system in the individual task of brushing teeth, a development of the system towards a pervasive assistance system is conceivable: pervasive assistance refers to assistance in multiple tasks taking place at the washstand such as washing hands or shaving. An extension to multiple tasks raises further research problems: firstly, the TEBRA system needs to distinguish between the tasks rapidly in order to provide appropriate assistance from the very beginning of a task. Secondly, the system needs to cope with concurrent and interleaved execution on both task and behavior level in the recognition and planning components.

LONGITUDINAL STUDIES The study with the target group users being assisted by the TEBRA system described in this thesis covered a period of five weeks in which users performed trials on nine different days. Hence, the study results are restricted to rather short-term effects in individual trials of users. Long-term effects using the TEBRA system

such as an increase in task performance for individual users over several months or years still need to be investigated in longitudinal studies in which a system is deployed for a longer period of time.

In this thesis, we have developed an ATC system for tooth brushing which is only one of many important basic ADLs in a person's daily routine. Besides further research with the TEBRA system as described in the previous list, extending research activities in the general field of ATC is highly recommendable, because ATC systems are able to contribute to the personal independence of individual users and to relieve the burden of caregivers - not only in the tooth brushing task but in numerous further areas of daily life. Additionally, the use of various ATC systems might contribute to reducing the enormous economic costs for healthcare provision by providing appropriate healthcare to persons with cognitive disabilities and to an increasing number of elderly. Hence, further research in the field of ATC is essential in order to deal with the future effects of the demographic shift to an aging population.



INTERVIEW MANUAL FOR THE STUDY WITH CAREGIVERS

Manual used in the interviews with the caregivers of *Haus Bersaba* about appropriate modalities of prompting. The results of the interviews are described in section 4.4.

Interviewleitfaden Promptings

Einstieg

Vielen Dank, dass Sie sich etwas Zeit für uns nehmen. Wie Sie wissen, entwickelt Christian Peters ein System, um das tägliche Zähneputzen geistig behinderter Menschen zu unterstützen und er hat die Möglichkeit, dieses System mit Ihrer Unterstützung zu entwickeln und bei Ihnen auch einzusetzen. Ein Teil dieses Systems besteht darin, dass Ihre Bewohner über einen Bildschirm und Lautsprecher Rückmeldungen darüber bekommen, ob Sie optimal putzen bzw. wo und wann sie etwas anders machen sollen. Zurzeit sind wir dabei diese Rückmeldungen zu entwickeln. Und in diesem Gespräch würden wir gern von Ihnen erfahren, für wie geeignet und verständlich Sie die von uns entwickelten Rückmeldungen halten.

Erster Fragenteil ohne Promptings

Doch bevor wir Ihnen unser Material zeigen, möchten wir gern von Ihnen wissen, wie Sie diese Rückmeldungen gestalten würden, was Sie für geeignet und verständlich halten.

1. Wie würden Sie dieses System und die Rückmeldungen gestalten? Was sind Ihre Ideen dazu?
2. Womit kann man die Aufmerksamkeit Ihrer Bewohner erlangen? Mit welchen Rückmeldungen, Bildern, Zeichen?
 - a. Ist es Ihrer Meinung nach hilfreich, die Bewohner bei den Rückmeldungen persönlich anzusprechen (mit Ihrem Namen)?
3. Wie sollten Ihrer Meinung nach die Anweisungen zum Zähneputzen aussehen (unterstützt mit Bildmaterial? Fotos? Videos? Oder nur Sprache?)? Sollten die Anweisungen eher kurz und knapp ausfallen oder längere Erklärungen beinhalten?
4. Sollte Lob eingesetzt werden? Wenn ja, wie kann dieses Lob aussehen (nur verbal?, mit Bildmaterial?, braucht es ein

Belohnungssystem?)? Wie setzen Sie Lob bei Tätigkeiten wie dem Zähneputzen ein?

Fragen zu den Promptings

A: Aufmerksamkeit erzeugen

Um den Bewohnern Hilfestellungen zum Zähneputzen geben zu können, muss die Aufmerksamkeit der Bewohner erlangt werden und das Zähneputzen muss kurz unterbrochen werden. Hierzu haben wir folgendes Material ausgewählt:

1. Welche dieser Rückmeldungen halten Sie für geeignet, die Aufmerksamkeit der meisten Bewohner zu erlangen?
2. Welche dieser Rückmeldungen würden Ihrer Meinung die meisten Bewohner eher ablenken oder verwirren?
3. Welche dieser Rückmeldungen können Ihrer Meinung nach von den meisten Bewohnern klar verstanden werden als Signal, das Zähneputzen zu unterbrechen?
4. Welche dieser Rückmeldungen können Ihrer Meinung nach von den meisten Bewohnern nicht verstanden werden?
5. Welche dieser Rückmeldungen würden bei den meisten Bewohnern positiv ankommen?
6. Welche dieser Rückmeldungen würden bei den meisten Bewohnern negativ ankommen?
7. Bei welchen dieser Rückmeldungen halten Sie es für wahrscheinlich, dass die Bewohner der Aufforderung das Zähneputzen zu unterbrechen folgen?
8. Fallen Ihnen weitere Bilder oder Rückmeldungen ein, die verwendet werden können, um das Zähneputzen zu unterbrechen?

B: Anweisungen zur Verbesserung des Zähneputzens

Hierbei sollen Rückmeldungen darüber erfolgen, was beim Zähneputzen verbessert oder nachgeholt werden soll (z.B. „bitte noch einmal unten putzen“ oder „bitte den Mund ausspülen“ etc.). Hierfür haben wir neben sprachlichen Hilfestellungen ebenfalls wieder Bildmaterial ausgewählt.

1. Welche dieser Rückmeldungen würden Ihrer Meinung nach die meisten Bewohner eher ablenken oder aus dem Konzept bringen?
2. Welche dieser Rückmeldungen können Ihrer Meinung nach von den meisten Bewohnern klar verstanden werden als Hinweise, z.B. zu spülen oder oben bzw. unten zu putzen?
3. Welche dieser Rückmeldungen können Ihrer Meinung nach von den meisten Bewohnern nicht verstanden werden?
4. Welche dieser Rückmeldungen würden Ihrer Meinung nach bei den meisten Bewohnern positiv ankommen?
5. Welche dieser Rückmeldungen würden Ihrer Meinung nach bei den meisten Bewohnern negativ ankommen?
6. Bei welchen dieser Rückmeldungen ist es Ihrer Meinung nach wahrscheinlich, dass die Bewohner der Aufforderung oder Anleitung (z.B. den Mund auszuspülen) folgen?
7. Fallen Ihnen weitere Bilder oder Rückmeldungen ein, die verwendet werden können, um das Zähneputzen zu unterbrechen?

C: Lob / positive Rückmeldung

Mit diesen Bildern und Rückmeldungen sollen die Bewohner gelobt bzw. belohnt werden, wenn Sie den Verbesserungsvorschlägen gefolgt sind.

1. Welche dieser Rückmeldungen können Ihrer Meinung nach von den meisten Bewohnern klar verstanden werden als Lob bzw. als positive Rückmeldung?
2. Welche dieser Rückmeldungen können Ihrer Meinung nach von den meisten Bewohnern nicht verstanden werden?
3. Welche dieser Rückmeldungen könnten Ihrer Meinung nach die meisten Bewohner motivieren, sich weiter anzustrengen?
4. Welche dieser Rückmeldungen würden Ihrer Meinung nach bei den meisten Bewohnern positiv ankommen?
5. Welche dieser Rückmeldungen würden Ihrer Meinung nach bei den meisten Bewohnern negativ ankommen?

Abschließende allgemeine Fragen:

1. Sollten die sprachlichen Rückmeldungen eher von einer weiblichen oder einer männlichen Stimme gesprochen werden?
2. Sollten die sprachlichen Rückmeldungen von einer Stimme gesprochen, die Ihre Bewohner kennen (z.B. von einem der Betreuer)? Oder wäre auch eine unbekannte Stimme ausreichend oder gar besser?
3. Sollten die eingesetzten Videos eine bekannte Person abbilden (z.B. einen Betreuer oder, wie bisher, Christian Peters)? Oder wäre auch eine unbekannte Person ausreichend oder gar besser?
4. Können auch negative Rückmeldungen bzw. Kritik eingesetzt werden? Wenn ja, in welcher Form ist solche Kritik möglich? Wie könnte diese aussehen?
5. Sollten bei den sprachlichen Rückmeldungen eher geduzt oder gesiezt werden?

DOCUMENTS FOR THE STUDY WITH TARGET GROUP USERS

B.1 SHEET OF INFORMATION

Sheet of information for the participants of the user study at *Haus Bersaba*.



**Center of Excellence Cognitive Interaction Technology (CITEC)
Von Bodelschwingsche Stiftungen Bethel, Bielefeld**

**PD Dr. Sven Wachsmuth
Dr. Thomas Hermann
Gabriele Hüttemann-Völker
Dr. Michelina Bocchicchio**

Informationsblatt für die Teilnehmer
zu der Studie

“Eine intelligente Waschtisch-Umgebung zur alltäglichen Unterstützung des Zähneputzens”

Liebe Studienteilnehmerin, lieber Studienteilnehmer,

wir möchten Sie bitten an unserer wissenschaftlichen Studie „Eine intelligente Waschtisch-Umgebung zur alltäglichen Unterstützung des Zähneputzens“ teilzunehmen. Die Studie wird gemeinsam vom Exzellenzcluster *Cognitive Interaction Technology* (CITEC) der Universität Bielefeld und der Wohnrichtung Haus Bersaba, v. Bodelschwingsche Stiftungen Bethel, Bielefeld durchgeführt.

Dieses Informationsblatt soll dazu dienen, Sie über Zweck und Ablauf der Studie aufzuklären.

Teilnahme

Die Teilnahme an der wissenschaftlichen Studie ist freiwillig. Sie können jederzeit und ohne Angabe von Gründen Ihre Studienteilnahme beenden, ohne dass Ihnen dadurch Nachteile entstehen.

Ziel der Studie

Am Exzellenzcluster *Cognitive Interaction Technology* (CITEC) wird eine interaktive Assistenztechnologie entwickelt, die Personen mit kognitiven Beeinträchtigungen bei der Einhaltung des Handlungsablaufes von Alltagstätigkeiten unterstützt. In der Studie wird ein erster Prototyp der Assistenztechnologie im Bereich Zähneputzen eingesetzt. Ziel der Studie ist die Auswertung des technischen Systems im Hinblick auf den Erfolg der Probanden und die Akzeptanz des Systems.

Ablauf der Studie

Wir möchten Sie bitten an unserer wissenschaftlichen Studie zum Einsatz von Assistenztechnologie für die Unterstützung beim Zähneputzen teilzunehmen: Sie putzen sich an einem mit Sensorik ausgestatteten Waschtisch die Zähne. Das System überwacht ihren Fortschritt in der Aufgabenausführung und gibt Ihnen, falls

nötig, audio-visuelle Hinweise.

Die Studie findet in der Wohneinrichtung Haus Bersaba, v. Bodelschwingsche Stiftungen Bethel, Bielefeld statt. Die Studiendauer umfasst 3 Wochen. Innerhalb des Studienzeitraums planen wir mit Ihnen die Aufnahme eines Putzvorgangs mit Unterstützung des Systems an jedem zweiten Werktag. An Wochenenden (Sa+So) möchten wir jeweils zwei Putzvorgänge mit Ihnen aufzeichnen. Daraus ergibt sich ein Zeitaufwand für die Studie von ca. 15 Minuten an jedem zweiten Werktag sowie jeweils ca. 30 Minuten an Wochenendtagen (Sa+So).

Beispielhaft könnte eine Aufzeichnung folgendermaßen ablaufen: Wenn Sie uns an dem entsprechenden Tag bei unserer Studie unterstützen wollen, werden Sie von einer Bezugspflegerkraft abgeholt und zu dem Raum begleitet, in dem unser Waschtisch aufgebaut ist. Dort putzen Sie sich nach einer kurzen Einführung durch den Studienleiter die Zähne und werden dabei von unserem System unterstützt. Während des Durchgangs wird die Bezugspflegerkraft vor Ort sein um die Assistenz im Falle eines Systemfehlers zu übernehmen. Nach dem Durchgang füllen Sie mithilfe der Bezugspflegerkraft einen Fragebogen aus, in dem Sie uns ihre Meinung über die Unterstützung durch das System mitteilen.

Datenschutz

Die Bestimmungen des Datenschutzes werden eingehalten. Alle Studienergebnisse werden streng vertraulich behandelt. Die Einzelergebnisse werden pseudonymisiert verwaltet und weiterverarbeitet. Wissenschaftliche Ergebnisse der Untersuchung sollen in Auszügen auf wissenschaftlichen Konferenzen und in Fachzeitschriften veröffentlicht werden. Das während der Studie aufgenommene Videomaterial wird nur durch die Studienleiter eingesehen und ausgewertet. Wenn Daten (auch Videomaterial) in Vorträgen verwendet werden soll, werden die Daten durch Pseudonymisierung so aufbereitet, dass eine Identifikation Ihrer Person nicht möglich ist.

Zugriff auf die bei der Studie erhobenen Daten haben nur die studienverantwortlichen Personen und die von Ihnen im Rahmen des Projekts eingesetzten wissenschaftlichen Mitarbeiter.

Während der Studie können Sie jederzeit Auskunft über die von ihnen erhobenen Daten erhalten. Nach der Studie möchten wir Ihnen - wenn Sie einverstanden sind - die Ergebnisse unserer Studie in einem persönlichen Gespräch mitteilen.

PD Dr. Sven Wachsmuth
Dr. Thomas Hermann
Gabriele Hüttemann-Völker
Dr. Michelina Bocchicchio

Kontaktadresse:

PD Dr. Sven Wachsmuth
Central Lab Facilities
Center of Excellence Cognitive Interaction Technology
Universität Bielefeld,
Universitätsstraße 21–23,
33615 Bielefeld
Tel.: +49-521-106-2937
E-Mail: swachsmu@techfak.uni-bielefeld.de

B.2 DECLARATION OF CONSENT

Declaration of consent including privacy policy which had to be signed by the participants of the study or their legal guardians. All participants of the study (users and caregivers) signed the declaration of consent.





**Center of Excellence Cognitive Interaction Technology (CITEC)
Von Bodelschwingsche Stiftungen Bethel, Bielefeld**

**PD Dr. Sven Wachsmuth
Dr. Thomas Hermann
Gabriele Hüttemann-Völker
Dr. Michelina Bocchicchio**

Einwilligungserklärung
zur Teilnahme an der wissenschaftlichen Studie

“Eine intelligente Waschtisch-Umgebung zur alltäglichen Unterstützung des Zähneputzens”

Hiermit erkläre ich, Frau/Herr _____, mich bereit, an der wissenschaftlichen Studie „Eine intelligente Waschtisch-Umgebung zur alltäglichen Unterstützung des Zähneputzens“ der Universität Bielefeld teilzunehmen. Ich bin mündlich und schriftlich umfassend über Inhalt, Zweck und Umfang der Studie informiert worden und habe keine weiteren Fragen zu der Studie.

Ich bin in einem persönlichen Gespräch durch eine studienverantwortliche Person

.....
Name der studienverantwortlichen Person

ausführlich und verständlich über den Ablauf der wissenschaftlichen Studie aufgeklärt worden. Ich habe darüber hinaus den Text der Patienteninformation sowie die hier nachfolgend abgedruckte Datenschutzerklärung gelesen und verstanden.

Die Teilnahme an der Studie ist **freiwillig**, und ich kann die Untersuchung **jederzeit** ohne Angabe von Gründen abbrechen, ohne dass mir dadurch irgendwelche Nachteile entstehen.

Datenschutzerklärung

Mir ist bekannt, dass bei dieser wissenschaftlichen Studie personenbezogene Daten über mich erhoben, gespeichert und ausgewertet werden sollen. Die Verwendung der Daten erfolgt nach gesetzlichen Bestimmungen und setzt vor der Teilnahme an der wissenschaftlichen Studie folgende freiwillig abgegebene Einwilligungserklärung voraus, das heißt ohne die nachfolgende Einwilligung kann ich nicht an der wissenschaftlichen Studie teilnehmen.

1. Ich erkläre mich damit einverstanden, dass im Rahmen dieser wissenschaftlichen Studie personenbezogene Daten über mich erhoben und auf elektronischen Datenträgern aufgezeichnet werden. Soweit erforderlich, dürfen die erhobenen Daten pseudonymisiert (verschlüsselt) weitergegeben werden:

- a) an den Auftraggeber oder eine von diesem beauftragte Stelle zum Zwecke der wissenschaftlichen Auswertung,
- b) im Falle unerwünschter Ereignisse: an den Auftraggeber und die zuständige Landesbehörde.

2. Ich bin darüber aufgeklärt worden, dass ich jederzeit die Teilnahme an der wissenschaftlichen Studie beenden kann. Beim Widerruf meiner Einwilligung, an der Studie teilzunehmen, habe ich das Recht, die Löschung aller meiner bis dahin gespeicherten personenbezogenen Daten zu verlangen.

3. Ich erkläre mich damit einverstanden, dass meine Daten und die Ergebnisse zu Forschungszwecken genutzt werden:

- a) generell zum Zwecke der wissenschaftlichen Auswertung,
- b) pseudonymisiert und in Auszügen für wissenschaftliche Vorträge,
- c) pseudonymisiert und in Auszügen für die Veröffentlichung auf wissenschaftlichen Konferenzen und in Fachzeitschriften.

Durch die Pseudonymisierung der Daten ist eine Identifikation Ihrer Person nicht möglich.

Ich erkläre mich gesondert damit einverstanden, dass die von mir im Rahmen der Studie erhobenen Videodaten in pseudonymisierter Form (Gesichter werden durch Verpixelung unkenntlich gemacht) in Auszügen für wissenschaftliche Vorträge genutzt werden können.

Bitte kreuzen Sie hier an, wenn Sie uns Ihr Einverständnis für die Benutzung der Videodaten in wissenschaftlichen Vorträgen geben.

4. Ich erkläre mich damit einverstanden, dass meine Daten nach Beendigung oder Abbruch der Studie mindestens zehn Jahre aufbewahrt werden. Danach werden meine personenbezogenen Daten gelöscht.

Ort, Datum

Unterschrift (Teilnehmer/in/gesetzliche(r) Betreuer(in))

Ort, Datum

Unterschrift (Studienleiter)

Ort, Datum

Unterschrift (Teamleitung Haus Bersaba)

B.3 QUESTIONNAIRE

Questionnaire used after each trial of the SYS scenario. The questions referred to users were asked by the caregivers who also judged the answers of the users with regard to the Likert scale.

VPN _____

Datum _____

FBN _____

Uhrzeit _____

PKNR _____

TrialNR _____

Vielen Dank für Ihre Teilnahme.

Um das Feedback des Assistenzsystems weiter verbessern zu können, möchten wir Sie als betreuende Pflegekraft bitten, den Benutzern des Systems folgende Fragen zu stellen und die Antworten in diesen Fragebogen auszufüllen. Dabei interessiert uns auch Ihre Einschätzung der Situation. Diesbezüglich sind die Fragen an die Benutzer und an Sie, die Pflegekraft, separat gekennzeichnet.

Selbstverständlich werden die Daten vertraulich behandelt und nicht an Dritte weitergeleitet.

Bitte tragen Sie hier die Antwort der Eingangsfrage an den Benutzer, „Wie geht es Ihnen heute?“ ein:

überhaupt nicht gut O O O O O sehr gut

Bitte stellen Sie die folgenden Fragen an den Benutzer und kennzeichnen Sie für die Antwort eins der jeweiligen Kringel/Bereiche/Kästchen:

Hat Ihnen das System geholfen?

überhaupt nicht geholfen O O O O O sehr geholfen

Hat es Ihnen gefallen das System zu benutzen?

überhaupt nicht gefallen O O O O O sehr gefallen

Die nun folgenden Fragen richten sich an Sie als Pflegekraft und Ihrer Einschätzung hinsichtlich auf den vom Benutzer aktuell durchgeführten Durchgang des Zähneputzens mit dem Assistenzsystem.

Wie gut sind die Zähne (vor einem möglichen Eingreifen von Ihnen) geputzt worden?

überhaupt nicht gut O O O O O sehr gut

Wie schätzen Sie die Hilfe des Assistenzsystems ein?

überhaupt nicht gut O O O O O sehr gut

Wie gut hat der Benutzer die Hinweise des Systems verstanden?

überhaupt nicht gut O O O O O sehr gut

Liegt hier die Antwort eher im negativen Skalenbereich: Was waren Ihrer Meinung nach die Gründe?

Wie gut hat sich der Benutzer beim Zähneputzen Mühe gegeben?

überhaupt nicht gut O O O O O sehr gut

Liegt hier die Antwort eher im negativen Skalenbereich: Was waren Ihrer Meinung nach die Gründe?

Vielen Dank für das Ausfüllen unseres Fragebogens.

BIBLIOGRAPHY

- [1] Aggarwal, J. K. and Ryoo, M. S. (2011). Human activity analysis: A review. *ACM Computing Surveys*, 43(3):1–43. (Cited on pages 17, 18, 19, and 21.)
- [2] Ambite, J. L. and Knoblock, C. A. (2001). Planning by rewriting. *Artificial Intelligence Research*, 15(1):207–261. (Cited on page 12.)
- [3] American Psychiatric Association (1987). *Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised*. American Psychiatric Association. (Cited on page 7.)
- [4] Amir, Y., Danilov, C., Miskin-Amir, M., Schultz, J., and Stanton, J. (2004). The spread toolkit: Architecture and performance. Technical report, Johns Hopkins University, Baltimore, USA. (Cited on page 47.)
- [5] Baffa, A. C. E. and Ciarlini, A. E. M. (2010). Modeling pomdps for generating and simulating stock investment policies. In *SAC'10, Symposium of Applied Computing*, pages 2394–2399. (Cited on page 24.)
- [6] Bao, L. and Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In *Pervasive'04, Int. Conf. on Pervasive Computing*, pages 1–17. (Cited on page 17.)
- [7] Bauchet, J., Giroux, S., Pigot, H., Lussier-Desrochers, D., and Lachapelle, Y. (2008). Pervasive assistance in smart homes for people with intellectual disabilities: a case study on meal preparation. *Assistive Robotics and Mechatronics*, 9(4):42–54. (Cited on pages 2, 12, 13, and 105.)
- [8] Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171. (Cited on page 19.)
- [9] Beigl, M. (2000). Memoclip: A location-based remembrance appliance. *Personal and Ubiquitous Computing*, 4(4):230–233. (Cited on page 26.)
- [10] Bellman, R. (1957). A markovian decision process. *Indiana University Mathematics Journal*, 6(4):679–684. (Cited on page 24.)
- [11] Bellman, R. and Kalaba, R. (1959). On adaptive control processes. *Automatic Control*, 4(2):1–9. (Cited on page 19.)
- [12] Bieber, G., Voskamp, J., and Urban, B. (2009). Activity recognition for everyday life on mobile phones. In *UAHCI '09, Int. Conf. on Universal Access in Human-Computer Interaction*, pages 289–296. (Cited on page 17.)

- [13] Blackburn, J. and Ribeiro, E. (2007). Human motion recognition using isomap and dynamic time warping. In *HumanMotion'07, Conf. on Human Motion - Understanding, Modeling, Capture and Animation*, pages 285–298. (Cited on page 19.)
- [14] Blair, C. E. (1999). Effect of self-care aids on self-esteem of intact nursing home residents. *Issues in Mental Health Nursing*, 20(6):559–570. (Cited on page 2.)
- [15] Bobick, A. F. and Davis, J. W. (2001). The recognition of human movement using temporal templates. *Pattern Analysis and Machine Intelligence*, 23(3):257–267. (Cited on page 17.)
- [16] Bondevik, M. and Skogstad, A. (1998). The oldest old, adl, social network, and loneliness. *Western Journal of Nursing Research*, 20(3):325–343. (Cited on page 6.)
- [17] Bourgeois, M. S., Camp, C., Rose, M., White, B., Malone, M., Carr, J., and Rovine, M. (2003). A comparison of training strategies to enhance use of external aids by persons with dementia. *Communication Disorders*, 36(5):361–378. (Cited on page 25.)
- [18] Brailsford, S. C., Potts, C. N., and Smith, B. M. (1999). Constraint satisfaction problems: Algorithms and applications. *European Journal of Operational Research*, 119(3):557–581. (Cited on page 22.)
- [19] Brand, M., Oliver, N., and Pentland, A. (1997). Coupled hidden markov models for complex action recognition. In *CVPR'97, Int. Conf. on Computer Vision and Pattern Recognition*, pages 994–999. (Cited on page 20.)
- [20] Buettner, M., Prasad, R., Philipose, M., and Wetherall, D. (2009). Recognizing daily activities with rfid-based sensors. In *UbiComp'09, Int. Conf. on Ubiquitous Computing*, pages 51–60. (Cited on page 17.)
- [21] Chavarriaga, R., del R. Millán, J., Sagha, H., Bayati, H., Lukowicz, P., Bannach, D., Roggen, D., Förster, K., Calatroni, A., Tröster, G., Ferscha, A., Kurz, M., and Hözl, G. (2010). Robust activity recognition for assistive technologies: Benchmarking ml techniques. In *NIPS-MLAT'10, Int. Workshop on Machine Learning for Assistive Technologies*. (Cited on page 17.)
- [22] Choudhury, T., Borriello, G., Consolvo, S., Haehnel, D., Harrison, B., Hemingway, B., Hightower, J., Klasnja, P. P., Koscher, K., LaMarca, A., Landay, J. A., LeGrand, L., Lester, J., Rahimi, A., Rea, A., and Wyatt, D. (2008). The mobile sensing platform: An embedded activity recognition system. *Pervasive Computing*, 7(2):32–41. (Cited on page 17.)
- [23] Colbry, D., Peintner, B., and Pollack, M. E. (2002). Execution monitoring with quantitative temporal dynamic bayesian networks. In

- AIPS'02, Int. Conf. on Artificial Intelligence Planning Systems*, pages 194–203. (Cited on page 12.)
- [24] Cole, E. (1999). Cognitive prosthetics: an overview to a method of treatment. *NeuroRehabilitation*, 12(1):39–51. (Cited on page 5.)
- [25] Darwiche, A. (2009). *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press. (Cited on page 56.)
- [26] Das, B., Cook, D. J., Schmitter-Edgecombe, M., and Seelye, A. M. (2012). Puck: an automated prompting system for smart environments. *Personal and Ubiquitous Computing*, 16(7):859–873. (Cited on page 10.)
- [27] Demchak, M. (1990). Response prompting and fading methods: A review. *American Journal on Mental Retardation*, 94(6):603–615. (Cited on pages 27 and 43.)
- [28] Efros, A. A., Berg, A. C., Mori, G., and Malik, J. (2003). Recognizing action at a distance. In *ICCV'03, Int. Conf. on Computer Vision*, pages 726–733. (Cited on page 17.)
- [29] Ferri, C. P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., Hall, K., Hasegawa, K., Hendrie, H., Huang, Y., Jorm, A., Mathers, C., Menezes, P. R., Rimmer, E., and Sczufca, M. (2006). Global prevalence of dementia: a delphi consensus study. *The Lancet*, 366(9503):2112–2117. (Cited on page 1.)
- [30] Fikes, R. and Nilsson, N. J. (1971). Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2(3/4):189–208. (Cited on page 22.)
- [31] Forney, G. D. (1973). The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278. (Cited on page 20.)
- [32] Gillespie, A., Best, C., and O'Neill, B. (2011). Cognitive function and assistive technology for cognition: A systematic review. *International Neuropsychological Society*, 18(1):1–19. (Cited on pages 1, 5, 7, 8, and 10.)
- [33] Giroux, S., Bauchet, J., Pigot, H., Lussier-Desrochers, D., and Lachappelle, Y. (2008). Pervasive behavior tracking for cognitive assistance. In *PETRA'08, Int. Conf. on PErvasive Technologies Related to Assistive Environments*, pages 1–7. (Cited on page 12.)
- [34] Gould, J. and Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3):300–311. (Cited on page 29.)
- [35] Grzes, M., Hoey, J., Khan, S. S., Mihailidis, A., Czarnuch, S., Jackson, D., and Monk, A. (2012). Relational approach to knowledge engineering for pomdp-based assistance systems as a translation of a psychological model. *Approximate Reasoning (in press)*. (Cited on page 14.)

- [36] Heckaman, K., Alber, S., Hooper, S., and Heward, W. (1998). A comparison of least-to-most prompts and progressive time delay on the disruptive behavior of students with autism. *Journal of Behavioral Education*, 8(2):171–201. (Cited on page 27.)
- [37] Hoey, J. (2006). Tracking using flocks of features, with application to assisted handwashing. In *BMVC'06, British Machine Vision Conference*, pages 367–376. (Cited on page 13.)
- [38] Hoey, J., Plötz, T., Jackson, D., Monk, A., Pham, C., and Olivier, P. (2011). Rapid specification and automated generation of prompting systems to assist people with dementia. *Pervasive and Mobile Computing*, 7(3):299–318. (Cited on pages 14 and 31.)
- [39] Hoey, J., Poupart, P., Bertoldi, A. v., Craig, T., Boutilier, C., and Mihailidis, A. (2010). Automated handwashing assistance for persons with dementia using video and a partially observable markov decision process. *Computer Vision and Image Understanding*, 114(5):503–519. (Cited on pages 2, 13, 17, 24, 27, and 105.)
- [40] Howard, R. A. (1960). *Dynamic Programming and Markov Processes*. MIT Press, Cambridge, MA. (Cited on page 24.)
- [41] Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1–2):99–134. (Cited on page 24.)
- [42] Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems*, 2(1):26–41. (Cited on page 38.)
- [43] Kirsch, N., Levine, S., Fallon-Krueger, M., and Jaros, L. (1987). The microcomputer as an 'orthotic' device for patients with cognitive deficits. *Journal of Head Trauma Rehabilitation*, 2(4):77–86. (Cited on page 5.)
- [44] Kluger, A., Gianutsos, J. G., Golomb, J., Ferris, S. H., George, A. E., Franssen, E., and Reisberg, B. (1997). Patterns of motor impairment in normal aging, mild cognitive decline, and early alzheimer' disease. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 52(1):28–39. (Cited on pages 29 and 50.)
- [45] Laptev, I. (2005). On space-time interest points. *Computer Vision*, 64(2–3):107–123. (Cited on page 18.)
- [46] Laptev, I., Caputo, B., SchÄEldt, C., and Lindeberg, T. (2007). Local velocity-adapted motion events for spatio-temporal recognition. *Computer Vision and Image Understanding*, 108(3):207–229. (Cited on page 17.)

- [47] Leo, M., D’Orazio, T., and Spagnolo, P. (2004). Human activity recognition for automatic visual surveillance of wide areas. In *VSSN’04: Int. Workshop on Video surveillance & sensor networks*, pages 124–130. (Cited on page 17.)
- [48] Leroy, G. (2011). *Designing User Studies in Informatics*. Springer, London, UK. (Cited on page 30.)
- [49] Lessmann, N., Kranstedt, A., and Wachsmuth, I. (2004). Towards a cognitively motivated processing of turn-taking signals for the embodied conversational agent max. In *AAMAS’04, Int. Workshop on Embodied Conversational Agents: Balanced Perception and Action*, pages 65–72. (Cited on page 42.)
- [50] Levine, S. P., Horstmann, H. M., and Kirsch, N. L. (1992). Performance considerations for people with cognitive impairment in accessing assistive technologies. *Head Trauma Rehabilitation*, 7(3):46–58. (Cited on page 7.)
- [51] Levinson, R. (1995). A general programming language for unified planning and control. *Artificial Intelligence*, 76(1–2):319–375. (Cited on page 11.)
- [52] Levinson, R. (1997). The planning and execution assistant and trainer (peat). *Head Trauma Rehabilitation*, 12(2):85–91. (Cited on pages 11, 27, and 105.)
- [53] Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140):1–55. (Cited on page 81.)
- [54] Lim, J.-H., Jang, H., Jang, J., and Park, S.-J. (2008). Daily activity recognition system for the elderly using pressure sensors. In *EMBC’08, Int. Conf. on Engineering in Medicine and Biology Society*, pages 5188–5191. (Cited on page 17.)
- [55] Lin, W., Sun, M.-T., Poovendran, R., and Zhang, Z. (2008). Human activity recognition for video surveillance. In *ISCAS’08, Int. Symp. on Circuits and Systems*, pages 2737–2740. (Cited on page 17.)
- [56] LoPresti, E. F., Bodine, C., and Lewis, C. (2008). Assistive technology for cognition - understanding the needs of persons with disabilities. *Engineering in Medicine and Biology Magazine*, 27(2):29–39. (Cited on page 7.)
- [57] LoPresti, E. F., Mihailidis, A., and Kirsch, N. (2004). Assistive technology for cognitive rehabilitation: State of the art. *Neuropsychological Rehabilitation*, 14(1-2):5–39. (Cited on pages 5, 6, and 7.)
- [58] Lynch, B. (2002). Historical review of computer-assisted cognitive retraining. *Journal of Head Trauma Rehabilitation*, 17(5):446–457. (Cited on pages 5 and 6.)

- [59] M, B. and EM, S. (2012). Caregiving burden, stress, and health effects among family caregivers of adult cancer patients. *American Medical Association*, 307(4):398–403. (Cited on page 2.)
- [60] McCarthy, C. E. and Pollack, M. E. (2002). A plan-based personalized cognitive orthotic. In *AIPS'02, Int. Conf. on Artificial Intelligence Planning Systems*, pages 243–252. (Cited on page 12.)
- [61] Mihailidis, A., Boger, J., Craig, T., and Hoey, J. (2008). The coach prompting system to assist older adults with dementia through handwashing: An efficacy study. *BMC Geriatrics (online)*, 8(28). (Cited on page 14.)
- [62] Mihailidis, A., Boger, J., Hoey, J., and Jiancaro, T. (2011). *Zero-Effort Technologies - Considerations, Challenges and Use in Health, Wellness and Rehabilitation*. Morgan & Claypool. (Cited on pages 9 and 10.)
- [63] Moeslund, T. B., Hilton, A., and Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2-3):90–126. (Cited on page 17.)
- [64] Murphy, K. P. (2000). A survey of pomdp solution techniques. Technical report, Massachusetts Institute of Technology, Cambridge, USA. (Cited on page 24.)
- [65] Murzynski, N. T. and Bourret, J. C. (2007). Combining video modeling and least-to-most prompting for establishing response chains. *Behavioral Interventions*, 22(2):147–152. (Cited on page 27.)
- [66] Natarajan, P. and Nevatia, R. (2007). Coupled hidden semi markov models for activity recognition. In *WMVC'07, Workshop on Motion and Video Computing*, pages 10–17. (Cited on page 20.)
- [67] Nau, D., Ghallab, M., and Traverso, P. (2004). *Automated Planning: Theory & Practice*. Morgan Kaufmann Publishers Inc. (Cited on page 25.)
- [68] Neapolitan, R. (2004). *Learning Bayesian networks*. Prentice Hall Series in Artificial Intelligence. Pearson Prentice Hall. (Cited on page 56.)
- [69] Nguyen, N., Phung, D., Venkatesh, S., and Bui, H. (2005). Learning and detecting activities from movement trajectories using the hierarchical hidden markov model. In *CVPR'05, Int. Conf. on Computer Vision and Pattern Recognition*, pages 955–960. (Cited on page 21.)
- [70] Oliver, N., Garg, A., and Horvitz, E. (2004). Layered representations for learning and inferring office activity from multiple sensory channels. *Computer Vision and Image Understanding*, 96(2):163–180. (Cited on page 21.)

- [71] Oliver, N. M., Rosario, B., and Pentland, A. P. (2000). A bayesian computer vision system for modeling human interactions. *Pattern Analysis and Machine Intelligence*, 22(8):831–843. (Cited on page 20.)
- [72] O’Neill, B., Best, C., Gillespie, A., and O’Neill, L. (2013). Automated prompting technologies in rehabilitation and at home. *Social Care and Neurodisability*, 4(1):17–28. (Cited on page 26.)
- [73] O’Neill, B. and Gillespie, A. (2008). Simulating naturalistic instruction: the case for a voice mediated interface for assistive technology for cognition. *Assistive Technologies*, 2(2):22–31. (Cited on pages 26 and 41.)
- [74] O’Neill, B., Moran, K., and Gillespie, A. (2010). Scaffolding rehabilitation behaviour using a voice-mediated assistive technology for cognition. *Neuropsychological Rehabilitation*, 20(4):509–527. (Cited on page 26.)
- [75] Panagiotakis, C., Ramasso, E., Tziritas, G., Rombaut, M., and Pellerin, D. (2006). Shape-motion based athlete tracking for multilevel action recognition. In *AMDO’06, Int. Conf. on Articulated Motion and Deformable Objects*, pages 385–394. (Cited on page 17.)
- [76] Parlak, S., Marsic, I., and Burd, R. S. (2011). Activity recognition for emergency care using rfid. In *BodyNets’11, Int. Conf. on Body Area Networks*, pages 40–46. (Cited on page 17.)
- [77] Patterson, D. J., Liao, L., Gajos, K., Collier, M., Livic, N., Olson, K., Wang, S., Fox, D., and Kautz, H. A. (2004). Opportunity knocks: A system to provide cognitive assistance with transportation services. In *UbiComp’04, Int. Conf. on Ubiquitous Computing*, pages 433–450. (Cited on page 26.)
- [78] Penberthy, S. J. and Weld, D. S. (1992). Ucpop: A sound, complete, partial order planner for ADL. In *KR’92. Int. Conf. on Principles of Knowledge Representation and Reasoning*, pages 103–114. (Cited on page 23.)
- [79] Pernecky, R., Pohl, C., Sorg, C., Hartmann, J., Komossa, K., Alexopoulos, P., Wagenpfeil, S., and Kurz, A. (2006). Complex activities of daily living in mild cognitive impairment: conceptual and diagnostic issues. *Age and Ageing*, 35(3):240–245. (Cited on page 6.)
- [80] Peters, C., Hermann, T., and Wachsmuth, S. (2011). Prototyping of an automatic prompting system for a residential home. In *RESNA/ICTA 2011 Conf. Proc. (online)*. (Cited on page 38.)
- [81] Peters, C., Hermann, T., and Wachsmuth, S. (2012). User behavior recognition for an automatic prompting system - a structured approach based on task analysis. In *ICPRAM’12, Int. Conf. on Pattern Recognition Applications and Methods*, pages 162–171. (Cited on page 50.)

- [82] Peters, C., Hermann, T., and Wachsmuth, S. (2013). Tebra - an automatic prompting system for persons with cognitive disabilities in brushing teeth. In *HealthInf'13, Int. Conf. on Health Informatics (in press)*. (Cited on page 65.)
- [83] Peters, C., Wachsmuth, S., and Hoey, J. (2009). Learning to recognise behaviours of persons with dementia using multiple cues in an hmm-based approach. In *PETRA'09, Proceedings of the 2nd Int. Conf. on Pervasive Technologies Related to Assistive Environments*, pages 1–8. (Cited on page 17.)
- [84] Pham, C., Plötz, T., and Olivier, P. (2010). A dynamic time warping approach to real-time activity recognition for food preparation. In *Aml'10, Int. Joint Conf. on Ambient Intelligence*, pages 21–30. (Cited on page 17.)
- [85] Philipose, M., Fishkin, K. P., Perkowitz, M., Patterson, D. J., Fox, D., Kautz, H. A., and Hähnel, D. (2004). Inferring activities from interactions with objects. *Pervasive Computing*, 3(4):50–57. (Cited on pages 10 and 17.)
- [86] Pigot, H., Savary, J.-P., Metzger, J. L., Rochon, A., and Beaulieu, M. (2005). Advanced technology guidelines to fulfill the needs of the cognitively impaired population. In *ICOST'05, Int. Conf. On Smart Homes and health Telematic*, pages 25–32. (Cited on page 13.)
- [87] Pollack, M. E., Brown, L. E., Colbry, D., McCarthy, C. E., Orosz, C., Peintner, B., Ramakrishnan, S., and Tsamardinos, I. (2003). Autominder: an intelligent cognitive orthotic system for people with memory impairment. *Robotics and Autonomous Systems*, 44(3–4):273–282. (Cited on pages 11 and 105.)
- [88] Poole, D. and Mackworth, A. K. (2010). *Artificial Intelligence - Foundations of Computational Agents*. Cambridge University Press. (Cited on page 22.)
- [89] Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286. (Cited on page 19.)
- [90] Rashidi, P. and Mihailidis, A. (2012). A survey on ambient assisted living tools for older adults. *IEEE Transaction on Information Technology in Biomedicine*, 17(3):579–590. (Cited on page 10.)
- [91] Richards, S. B., Taylor, R. L., Ramasamy, R., and Richards, R. (1998). *Single Subject Research and Designs: Applications in Educational and Clinical Settings*. Singular Publ. Group. (Cited on page 84.)
- [92] Robson, C. (2002). *Real world research: a resource for social scientists and practitioner-researchers*. John Wiley & Sons. (Cited on page 84.)

- [93] Ross, S., Chaib-Draa, B., and Pineau, J. (2008). Bayesian reinforcement learning in continuous pomdps with application to robot navigation. In *ICRA'08, Int. Conf. on Robotics and Automation*, pages 2845–2851. (Cited on page 24.)
- [94] Russell, S. J. and Norvig, P. (1996). *Artificial intelligence: a modern approach*. Prentice-Hall, Inc. (Cited on pages 22 and 23.)
- [95] Ryoo, M. S. and Aggarwal, J. K. (2009). Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities. In *ICCV'09, Int. Conf. on Computer Vision*, pages 1593–1600. (Cited on page 18.)
- [96] Ryu, H. and Monk, A. (2009). Interaction Unit Analysis: A New Interaction Design Framework. *Human-Computer Interaction*, 24(4):367–407. (Cited on pages 30 and 31.)
- [97] Scherer, M. J., Hart, T., Kirsch, N., and Schulthesis, M. (2005). Assistive technologies for cognitive disabilities. *Critical Reviews in Physical and Rehabilitation Medicine*, 17(3):195–215. (Cited on pages 2, 6, and 29.)
- [98] Schilit, B., Adams, N., and Want, R. (1994). Context-aware computing applications. In *WMCSA'94, Workshop on Mobile Computing Systems and Applications*, pages 85–90. (Cited on page 9.)
- [99] Seelye, A. M., Schmitter-Edgecombe, M., Das, B., and Cook, D. J. (2012). Application of cognitive rehabilitation theory to the development of smart prompting technologies. *IEEE Reviews in Biomedical Engineering*, 5:29–44. (Cited on pages 25, 26, and 41.)
- [100] Senin, P. (2008). Dynamic Time Warping Algorithm Review. Technical report, University of Hawaii, Honolulu, Hawaii. (Cited on page 19.)
- [101] Shin, K. G. and Ramanathan, P. (1994). Realtime computing: A new discipline of computer science and engineering. 82(1):6–24. (Cited on page 47.)
- [102] Siepmann, F., Ziegler, L., Kortkamp, M., and Wachsmuth, S. (2012). Deploying a modeling framework for reusable robot behavior to enable informed strategies for domestic service robots. *Robotics and Autonomous Systems (in press)*. (Cited on page 52.)
- [103] Sohlberg, M. M., Ehlhardt, L., and Kennedy, M. (2005). Instructional techniques in cognitive rehabilitation: A preliminary report. *Seminars in Speech and Language*, 26(4):268–279. (Cited on page 25.)
- [104] Spector, W. D., Katz, S., Murphy, J. B., and Fulton, J. P. (1987). The hierarchical relationship between activities of daily living and instrumental activities of daily living. *Journal of Chronic Diseases*, 40(6):481–489. (Cited on page 7.)

- [105] Sprunt, B., Sha, L., and Lehoczky, J. (1989). Aperiodic task scheduling for hard-real-time systems. *Real-Time Systems*, 1(1):27–60. (Cited on page 47.)
- [106] Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT press, Cambridge, USA. (Cited on page 24.)
- [107] Swain, M. J. and Ballard, D. H. (1990). Indexing via color histograms. In *ICCV'90, Int. Conf. on Computer Vision*, pages 390–393. (Cited on page 52.)
- [108] Szymkowiak, A., Morrison, K., Gregor, P., Shah, P., Evans, J., and Wilson, A. (2005). A memory aid with remote communication using distributed technology. *Personal Ubiquitous Computing*, 9(1):1–5. (Cited on page 26.)
- [109] Thielke, S., Harniss, M., Thompson, H., Patel, S., Demiris, G., and Johnson, K. (2012). Maslow’s hierarchy of human needs and the adoption of health-related technologies for older adults. *Ageing International*, 37(4):470–488. (Cited on page 9.)
- [110] Thomson, B. and Young, S. (2010). Bayesian update of dialogue state: A pomdp framework for spoken dialogue systems. *Computer Speech & Language*, 24(4):562–588. (Cited on page 24.)
- [111] Tsui, K. M. and Yanco, H. A. (2010). Prompting devices: A survey of memory aids for task sequencing. In *QLoT'10, Int. Symp. on Quality of Life (QOL) Technology: Intelligent Systems for Better Living (held in conjunction with RESNA 2010)*. (Cited on page 10.)
- [112] Turaga, P., Chellappa, R., Subrahmanian, V. S., and Udrea, O. (2008). Machine recognition of human activities: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(11):1473–1488. (Cited on page 17.)
- [113] Veeraraghavan, A., Chellappa, R., and Roy-Chowdhury, A. K. (2006). The function space of an activity. In *CVPR'06, Int. Conf. on Computer Vision and Pattern Recognition*, pages 959–968. (Cited on page 19.)
- [114] Wienke, J. and Wrede, S. (2011). A middleware for collaborative research in experimental robotics. In *SII'11, Int. Symp. on System Integration*, pages 1183–1190. (Cited on page 47.)
- [115] Wilmoth, J. (2000). Demography of longevity: past, present, and future trends. *Experimental Gerontology*, 35(9–10):1111–1129. (Cited on page 1.)
- [116] Wilson, B., Evans, J., Emslie, H., and Malinek, V. (1997). Evaluation of neuropage: a new memory aid. *Neurology, Neurosurgery, and Psychiatry*, 63(1):113–115. (Cited on page 26.)

- [117] World Health Organization (2000). *Ageing and Intellectual Disabilities - Improving Longevity and Promoting Healthy Ageing: Summative Report*. Geneva, Switzerland. (Cited on page 1.)
- [118] Wu, J., Osuntogun, A., Choudhury, T., Philipose, M., and Rehg, J. M. (2007). A scalable approach to activity recognition based on object use. In *ICCV'07, Int. Conf. on Computer Vision*, pages 1–8. (Cited on page 20.)
- [119] Yamato, J., Ohya, J., and Ishii, K. (1992). Recognizing human action in time-sequential images using hidden markov model. In *CVPR '92, Int. Conf. on Computer Vision and Pattern Recognition*, pages 379–385. (Cited on page 20.)
- [120] Zhang, S. and Sridharan, M. (2010). Vision-based scene analysis on a mobile robot using layered pomdps. In *ICAPS'10, Workshop on POMDP Practitioners*. (Cited on page 24.)
- [121] Zhu, C. and Sheng, W. (2009). Human daily activity recognition in robot-assisted living using multi-sensor fusion. In *ICRA'09, Int. Conf. on Robotics and Automation*, pages 3644–3649. (Cited on page 17.)
- [122] Zhu, C. and Sheng, W. (2010). Recognizing human daily activity using a single inertial sensor. In *WCICA'10, World Congress on Intelligent Control and Automation*, pages 282–287. (Cited on page 17.)
- [123] Zouba, N., Boulay, B., Bremond, F., and Thonnat, M. (2008). Monitoring activities of daily living (adls) of elderly based on 3d key human postures. In *Lecture Notes in Computer Science - Cognitive Vision*, volume 5329, pages 37–50. (Cited on page 17.)
- [124] Zouch, M., Yeung, T., and Castanier, B. (2011). Optimal resurfacing decisions for road maintenance: A pomdp perspective. In *RAMS'11, Reliability and Maintainability Symposium*, pages 1–6. (Cited on page 24.)