Advancing the COACH automated prompting system toward an unsupervised, real-world deployment

by

Stephen Czarnuch

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Institute of Biomaterials and Biomedical Engineering
University of Toronto

© Copyright by Stephen Czarnuch 2014
Advancing the COACH automated prompting system toward an unsupervised, real-world deployment

Stephen Czarnuch
Doctor of Philosophy
Institute of Biomaterials and Biomedical Engineering
University of Toronto
2014

Abstract

The COACH is a prototype automated prompting system designed to encourage independent participation in daily tasks by older adults with dementia. Supervised, clinical trials with the COACH were promising, yet little was known about the capabilities of the system in an unsupervised, real-world environment or the needs of the system’s users in such a deployment. To continue development of the COACH toward an unsupervised, real-world in-home deployment I employed an informed approach. First, the needs of the users of the system were identified through a national questionnaire completed by family caregivers of older adults with dementia and validated using a follow-up focus group. Participants indicated that the COACH should provide unobtrusive support of private or personal tasks by leveraging the remaining abilities of the older adults with dementia using the system, promoting independent task completion. Next, the performance of the COACH was evaluated in an unsupervised, real-world clinical deployment. One of the COACH modules responsible for tracking the interactions between the system users and objects in the environment was implicated as the cause of the majority of the failures, successfully identifying only 46.9% of the task steps completed by the trial participants. The technical factors limiting system performance in the real-world clinical
deployment were synthesized with the needs of the users to develop technical design criteria which were used to guide the development of an improved COACH prototype. The performance of the new COACH prototype was evaluated in a second supervised deployment, resulting in the successful identification of 93.7% of the task steps completed by trial participants. Encouraged by these results, the COACH system is again ready for evaluation in an unsupervised, real-world environment.
Acknowledgments

First and foremost, I would like to express my sincerest appreciation and thanks to my supervisor, Dr. Alex Mihailidis. I could not have navigated the many personal and professional challenges without his leadership, guidance, support, patience and encouragement and cannot say enough about his knowledge, advice and insight. I would also like to thank my committee members, Dr. Jill Cameron, Dr. Elizabeth Rochon, and Dr. Tom Chau, for their irreplaceable role as my academic mentors and role models. I would also like to thank my examiners, Dr. Elaine Biddiss and Dr. Diane Cook, for their insightful appraisals, constructive feedback and encouragement.

I would like to thank my parents, Stan and Regina, who taught me to value asking questions, have the confidence to seek answers and most importantly to be able to make hard decisions. I would also like to thank Giovanni and Antonia Ricciardelli for showing the value of hard work, not only through their words but through their actions. Most of all, I offer my heart as well as my thanks to my family. My wife, Rose, is my leader, my love, and my motivation. Without her I would never have considered the Ph.D. and her contributions penetrate every aspect of my work and my life. And finally, I thank my daughter, Sadie, and sons Nate and Johny. Sadie, I thank for keeping me calm during my worst storms with her simple, yet utterly profound questions and comments that somehow make the most complicated circumstances seem trivial and tractable. Nate, I thank for his ability to show me, by example, that one can come out of any trial with a smile on their face, but more importantly, with a long memory and a lesson learned. Johny, I thank for showing me that despite being an underdog by all standard measures, one can always achieve a goal without losing one’s style, wit, morals and sense of self. Most of all, I thank my family for showing me, whether I liked it or not, that I could be a good academic, a good father, and a good husband all at the same time.
# Table of Contents

## Contents

Acknowledgments .......................................................................................................................... iv  
Table of Contents ...................................................................................................................... v  
List of Abbreviations ............................................................................................................... x  
List of Tables ............................................................................................................................ xv  
List of Figures .......................................................................................................................... xvii  
List of Appendices .................................................................................................................. xviii  
Chapter 1 ................................................................................................................................... 1  
1 General Introduction ............................................................................................................. 1  
1.1 Statement of Problem and Rationale .............................................................................. 1  
1.2 The Current COACH Automated Prompting System .................................................. 4  
1.3 The History of the COACH ........................................................................................... 5  
1.4 Dissertation Objectives ................................................................................................. 7  
1.4.1 Research Questions .................................................................................................. 7  
1.5 Dissertation Overview ................................................................................................. 8  
1.5.1 Chapter 2: The Design of Intelligent in-home Assistive Technologies: Assessing the Needs of Older Adults with Dementia and Their Caregivers .... 9  
1.5.2 Chapter 3: A Real-world Deployment of the COACH Prompting System ........... 10  
1.5.3 Chapter 4: Depth Image Hand Tracking from an Overhead Perspective Using Partially Labeled, Unbalanced Data: Development and Real-world Testing .... 11  
Chapter 2 ................................................................................................................................ 15  
2 The Design of Intelligent in-home Assistive Technologies: Assessing the Needs of Older Adults with Dementia and Their Caregivers ........................................... 15  
2.1 Publication Citation ........................................................................................................ 15  
2.2 Abstract .......................................................................................................................... 15
### 3.3.1 Supporting Daily Activities with Assistive Technology for Cognition: Clinical Trials ................................................................. 36

### 3.3.2 Current Study ........................................................................................................ 37

### 3.4 Methods ..................................................................................................................... 38

#### 3.4.1 Participants ........................................................................................................ 38

#### 3.4.2 Device (COACH) .................................................................................................. 38

#### 3.4.3 Procedure ........................................................................................................... 40

#### 3.4.4 Data analysis ..................................................................................................... 42

### 3.5 Results ........................................................................................................................ 45

#### 3.5.1 Participants ........................................................................................................ 45

#### 3.5.2 Trials .................................................................................................................. 45

#### 3.5.3 System Performance .......................................................................................... 46

### 3.6 Discussion .................................................................................................................. 49

#### 3.6.1 Participants ........................................................................................................ 49

#### 3.6.2 System Performance .......................................................................................... 50

#### 3.6.3 System component analyses .............................................................................. 53

#### 3.6.4 Limitations .......................................................................................................... 56

#### 3.6.5 Future work and required areas of improvement ............................................... 57

### 3.7 Conclusions ............................................................................................................. 58

#### 3.7.1 Acknowledgements ........................................................................................... 60

### Chapter 4 ....................................................................................................................................... 61

#### 4 Development and evaluation of a hand tracker using depth images captured from an overhead perspective ......................................................................................... 61

#### 4.1 Publication Citation ................................................................................................. 61

#### 4.2 Abstract .................................................................................................................... 61

#### 4.3 Introduction .............................................................................................................. 61
4.3.1 Related Work ................................................................. 63
4.3.2 Existing Tracking Approaches and the COACH .................. 64
4.3.3 Contributions ................................................................. 65
4.4 Method ........................................................................... 66
  4.4.1 Data Collection ............................................................. 66
  4.4.2 Intermediate Multiclass Classification ......................... 68
  4.4.3 Aggregate Part Proposals ............................................. 69
  4.4.4 Task-based Hand Positions ........................................ 71
4.5 Results ........................................................................... 72
  4.5.1 Data Collection ............................................................. 72
  4.5.2 Intermediate Multiclass Classification ......................... 72
  4.5.3 Aggregate Part Proposals ............................................. 73
  4.5.4 Task-based Hand Positions ........................................ 74
4.6 Discussion and Conclusion ............................................. 75
  4.6.1 Acknowledgements ....................................................... 80
Chapter 5 ............................................................................... 81
5 Conclusion .......................................................................... 81
  5.1 Dissertation Contributions ............................................. 81
    5.1.1 Research Contributions .............................................. 81
    5.1.2 Engineering Contributions ......................................... 90
  5.2 Overall Limitations ....................................................... 90
  5.3 Future Work ................................................................. 92
  5.4 Final Remarks .............................................................. 95
6 References ................................................................. 97

Appendix 1: House of Quality used to determine required changes for the three major COACH modules: (1) hand tracking, (2) policy, and (3) prompting ................................. 109
List of Abbreviations

ADL

activities of daily living ........................................................................................................... 1

AP

average precision ...................................................................................................................... 70

API

application programming interface ........................................................................................ 64

AT

assistive technology ................................................................................................................ 1

ATC

assistive technology for cognition .......................................................................................... See IAT

COACH

Cognitive Orthosis for Assisitng with aCtivities in the Home ................................................. 4

CV

critical value ............................................................................................................................ 24

DSM IV-TR

Diagnostic and Statistical Manual of Mental Disorders, 4th edition, text revision ............... 38

DTW

dynamic time warping ............................................................................................................. 45

EER
equal error rate .......................................................................................................................... 70

FN
false negative .................................................................................................................................. 70

FP
false positive .................................................................................................................................. 70

GPS
global positioning system ........................................................................................................... 17

HCI
human-computer interaction ........................................................................................................ 93

HOQ
House of Quality .......................................................................................................................... 11

IADL
instrumental activities of daily living ........................................................................................ 18

IAT
intelligent assistive technology .................................................................................................. 1

ICH-GCP
International Conference on Harmonisation of Technical Requirements for Registration of
Pharmaceuticals for Human Use - Good Clinical Practice guidelines ....................................... 38

IRB
Institutional Review Board .......................................................................................................... 41

KMO
Kaiser-Meyer-Olkin measure of sampling adequacy ................................................................. 24

mAP

mean average precision ........................................................................................................... 70

MDP

Markov decision process ...................................................................................................... 5

MMSE

Mini Mental State Exam ......................................................................................................... 32

OSR

Overall Success Rate ............................................................................................................. 69

PADL

personal activities of daily living .......................................................................................... 18

PDF

probability density functions ................................................................................................. 13

PO

Primary objective ..................................................................................................................... 7

POMDP

partially observable Markov decision process ..................................................................... 39

PR

precision-recall ....................................................................................................................... 70

QFD
<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality Function Deployment</td>
<td>19</td>
</tr>
<tr>
<td>REB</td>
<td>23</td>
</tr>
<tr>
<td>Research Ethics Board</td>
<td>23</td>
</tr>
<tr>
<td>RFID</td>
<td>17</td>
</tr>
<tr>
<td>radio frequency identification</td>
<td>17</td>
</tr>
<tr>
<td>RGB</td>
<td>66</td>
</tr>
<tr>
<td>red, green, blue</td>
<td>66</td>
</tr>
<tr>
<td>S02</td>
<td>7</td>
</tr>
<tr>
<td>Secondary objective 2</td>
<td>7</td>
</tr>
<tr>
<td>SDK</td>
<td>64</td>
</tr>
<tr>
<td>software development kit</td>
<td>64</td>
</tr>
<tr>
<td>SDT</td>
<td>44</td>
</tr>
<tr>
<td>signal detection theory</td>
<td>44</td>
</tr>
<tr>
<td>SNAP</td>
<td>58</td>
</tr>
<tr>
<td>Syndetic Assistance Process</td>
<td>58</td>
</tr>
<tr>
<td>SO1</td>
<td>7</td>
</tr>
<tr>
<td>Secondary objective 1</td>
<td>7</td>
</tr>
<tr>
<td>TMP</td>
<td>37</td>
</tr>
<tr>
<td>Toronto Memory Program</td>
<td>37</td>
</tr>
<tr>
<td>TN</td>
<td></td>
</tr>
</tbody>
</table>
true negative ........................................................................................................................................ 70

TP

ture true positive ........................................................................................................................................ 70

UAR

unweighted average recall .......................................................................................................................... 69

UCD

User centred design ........................................................................................................................................ 19

UI

user interface .................................................................................................................................................. 93
List of Tables

Table 2.1: Factorability statistics, factors satisfying each of the three inclusion criteria and optimal rotation techniques for each of the four survey sections ................................................. 25

Table 2.2: Independent ADL completion by person with dementia ........................................... 26

Table 2.3: Caregivers assisting completion of ADL ..................................................................... 27

Table 2.4: Supporting ADL with Assistive Technology .............................................................. 28

Table 2.5: Features and functions of an in-home IAT ................................................................. 30

Table 2.6: Reliability Statistics of each factor for each of the four survey sections, calculated Using Cronbach’s Alpha ........................................................................................................... 30

Table 3.1: User behaviours and scoring rules for the first stage of analyses .............................. 42

Table 3.2: Behavioural modifiers for the six task steps in the first stage of analyses ...................... 42

Table 3.3: Binary classification function for the task step performance of the COACH ............... 43

Table 3.4: Signal Detection Theory analysis for the task assistance performance of the COACH ..................................................................................................................................................... 44

Table 3.5: Step-based effectiveness of the COACH at identifying a completed or incomplete user step .................................................................................................................................................... 47

Table 3.6: Breakdown of system module failure modes for steps labeled as Misses or False Alarms .............................................................................................................................................. 47

Table 3.7: User performance identification during live trials ..................................................... 47

Table 3.8: Task assistance effectiveness of the COACH using Signal Detection Theory ............... 48

Table 3.9: Prompt performance based on prompt characteristics of sequencing and timing ....... 48
Table 3.10: Frame-by-frame dynamic time warping hand tracker analyses between trial data and ground truth data for 5, 10 and 21 random trials ................................................................. 49

Table 3.11: User performance identification using ideal hand tracker input to COACH .......... 55

Table 4.1: Aggregate part proposal performance parameters and outcomes ....................... 73

Table 4.2: Mean final part proposal $F_{0.5}$ measure of performance over the set of validation trials, categorized by action and body part $p$; number of validation and training frames within each action category ................................................................................................................................. 74

Table 4.3: The performance of the COACH while monitoring the task-based activities of the participants over the set of validation trials. ........................................................................... 75

Table 5.1: Comparison of step tracking performance between depth-based hand tracker and colour-based hand tracker ........................................................................................................ 89
List of Figures

Figure 3.1: COACH System conceptual overview ................................................................. 39

Figure 3.2: COACH at Toronto Memory Program............................................................... 41

Figure 4.1: Sample images captured from the overhead perspective. Top left: background image; Top right: depth image; Bottom left: labeled foreground image; Bottom right: RGB image. ..... 67

Figure 4.2: Unweighted average recall (UAR) versus training parameter. (a) Depth of decision tree. (b) Number of training images per forest. (c) Minimum cutoff gain for leaf node. (d) Offset vector maximum offset (in pixel-meters). (e) Number of training sample pixels per image: 1723 and 4971 training images per forest. (f) Number of trees in forest (T). ................................. 73
List of Appendices

Appendix 1: House of Quality used to determine required changes for the three major COACH modules: (1) hand tracking, (2) policy, and (3) prompting ................................................................. 109

Appendix 2: House of Quality used to identify features and functions deemed essential to the successful installation and use of COACH in a home environment .............................................. 110

Appendix 3: Doctoral Publication List .................................................................................................. 111
Chapter 1

1 General Introduction

1.1 Statement of Problem and Rationale

Over 35 million people worldwide and at least 500,000 people in Canada are estimated to be living with dementia [1]. The prevalence of dementia increases with age, where in Canada 1 in 14 people over the age of 60, and 1 in 2 people over the age of 90 currently have dementia [2]. Globally, these numbers are expected to double every twenty years at a rate of almost eight million new cases per year [2, 3]. Presently, the total estimated worldwide cost of dementia is over $6 billion USD, represented almost entirely by the costs associated with informal (unpaid) and social care [3]. The onset of dementia negatively affects one’s health-related quality of life; specifically when a key component of quality of life, activities of daily living (ADL) [2] become challenging or even impossible to complete independently. Furthermore, an estimated 55% of all Canadians with dementia live at home, and this estimate is expected to rise to 62% by 2032 [2]. Since care for older adults with dementia is often provided by informal (unpaid) caregivers, as opposed to formal (paid) caregivers, in the home environment, the burden associated with helping with ADL completion is often shifted onto the informal caregiver as the dementia progresses [4]. Informal caregivers are generally family members of the care recipient (typically an older female child or spouse) though they can be others such as friends or neighbours [5-7].

Assistive technologies (ATs) have been developed to support a range of deficits experienced by older adults with dementia, including memory loss, executive dysfunction, social dysfunction, visuospatial impairment, behavioral and mood disturbances, disorientation, the inability to complete daily activities, and the inability to ensure personal safety [8-10]. A subset of these assistive technologies, called intelligent assistive technology (IAT), are able to identify the context of a user’s activity and determine if assistance is required [8]. Bharucha et al. [8] categorized IATs by their primary intended purpose within one of four categories: cognitive aids, physiological sensors, environmental sensors, and advanced integrated sensor systems. According to this classification, IATs designed to support the completion of ADL are by necessity considered advanced integrated sensor systems because of the challenges associated with activity detection, task progression and the provision of assistance. Examples of IATs
designed to support ADL completion are the COACH [11, 12], Autominder [13, 14], the Assisted Cognition Project [15] and PROACT [16]. These IATs are able to monitor the task progression or task compliance of older adults during ADL such as hand washing, toileting, meal preparation, and medication adherence. Furthermore, these IATs are designed with the intention of providing appropriately timed and useful automated prompts and reminders, among other forms of assistance, as necessary.

However, despite the potential help IATs offer, studies have shown that IATs designed to support people with dementia are simply not translating into commercial products [8]. Indeed, few IATs ever reach the stage of clinical trials [8], let alone real-world testing, resulting in limited knowledge of the effectiveness of IATs in real-world settings. In a systematic review including fifty-eight (58) IATs designed to support people with dementia, Bharucha et al. [8] reported that published clinical trial results were available for only three (3) IATs: the COACH [11], CareWatch [17, 18] and CareMedia [19]. The same study by Bharucha et al. reported that results of technical prototype testing and case studies were available for an additional twenty-one (21) IATs. However, although a strong argument can be made in favour of the value of case studies (and prototype testing), and even to the fact that they can be generalizable [20], each study was designed to test a specific hypothesis covering a range of purposes – some of which are not relevant to the IATs designed to support ADL (e.g., navigation aids [21], fall detectors [22-24], balance improvement aids, [25] and physiological monitors [26].

The dearth of IATs reaching the stage of real-world testing only provides a partial explanation for the poor acceptance and adoption of IATs for people with dementia. One of the fundamental purposes of real-world testing is to evaluate whether an IAT meets the broad needs of the end user (e.g., task-based needs, financial restrictions, features and functions of an IAT). Even if more IAT were ready for real-world testing, little is known about the needs of the users of such systems beyond their task-based needs [9, 27]. In particular, the place of residence of the majority of all Canadians living with dementia, has trended from institutional care to the home and is expected to continue [2]. Accordingly, the needs of IAT users now include those inherent to the task-based challenges they face as well as the features and functions of an IAT that would be integrated into a home environment [28].
A review conducted by Topo [27] highlighted the limited focus AT developers place on home settings. Topo’s review included sixty-six (66) assessment or intervention studies focusing on technology to support people with dementia and their caregivers; ten (10) studies included in the analysis had the primary intention of supporting the independence of people with dementia and their caregivers in the home. Additionally, five (5) studies included the person with dementia as an active user of the AT, while the other studies were focused on other users (e.g., formal or family caregiver). Topo’s review revealed that need assessments almost always prioritized the needs of caregivers, seldom including the needs of people with dementia who are often primary users of the interventions. Furthermore, the review also identified that limited information was provided about the environment ATs were developed to be used within [27]. This was despite the known impact environment-related factors have on the perceived usefulness of ATs [29]. The limited knowledge of user needs – particularly with respect to the features and functions of an in-home AT – including the broad needs of users in all stages of AT development is increasingly emphasized in the work of AT developers [8, 27], yet remains under-examined; representing a gap in knowledge and literature.

The restricted scope of existing needs assessments and the small number of real-world trials highlight three significant limitations obstructing the development of functional IATs designed to support people with dementia through ADL in a home environment. First, the complete needs of older adults with dementia and their caregivers for IATs in the home environment remain unknown. Second, without real-world testing it is not possible to assess whether IATs are satisfying the actual needs of the end users in the target environment. Third, without knowing how effectively an IAT satisfies user needs, it is not possible to identify areas of potential improvements for the IAT beyond anecdotal evidence.

Notwithstanding the limitations of existing IAT developments, IATs exhibit significant potential as tools to provide support for an aging global population [8, 30]. IATs have “the potential to improve the quality-of-care and quality-of-life of [people with dementia] and their family caregivers” [8]. This potential support is particularly relevant as support for people with dementia increasingly becomes the responsibility of family caregivers [31]. In order for IATs to achieve this potential, IAT designers must understand and incorporate the task-based and functional needs of older adults with dementia and their caregivers, and evaluate the effectiveness of the developed devices [28]. To ensure user needs are reflected in the design of
IATs, a user-centered design (UCD) philosophy [32] has emerged among IAT developers [e.g., 33, 34, 35] that is intended to increase the likelihood of device acceptance and adoption among users [8]. User needs form the foundation of the UCD philosophy and necessarily include the characteristics of the users, their environments, and the task needs motivating the design of an IAT [8, 32]. These needs are integrated into an iterative cycle of development and evaluation in a controlled environment, ultimately resulting in a prototype device suitable for real-world evaluation [32]. This approach arguably increases the overhead associated with the development of a device, but ultimately improves the likelihood of user adoption and acceptance [8, 36] and has direct application to the development of IATs for older adults with dementia.

1.2 The Current COACH Automated Prompting System

The Cognitive Orthosis for Assisting with aCtivites in the Home (COACH) is a prototype IAT designed to promote independence in older adults with dementia during ADL completion (e.g., hand washing) in a home environment [11, 37-47]. The COACH system, conceptually, is composed of three functionally independent modules: hand tracking, planning, and prompting [40]. Hand tracking processes images captured using an overhead camera mounted above the sink in a washroom. The hand positions, located in each captured video frame, are used to identify the completion of hand washing actions [45, 46] – interactions between the hands and environmental objects or regions. The action observations are passed to the planning module which is responsible for translating the actions into hand washing task steps. Here, the system is able to monitor the user’s progression through the hand washing task. If necessary, the planning module can indicate that a user requires assistance when task progression has ceased.

Furthermore, the planning module continuously updates an estimate of the user’s current lucidity and responsiveness to assistance. If directed by the planning module, the prompting module will provide one of three levels of assistance: an audible prompt with general instructions (e.g., can you turn on the water), an audible cue with specific instructions (try turning the silver faucets), or a combined audiovisual prompt (e.g., a cue along with a visual demonstration) [11]. The appropriate level of assistance is determined by an estimate of current user’s lucidity and responsiveness provided by the planning module.
1.3 The History of the COACH

Development of the COACH has progressed through three major revisions [see 11 for a short summary]. The first prototype of COACH [48] employed a pattern-matching algorithm in conjunction with a marked bracelet worn by a system user to identify the positions of each hand. A probabilistic neural network classified the filtered Cartesian coordinates of the hand positions into task step estimates (e.g. hands at taps), which were disambiguated (e.g., turning on the water versus turning off the water) by a logic module that accounted for step prerequisites. The task plan (sequence of task steps) the user was completing was predicted (if necessary) by a plan recognition algorithm that would also initiate a verbal cue for each task step if the user had deviated from all possible plans. Using the first COACH prototype, an efficacy study including ten (10) participants over sixty (60) days concluded that the COACH increased the independent completion of hand washing task steps by an average of 25% [37].

Encouraged by these results, a second prototype was developed to address two potential limitations inherent in the first COACH prototype; that (1) marker-based tracking was obtrusive and (2) uncertainty existed in the effects of the actions of the planning module. A new markerless, colour-based hand tracker was then developed to address this first limitation. The hand tracker identified skin-coloured pixels in each image from the foreground image (i.e., regions of interest in each image that are not part of the background scene) [42]. A Markov decision process (MDP) was developed to address the second limitation, replace the original planning module [40]. An MDP, modeling the hand washing task, is able to account for situations where a user completes a task step that is different than the current step in the task plan (e.g., a user turns on the water when he/she was prompted to get the soap). Additionally, the MDP can effectively balance multiple potentially-conflicting objectives, such as promoting user independence by not intervening while encouraging task completion through the provision of a prompt when necessary [40].

Development of the third (current) COACH prototype focused on further hand tracking and planning improvements, along with an improved understanding of factors affecting communication between the COACH and its users. The colour-based hand tracker developed in the previous prototype was improved by connecting the skin-coloured pixels using a flock of colour features which effectively reduced the influence of occlusions, shape variations and
potential distracters (other skin coloured objects) [45, 46]. Similarly, the MDP planning module of the previous prototype was re-implemented by using a partially observable Markov decision process (POMDP) [40, 45, 49]. A POMDP is a decision-theoretic (i.e., planning under uncertainty) framework suitable for real-world sequential processes [39]. The POMDP effectively models the hand washing task while incorporating the various sources of uncertainty inherent in the COACH environment (e.g., the uncertainty in the hand tracker observations, the effects of any actions the COACH may take) [39]. The POMDP, in addition, allows for regression in the completion of the hand washing task steps (i.e., turning the water off after turning it on and wetting hands, but before washing and rinsing) and, thus, more accurate modeling of the various ways user’s may complete a task [39]. The planning module attempts to maximize user independence (thereby achieving a maximum reward) by passively monitoring user performance [39, 40]. Additionally, the prompting module was extended to provide three different levels of prompting [43, 47] based on an estimate (provided by the planning module) of the system user’s degree of difficulty with each step. Controlled clinical trials were performed with the current prototype in a supervised hospital setting including six (6) participants with moderate to severe dementia [11]. Participant interactions with formal caregivers decreased by 60% compared to baseline performance and independent step completion increased by 11%.

The aforementioned stages of development suggest that older adults with dementia complete the hand washing task more independently with assistance from the COACH in a clinical, supervised setting (i.e., a member of the research team is present). Additionally, the COACH reduces the burden of care experienced by caregivers supporting the care recipients in the same environment through a reduction in the number of interactions between caregivers and care recipients during the task. However, while these findings support a clinically relevant impact on the hand washing task (e.g., increased number of independent task steps completed), little research has been published exploring the technical performance of the COACH during these deployments. Specifically, the capability of the COACH to correctly infer task step progression is unknown (i.e., how well the system correctly infers that task steps completed by users are complete, while steps not completed by users are identified as incomplete). In the previous supervised COACH trials the system was enabled at the beginning of the trials and disabled at the end, potentially interfering with future step inference and inhibiting future prompts or system interactions. Furthermore, the performance of the system was evaluated in terms of the ability of the system’s
users to complete the task (with or without COACH) and on the prompts provided (or not provided) during the trials. The internal state of the COACH was largely unexplored during these trials. Additionally, since the COACH’s inception, little work has been done to verify the task-based needs of older adults with dementia in a real-world deployment. Furthermore, the features and functions of an IAT designed to be integrated into the homes of older adults with dementia and their family caregivers are unknown.

1.4 Dissertation Objectives

This dissertation looks to build upon the results of three iterations of development and supervised clinical trials with the COACH system. The primary objective (PO) is to continue development of the COACH toward an unsupervised, real-world in-home deployment. To accomplish this objective I employ an informed, empirical approach, determining the direction of development of the COACH through the analysis of data obtained by identifying the needs of the users of the system, and the technical factors limiting a real-world COACH deployment. Accordingly, the secondary objectives are: (SO1) to identify the needs older adults with dementia and their family caregivers have for IAT designed to support the completion of ADL in a home environment; and (S02) to identify factors limiting the technical performance of the COACH in an unsupervised, real-world deployment in a clinical setting.

1.4.1 Research Questions

To satisfy the primary and secondary objectives of this research, this dissertation seeks to answer the following research questions:

SO1: to identify the needs older adults with dementia and their family caregivers have for IAT designed to support the completion of ADL in a home environment

R1. *What activities of daily living are most challenging for older adults with dementia and their family caregivers in a home environment?*

R2. *What features and functions are required for the successful integration of an intelligent assistive technology that is intended to support the completion of daily activities by older adults with dementia and their family caregivers in a home environment?*
SO2: to identify factors limiting the technical performance of the COACH in an unsupervised, real-world deployment in a clinical setting

**R3.** How effectively does the COACH infer the progression of older adults with dementia through the task of hand washing in an unsupervised, real-world clinical deployment?

**R4.** What technical factors explain occurrences where COACH failed to infer the progression of older adults with dementia through a hand washing task in an unsupervised, real-world clinical deployment?

PO: to continue development of the COACH toward an unsupervised, real-world in-home deployment

**R5.** What technical design criteria, used to develop an improved COACH prototype, can be identified by synthesizing the needs of older adults with dementia and their family caregivers for assistive technologies in a home environment with the capabilities of the COACH in an unsupervised, real-world deployment?

**R6.** How effectively do(es) the COACH module(s) developed satisfy the most relevant technical design criteria identified in R5 perform in a real-world setting compared to the module(s) implemented in the current prototype?

### 1.5 Dissertation Overview

This dissertation is composed of three methodologically diverse yet thematically related studies. Each study is presented as a chapter in the dissertation, and is composed as an independent publication. The first chapter, published in *Gerontechnology*, identifies the needs of older adults with dementia and their family caregivers, addressing research questions one (R1) and two (R2). The second chapter, published in the *Journal of Ambient Intelligence and Smart Environments, Thematic Issue on Designing and Deploying Intelligent Environments*, evaluates the efficacy of the COACH in an unsupervised, real-world clinical deployment. Furthermore, this chapter identifies the factors influencing the overall performance of the system in this environment, addressing research questions three (R3) and four (R4). The third chapter, under review in the *Journal of Biomedical and Health Informatics*, begins with the identification of technical design criteria for a new prototype system. The paper then outlines the methodology used to address the
design criteria, and concludes with a thorough evaluation of the performance capabilities of the new prototype in a real-world deployment, answering research questions five (R5) and six (R6).

1.5.1 Chapter 2: The Design of Intelligent in-home Assistive Technologies: Assessing the Needs of Older Adults with Dementia and Their Caregivers

Participation in ADL, undoubtedly, is instrumental in maintaining individual health and happiness. This is particularly true for older adults with dementia, who report that independent completion of ADL is critical to their quality of life [50]. The onset of dementia, however, significantly affects an older adult’s ability to independently complete ADL [10]. Family caregivers often assume the responsibility of supporting the continued participation of an older adult with dementia in ADL amounting to an average of 3.6 hours of support per day [3]. Unfortunately, despite the importance older adults with dementia report they place on independent ADL completion [51], generalizable results identifying which ADL are most challenging to complete, particularly in a home environment, are unavailable [27]. ADL needs assessments are also almost entirely restricted to task-based needs, rarely focusing on the needs associated with an IAT integrated into a home environment [27]. Regardless, several validated questionnaires have been developed to identify the daily task needs of people with cognitive disabilities in general (including older adults with dementia) [e.g., 52, 53, see 54 for a review of ADL scales]. Unfortunately, none has considered the needs of older adults with dementia living in a home environment within the scope of the IATs designed to support those needs.

In this manuscript, I address the paucity of knowledge and literature reflecting the IAT needs of older adults with dementia and their family caregivers in a home environment; with a specific focus on ADL. The primary contribution in this study, achieved through the development and validation of a questionnaire specifically designed to elicit the opinions of family caregivers, is an understanding of: the ADL older adults with dementia struggle to complete at home; the ADL family caregivers struggle to support at home; the role of IATs in supporting ADL at home; and the required features and functions of an in-home IAT. The 94-item exploratory questionnaire was constructed based on prevalent themes drawn from the literature [33, 55, 56], and by adapting existing validated scales relevant to the study [e.g., 52, 53]. A total of 106 questionnaires were fully completed by respondents who voluntarily participated in the study. Afterwards to follow-up on the knowledge generated, a focus group was conducted with six (6)...
family caregivers of older adults with dementia living in the Greater Toronto Area, Canada to clarify any ambiguities in the survey data and to provide content validity to the results.

The findings of this study answer the first two research questions of the dissertation: “What activities of daily living are most challenging for older adults with dementia and their family caregivers in a home environment?” and “What features and functions are required for the successful integration of an intelligent assistive technology that is intended to support the completion of daily activities of older adults with dementia and their family caregivers in a home environment?”.

Recognizing that a key finding in Topo’s [27] review of technology studies, designed to meet the needs of people with dementia and their caregivers, was that most studies focused on the needs of the caregivers rather than the needs of people with dementia I sought to elicit responses toward filling this literary lacuna. Thus, despite family caregivers being surveyed, questionnaire items relevant to the care recipients were specifically constructed or adapted from existing scales to properly capture the views of older adults with dementia. The responses were then statistically validated through the process of identifying dominant themes, reviewed by the research team to ensure face validity, and discussed in the follow-up focus group to ensure content validity. Accordingly, this study contributes a validated assessment of the needs of older adults with dementia and their family caregivers for IATs integrated in a home environment.

1.5.2 Chapter 3: A Real-world Deployment of the COACH Prompting System

Despite the substantial number of IATs designed to support the challenges experienced by older adults with dementia and their caregivers, only a small number have undergone clinical trials [8], precluding real-world testing of these systems. Without real-world testing it is difficult to identify the challenges restricting the ultimate acceptance and adoption of these IATs by the intended users. In this paper I describe the deployment of the COACH in an unsupervised real-world, clinical environment. The study was designed to identify technical challenges the system experiences in the real world clinical environment; challenges that once overcome will guide the development of a new COACH prototype. Specifically, the study empirically evaluated the performance of the hand tracking, policy and prompting modules during forty-one (41) hand washing trials completed by 27 participants over a four month period.
This unsupervised deployment of the COACH was, at the time of the study, the first unsupervised real world trial of an IAT designed to support older adults with dementia through an ADL. Accordingly, a methodology for evaluating an IAT in this environment did not exist. However, the COACH system had been tested in several supervised clinical deployments [11, 37, 41]. The methodologies of the supervised clinical deployments provided the foundation for developing the methods used to evaluate the performance of the COACH in this unsupervised deployment. The COACH system’s technical performance was assessed in three stages: 1) the system’s ability to correctly identify individual task steps as each step in the task was completed by a participant (e.g., get soap, turn on water); 2) the system’s ability to correctly identify participants’ overall task performance (i.e., correctly infer the task steps completed by participants); and 3) the system’s ability to provide appropriate assistance through prompts to aid the participant in task completion as and when required. This approach allowed the system’s efficacy to be empirically evaluated in an unsupervised, real-world clinical environment. It further allowed the evaluation of the hand tracking, policy and prompting modules to be evaluated independently. Thus, any occurrence where COACH failed to correctly identify the participant’s progression through the hand washing task could be attributed to a specific module and explored to provide insight into why the module failed.

The findings of this study answer the third and fourth research questions of my dissertation, “How effectively does the COACH track the progression of older adults with dementia through the task of hand washing in an unsupervised, real-world clinical deployment?” and “What technical factors explain any occurrences where COACH failed to track the progression of older adults with dementia through an unsupervised, real-world clinical deployment?”. These findings, moreover, provide the basis for continued development of the COACH system toward an unsupervised, in-home deployment by identifying technical design criteria based on user needs and system capabilities and improving the system to meet these criteria.

1.5.3 Chapter 4: Depth Image Hand Tracking from an Overhead Perspective Using Partially Labeled, Unbalanced Data: Development and Real-world Testing

Technical design criteria for development of a new COACH prototype were determined using the House of Quality (HOQ) approach [57, 58]. The HOQ is a systematic methodology used to convert customer needs and system capabilities into design specifications. The needs of older
adults with dementia and their family caregivers were determined through the needs assessment questionnaire (see Chapter 2: The design of intelligent in-home assistive technologies: Assessing the needs of older adults with dementia and their caregivers). The current capabilities of the COACH were determined through real-world trials (Chapter 3: A real-world deployment of the COACH prompting system). A prioritized list of technical design criteria for development was then determined by weighing COACH user needs against the capabilities of the system in an unsupervised, real-world environment.

Two HOQ matrices were constructed. The first was designed to determine required changes for the three major COACH modules: (1) hand tracking, (2) policy, and (3) prompting. The second HOQ matrix was designed to identify features and functions deemed essential to the successful installation and use of COACH in a home environment. Required changes were grouped into three categories: (1) user interface, (2) physical appearance, and (3) cost. These HOQ matrices allow me to answer my fifth research question, “What technical design criteria, used to develop an improved COACH prototype, can be identified by synthesizing the needs of older adults with dementia and their family caregivers for assistive technologies in a home environment with the capabilities of the COACH in an unsupervised, real-world deployment?”.

The HOQ analyses identified that the overall performance of the COACH in an unsupervised, real-world environment would be improved by developing a new hand tracker that is lighting and colour invariant and utilizes depth imaging to develop a skeleton model of the users, improving detection of user-object interactions. In this paper, I describe the development and testing of a tracking algorithm that is lighting, shape, texture and colour invariant. Through this approach the tracker is capable of classifying at least 30 different body parts or joints [see 59] from a single depth image. These body parts can subsequently be connected through a kinematic model to provide a skeleton representation of the users. The tracker utilizes a depth camera to provide uninitialized (i.e., frame-by-frame) body part tracking locations from an overhead perspective through two stages. First, an intermediate classification is performed for each of a set of random pixels sampled across the depth image. The intermediate classification provides a probability density function for each sampled pixel (i.e., the likelihood each pixel will be each body part). Final part positions are then proposed using a local mode find approach that aggregates the information contained in the underlying probability density function provided by the intermediate classifier. Using the final hand positions, the users’ hand washing task actions are
then determined based on proximity to hand washing objects in the environment (e.g., getting soap, turning on the water, washing hands).

The tracker was implemented with three random decision trees used together in an ensemble – a single random decision forest classifier [60]. Trees were trained on a manually labeled and partially labeled data set of up to 5,000 real-world hand-washing trial images captured in the HomeLab [61] at the Toronto Rehabilitation Institute [62] in Toronto, Canada. Training parameters were first optimized by iteratively developing random decision forests by independently varying several critical parameters (e.g., tree depth, number of training images). The performance of each forest was evaluated for both the intermediate classification and aggregate part position proposals using a holdout set of 854 manually labeled ground-truth images to identify optimal training parameters. The optimal training parameters were then used to train the decision forest. The intermediate multiclass classification accuracy of the final decision forest was 78.34%, meaning that the most likely body part within the probability density functions (PDF) assigned to each sampled pixel was the same as the ground-truth part 78.34% of the time. Similarly, the performance of the aggregate part position proposals was evaluated, resulting in a mean average precision of 0.846, where a score of zero (0) indicates the worst possible performance and a score of one (1) represents perfect performance. Confident in the results of the classifier performance metrics, the new part tracker was integrated into the COACH and was validated using a manually labeled set of approximately 24,000 real-world hand washing images captured over forty-one (41) trials. The new depth-based hand tracker correctly identified 192 of 205 steps, answering my sixth research question, “How effectively does the COACH module(s) developed to satisfy the most relevant technical design criteria identified in research question 5 perform in a real-world setting compared to the module(s) implemented in the current prototype?”.

Thus, through my needs assessment study I have identified that daily tasks involving an invasion of privacy or tasks that are typically completed independently are most relevant to IATs designed to support ADL completion in a home environment. Furthermore, through my study designed to identify the performance capabilities of the COACH in an unsupervised, real-world clinical deployment I determined that the COACH could not effectively track the hands of users. By synthesizing the needs of IAT users with the capabilities of the COACH in a real-world clinical environment, I developed technical design criteria outlining the improvements required to
overcome the hand tracking deficiencies. My final study contributed the successful development of a new depth-based hand tracker, overcoming the tracking challenges experienced in my real-world clinical deployment and satisfying the design criteria. Now, with this new hand tracker we are able to test the performance of the COACH in an unsupervised, real-world deployment anew, ultimately toward a real-world, in home deployment.

My diverse studies contribute to theoretical knowledge in areas investigating the IAT needs of vulnerable populations in a real-world home environment, as well as effective performance testing methodologies for IAT in such settings. I do so by providing some of the first data in each field respectively. Furthermore, my work extends the existing theoretical knowledge in computer vision-based part classification from single overhead depth images by successfully implementing a part tracker using partially labeled data captured from an overhead perspective. My work empirically contributes a validated scale designed to explore the IAT needs of older adults with dementia and their family caregivers living in a home environment. I contribute empirical data captured from one of the first, and largest, real-world clinical deployments of an IAT designed to support older adults with dementia, and the largest experimental, real-world data set used to develop and evaluate a body part tracker using single depth images. Finally, this work is strongly interdisciplinary in its scope and reach, with the ability to impact practices in engineering, computer science, social science, gerontology and rehabilitation, but solidly situated in the discipline of biomedical engineering. Thus, my primary contributions to the field of biomedical engineering are: 1) the evaluation of the performance of the COACH in an unsupervised, real-world clinical deployment; 2) the assessment of the needs of COACH users for an IAT in this environment; and 3) the development of a new COACH prototype designed to overcome challenges identified in the real-world clinical deployment, informed by the needs of COACH users.
Chapter 2

2 The Design of Intelligent in-home Assistive Technologies: Assessing the Needs of Older Adults with Dementia and Their Caregivers

2.1 Publication Citation


2.2 Abstract

Objective: To determine the needs of older adults with dementia and their family caregivers during Activities of Daily Living (ADL), and the role of intelligent assistive technology (IAT) in supporting these needs.

Methods: An 85 item questionnaire was administered to family caregivers of older adults with dementia exploring: (i) challenging ADL for an older adult with dementia to complete independently, (ii) difficult ADL for a caregiver to assist, (iii) the role of IAT supporting ADL completion, and (iv) the features and functions of in-home IAT designed to support ADL.

Results: Respondents (n=106) indicated the person they care for has partial ability to complete ADL, that private tasks (e.g., showering) are difficult to assist, and that IAT designed to support ADL must be autonomous, familiar, simple and unobtrusive. Respondents also showed little knowledge of existing IAT that support ADL.

Conclusions: Designers of IAT should focus on supporting caregivers and older adults with dementia in the completion of private and personal ADL.

Keywords: needs assessment, family caregiver needs, Activities of Daily Living
2.3 Introduction

The world’s population is aging[63] with the fastest growing segment being those aged 80 or over. Age is one of the leading risk factors associated with Alzheimer’s disease – a form of dementia[64]. Thus, as the world’s population ages the number of cases of Alzheimer’s disease will also increase. It is estimated that 4.3 million people in North America and 35.6 million people worldwide had Alzheimer’s disease in 2010[4]. By 2050, researchers estimate 11 million people in North American and over 115 million people worldwide will be living with Alzheimer’s Disease[4]. Based on these estimates, the health care costs associated with dementia will soon become unsustainable and the need for informal care will increase considerably. In 2010, the cost of dementia globally was estimated at $604 billion[1]. Based on predicted increases in prevalence rates alone, by 2030 costs are expected to increase by 85%[1]. However, there remains a need to preserve the independence, autonomy and quality of life of older adults with dementia, and to relieve the burden of care experienced by informal caregivers by addressing the functional limitations of older adults with dementia, supporting the current health care system without sacrificing quality of care, and maintain the standard of living of both the person with dementia and the caregiver. One potential solution is to support older adults with dementia and their caregivers with technology, though the actual use of technology tends to shift from the older adult to the informal caregiver over time[65], reducing the utility of the device. However, a substantial amount of interest has recently been directed toward the development of intelligent technologies that support aging in place – devices that support cognitive deficits while reducing caregiver burden[8] – increasing the likelihood of their use.

2.3.1 Supporting Cognitive Deficits

Emerging technologies have been designed to support individuals with decreased cognitive abilities resulting from dementia. Such technologies range in function from low technology aids (e.g., medication pill organizers, schedules and notes) to higher technology aids (e.g., intelligent assistive devices that are contextually aware and can provide help when appropriate)[8]. Advances in computer hardware and software, particularly in the domains of computer science and engineering have improved the reliability, affordability and capability of intelligent devices. Accordingly, intelligent devices are increasingly investigated by researchers as potential tools to assist people with dementia who struggle with a range of cognitive disabilities[8].
Intelligent assistive technologies (IAT) have been created to compensate for the loss of memory[13, 66-70]; and executive function[67-69, 71], and show potential for application to older adults with dementia. Intelligent AT such as the Autominder[13, 66] and the Memory Glasses[70] are context-aware memory aids because they provide timed reminders to users as determined necessary by the systems. These systems employ artificial intelligence to detect when a user may have forgotten a required activity and if a reminder should be issued. In addition to time-based reminders, the Memory Glasses also aid in memory recall deficits (e.g. amnesia, Alzheimer’s disease, agnosia or prosopagnosia) that cannot always be corrected with simple scheduling. The ISAAC system[67] acts as a problem solving aid during scheduled reminders through the use of a sequential checklist that outlines the steps required to successfully complete an activity. Systems such as the PEAT[68] and Essential Steps[71] go beyond contextually sensitive memory aids by providing planning and organizing support[68, 71]. Recently, the PEAT was further extended[69] to include sensor data from global positioning systems (GPS), radio frequency identification (RFID) tags and pressure mats to reduce the amount of explicit, user-initiated feedback required by the original system and to allow the system to infer user activities.

Intelligent assistive devices have also been used to support independent navigation[15, 21, 72]. Commercially available positioning and navigation systems (e.g., GPS) provide the same type of directions (e.g., small visual display, audio prompts) regardless of the users capabilities, and typically cannot provide the specific assistance required by a person with dementia. Systems such as the navigation tools developed as part of the Assisted Cognition project[15, 72] provide GPS functionality to cognitively impaired users. Customized user interfaces and artificially intelligent decision making enable people with cognitive challenges to navigate outdoors without requiring implicit user input. Other devices such as the Robotic Walker[21] and intelligent anti-collision wheelchair[73] provide indoor navigation assistance, automatically determining the user’s position and alerting the user about obstacles or hazards.

2.3.2 Activities of Daily Living

Arguably one of the most desirable applications of intelligent assistive technology is as a tool to support an older adult with dementia through activities of daily living (ADL), particularly because support for these tasks is often provided by informal, unpaid care[65]. Logsdon,
McCurry & Teri[50] summarized several studies where people with dementia reported the ability to independently complete ADL significantly affected his or her quality of life. Wimo, Winblad & Jonsson[74] found that caregivers provide care by providing supervision and supporting ADL. They separated ADL support further into two categories: personal activities of daily living (PADL) or fundamental tasks, and instrumental activities of daily living (IADL) or more advanced daily tasks. A comprehensive review of 27 studies on caregiver burden revealed that in high income countries caregivers of people with dementia spent an average of 3.7 hours per day supporting ADL (1.6 hours for PADL and 2.1 hours for IADL)[74]. The effect of this demand on caregivers is increasingly evident in the costs associated with the provision of informal care[65] and through reports of caregiver strain (e.g., physical illness, psychological problems and depression)[4].

Accordingly, the development of IAT for recognizing and supporting ADL has become a central research area with the goal of maintaining a person’s ability to independently complete ADL as well as reducing the burden experienced by his or her caregiver. An example of such a system is the COACH[11, 49], an intelligent AT designed to help a person with dementia progress through the task of handwashing without assistance from a caregiver. An overhead camera unobtrusively tracks the positions of a user’s hands as well as environmental objects such as the towel. The system uses artificial intelligence to determine whether an intervention is required based on the user’s progress through the task, taking into account uncertainty in the observations provided by the camera as well as estimations of the user’s cognitive state. Interventions are provided through a combination of audio and/or video prompts that vary in specificity according to the user’s responsiveness to the system. A second intelligent AT, a type of technology assistant[75] was designed to support older adults using a blood glucose meter. This device uses a camera to track the user’s hands, blood glucose meter, testing strips and liquid bottle. The camera information is interpreted as progress through the explicitly coded task, and feedback is provided if the user makes any mistakes. A third system, the Ambient Kitchen[76] employs an array of physical inputs (e.g., RFID tags, accelerometers and pressure sensors) and six integrated cameras to detect kitchen activity and provide assistance as needed. Both audible and visual cues are provided throughout the kitchen using speakers and video projectors. The Ambient Kitchen provides an extensive test bed for developing and testing various intelligent AT.
2.3.3 User-Centred Design

Despite the wide range of IAT in development to support people with dementia or their caregivers, few of these intelligent AT have progressed beyond the initial development stage (e.g., a recent review of 58 technologies highlighted only three devices that have undergone clinical trials and none had entered real-world testing)[8]. As such, the appropriateness of many of these devices as tools to support the needs of people with dementia and their caregivers remains in question. To facilitate the process of developing usable IAT, we need to understand the needs of cognitively disabled users[77] – fundamental if the end goal is to develop tools that support older adults with dementia[78].

Recently a small number of studies have looked specifically at IAT as tools to support ADL. Wherton & Monk[55] conducted two small-sample interview studies investigating which daily activities in the home were most important to caregivers and the person they care for. The first study investigated opinions of professional caregivers with semi-structured interviews (n=9) and a follow-up focus group (n=20); the second study interviewed caregiver–person with dementia pairs (n=9). A grounded qualitative analysis identified the category ‘daily activities’ within the core theme ‘Problems in the Home’ which found that the ADL most in need of support were dressing, taking medication, personal hygiene, food and drink tasks, and toileting. In a cross-sectional post-analysis of a Medicare beneficiary survey in the United States, Dudgeon, et al.[56] found that people with dementia needed support with heavy housework, walking, shopping and money management while caregivers desired help assisting with walking, bathing and toileting. The varying results of these studies suggest that more information is required to generalize these needs into a foundation for the development of intelligent AT for older adults with dementia.

Determining the needs of technology users for developing technological products has a well-established history in industry. Industrial tools such as Quality Function Deployment (QFD)[57], the stage-gate model[79] and Design for Six Sigma[80] are developed principles promoting solutions driven by end users of a product or process. In the development of IAT an analogous approach (though much less developed or systematic) called User Centred Design (UCD), has recently emerged[33-35]. The UCD approach involves actively identifying the needs of users, which are then synthesized into technical design criteria using tools such as QFD and used to develop a functional, testable prototype. The efficacy of the prototype is later tested in a real-
world environment, with the results iteratively cycled back into the design criteria and development stages. The main difference between the UCD process and more typical design philosophies used by IAT developers is that the end users are involved in all stages of the process, not just during the needs assessment stage.

Recent work has seen the first iterations of a UCD approach toward the successful development and deployment of IAT to support people with dementia. The Keep In Touch Everyday project conducted a needs assessment of people with dementia and their caregivers, developed two prototype devices to help increase independence for people with dementia, and conducted two simple trials with participants of the UCD[35]. Kinney, et al.[81] also examined how IAT can assist caregivers of people with dementia, finding safety was their primary concern followed by the person with dementia maintaining their pre-illness lifestyle. Based on these two needs a commercially available internet-based monitoring system was installed in a trial home and tested by the participants with generally positive evaluations. The Bath Institute of Medical Engineering recently completed an entire iteration of the UCD project to design and evaluate entire smart-home systems for people with dementia[82]. The study identified safety, task guidance (through prompting) for cooking, toileting and bathing, and social connectivity for family and friends as most desired. Familiarity, caregiver emulation and user control were also identified as critical for successful adaptation. Several device prototypes were developed and tested, revealing that device usability and robustness, as well as careful user-interface considerations were critical to device acceptance. Perhaps the most substantial project to employ a UCD approach is the COGKNOW project[33] which, through focus groups and workshops, found that caregivers and people with dementia desired IAT that reinforced memory, socialization, ADL support and safety. Emphasizing the development of a commercial product, the study followed a UCD approach to produce a complete system to support these needs. The system was installed and tested in the homes of 16 people with dementia, highlighting several successes and failures that will be integrated into additional UCD cycles. However, these projects are few in number and provide data from a small number of participants making it difficult to generalize the needs into something useable by intelligent AT developers.
2.3.4 The Current Study

The market for technology to support older adults with dementia is rapidly growing, largely because the prevalence of dementia is rising dramatically while the ratio of caregivers to older adults with dementia is becoming less favorable. Yet technology developers have little data beyond preliminary user needs assessments to identify the technology needs of older adults with dementia, let alone to determine the efficacy of developed devices. This research presents the results of a user needs assessment – the first stage of a user-centered design process looking to extend the COACH beyond clinical trials into an intelligent AT that works in the homes of older adults with dementia. Toward this end, the study contributes to the extension of the COACH and to the development of other intelligent AT by identifying the needs of older adults with dementia and their caregivers during ADL completion and determining valued features and functions of in-home IAT. In doing so, this paper addresses four key research questions:

What ADL do people with dementia struggle with while trying to complete independently?

What ADL do informal caregivers of people with dementia struggle with to support?

Can intelligent assistive technology play a role in supporting ADL completion?

What features and functions are required for an in-home intelligent assistive technology in order to facilitate its acceptance?

2.4 Methods

2.4.1 Purpose

A pilot questionnaire was designed to explore which ADL are challenging for an older adult with dementia to complete independently, which ADL are difficult for a caregiver to assist, the role of intelligent AT as a tool to support ADL completion, and the features and functions of an in-home IAT designed to support the completion of ADL. The study sought to elicit views of family caregivers, defined as any person providing care without financial compensation, as well as to identify the needs of older adults with dementia from the perspective of the caregiver.
2.4.2 Participants

Participants were recruited in five ways. First, information about the study and online questionnaire was posted on internet message and discussion boards serving caregivers of older adults with dementia (e.g., American and Canadian Alzheimer Associations). Second, emails were sent out through caregiver email lists. Third, caregivers were contacted through formal advocacy and care organizations located in the Greater Toronto Area (Ontario, Canada). Fourth, ads were posted in caregiver resources such as newspapers and newsletters for caregivers. Lastly, respondents were informed of the study via word-of-mouth from respondents contacted through the previous methods. Respondents were included in the study only if they were currently a primary, informal caregiver of a person with dementia.

2.4.3 Materials

The exploratory 94 item online questionnaire was constructed, including demographic questions, based on our research questions [83]. ADL were explored through 54 items constructed based on the results obtained in other similar studies[33, 55, 56] in three categories (18 items in each): Independent ADL completion by a person with dementia; Caregivers assisting with completion of ADL; and Supporting ADL with intelligent Assistive Technology. In addition, features and functions of IAT as tools in the home were explored through 24 items constructed using face validity as other studies or scales to measure this don’t exist. In this category multiple items were constructed to facilitate post-analysis validity and reliability tests[84]. Four additional items (e.g., ‘What is your relationship to the person you are caring for?’) identified the respondent’s relationship to the older adult with dementia. The questionnaire was initially reviewed by an expert in the field of intelligent AT designed to support ADL completion, validated by members of the research team, piloted by students in an academic research lab and then piloted by three caregivers of older adults with dementia.

Caregiver’s opinions about their own role in daily tasks as well as the abilities of the person they were caring for were considered in three separate sections: Independent ADL completion by a person with dementia; Caregivers assisting with completion of ADL; and Supporting ADL with intelligent Assistive Technology. Within each section 18 close-ended items presented daily tasks within common PADL and IADL categories[53, 85]. For the first section respondents were asked: ‘Typically, how easily can the person you’re caring for complete the following tasks’.
each of the 18 items respondents were asked to choose the response that best reflected their belief on a 5 point Likert-like scale ranging from ‘Cannot complete at all on his/her own’ to ‘Can easily complete on his/her own’.

For the second section respondents were asked: ‘Typically, how hard is it to assist the person you’re caring for to complete the following tasks’. For each of the 18 items respondents were asked to choose the response that best reflected their belief on a 5 point Likert-like scale ranging from ‘Very hard to help’ to ‘Very easy to help’. For the final section respondents were presented with the statement: ‘Assuming technology exists that can help you care for a person with dementia, for the following list of common tasks please circle the answer that best describes your opinion about the technology supporting your care efforts’. For each of the 18 items respondents were asked to choose the response that best reflected their belief on a 5 point scale ranging from ‘Technology can’t help me support this’ to ‘Technology can help me support this’.

Caregiver’s opinions about the features and functions of an in-home IAT were assessed through 24 close ended items. Items were constructed within three conceptual categories: physical attributes, functionality, and device cost. For each item respondents were asked to select the response that best represented their belief from the possible responses of ‘Strongly Disagree’, ‘Disagree’, ‘Neutral’, ‘Agree’ and ‘Strongly Agree’.

2.4.4 Procedure

Ethics approval was granted for the questionnaire by University of Toronto (REB #24637). The questionnaire was posted online on December 14, 2009, and paper copies were distributed on April 30, 2010. Informed consent was required for online respondents in order to continue to the survey, or collected in person before completion of the questionnaire.

2.4.5 Data Analysis

The questionnaire data were analyzed using an exploratory factor analysis. Factor analysis is a statistical process that reduces a large number of observed independent variables (survey items) into a smaller number of inferred hypothetical variables called factors[86]. Exploratory factor analysis first determines underlying factors represented in the questionnaire items by assuming that each item may be related to each factor. Each factor is reported as explaining a percentage of the total variance in the data. In other words, each factor represents an underlying quality or
belief that may not be intuitively obvious in the original items. Factors were extracted using Principal Axis Factoring, a technique recommended for exploratory analyses[87]. Following factor extraction rotation was used to simplify the data structure revealing a more clear relationship between each item and the factors. For each factor analysis oblique (Promax) and orthogonal (Varimax) rotation methods were compared. Factors were described by the rotated factor matrix, which shows the factor loadings of each item on each factor. Factors were retained if they satisfied Kaiser’s criterion (eigenvalue > 1); were above the elbow of the scree plot; and had at least three items[88].

Items were associated with a factor if the factor loading was greater than the critical value (CV) of 0.505, calculated using the formula $CV = \frac{5.152}{\sqrt{(n-2)}}$ and a sample size of $n = 106$ [89]. Items were deleted from analysis if they loaded below the critical value or loaded above the critical value on more than one factor after rotation. The reliability of the resulting factors was measured for internal consistency using Cronbach’s $\alpha$ [90]. Internal consistency in this context measures whether different scale items assigned to a factor provide the same results.

### 2.5 Results

One hundred and six (106) respondents including professionals, executives, educators, caregivers, members of the military, artists, administrators, self-employed, service workers and retirees participated in the study. The age of respondents ranged from 21 to 77 with an average age of 56, and years of caregiving ranged from six months to sixteen years, with an average of 4.5 years. The majority of the sample reported their relationship to the person they care for as either their child (n = 37), family member (n = 36) or partner (n = 30). Two respondents reported their relationship as friend (n = 1) and other (n = 1). One respondent did not state a relationship. The sample reported that the person with dementia either shared a residence with them (n = 63), lived alone (n = 28) or lived in a long-term care facility (n = 8). Seven respondents did not state where the person they care for lived relative to themselves.

An initial factor solution was obtained for each of the four survey sections: (i) independent ADL completion by a person with dementia; (ii) caregivers assisting with completion of ADL; (iii) supporting ADL with intelligent Assistive Technology; and (iv) features and functions of an in-home IAT. Large Kaiser-Meyer-Olkin (KMO) measures of sampling adequacy (greater than 0.7)[89] and significant Bartlett’s Test of Sphericitys (p < 0.05)[89] in all four sections
substantiate the use of factor analysis on the items. The factors that satisfied the eigenvalue and Scree plot conditions were identified. The rotation method that produced the least complex items and the most items loaded on factors was selected as the optimal technique[88] and items that were complex (loaded on multiple items) or below the critical value (CV) were removed from the data (Table 2.1). For each section a final factor solution was forced on the items satisfying the inclusion criteria with the number of factors satisfying the three conditions.

<table>
<thead>
<tr>
<th>Table 2.1: Factorability Statistics, Factors Satisfying Each of the Three Inclusion Criteria and Optimal Rotation Techniques for Each of the Four Survey Sections</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KMO measure of sampling adequacy</strong></td>
</tr>
<tr>
<td><strong>Bartlett’s Test of Sphericity</strong></td>
</tr>
<tr>
<td>Eigenvalues &gt; 1</td>
</tr>
<tr>
<td>Factors above the “elbow” in scree plot</td>
</tr>
<tr>
<td>Factors with &gt; 3 significant items</td>
</tr>
<tr>
<td>Factors satisfying all three criteria</td>
</tr>
<tr>
<td>Rotation technique with best results</td>
</tr>
</tbody>
</table>

2.5.1 Independent ADL Completion

In trying to identify which ADL were difficult for a person with dementia to complete using exploratory factor analysis two factors clearly emerged as well-documented ADL categories: \( PADL \) (fundamental tasks)[85] and \( IADL \) (advanced daily tasks)[53] (Table 2.2). Looking at the factor means suggest that caregivers believe the person they care for still has some ability to independently complete \( PADL \) (\( M = 3.52, SD = 1.22 \)) but almost no ability to complete \( IADL \) (\( M = 1.58, SD = 0.94 \)). The individual item means give some insight into the specific tasks people with dementia struggle with. For example the mean for the PADL ‘Getting dressed’ (\( M = 2.7 \)) shows that respondents believe the person they care for has some difficulty completing this task. Similarly, low item means for all ADL within the IADL factor indicate that people with dementia struggle with ‘Preparing simple meals’, ‘Cleaning the house’ and ‘Preparing complex meals’. Conversely, based on higher item means within the PADL factor respondents’ believe the person they care for has the ability to partially complete the tasks ‘Eating finger foods’, ‘Drinking’, ‘Eating with cutlery’, and ‘Using the bathroom’. For intelligent AT designers this shows that a user’s abilities need to be considered when targeting ADL with technologies. For
example, devices supporting PADL completion should leverage a user’s remaining abilities while devices targeting IADL completion likely need to be more autonomous and provide additional functionality to compensate for the user’s lost abilities.

**Table 2.2: Independent ADL Completion by Person with Dementia**

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean²</th>
<th>Rotated matrix factor loadings¹</th>
<th>Factor 1 (PADL)</th>
<th>Factor 2 (IADL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eating finger foods</td>
<td>4.2</td>
<td>.847</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drinking</td>
<td>4.1</td>
<td>.900</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eating with cutlery</td>
<td>3.5</td>
<td>.770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using the bathroom</td>
<td>3.1</td>
<td>.744</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Getting dressed</td>
<td>2.7</td>
<td>.645</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preparing simple meals (e.g., Sandwich, salad)</td>
<td>2.0</td>
<td>.813</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleaning the house</td>
<td>1.4</td>
<td>.667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preparing complex meals (e.g., Using a stove)</td>
<td>1.3</td>
<td>.721</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>4.84</td>
<td>1.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of variance</td>
<td>42.99</td>
<td>7.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of rotated variance</td>
<td>52.93</td>
<td>11.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor Mean</td>
<td>3.52</td>
<td>1.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.22</td>
<td>.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability (Cronbach’s Alpha)</td>
<td>.92</td>
<td>.78</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Factor loadings below 0.505 are not shown

¹Varimax Rotation
²Mean ratings on a scale of 1 (Cannot complete at all on his/her own)-5 (Can easily complete on his/her own)

### 2.5.2 Caregivers Assisting with ADL

When looking at the challenges faced by caregivers while assisting with ADL, the same two factors emerged from the data: PADL and IADL (Table 2.3). In contrast to the previous section which looked at independent task completion the factor means here indicated that, in general, caregivers do not find it difficult to provide support for PADL (M = 3.23, SD = 1.19) and IADL (M = 3.42, SD = 1.42) completion. Looking at the individual item means shows that supporting a person with dementia with ‘Remembering to take medication’, ‘Preparing simple meals’, ‘Eating with finger foods’, ‘Drinking’ and ‘Eating with cutlery’ was rather easy. Additionally, the item means for tasks such as ‘Cleaning the house’, ‘Having a conversation’ and ‘Preparing complex meals’ showed a somewhat neutral response. However, the item means for more private tasks like ‘Getting dressed’, ‘Washing hands’, ‘Brushing teeth’ and ‘Using the bathroom’ show they were more difficult to support. This is not surprising as tasks that involve an invasion of privacy (e.g., getting dressed or using the bathroom) or tasks that are typically performed
independently throughout one’s lifetime (e.g., washing hands or brushing teeth) can be difficult to help a person complete. As such, from a caregiver’s perspective tasks that are private and personal are perhaps most desirable for intelligent AT to support.

Table 2.3: Caregivers assisting completion of ADL

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean²</th>
<th>Rotated matrix factor loadings¹</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Factor 1 (IADL)</td>
</tr>
<tr>
<td>Remembering to take medication</td>
<td>3.8</td>
<td>.741</td>
</tr>
<tr>
<td>Preparing simple meals (e.g., Sandwich, salad)</td>
<td>3.5</td>
<td>.887</td>
</tr>
<tr>
<td>Paying bills</td>
<td>3.4</td>
<td>.730</td>
</tr>
<tr>
<td>Cleaning the house</td>
<td>3.3</td>
<td>.800</td>
</tr>
<tr>
<td>Having a conversation with someone</td>
<td>3.3</td>
<td>.560</td>
</tr>
<tr>
<td>Preparing complex meals (e.g., Using a stove)</td>
<td>3.2</td>
<td>.848</td>
</tr>
<tr>
<td>Eating finger foods</td>
<td>3.9</td>
<td>.767</td>
</tr>
<tr>
<td>Drinking</td>
<td>3.8</td>
<td>.822</td>
</tr>
<tr>
<td>Eating with cutlery</td>
<td>3.7</td>
<td>.708</td>
</tr>
<tr>
<td>Getting dressed</td>
<td>3.0</td>
<td>.735</td>
</tr>
<tr>
<td>Washing hands</td>
<td>2.9</td>
<td>.694</td>
</tr>
<tr>
<td>Brushing teeth</td>
<td>2.8</td>
<td>.591</td>
</tr>
<tr>
<td>Using the bathroom</td>
<td>2.5</td>
<td>.643</td>
</tr>
<tr>
<td>Eigenvalue</td>
<td></td>
<td>6.02</td>
</tr>
<tr>
<td>Percentage of variance</td>
<td></td>
<td>38.49</td>
</tr>
<tr>
<td>Percentage of rotated variance</td>
<td></td>
<td>43.39</td>
</tr>
<tr>
<td>Factor Mean</td>
<td></td>
<td>3.42</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td></td>
<td>1.42</td>
</tr>
<tr>
<td>Reliability (Cronbach's Alpha)</td>
<td></td>
<td>.91</td>
</tr>
</tbody>
</table>

Note: Factor loadings below 0.505 are not shown
¹Varimax Rotation
²Mean ratings on a scale of 1 (Very hard to help)-5 (Very easy to help)

2.5.3 Supporting with Intelligent Devices

When considering the role of assistive technology as a tool to support the completion of ADL, three factors were revealed: Hygiene & Personal Care; Food & Nourishment; and Medication & Housekeeping (Table 2.4). The Hygiene & Personal Care factor represented activities typically performed independently and/or in private (e.g., washing hands, bathing, getting dressed). The Food & Nourishment factor was comprised of daily tasks related to food preparation and consumption – tasks that are often performed in more public places and/or with others present. The Medication & Housekeeping factor was comprised of tasks that span a relatively long period of time, requiring longer-term memory and focus. The factor means were all notably low –
particularly Hygiene & Personal Care (M = 2.54, SD = 1.19) and Food & Nourishment (M = 2.26, SD = 1.25) though Medication & Housekeeping (M = 2.92, SD = 2.92) was also relatively low. The individual item means for the factor Hygiene & Personal Care were all consistent with the factor mean showing that caregivers think it is unlikely that intelligent AT can support a person with dementia through the ADL ‘Brushing teeth’, ‘Using the bathroom’, ‘Washing hands’, ‘Bathing’, ‘Getting dressed’ and ‘Showering’. Within the factor Food & Nourishment the individual item means suggest that caregivers also think it is unlikely that intelligent AT can support ‘Drinking’ and ‘Eating with cutlery’, and that ‘Eating finger foods’ and ‘Preparing complex meals’ cannot be supported with intelligent AT. Notably, the factor mean for Medication & Housekeeping (M = 2.92, SD = 2.92) did not represent the item means well. Specifically, the item mean for ‘Remembering to take medication’ (M=3.4) was higher than the factor mean (M=2.92) and was the only item on in the survey section with a response above average – the only activity caregivers believed intelligent AT could assist.

### Table 2.4: Supporting ADL with Assistive Technology

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>Factor 1 (Hygiene &amp; Personal Care)</th>
<th>Factor 2 (Food &amp; Nourishment)</th>
<th>Factor 3 (Medication &amp; Housekeeping)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brushing teeth</td>
<td>2.7</td>
<td>.721</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using the bathroom</td>
<td>2.7</td>
<td>.665</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washing hands</td>
<td>2.6</td>
<td>.665</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bathing</td>
<td>2.5</td>
<td>.827</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Getting dressed</td>
<td>2.5</td>
<td>.561</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Showering</td>
<td>2.4</td>
<td>.838</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drinking</td>
<td>2.5</td>
<td>.773</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eating with cutlery</td>
<td>2.5</td>
<td>.854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eating finger foods</td>
<td>2.1</td>
<td>.822</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preparing complex meals (e.g., Using a stove)</td>
<td>1.9</td>
<td>.526</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remembering to take medication</td>
<td>3.4</td>
<td>.818</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleaning the house</td>
<td>2.8</td>
<td>.812</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paying bills</td>
<td>2.5</td>
<td>.625</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Eigenvalue                          | 6.87  | 1.77                               | 1.23                          |
Percentage of variance              | 43.94 | 9.54                               | 5.92                          |
Percentage of rotated variance      | 5.60  | 11.36                              | 7.09                          |
Factor Mean                         | 2.54  | 2.26                               | 2.92                          |
Standard Deviation                  | 1.19  | 1.25                               | 1.35                          |
Reliability (Cronbach's Alpha)      | .92   | .88                                | .83                           |

Note: Factor loadings below 0.505 are not shown
2.5.4 In-home IAT

Investigating the features and functions required for an in-home IAT identified three factors: Appearance & Usability; Familiarity & Autonomy; and Technical Expertise (Table 2.5). The Appearance & Usability factor (M = 3.04, SD = 1.00) represents appearance items (e.g., visibility of computers and wires) as well as usability items (e.g., simple, intuitive technologies). The Familiarity & Autonomy factor (M = 3.60, SD = 0.91) reflects caregiver’s opinions on how the device should blend in with other household items (e.g., be familiar), and how it should adapt to the different members of the home without taking away the user’s sense of control. The Technical Expertise factor (M = 2.57, SD = 0.91) represents the knowledge and experience our caregivers have with technology.

Looking at the individual item means within the Appearance & Usability factor demonstrates that our caregivers want IAT that are simple to use and intuitive because they are busy and do not have time to dedicate to learning how to use complicated devices (‘I’m busy so I like things to be simple’, ‘I don’t have time to play with it’), even though they believe they have the capability to learn (‘I’m past the point where I want to learn new things’). Furthermore responses to items such as ‘It’s important I don’t see any wires’ and ‘It’s important I don’t see any computers’ suggests that caregivers prefer less physically obtrusive IAT – an important feature to consider when designing IAT. Within the Familiarity & Autonomy factor high item means for the items ‘The visual display needs to be large enough to see clearly’ and ‘I want things to look as familiar/normal as possible’ highlight a need for in-home IAT – especially their user interfaces – to look similar to other devices commonly found in the homes of older adults. High means within the factor for the three items ‘It needs to recognize different users on its own’, ‘I want to know that I’m in control’ and ‘I envision something revolutionary’ emphasize that an in-home IAT must be able to autonomously adapt to multiple users without infringing on the users’ sense of control. The Technical Expertise factor suggests that caregivers are not overly technical people and have little time to spend learning new technology based on low item means for ‘I’m a very technical person’, ‘I love to play with cutting-edge gadgets’ and ‘I’m always buying new toys and gizmos’. This is likely an expression of the burden and time demands experienced by caregivers of older adults with dementia.
### Table 2.5: Features and Functions of an In-Home IAT

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1 (Appearance &amp; Usability)</th>
<th>Factor 2 (Familiarity &amp; Autonomy)</th>
<th>Factor 3 (Technical Expertise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I’m busy so I like things to be simple</td>
<td>3.4</td>
<td>.721</td>
<td></td>
</tr>
<tr>
<td>It’s important that I don’t see any wires</td>
<td>3.3</td>
<td>.790</td>
<td></td>
</tr>
<tr>
<td>I don’t have time to play with it</td>
<td>3.1</td>
<td>.608</td>
<td></td>
</tr>
<tr>
<td>It’s important that I don’t see computers</td>
<td>3.0</td>
<td>.884</td>
<td></td>
</tr>
<tr>
<td>I’m past the point where I want to learn new things</td>
<td>2.4</td>
<td>.646</td>
<td></td>
</tr>
<tr>
<td>The visual display needs to be large enough to see clearly</td>
<td>4.0</td>
<td>.689</td>
<td></td>
</tr>
<tr>
<td>It needs to recognize different users on its own</td>
<td>3.8</td>
<td>.526</td>
<td></td>
</tr>
<tr>
<td>I want things to look as familiar/normal as possible</td>
<td>3.7</td>
<td>.638</td>
<td></td>
</tr>
<tr>
<td>I want to know that I’m in control</td>
<td>3.4</td>
<td>.725</td>
<td></td>
</tr>
<tr>
<td>I envision something revolutionary</td>
<td>3.2</td>
<td>.563</td>
<td></td>
</tr>
<tr>
<td>I’m a very technical person</td>
<td>2.9</td>
<td>.794</td>
<td></td>
</tr>
<tr>
<td>I love to play with cutting-edge gadgets</td>
<td>2.7</td>
<td>.973</td>
<td></td>
</tr>
<tr>
<td>I’m always buying new toys and gizmos</td>
<td>2.1</td>
<td>.522</td>
<td></td>
</tr>
</tbody>
</table>

Eigenvalue: 4.13 2.83 1.17
Percentage of variance: 24.67 15.12 8.98
Percentage of rotated variance: 28.29 19.07 5.53
Factor Mean: 3.04 3.60 2.57
Standard Deviation: 1.00 .91 .91
Reliability (Cronbach’s Alpha): .85 .76 .80

Note: Factor loadings below 0.505 are not shown

1Promax Rotation
2Mean ratings on a scale of 1-5

### 2.5.5 Reliability

The internal consistency, or degree to which the different items within each factor represent their common factor, was measured for each factor in each section (Table 2.6). High values indicate good reliability, suggesting the items are measuring the common factor.

### Table 2.6: Reliability Statistics of Each Factor for Each of the Four Survey Sections, Calculated Using Cronbach’s Alpha

<table>
<thead>
<tr>
<th>Survey Section</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent ADL completion by person with dementia</td>
<td>.921</td>
<td>.779</td>
<td>-</td>
</tr>
<tr>
<td>Caregivers assisting completion of ADL</td>
<td>.905</td>
<td>.887</td>
<td>-</td>
</tr>
<tr>
<td>Supporting ADL with Assistive Technology</td>
<td>.920</td>
<td>.884</td>
<td>.830</td>
</tr>
<tr>
<td>Features and functions of an in-home IAT</td>
<td>.845</td>
<td>.762</td>
<td>.798</td>
</tr>
</tbody>
</table>
2.6 Discussion

Given these results it appears that caregivers of people with dementia believe the person they care for has at least partial ability to complete many personal activities of daily living (PADL or fundamental tasks) but not instrumental activities of daily living (IADL or advanced daily tasks). They also indicate that assisting with the completion of most PADL and IADL is not overly difficult but that PADL typically performed privately (e.g., getting dressed, using the bathroom, washing hands, brushing teeth) are most challenging to assist. Accordingly our findings imply that intelligent assistive technologies that determine when assistance with a PADL is required, particularly more private PADL, are likely to both reduce the burden on the caregiver and increase the independence and quality of life of the person with dementia. This findings are of fundamental importance for intelligent AT developers as many interventions – both technical and non-technical – tend to shift the burden of care from one individual to another (e.g., from the person with dementia to the caregiver) or reduce the quality of life or independence of the user (e.g., complete a task rather than allow the user to attempt the task).

When asked to assume ‘technology exists that can help you care for a person with dementia’ our participants overwhelmingly indicated they did not feel technology could help with any of the ADL presented except ‘Remembering to take medication’. As such, caregivers appear to know very little about existing or emergent intelligent AT that have been shown to help with various ADL[11, 33, 35, 82]. However, intelligent AT developers need to consider the role of users in the entire process of device development; not simply in terms of determining user needs to ensure the product is developed appropriately but also to increase the chance of device adoption by users. Central to this objective is adopting a User-Centred Design (UCD) approach [33-35], where users are included in all stages of device development including: needs assessment; idea generation; device prototyping; and efficacy testing. Considering the lack of knowledge caregivers showed about available or emerging intelligent AT, our findings suggest that developers of intelligent AT interested in using a UCD approach must revise existing strategies of recruitment and dissemination of information (which are arguably not working) if caregiver participation and device acceptance is desired. For example, the use of caregiving resources (e.g., caregiver newsletters, websites) or more public magazines (e.g., Popular Science, AARP The Magazine) may spread the word more effectively. Indeed, the lack of penetration of intelligent
AT into the general knowledge of the public warrants a more comprehensive study to understand the barriers preventing this knowledge translation.

We also investigated needed features and functions for intelligent assistive technologies designed to operate in the homes of people with dementia. Our caregivers emphasized that in-home intelligent AT must be able to autonomously adapt to the changing capabilities of a person with dementia and to the different users of the technology. This was substantiated by the fact that caregivers indicated having little time to interact with any sort of IAT and were not interested in learning new technologies (even though they thought they were still capable) – again, a reflection of the burden associated with being a caregiver. The caregivers also wanted any in-home IAT to be familiar and unobtrusive, presenting an interesting challenge for intelligent AT designers; acquire adequate sensor information and provide useful and intuitive user interfaces without compromising the user’s home atmosphere.

This study provides insight into the general needs of older adults with dementia and their caregivers as they participate in ADL together and investigates the role of IAT in this relationship, but is important to remember that caregivers in the study acted as proxies for the person they care for. Studies directly involving older adults with dementia as well as studies with a larger sample would increase the generalizability of the findings. Additionally, the exploratory nature of this research compels caution when interpreting these results to rank individual ADL and design features. The exploratory approach was necessitated by a lack of existing scales to measure: the challenges faced by older adults with dementia during ADL completion; the corresponding support challenges faced by their caregivers; and features and functions required by in-home IAT. Future studies that adapt existing ADL scales to older adults with dementia [53, 85], or the development of new scales specifically designed to measure the burden experienced by caregivers supporting ADL can provide more specific information about individual tasks. The effect of other factors on these results need to be considered, such as the level of dementia of the user, where the user lives relative to the caregiver (e.g., in the same home or alone), the relation (e.g., partner, family, friend) and the genders of both the person with dementia and the caregiver. Dementia level was not a factor included in this study because our respondents acted as proxies for the person they cared for; obtaining this information would either compromise ethical approval (e.g., if MMSE scores existed for the older adult with dementia) or be subjective (i.e., the opinion of the caregiver). The other factors were obtained in this study but could only be
included in the analyses with a larger sample size. Larger studies investigating user needs as shaped by these factors, as well as other social factors (e.g., culture) could provide more context and understanding of user needs.

2.7 Conclusions

The ability to independently complete ADL is a critical component of our sense of self and quality of life. However, the loss of cognition associated with dementia compromises one’s ability to perform ADL necessitating support from a caregiver; a role typically filled by a family member or friend. Accordingly, interest in the development of intelligent AT designed to support people with dementia and their caregivers has increased, but the specific needs of these users are not well known. This study investigated the needs older adults with dementia and their caregivers have for intelligent assistive technologies. Findings suggest that older adults with dementia still have at least partial ability to participate and complete ADL, that caregivers find private tasks (e.g., showering) are difficult to assist, and that in-home intelligent assistive technology must be autonomous, familiar, simple and unobtrusive. Based on these results, intelligent assistive technology developers should focus on devices that can support caregivers and older adults with dementia in the completion of private and personal tasks – where the most help is necessary.

2.8 Acknowledgements/Conflicts/Funding Sources

This study was funded by the Canadian Institutes of Health Research (Operating Grant) and the Alzheimer Association (ETAC Program).
Chapter 3

3 A Real-world Deployment of the COACH Prompting System

3.1 Publication Citation

3.2 Abstract
The loss of cognition associated with dementia affects an individual’s ability to participate in even the most fundamental activities of daily living (ADL). Assistive technology for cognition (ATC) has shown potential to support ADL completion, but few devices have undergone real-world deployments. The COACH is an existing ATC that has been shown in supervised trials to support older adults with dementia through the ADL of hand washing. This paper presents the results of a study of the COACH in a long-term, real-world, community-based deployment.

The COACH was installed in a washroom at the Toronto Memory Program. The COACH was configured to run in an unsupervised state, interacting with users when they were not progressing through the task. Video was collected from an overhead camera, and was manually annotated to determine the system’s capabilities. The trials were conducted from February to May, 2012.

Twenty participants contributed forty-one hand washing trials. Results suggest that the COACH was able to identify completed task steps with 46.6% accuracy and the participants’ true performance with 54.9% accuracy. This study clarifies the need for more robust, accurate and generalized user tracking, including the use of three-dimensional tracking, gesture, grip and rotation detection.
Keywords: Assistive technology, dementia, real-world deployment, activities of daily living

3.3 Introduction

Worldwide, over 35 million people are living with dementia [4] at an estimated cost of $604 billion per year [1]. The aging of the global population is expected to continue to contribute to an increase in the incidence of dementia worldwide. The prevalence of dementia is expected to double every 20 years to approximately 66 million in 2030 and 115 million by 2050. The costs of dementia are enormous and affect those with the illness, their family members, and health care and social systems [1]. The increase in prevalence of dementia will further increase the related costs while the loss of cognition associated with dementia impacts an individual’s ability to participate in even the most fundamental activities of daily living (ADL). The burden associated with caring for an individual with dementia often falls on informal caregivers, who are estimated to spend an average of 7.4 hours per day providing care, of which 3.7 is associated with supporting ADL [51].

Accordingly, Assistive Technologies for Cognition (ATC), defined as “any technology which assists cognitive function during task performance” [91] have shown potential to support both caregivers and people with dementia [8, 92]. Nonetheless, little has been done to progress these devices beyond the developmental stages [8]. ATC designed to support ADL completion [for examples, see 11, 16, 33, 40, 67, 68, 69, 93], have relied on individual case studies and clinical trials with a small number of representative users. In 2009, a systematic review of 58 ATC for dementia support revealed that only three had undergone even small clinical trials with people with dementia [8]. Only one of the three systems, the COACH [11], supported ADL completion. Recently, two other ATC that have the ability to support ADL, the Proactive Activity Toolkit [16, 94] and the COGKNOW Project [33], have also undergone preliminary clinical trials. Despite the potential that these ATC offer to clinical practice, more evidence-based research is necessary to understand the capabilities of these devices in real life situations and environments in order to improve their acceptance [95].
3.3.1 Supporting Daily Activities with Assistive Technology for Cognition: Clinical Trials

The initial prototype of the Proactive Activity Toolkit [16] used short-range Radio Frequency Identification (RFID) tag tracking, and was further developed [94] to use long-range RFID tag tracking to follow human activities. RFID tags are embedded in items of interest in the environment and are read using RFID readers placed in the walls and ceiling. A Hidden Markov Model infers daily activities from the sensor data. A trial was conducted in an instrumented apartment in the United States with 10 participants who performed 14 activities. The system was able to recognize tasks with approximately 90% precision (sensor reading) and 91% recall (activity detection), compared to their previous system which provided 95% precision and 60% recall [16]. However, discussion was limited largely to the technical device performance with little focus on additional factors such as the different ways in which the fourteen tasks could be completed, as well as how the system performed with task variability. Additionally, discussion on the nature of the participants and the trial protocol (e.g., when the system was enabled and disabled) was limited.

The COGKNOW Project [33] was also initiated in part to support ADL completion, as well as to provide cognitive reinforcement for memory loss, social interaction and safety. Prototypes of the system were deployed for between half a day and two days in the homes of 16 participant-caregiver dyads in test sites in the Netherlands, Sweden and Northern Ireland. Inhome testing of the system was almost entirely restricted to technical device performance and user-interface considerations, though semi-structured interviews were conducted with the participants during and after the trials. Technical results were positive, encouraging progress to a larger study, while findings related to ADL were limited to suggestions of improved personalization.

The COACH [39, 44, 48] was developed to monitor and track the actions of a user unobtrusively with an overhead camera. The system employs a partially observable Markov decision process to make decisions about the progress of the user under conditions of uncertainty (currently through the task of hand washing) and offer audiovisual prompts similar to those of a caregiver if the user needs assistance. Clinical trials were conducted with six people with moderate to severe dementia, showing an increase in the number of hand washing step completed with the COACH compared to the number of steps completed independently. The system was individually calibrated for each participant at the beginning of the study and the system was enabled at the
beginning of the hand washing task and disabled at the end. This precluded an understanding of what may have happened between trials.

These trials provide evidence to suggest that ATC may be useful for both ADL detection and as tools to support people with dementia. However, they suffer from small sample sizes which limit the generalizability of the findings. Furthermore, from the perspective of ADL completion, a small number of trials does not necessarily account for the normal variation in task performance and the ability of the systems to perform under this variability. Still further, the fact that these systems are typically calibrated to run with each specific user, inhibits their application to a real-world setting where the current user of the system is often not known, and individual calibration often requires expert knowledge and time. Finally, the short durations of trials, from study start to study finish and the duration of the actual technical trials, do not necessarily reflect real-world use.

### 3.3.2 Current Study

In the case of individuals with dementia, experimental protocols such as Single Subject Research Designs [96] or observational studies without a control group have been used in ATC studies largely because of: the exploratory nature of ATC research; the challenges in recruiting adequate sample sizes [11]; and because the variability associated with dementia prevents the definition of a baseline control group. This paper presents the results of an observational study of the latest prototype of the COACH system which was deployed at Toronto Memory Program (TMP), a multidisciplinary, community-based, medical facility in Toronto, Canada, specializing in the diagnosis and treatment of Alzheimer’s disease and related disorders [97]. This research improves on previous studies investigating the capabilities of ATC designed to support people with dementia through ADL completion in the following ways: 1) participants in this study interacted with the system without any instruction or equipment other than pressing a button to activate the system; 2) participants were free to use the system whenever needed rather than at prescribed times; 3) the system was tested in a completely unsupervised state, continuously running for four months without any technical intervention; 4) the system implemented a completely unobtrusive activity detection methodology, requiring zero effort from the clinical staff and participants; 5) the system was not calibrated or adjusted to any individual user; and 6)
the system was deployed in a real-world environment rather than a simulated environment. Based on these contributions, we sought to answer the following research questions:

1. How accurately can the COACH track the actions of multiple older adults with dementia through the task of hand washing without individual calibration or configuration in an unsupervised clinical environment?

2. How appropriately does the COACH respond to the actions of older adults with dementia while supporting hand washing in an unsupervised clinical environment?

3. What technical challenges will the COACH face running continuously in an unsupervised state in a real-world clinical environment?

Although the COACH is theoretically extendable to support a variety of ADL, the current study focused on the task of hand washing because of the availability of existing data for this task. Results will guide the future development and engineering of the COACH toward long-term deployment in the homes of people with dementia.

3.4 Methods

3.4.1 Participants

Study participants were recruited from patients with dementia attending and receiving medical care at TMP. These patients were already known to, and well characterized by, TMP’s research staff. Patients were invited to participate in the COACH study during their routine visits to the centre. Informed consent was obtained by TMP staff in manner compliant with ICH-GCP [98]. All subjects in the current study had a diagnosis of Dementia of the Alzheimer’s Type according to DSM IV-TR [99], were able to speak and understand English adequately, were able to see and hear audiovisual prompts, and were well enough to attempt the task of hand washing.

3.4.2 Device (COACH)

A conceptual representation of the COACH system is presented in Figure 3.1. The user interface allowed configuration of the system and user reports, though in this study the configuration was constant throughout the trials. The COACH system was represented by three fundamental components: the hand tracker, policy and prompter. The hand tracker employed an overhead 2D
webcam (Sony PS3 Eye) and was responsible for detecting the users’ hands interacting with the taps, water, soap and towel. The policy, or belief monitor, processed the data provided by the hand tracker and estimated the users’ progress through the hand washing task using a partially observable Markov decision process (POMDP). The POMDP represented the task as a sequence of steps [see 100 for POMDP details], inherently dealing with uncertainty in both the hand tracker observations and the effects of any actions the system took. The users simply interacted with the environment, in this case washing their hands, and the system unobtrusively monitored their progress. If the policy determined the users needed assistance, the prompter provided assistance through audio, visual or a combination of audiovisual prompts using speakers and a computer monitor.

![COACH System conceptual overview](image)

The COACH system used in this study incorporated five significant changes from the system used in previous trials [11]: 1) user tracking was calibrated for a generic user; 2) the system was modified to run continuously without intervention from the research team; 3) towel tracking was restricted to a single region; 4) physical remote start capability was added; and 5) verbal and visual prompts that were suitable for multiple users were recorded. The computer vision used to track the users’ hands in this study and in previous trials was a colour flocking technique developed for the COACH system [46]. The flocking method combines a Bayesian sequential
estimation technique with a colour skin model and was shown to be effective at dealing with tracking issues such as occlusion or lighting changes [46, 101]. Tracking in previous systems was calibrated to each user, and in addition, a calibration file was loaded for the specific user before each trial. For the TMP trials, a generic colour calibration file was developed by sampling several different skin colours and computing an average skin colour for tracking. The COACH was also configured to run continuously by detecting when a user engaged the sink area rather than requiring a member of the research team to initiate the system before a trial. Similarly, the COACH would automatically stop either when the user completed the task or when the system decided the user was not progressing and could no longer be provided with assistance. Towel tracking was modified to a single region similar to the soap, taps and faucet to account for a paper towel dispenser used in the clinic. Previous COACH trials had used a cloth towel, which was tracked around the sink area using flocking. Additionally, a presentation remote was modified and mounted in an enclosure with a push button to provide a physical method of initiating the software recording when a consenting participant entered the washroom. Finally, the audio prompts used in previous COACH trials [11] were modified to remove the user-specific content and to simplify language content while the video prompts were re-recorded to reflect the TMP washroom used in the study.

3.4.3 Procedure

The COACH was installed in a washroom at TMP in February of 2012 (Figure 3.2). This washroom was frequented by patients of TMP throughout their routine visits to the centre. The installation included an overhead video camera, processing components (computer), user interfaces (configuration screen for the research team, prompting screen and speakers for the study participants), cabling, and an activation pushbutton. The camera was mounted such that only the wash basin and sink area could be seen. The washroom was selected because it was large enough to provide adequate spacing between the toilet and sink (i.e. the camera could not capture video near the toilet). An activation button was mounted inside the washroom so that participants, their caregivers, or staff could enable the system before a consented patient entered the washroom. To participate in the study, users enabled the system by pressing the activation button. Once the button was pushed, the system monitored the sink area until the user began washing. When the hand washing task began, the COACH began recording video and monitoring the user’s progress. If required, the system offered assistance through verbal and
visual prompts. If the COACH identified that the user had finished the hand washing task or if the user did not begin the task within ten minutes after pressing the activation button, the system would turn off and wait for the next user. If the system determined that the user was stuck at a particular step in the task and was not responding to assistance provided by the system, the system turned off. A participant log was maintained by TMP staff and recorded each subject’s identification number, last Mini-mental State Examination (MMSE) [102] score, date and time of entering the washroom. Any video captured during a time that was not on the log was attributed to a non-study participant and was removed. This ensured that only participants who had consented to the study provided video data. Ethics approval for this study was granted by the University of Toronto (REB #25953) and IRB Services.

Figure 3.2: COACH at Toronto Memory Program
3.4.4  Data analysis

3.4.4.1  System Performance

Video data were manually annotated in two stages using The Observer XT 10 behavioral observation software [103]. The first stage assessed the COACH’s ability to identify individual tasks steps completed by the users as well as the users’ overall task performance. The second stage assessed the COACH’s ability to provide appropriate assistance when the user needed help or to continue observing unobtrusively while the user was progressing independently through the task.

3.4.4.2  Task Step Identification

Using The Observer XT 10 a coding scheme was developed consisting of discrete participant behaviours corresponding to six hand washing task steps (Table 3.1). Scoring indicated the occurrence of a behaviour, where each behaviour was defined based on an observable user action and the presence (or absence) of an observable outcome. Prerequisite behaviours were defined to identify when behaviours were considered valid. A set of behavioural modifiers and associated mutually exclusive modifier states common to all behaviours was then defined to capture relevant states of the COACH while the participant exhibited an associated behavior, shown in Table 3.2. Whenever a participant exhibited a relevant behaviour during video annotation, the behaviour was scored as complete along with one modifier state for each of the four behavioural modifiers.

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Scoring Rule</th>
<th>Prerequisite Step(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn on Water</td>
<td>Hand(s) start to move a tap to the on position, water flows</td>
<td>None</td>
</tr>
<tr>
<td>Use Soap</td>
<td>Hand(s) stop moving above the soap dispenser, soap on hands</td>
<td>None</td>
</tr>
<tr>
<td>Wash Hands</td>
<td>Hand(s) first touch the water after scrubbing, hands wet</td>
<td>Turn on Water, Use Soap</td>
</tr>
<tr>
<td>Turn off Water</td>
<td>Hand(s) start to move a tap to the off position, water stops</td>
<td>Wash Hands</td>
</tr>
<tr>
<td>Take Towel</td>
<td>Hand(s) begin to tear a towel off the towel holder</td>
<td>Wash Hands</td>
</tr>
<tr>
<td>Dry Hands</td>
<td>Hand(s) begin to dry using towel</td>
<td>Turn off Water, Take Towel</td>
</tr>
</tbody>
</table>

Table 3.2: Behavioural modifiers for the six task steps in the first stage of analyses

<table>
<thead>
<tr>
<th>Modifier</th>
<th>Modifier State</th>
<th>Modifier State Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hand tracker correct</td>
<td>Yes</td>
<td>Hand tracker provided correct hand position</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Hand tracker did not provide correct hand position</td>
</tr>
<tr>
<td>Behaviour correct</td>
<td>Yes</td>
<td>Belief monitor indicated correct behaviour</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Belief monitor did not indicate correct behaviour</td>
</tr>
<tr>
<td>State change</td>
<td>Yes</td>
<td>Belief state changed correctly</td>
</tr>
</tbody>
</table>
COACH’s step identification performance was categorized using common categories from signal detection theory [104](Table 3.3). For each step of each trial the COACH’s performance was scored based on the users’ performance and the associated system outcome. Accordingly, when a user completed a step, a True Positive was defined as a correct state change, while a False Negative was anything other than a correct state change. Similarly, when the user did not complete a step, a False Positive was scored if the COACH indicated a state change and a True Negative was scored if the system did not indicate a state change.

**Table 3.3: Binary classification function for the task step performance of the COACH**

<table>
<thead>
<tr>
<th>Participant Performance</th>
<th>System Outcome</th>
<th>Identified step complete</th>
<th>Identified step incomplete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step Completed</td>
<td>True Positive</td>
<td>False Negative</td>
<td></td>
</tr>
<tr>
<td>Step not Completed</td>
<td>False Positive</td>
<td>True Negative</td>
<td></td>
</tr>
</tbody>
</table>

Three common categories of system performance measures were calculated using data collected in Table 3.3. Equation 1 (Sensitivity) presents the likelihood the COACH will correctly identify a step completed by the user. Equation 2 (Specificity) presents the likelihood the COACH will correctly identify a step not completed by the user. Equation 3 (Accuracy) represents the likelihood that the COACH will measure the true participant performance.

\[
\text{Sensitivity} = \frac{TP^1}{TP^1 + FN^4} \quad (1) \\
\text{Specificity} = \frac{TN^3}{TN^3 + FP^2} \quad (2) \\
\text{Accuracy} = \frac{(TP^1 + TN^3)}{(Total \ Outcomes)} \quad (3)
\]

\begin{align*}
^1TP &= \text{True Positives} \\
^2FP &= \text{False Positives} \\
^3TN &= \text{True Negatives} \\
^4FN &= \text{False Negatives}
\end{align*}

**3.4.4.3 Identifying User Performance**

The COACH was scored on how well it was able to identify the users’ actual performance. User data were divided into the number of hand washing steps completed by users during the trial, and within each division three metrics were scored: 1) the number of trials where COACH identified all steps completed by users; 2) the most completed steps identified; and 3) the average number of steps identified.
### 3.4.4.4 Task Assistance

COACH’s task assistance performance was categorized using Signal Detection Theory (SDT) [104], presented in Table 3.4, similar to previous COACH trials [11]. The users’ actions were categorized as either stuck on a step or performing the wrong step, or progressing independently through the task. The system’s actions were categorized as either providing a prompt or observing. Accordingly, when the user was stuck on a step or performed an erroneous step and the COACH provided a prompt, the system was scored with a Hit, whereas if the system reached a decision point and a prompt was not provided, the system was scored with a Miss. A decision point was a point during the trial where the system updated its policy based on hand tracker information or enough time passed that the system forced itself to take an action, even if the action was to do nothing. When the user was progressing through the task independently, if the COACH provided a prompt the system was scored with a False Alarm. If the system reached a decision point and did not provide a prompt the system was scored with a Correct Reject.

<table>
<thead>
<tr>
<th>System Action</th>
<th>Prompt provided</th>
<th>Prompt not provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant Action</td>
<td>Stuck/error on step</td>
<td>Hit</td>
</tr>
<tr>
<td>Action</td>
<td>Progressing through task</td>
<td>False Alarm</td>
</tr>
</tbody>
</table>

Three common categories of task assistance performance measures were calculated using data collected in Table 3.4. Equation (4) (Sensitivity) presents the likelihood the COACH will provide assistance when a user is stuck or has performed an error on a step. Equation (5) (Specificity) presents the likelihood the COACH will continue to observe without prompting when a user is progressing through the task. Equation (6) (Accuracy) represents the likelihood that the COACH will take an appropriate action based on the users’ performance.

\[
\text{Sensitivity} = \frac{\text{Hit}}{\text{Hit} + \text{Miss}} \quad (4)
\]

\[
\text{Specificity} = \frac{1^{\text{CR}}}{1^{\text{CR}} + 2^{\text{FA}}} \quad (5)
\]

\[
\text{Accuracy} = \frac{(\text{Hit} + 1^{\text{CR}})}{(\text{Total Actions})} \quad (6)
\]

\[
1^{\text{CR}} = \text{Correct Reject}
\]

\[
2^{\text{FA}} = \text{False Alarm}
\]

Prompt performance was also analyzed in terms of sequencing and timing. Prompt sequencing identified whether the user was actively struggling with a step when a relevant prompt (i.e., the user had not yet completed the step) was provided or was on a different step. Prompt timing was scored relative to the users’ temporal progress through the task step as either too soon,
appropriate, or too late. These classifications were assigned after the entire trial was observed to fully understand the user’s capabilities, such as how quickly each step was completed and how effectively the hands were actually washed. This grounded approach [105], or the discovery of each user’s relative performance capability through analysis of the data, provided an understanding of the temporal limits that defined when each user would begin or complete a task step independently. ‘Too soon’ was defined as any prompt provided before the user would independently begin the task step. ‘Too late’ was scored as any prompt provided after the user would independently complete the task step. ‘Appropriate’ was defined as any prompt provided that was neither too soon nor too late.

3.4.4.5 Hand Tracker Validation

Ground-truth hand tracker outputs were created by manually reviewing video data for each trial and labelling the hand positions to reflect the users’ actual hand positions. Overall hand tracking performance was validated by comparing the hand tracker output captured during the trials to the ground-truth data for each trial using Dynamic Time Warping (DTW) [106, 107]. The resultant shortest path between each pair of time series provided a count of the number of hand tracker video frames that were erroneous compared to the ground truth while allowing for variability in the manual labeling of the start frame and duration of the hand positions. Only video frames recorded by the system while the users were washing their hands were included in the hand tracker validation. For example, if the system continued to record video after the participant completed the hand washing task and left the sink area, only frames up to and including the last frame where the users’ hands were visible were included in the analyses.

3.5 Results

3.5.1 Participants

Twenty-seven patients with MMSE scores ranging from 9 to 28 (Mean = 18.9) consented to participate in the study between February 2012 and May 2012. One participant consented to participate in the study but did not use the washroom and hence did not contribute a trial.

3.5.2 Trials

A total of fifty-eight trials were generated by the 27 participants as documented on the participant log. Forty-one trials, provided by 20 participants, were used in the analyses. Trials
documented on the log were not used in the analyses when: the start button was not pressed (9 trials); the start button was pressed but the computer failed to respond (4 trials); a software error caused the computer’s hard drive to appear full (3 trials); and an incomplete log entry was made (1 trial). Of the forty-one trials included in the analyses, each participant contributed an average of 2.05 trials with two participants completing six trials each.

3.5.3 System Performance

3.5.3.1 Task Step Identification

System performance was measured for each of the six steps that comprise the task in all forty-one trials yielding a total of two hundred forty six (246) steps in the analyses. Data representing the system’s effectiveness at identifying whether users completed or did not complete steps are presented in Table 3.5. Across the 246 steps users completed 206 (83.7%), of which 96 (39.0%) were successfully identified as completed by the system (True Positives) and 110 (44.7%) were not successfully identified (False Negatives). For each of the six steps within the task, the number of times the steps were completed by the users is also reported as the number of times the COACH was successful (True Positive) and unsuccessful (False Negative) at identifying the step as completed. A total of 40 (16.3%) of the 246 steps were not completed by the users. The COACH successfully identified 39 incomplete steps properly (True Negative), but reported one incorrectly as a completed step (False Positive). For the six individual steps within the task, the number of times the steps were not completed by the users is also reported, along with the number of times the COACH was successful (True Negative) and unsuccessful (False Positive) at identifying the step as not completed. Based on the performance statistics presented in Table 3.5, the COACH was likely to correctly identify a completed step 46.6% of the time (Sensitivity, Eq. (1)), and correctly identify an incomplete step as not completed 97.5% of the time (Specificity, Eq. (2)). Additionally the system was likely to measure the true participant performance 54.9% of the time (Accuracy, Eq. (3)). There was a significant difference between the task step and the ability of the COACH to detect the user completing the step at the p < .05 level for the six task steps [F(5, 201) = 7.069, p = .000. Post-hoc comparisons using the Tamhane’s T2 test indicated that the mean score for the “turn on water” (M = .6098, SD = .494), “get soap” (M = .8148, SD = .396) and “rinse hands” (M = .5926, SD = .501) were significantly different than the final three task steps.
In all cases where the COACH incorrectly identified the users’ real performance (False Negative and False Positive), the error was classified as either a system failure or a case where necessary preconditions were not met (i.e., a required previous step had not been completed). The results, presented in Table 3.6, show the aggregate for all steps by failure mode with a breakdown by individual task step.

### Table 3.6: Breakdown of System Module Failure Modes for Steps Labeled as Misses or False Alarms

<table>
<thead>
<tr>
<th>Error classification</th>
<th>All Steps</th>
<th>Turn on water</th>
<th>Use soap</th>
<th>Rinse hands</th>
<th>Turn off water</th>
<th>Take towel</th>
<th>Dry hands</th>
</tr>
</thead>
<tbody>
<tr>
<td>System failure</td>
<td>55 (49.5)</td>
<td>16</td>
<td>5</td>
<td>7</td>
<td>12</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>Preconditions not met</td>
<td>56 (50.5)</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>14</td>
<td>13</td>
<td>25</td>
</tr>
<tr>
<td>Total steps in trials</td>
<td>111</td>
<td>16</td>
<td>5</td>
<td>11</td>
<td>26</td>
<td>25</td>
<td>28</td>
</tr>
</tbody>
</table>

#### 3.5.3.2 Identifying user performance

Summary data on the system’s ability to identify the users’ performance are presented in Table 3.7. Of the forty-one trials, participants completed all relevant steps and successfully washed their hands 26 times (63.4%). Of these trials, the system correctly identified the task as complete eight times (30.8%), failing to identify the task as complete for the other 18 trials (69.2%). Participants did not complete all steps in the task for 15 trials (36.6%) and the COACH did not track all the users’ completed steps in any of these cases. Overall, participants completed an average of 5.05 out of 6 steps and the COACH correctly identified a significantly lower average of 2.37 steps, \( t(41) = 6.343, p < .001 \).

### Table 3.7: User performance identification during live trials

<table>
<thead>
<tr>
<th>Steps completed by participants</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>5.05</td>
</tr>
</tbody>
</table>
3.5.3.3 Task assistance

The task assistance performance of the system was analyzed in terms of its ability to assist the user in task completion. Following the Signal Detection Theory (SDT) outlined in Table 3.4, the prompting module performance was scored as a Hit, Miss, False Alarm or Correct Reject. Table 3.8 shows the results of the SDT analysis for the prompter.

Based on the performance statistics presented in Table 3.8, the COACH was likely to provide assistance when required 48.9% of the time (Sensitivity, Eq. (4)), and continue observing when assistance was not required 81.4% of the time (Specificity, Eq. (5)). Additionally the system was likely to take an appropriate action based on the true participant performance 74.4% of the time (Accuracy, Eq. (6)).

Table 3.8: Task assistance effectiveness of the COACH using Signal Detection Theory

<table>
<thead>
<tr>
<th>COACH's action</th>
<th>Prompt</th>
<th>No prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant</td>
<td>Step not Completed</td>
<td>45 (10.5%) (Hit)</td>
</tr>
<tr>
<td>Performance</td>
<td>Step Completed</td>
<td>63 (14.6%) (False Alarm)</td>
</tr>
</tbody>
</table>

Table 3.9 presents a summary of the prompting performance of the system over the 41 trials. The COACH provided a total of 108 prompts throughout the trials, where 45 (41.7%) prompts were classified as Hits and 63 (58.3%) were classified as False Alarms. Of the 45 hits, 23 prompts were provided for the step the user was actively trying to complete while 22 prompts were provided for a step that was different than the one the user was currently completing. Only 10 (22.2%) of the hits were appropriately timed, while 23 (51.1%) of the prompts were too late and 12 (26.7%) were too soon. For all of the 63 prompts classified as false alarms, the user was on a different step than the prompted step provided by the COACH. Only 4 (6.3%) of these prompts were appropriately timed, while 48 (76.2%) were too late and 11 (17.5%) were too soon. Overall, out of 108 total prompts the COACH’s prompts were appropriately timed 14 (13%) times, were too late 71 (65.7%) times and were too soon 23 (21.3%) times.

Table 3.9: Prompt performance based on prompt characteristics of sequencing and timing
### 3.5.3.4 Hand tracker validation

The results of the frame-by-frame hand tracking analyses are presented in Table 3.10 for 5 (12.2% of total), 10 (24.4% of total) and 21 (51.2% of total) random trials selected from the 41 trials. Results for the average fail rate for both the left and right hand analyses suggesting that it is not necessary to manually label and analyze the remaining trials. Based on the largest random sample (21 trials), the hand tracker had a fail rate of 27.32% tracking the left hand and 26.11% tracking the right hand.

<table>
<thead>
<tr>
<th>Prompt Timing (%)</th>
<th>Too late</th>
<th>Appropriate</th>
<th>Too soon</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>User on step</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>User on different step</td>
<td>18</td>
<td>1</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td><strong>Total Hits</strong></td>
<td>23 (51.1)</td>
<td>10 (22.2)</td>
<td>12 (26.7)</td>
<td>45 (41.7)</td>
</tr>
<tr>
<td>False Alarm</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>User on step</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>User on different step</td>
<td>48</td>
<td>4</td>
<td>11</td>
<td>63</td>
</tr>
<tr>
<td><strong>Total False Alarms</strong></td>
<td>48 (76.2)</td>
<td>4 (6.3)</td>
<td>11 (17.5)</td>
<td>63 (58.3)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>71 (65.7)</td>
<td>14 (13)</td>
<td>23 (21.3)</td>
<td>108</td>
</tr>
</tbody>
</table>

### 3.6 Discussion

#### 3.6.1 Participants

Hand washing task performance as assessed by the COACH (both the number of task steps identified and overall task performance) remained relatively consistent across all participants. There was no significant difference between the number of steps completed by individuals with lower MMSE scores compared to those with higher MMSE scores, $F(3, 37) = .605, p = .616$. The lack of difference in performance among those with different severities of dementia is counterintuitive. This was likely not a reflection of the participants’ actual ability to complete the task of hand washing but rather due to the fact that the scoring system used to measure the
capabilities of the COACH did not adequately detect differences in hand washing performance between the groups.

Of interest, some participants who completed multiple trials did show variability in their ability to complete the task. For example, one participant who completed four trials was able to complete all six tasks steps during two trials, but only completed two task steps in the other two trials. This may be a reflection of both the day to day variability of task performance as well as the fact that participants were not forced to complete all the task steps before exiting the washroom so in some cases the participants simply did not wash their hands.

### 3.6.2 System Performance

In this study, system performance represents the COACH’s technical effectiveness at correctly identifying the users’ actions, regardless of whether the actions are correct or not, and providing appropriate assistance based on the users’ actual progress through the task. The following discussion first considers the COACH’s effectiveness at identifying the individual task steps, followed by the system’s ability to adapt to potential step-based errors in order to correctly identify the users’ overall task performance. Next, the appropriateness of the system’s decisions based on the users’ actual performance (as opposed to the system’s belief state) and the relevance of any prompts provided is discussed. Finally, the validation of the system’s core modules is considered.

#### 3.6.2.1 Task Step Identification

An indication of the overall effectiveness of the COACH is the system’s ability to identify completed steps as complete or incomplete steps as incomplete. Throughout the trials, the system effectively identified 39 of 40 incomplete steps as incomplete (97.5%), and correcting for unmet preconditions identified 152 (96 directly identified + 56 identified but failed for unmet preconditions) of 206 completed steps as complete (73.8%). However, system performance varied substantially across different task steps. The system was most successful at identifying the first three task steps. This was not surprising considering, for example, the location of the soap was isolated from the rest of the sink and the pump-style dispenser ensured users’ hands remained in the hand tracking region for an ample amount of time. Additionally, the *Turn on water* was the first step and did not have any preconditions. The system was least successful at
identifying the last three steps which is again not surprising considering the reliance these steps have on previous steps being completed. The impact of the necessary preconditions defined within the task were perhaps most obvious when comparing the success rates of Turn on water (25 of 41, 61.0%) and Turn off water (15 of 41, 36.6%). Considering the underlying mechanisms for tracking the two steps were identical (i.e., the same hand tracking and region definition), the only remaining variable is the policy or more specifically the fact that the previous steps had not been completed.

One noteworthy step is the single False Alarm that occurred during the Dry hands step. During the trial, the user turned on the water, wet his hands while repeatedly adjusting the taps, moved his hands around the sink area, and then turned off the water. By moving his hands around the sink area and adjusting the water, the system had assigned a high enough probability that all the prerequisite steps were complete that when he dried his hands the system incorrectly indicated he had completed all the task steps. This was a reflection of the restriction of the hand tracker to two-dimensional image processing (as opposed to 3D tracking), where the system could not differentiate between a hand passing over an object and a hand interacting with the object.

3.6.2.2 Identifying user performance

The system’s ability to identify the users’ overall task performance independent of the step-based performance is also a good measure of system effectiveness. In this study, even if a participant did not successfully wash his/her hands, the system was considered to have worked if the COACH correctly identified the user’s actual performance. As presented in Table 3.7, the system correctly identified the users’ performance in 8 trials wherein the users completed all task steps. For two trials in which the users completed all steps the system did not detect the “Dry hands” step because the towel holder was not correctly in the towel region. In an additional two trials the “Dry hands” step was not detected because of modifications made to the system to track the towel by region as opposed to using flocking. If a user’s hands did not leave the towel region before a belief state timeout the system would lose confidence that the step was complete and could not recover. For the remaining 14 trials at least one step was missed and the system was not able to adjust its belief state to accommodate.

In all cases where the users did not complete at least one task step, the COACH failed to correctly identify the users’ task performance. Reviewing the data and recorded videos shows
clearly that all the two step trials were users turning on the water, quickly rinsing their fingertips and turning off the water, and the four step trials were users turning on the water, rinsing their hands, turning off the water and drying their hands. In all of these cases the users completed each step very quickly such that the hand tracker was not able to detect the users’ hands in the relevant regions longer than the minimum threshold.

Based on these results, it is clear that the system is much more effective at identifying user performance when users wash their hands slowly as well as when participants complete all relevant steps. This is largely because when users performed each step for a longer period of time, they more frequently created a temporal delay between steps. This temporal component is critical for the COACH in order to allow enough time to properly detect user actions. In previous COACH trials, users had more severe diagnoses of dementia creating a naturally slower progression through the task.

3.6.2.3 Task assistance

The assistance the COACH can give is simply to provide a prompt or not provide a prompt. Based on Eq. (4) (system sensitivity), the system was likely to provide a prompt 48.9% of the time when a user needed assistance. However, based on Eq. (5) (system specificity) the system was likely to decide to remain in observation 81.4% of the time when a user was progressing appropriately. Summing all the Hits, Misses, False Alarms, and Correct Rejects indicates that the COACH made a total of 430 decisions about the provision of assistance during the trials. Accordingly, Eq. (6) (system accuracy) indicates that overall the COACH provided appropriate assistance 74.4% of the time. The dependency of the task assistance on the hand tracking and policy suggest that this is appropriate considering the overall system performance.

The low system sensitivity (48.9%) and high specificity (81.4%) suggest that the system was biased toward underprompting (i.e., not providing prompts). This bias was related almost entirely to the hand tracker, rather than the policy, or decision making. The hand tracking was frequently missing the completion of user steps and giving erroneous hand locations, resulting in the policy often underestimating the users’ progression and remaining in a probabilistic state of uncertainty. The result is that the system often did nothing, contributing to the high specificity. Additionally, the system’s state of uncertainty also inhibited potential False Alarms, but increased the number of Misses. However, almost all the participants in this study performed the hand washing task
well, with an average of 5.05 steps completed out of 6. Thus, a system that was biased to doing nothing would naturally perform well at providing task assistance, reflecting the relatively high overall system accuracy (74.4%).

3.6.3 System component analyses

Each system module’s interaction with the COACH as a whole was isolated and analyzed independent of the rest of the system.

3.6.3.1 Hand tracker

The use of dynamic time warping (DTW) to measure the hand tracking performance provided informative results in terms of detecting the presence (or absence) of critical task events. For example, DTW effectively highlighted intermittent hand location observations. During manual review of the trial videos it was clear that in some instances the users’ hands were steadily in a region (e.g., the water) but the hand tracker was reporting the hands as toggling in and out of the region. This type of unstable observation had an unexpectedly negative influence on the systems’ performance because observations needed to stay consistent for longer than a minimum time threshold (from 0.3 to 1 second depending on the region of interest) before they were passed to the policy. The result is that user steps where the observations didn’t pass the minimum time threshold were considered incomplete by the system.

The DTW method for evaluating the hand tracker penalized intermittent observations by considering only the first instance of the toggling sequence as a relevant observation, with every consecutive observation counting as a sequence of failed frames. For example, consider a scenario where the ground truth labeling of a user’s hands were at the soap for 200 frames. For the same task step the hand tracker captured the user’s hands toggling between the soap region and the sink region cyclically for 10 frames at the soap followed by 10 frames at the sink over the 200 frames (a somewhat real-life scenario). The hand tracker would have been scored with 10 frames correct and 190 frames incorrect.

Similarly, DTW accounted well for temporal differences between manual labeling and the hand tracker outputs. In almost all cases the manual labeling of a video will not temporally align with the actual hand tracker output. For example, the hand tracker may report that a user’s hands entered the soap region at frame 150 but the manual labeling of the task may indicate that the
soap was used at 160 frames. The temporal shifting ability of the DTW method would account for this shift and not penalize the hand tracker.

The DTW analyses revealed that the hand tracker was only 27.32% effective tracking the left hand and 26.11% effective tracking the right hand. Manual review of frames that were not properly tracked revealed that the hand tracker failed when users: were bald (mistook head for hands); rolled their sleeves up (didn’t identify the ends of the hands properly); had darker toned skin; moved the soap out of the soap region; removed the towel from the towel region; and moved their hands over various regions without interacting with them. The first three failure modes (bald users; rolled up sleeves; and dark skin tone) could have been reduced if the hand tracker didn’t predominantly rely on skin colour. In the case of the towel, previous versions had tracked the towel as well as the hands which may have reduced tracking errors. For this trial the sink had a paper towel dispenser as opposed to a cloth towel which prevented the use of towel tracking. For tracking errors that occurred when the users moved the soap, similar tracking used in previous COACH trials for the towel could have helped reduce errors. Tracking errors that occurred when the users moved their hands over the regions without interacting could have been reduced if the hand tracker knew where the hands were in beyond two dimensions (i.e., depth).

3.6.3.2 Policy

As the results showed in Table 3.7, the policy did not perform well with respect to determining the correct state of the person. However, it is unclear whether these poor results are because the policy itself was not tuned appropriately in these trials, or because it was receiving poor data from other components of the system, specifically the hand tracker. As such, it is important to validate the policy independent of the hand tracker. To accomplish this, the ground truth hand tracker outputs used in the hand tracking validation were implemented as a sequence of video frames for each trial. The COACH system was rerun by replacing the live camera input with the ground truth hand tracker videos for all 41 trials. Videos of the ground truth trials were annotated using The Observer XT 10 and the developed coding scheme. The COACH’s performance was again measured in terms of the system’s ability to identify individual steps as well the actual users’ performance in the same way as for the raw trial videos.

Statistics on the system’s ability to identify the users’ performance based on the simulated ground truth tracker data are presented in Table 3.11, similar to results presented in Table 3.7
using the actual hand tracker data. With the ground truth data, the system correctly identified 23 of 26 completed tasks as complete (88.5%). Of the 15 trials where the users did not complete all the steps, the system correctly identified all the steps the users did complete in 14 trials, only failing to identify one step in one trial where the user only completed two steps. Using the ground truth data the system correctly identified all 40 of 40 (100%) incomplete steps as incomplete and identified 201 of 206 (97.6%) of completed steps as complete.

<table>
<thead>
<tr>
<th>Steps completed by participants</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>5.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of participants</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>26</td>
<td>-</td>
</tr>
<tr>
<td>Trials where COACH identified all user steps</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>11</td>
<td>-</td>
<td>23</td>
<td>-</td>
</tr>
<tr>
<td>Most steps identified by COACH</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Average steps identified by the COACH</td>
<td>1</td>
<td>1.67</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td>5.85</td>
<td>4.93</td>
</tr>
</tbody>
</table>

When using the ground truth hand tracker data, the system’s ability to correctly identify completed task steps improved dramatically from 73.7% (trial data) to 97.6% (ground truth data). Additionally, while users completed an average of 5.05 steps per task, using ground truth data improved the system’s ability to correctly identify the users’ progress from an average of 2.37 (46.9%) steps to 4.93 (97.6%). This difference in actual user step completion versus system step detection was no longer significantly different, \( t(41) = .385, p = .701 \). Finally, the system correctly identified the users’ overall task performance for 8 of 41 (19.5%) trials using the raw camera data but was able to correctly identify 37 of 41 (90.2%) trials with the ground truth tracker data. This represents a significant improvement in overall policy performance based reliable input to the system.

### 3.6.3.3 Prompter

Prompting was validated by looking at the characteristics of the prompts provided by the system, identified as Hits or False Alarms in Table 3.9. The vast majority of the 108 prompts provided by the system were either too late (71, 65.7%) or too soon (23, 21.3%). Only 14 (13.0%) of the prompts were appropriately timed. Additionally, 85 (78.7%) of the prompts were promoting the completion of a step that the user was not actively trying to complete, compared to 22 (21.3%) of the prompts that corresponded to the step the user was trying to complete. These results are not surprising considering both the timing and the sequencing of the prompts are largely dependent...
on capabilities of both the hand tracker and the policy, both of which underperformed to expectations during the live trials. The mode of prompt delivery – audiovisual prompts provided through speakers mounted in the ceiling and a computer display on the wall – may also have been ineffective.

In general the participants ignored prompts that were inappropriately timed. Similarly, prompts that suggested the completion of incorrect steps were also ignored by the participants. The provision of incorrect or inappropriately timed prompts was not observed to have confused or distressed the users. Some users notified the research staff about the prompt delays but were not observed to be concerned about this. The performance of the COACH’s individual system components heavily implicates the hand tracker as the source of the majority of system failures during the trials. When considering previous trials performed with ATC for ADL support it is not surprising that these systems reported better overall performance than the COACH system used in this study, as the systems were modified and calibrated for the individual user or environment. Indeed, when looking at the performance of the task detection (COACH policy), system performance was extremely high when given ideal hand tracker data.

### 3.6.4 Limitations

During the course of the study, several limitations were noted. Firstly, the relatively small sample size makes it difficult to generalize results to the larger population. Additionally, the system was deployed in a single washroom in a community-based clinic in North Toronto which, despite being a real world washroom, cannot predict all of the variables that will need to be addressed in progressing to other community and in home settings.

In this study, the ability of the COACH to run continuously without being turned on by pressing a button at the outset of each trial was not assessed. Furthermore, the lag time between activating the COACH and the participant engaging the sink area was not assessed and may have introduced additional errors. Finally, though the visual prompts were clearly visible on the computer monitor mounted beside the mirror, the audio prompts were played through speakers mounted in the drop ceiling of the washroom. The exhaust fan that turned on at the same time as the light may have prevented some participants from hearing the audio prompts. Previous COACH trials had been conducted in a hospital setting where the washrooms did not have exhaust fans. Though the current study focused primarily on the technical capabilities of the
COACH, subsequent trials that explore the system’s ability to promote task completion must address such physical variables

3.6.5 Future work and required areas of improvement

Based on our previous research investigating the user needs for ATC which support ADL completion [28] and the results of this study, future work will be guided by a House of Quality approach [57, 58] – a systematic methodology used to convert customer needs into design specifications. The results presented here suggest that a real-world, community-based deployment of an unobtrusive ATC in general, and the COACH system in particular, requires robust and accurate user tracking without individual user calibration. Individual user calibration typically requires expert knowledge of the ATC preventing general users, typically informal caregivers, from effectively setting up a system. Additionally, calibration requires additional time and maintenance, potentially shifting the burden of caring for a person with dementia from the actual care task to the support of a technological device. Critical to the concept of robust and accurate user tracking, and based on the current capabilities of the COACH system presented in this study, is the understanding that using skin colour to track a user’s hands in space may not be sufficient to properly identify the user’s interactions with the environment.

Accordingly, the addition of a third dimension to capture image depth will be examined, as well as methods of understanding more subtle motions of the user’s hands such as gestures, grip, and rotation. Modifications to the Kinect sensor software development kit [108] skeletal tracking in order to track arms, hands and fingers from an overhead camera independent of colour is a key method that will be explored. In particular, the use of the open source OpenNI module and PrimeSense NITE middleware [109] allow full access to the Kinect Hardware, and have comparable software libraries that mirror the capabilities of the Kinect software development kit. Access to the Kinect webcam will allow continued use and development of the existing COACH flocking techniques, as well as different 2D approaches such as optical flow and skeletal tracking. The addition of a depth map from the Kinect sensor will allow the exploration of 3D methods of hand gesture recognition, as well as the integration of our existing 2D methods with depth to know, for example, when the users’ hands are moving over an object versus actually touching an object. These hand tracker modifications will also help improve the system’s ability
to detect user interactions that occur very quickly, and will reduce the system’s dependence on threshold timers used to detect the occurrence of an interaction.

Prompts were often poorly timed, and suggested completion of incorrect task steps. Yet users had almost no response to the prompts – positive or negative. The content and delivery method of the prompts will be investigated to improve the system’s ability to interact with the users. The limited response participants had for any provided prompts will also be investigated.

The system’s ability to support additional daily tasks, such as tooth brushing, will also be considered, as well as the COACH’s capability to simultaneously support these multiple tasks. However, specification of the POMDP policy represents a substantial commitment of time and expert knowledge, and the simultaneous management of these policies poses a challenging problem. Accordingly, the continued development of a rapid POMDP policy specification technique [110], called a Syndetic Assistance Process (SNAP), is critical to the extension of the COACH beyond hand washing. Furthermore, the development of a hierarchical POMDP controller will also be required to manage simultaneous task monitoring. These improvements are considered vital to the project’s overall goal of testing the efficacy of the system in several long-term deployments in the homes of older adults with dementia.

### 3.7 Conclusions

This study represents the first attempt at assessing the COACH system outside of the controlled lab or hospital setting and reflects the largest number of patients with dementia interacting with the COACH system in an unsupervised state to date. The questions posed at the outset of the project were sufficiently addressed by this study to inform next steps in technical refinements of the system which will pave the way toward successful implementation in the home. The majority of patients approached regarding the current study were interested in participating. They accepted the COACH washroom set up without any concern or negative reaction to the presence of a computer monitor beside the sink or a camera above the sink. Similarly, non study participants using the same washroom were not troubled by the COACH apparatus. These findings alone suggest that installation of the COACH in the home setting would be well accepted by the target population and their caregivers. However, this first real-world deployment of the COACH, an ATC designed to support ADL completion, revealed that the system requires
substantial modification in order to improve its performance. This study has allowed us to answer our initial research questions:

1. *How accurately can the COACH track the actions of multiple older adults with dementia through the task of hand washing without individual calibration or configuration in an unsupervised clinical environment?* Based on these results, a general configuration, compared to previous studies using user-specific configurations, can be used to track multiple users’ physical hand positions. However, simply knowing the position of the users’ hands is likely not sufficient to understand what actions they are performing and what objects in the environment they are interacting with. The result is that more contextual and subtle information is required to understand what the users’ hands are actually doing within the environment, such as gesture recognition, grip and rotation of the hands.

2. *How appropriately does the COACH respond to an older adult with dementia’s actions while supporting hand washing in an unsupervised clinical environment?* Based on system performance evaluated using raw hand tracker data, the COACH did not track user progress well and did not provide appropriately timed prompts. The poor performance was almost entirely due to the user tracking, which in previous studies was calibrated to individual users whereas in this study the system was calibrated to a general user. Accordingly, with its current hand tracker the COACH does not respond appropriately to general users’ actions in an unsupervised, real-world deployment. However, when the user tracking data captured during the trials were replaced with manually labeled ground-truth data the system identified the users’ progress through the hand washing task with near perfect accuracy. This suggests that with a more robust and accurate hand tracker the COACH will respond more appropriately to users’ actions, even in an unsupervised real-world environment.

3. *What technical challenges will the COACH face running continuously in an unsupervised state in a real-world clinical environment?* Throughout the four month trial the system encountered two technical difficulties: faulty wiring in the study activation button, and a software error with an automatic backup utility that temporarily consumed the free computer memory effectively shutting down the
system. Both of these technical issues, though they indeed affected the potential number of trials that could have been included in the study, were not directly related to the core COACH system suggesting that the device is technically ready for long-term, in-home deployment.

In general the results presented in this paper further support the application of ATC with people with dementia. However, based on comparisons with previous COACH trials which were similar to other trials conducted with this class of ATC, significant challenges may exist in a real-world deployment such as variations in different ways activities can be performed, and the inability to know exactly who is using the system. Accordingly, ATC developers must conduct long-term trials in real-world settings to fully understand the capabilities of these devices in order to improve their utility for people with dementia.

3.7.1 Acknowledgements

The study was funded by grants from the Alzheimer Association (ETAC), Canadian Institute of Health Research (Operating Grant), and student funding from the NSERC CREATE CARE program.
Chapter 4

4 Development and evaluation of a hand tracker using depth images captured from an overhead perspective

4.1 Publication Citation


4.2 Abstract

We present the development and evaluation of a hand tracking algorithm based on single depth images captured from an overhead perspective for use in the COACH prompting system. We train a random decision forest body part classifier using ~5,000 manually labeled, unbalanced, training images. The classifier represents a random subset of pixels in each depth image with a learned probability density function across all trained body parts. A local mode-find approach is used to search for clusters present in the underlying feature space sampled by the classified pixels. In each frame, body part positions are chosen as the mode with the highest confidence. User hand positions are translated into hand washing task actions based on proximity to environmental objects. We validate the performance of the classifier and task action proposals on a large set of ~24,000 manually labeled images.

Index Terms—Ambient intelligence, Assistive technology, Context-aware sensing, Decision support systems, Smart homes.

4.3 Introduction

In this paper we describe the development and testing of a new hand tracker for the COACH automated task prompting system [12, 44] using depth images captured using a Kinect [108] sensor mounted above a washroom sink. The COACH [11, 12] is an automated prompting
system designed to facilitate independent completion of daily activities (e.g., hand washing) by older adults with dementia. The system employs an automated planning algorithm, called a partially observable Markov decision process (POMDP) [100] to infer task progression. The hand washing task is divided into five fundamental steps or behaviors [12], where some steps must be completed in an ordered sequence while others can be completed in any order. The POMDP translates the primitive actions of users into a measure of their progression through the task, allowing for ordered and unordered step completion, while accounting for the partial observability, or unreliability, of the sensor readings identifying the users’ actions.

For the task of hand washing, data input to the POMDP planning module is the user’s hand locations relative to environmental objects. Hand locations can take on one of six positions: in the sink, adjusting the faucets, rinsing in the water, using the soap, using the towel, or away from the sink [44]. To determine these positions, the COACH employs an overhead computer-vision tracking algorithm that identifies the hand positions in two dimensions [49]. Prior to using the system each of the six environmental positions is statically defined. The hand positions are then translated to one of the six environmental positions based on their proximity to the predefined environmental regions in the image.

The COACH hand-tracking system has progressed through three technical generations of development [11, 37-47] to its current state that employs a color-based flocking algorithm [46]. A controlled, clinical evaluation of the COACH system, utilizing the color flocking algorithm, suggested that color-based hand tracking was effective and that the system could follow the actions of its users through the hand washing task [11]. These results motivated a needs assessment study to understand the role of the COACH in the lives of older adults with dementia and their caregivers, toward an ultimate real-world deployment [28]. To ensure the COACH was satisfying these needs, the system was deployed in an unsupervised state in a dementia treatment facility in Toronto, Canada[12]. This deployment revealed that color-based hand tracking could not accurately track the actions of a user if the user was not known a priori. Additionally, the study showed that color-based tracking could not easily separate different body parts that were similar in color. Examples of this include times when a person was wearing skin-colored or sleeveless clothing, or was partially or fully bald. These conditions are particularly challenging when considering that the images are captured from an overhead perspective, where often only the head, shoulders, arms and hands of the users are visible. Most notable, however, was that the
color-based hand tracking could not disambiguate the hands passing over objects in the environment (e.g., taps, soap dispenser) versus the hands actually interacting with the objects. Unreliable hand tracking significantly affected the overall performance of the COACH system as a whole [12]. However, the trials were simulated using manually constructed ground-truth hand tracking data. The simulations resulted in excellent overall system performance [12], suggesting that the color-based hand tracking was restricting the overall performance of the COACH.

To ensure that improvements to the hand tracking would ultimately fulfill the needs of the users of COACH a House of Quality (HOQ) approach [57, 58] was employed. HOQ is a systematic methodology used to convert customer needs into design specifications. Using an HOQ approach, the needs of the users of COACH [12] were weighed against the technical capabilities of the COACH in a real world deployment [28]. The results of the HOQ analysis motivated the development of a more accurate and reliable tracking approach that would improve the overall performance of the COACH system.

4.3.1 Related Work

The challenge of obtaining reliable, unobtrusive human body tracking data is not unique to the COACH system, affecting diverse areas such as gaming, human-computer interaction, and health care [59]. The development of the Kinect [108] sensor and software, largely in response to increased gaming requirements, provided an approach to human posture recognition and tracking that overcame a substantial number of limitations experienced by previous tracking systems [59]. The approach utilized depth imaging to accurately provide uninitialized (frame by frame), three-dimensional body part proposals that were largely color, texture, shape and lighting invariant [111]. The methodology forms a core component of the Microsoft Kinect [108] gaming platform which has been successfully deployed in the homes of many users around the world.

The methodology of Shotton et al. [59] has been used directly to classify full-frame hand and finger poses [112]. Furthermore, proposed improvements to various elements of the approach, geared toward increased classification accuracy and speed, have been discussed [113, 114]. Kohli and Shotton [115] have even extended the earlier work of Shotton et al. [59] to remediate problems associated with the approach.
Depth-based tracking has not been restricted to extensions of the methodology of [59]. Tracking using range cameras (e.g., structured light like the Kinect [108] and time of flight [see 116 for details]) have been used to successfully identify human poses and positions [117, 118], and distinguish hand poses from a fixed frontal [112, 119] and egocentric [120] perspective. Depth-based tracking has also been used from a fixed frontal perspective to identify hand and head positions [121], body positions [122], upper body segmentation [114], and full joint and body part positions [111, 113]. The task of body part classification from a frontal perspective became significantly easier through the release of the Kinect SDK and the development of open source APIs [e.g., 109].

4.3.2 Existing Tracking Approaches and the COACH

Three significant limitations prevent the application of existing depth-based hand and body tracking approaches to the COACH. The first limitation is that the COACH system utilizes an overhead, birds-eye tracking perspective to ensure an unobstructed view of its users and unobtrusive installation. The COACH system is targeted at older adults with dementia who are cognitively compromised requiring that any installed hardware remain out of reach. Furthermore, caregivers of older adults with dementia have indicated that any assistive technologies must integrate into the environment to reduce the likelihood of stigmatization [28]. To our knowledge, existing approaches have not attempted part tracking and recognition from a fixed overhead perspective using only depth imaging. Lei et al. [123] provided a proof-of-concept of a combined colour and depth approach to a kitchen task using an overhead depth camera, reporting preliminary performance metrics such as a 60% object recognition accuracy, 80% object functionality detection, and 75% action recognition (based on hand trajectories). Rather, most approaches utilize a frontal perspective [59, 113, 114, 117-119, 122], and/or perspectives unique to a particular application [112, 114, 119]. The work of Shotton et al. [111] is most similar to our approach, and achieved a part tracking accuracy of 73.1% over 31 body parts and 900k synthetic video frames. In [114], 7 body parts were tracked with 87.9% accuracy using 500 validation frames, and [124] reported 75.4% accuracy tracking 10 parts over 9000 frames. The second limitation is that many existing tracking approaches require the entire object to be in the scene in order to initialize the tracking model. In the case of the COACH system, users are often only partially in the scene when the hand washing task begins, specifically when washrooms are small. The final limitation preventing the use of existing tracking approaches is that existing
approaches implement tracking from either a global or a local perspective. Global perspectives require fully labeled training data over all body parts and typically require substantial training sets to provide reliable part tracking over the entire range of possible human motions [e.g., 59]. The generation of large, fully labeled data sets is time consuming and prone to labeling error. Local perspectives, on the other hand, focus on specific challenges such as hand or finger poses, or the reduction of all possible body positions to a small number of predefined poses [e.g., 117, 118]. The task of hand washing, moreover, is inherently variable, preventing the direct use of poses or posture as an indication of task progression. However despite existing limitations, the work proposed by Shotton et al. [59] presented a general framework for training a body part and joint classifier using single depth imaging.

4.3.3 Contributions

To our knowledge, existing tracking by detection approaches have not attempted uninitialized frame-by-frame part tracking from a fixed overhead perspective using only depth images. Our main contribution is the development of an uninitialized frame-by-frame overhead hand tracking methodology using only single depth imaging with application to the COACH prompting system. In the process, we have developed a method of training an uninitialized, frame-by-frame classifier, based on the work of Shotton et al. [59], using unbalanced training data captured from an overhead perspective. These unbalanced training data consist of some body parts with a small number of labeled image pixels relative to the total number of labeled pixels in the training image. We then show that the classification accuracy of the part tracking is sufficient to integrate directly into the COACH system as a hand tracker without a temporal or kinematic model. Finally, we validate the COACH’s ability to track the task-based activities of the participants through the hand washing task using the depth-based hand tracker on a large set of supervised trials completed in a functional washroom.

Through this work we seek to answer the following four research questions related to the new depth-based overhead hand tracker trained using unbalanced training data:

1. What classification performance can we achieve using a random decision forests classifier trained with unbalanced training data compared to the previous colour-based tracker?

2. What training parameters impact the overall performance of the classifier?
3. What is the mean average precision of an optimal classifier when proposing hand positions in three dimensions compared to ground-truth hand positions?

4. How accurately can the COACH track the completion of task-based activities of users through the task of hand washing in a real washroom?

**4.4 Method**

The research of Shotton et al. [59] provided the basis for the approach we use to classify individual body parts from a single overhead depth image on a per-frame basis. The acquisition and development of our unique training and validation data are captured from a fixed overhead perspective. We first generate a random decision forest using a simple depth feature to provide intermediate multiclass probability density functions (PDF) for each sampled image pixel. We then propose final body part positions by aggregating the information contained in the underlying PDF. We evaluate the performance of the intermediate and aggregate classifiers and optimize key training parameters. The optimal parameters are used to train a final decision forest resulting in a new depth-based hand tracker. We evaluate the final hand tracker against ground-truth hand positions and integrate the tracker into the COACH system. Finally, we evaluate the COACH’s ability to track the task-based activities of the participants using a set of validation images.

**4.4.1 Data Collection**

Study participants were recruited from a pool of researchers associated with Toronto Rehabilitation Institute [62] on a voluntary basis. An overhead Kinect sensor recorded depth and RGB images of participants washing their hands in a fully functional washroom located in Toronto Rehabilitation Institute’s HomeLab [61]. Both depth and RGB images were captured at 30 frames per second and a resolution of 640 x 480. A random subset of the recorded trials was removed from the data set and used as training data for the classifier. A second random subset, independent of the first, was also removed from the data set and used as a holdout image set to measure the performance of the classifier. The remaining data were used to validate the performance of the COACH system.

Before each trial a static background mask was created by calculating the per-element average of all valid pixels (depth value returned by sensor) in the twenty depth frames immediately
preceding the trial. The foreground from each depth image in the training data set was then isolated from the background. Any invalid pixels (depth value not returned by sensor) in the training image set remained invalid in the foreground image. Any pixels that were valid in the training set but invalid in the background mask remained valid in the foreground image. Each foreground image pixel was assigned to a class $p = \{\text{left hand}; \text{ right hand}; \text{ head}; \text{ body}\}$. Only the left hand, right hand and head were directly labeled; any remaining pixels in the foreground were assumed to be body. The labeled images composed the set of images used to train the classifier. The same foreground extraction and labeling methodology was employed to create the independent set of holdout images used to evaluate the classifier performance. The ground-truth centers for each part in each holdout image were calculated as the center of mass of each body part. Figure 4.1 shows a sample set of captured and processed images.

![Sample images](image)

**Figure 4.1:** Sample images captured from the overhead perspective. Top left: background image; Top right: depth image; Bottom left: labeled foreground image; Bottom right: RGB image.

For each validation image, the participant’s action was assigned one of a mutually exclusive set of $action = \{\text{walking}; \text{ washing hands}; \text{ drying hands}; \text{ turning}\}$. Walking was defined as any frame where the participant was approaching or leaving the washing area along the normal path from the door of the washroom to the sink area. Washing hands was defined as any frame where the participant was facing the sink area with any body part over the counter. Drying hands was defined as any frame in which the participant’s hand was in contact with the towel and not turning. Turning was defined as any frame where the participant was not facing the sink area, and not walking. Furthermore, for each image, the participant’s task activity was identified as one of $activity = \{\text{away}; \text{ soap}; \text{ tap}; \text{ water}; \text{ sink}; \text{ towel}\}$ according to our previous work [44].
Finally, for each image the center of each part was labeled. An annotation tool was developed to label the part centers in the validation images. To reduce annotation time, the center of each part was initially estimated by the classifier and manually corrected if required.

### 4.4.2 Intermediate Multiclass Classification

An intermediate multiclass classification was implemented, based on the work of Shotton et al. [59], which represented each of a random subset of pixels across a depth image $I$ as a learned probability density function across all body parts. The background segmented and labeled training images $I_t$ were used to train a set of random decision trees. A set of training samples $S = \{I_t, x\}$ were determined by randomly proposing a set of $N$ pixels $x$ from each training image. Then, a set of offset vectors $\theta = (u, v)$ and thresholds $\tau$ were randomly generated with maximum magnitude $\theta_{max}$ and $\tau_{max}$ respectively. Together, the offset vectors and thresholds created a set of splitting criteria $\phi = (\theta, \tau)$. A simple depth feature from [59] was then defined as

$$f_\phi(I_t, x) = d_{I_t} \left( x + \frac{u}{d_{I_t}(x)} \right) - d_{I_t} \left( x + \frac{v}{d_{I_t}(x)} \right)$$

where $d_{I_t}(x)$ was the depth at pixel $x$ in image $I$. The resulting depth feature was then calculated as the difference between depth probes at the points identified by each offset vector $u$ and $v$. To ensure that the offset vectors were invariant to depth, vectors $u$ and $v$ were normalized by the depth of pixel $x$ providing the same world-space offset regardless of the distance of the pixel from the sensor. In the case where the pixel identified by an offset probe was located outside on the background or off the image, the depth probe returned a large positive value greater than the maximum depth possible from a valid pixel.

A set of binary decision trees was trained using training sets unique to each tree. Training started at the root node with all training samples, and each node $i$ of the tree was determined according to the following steps:

1. Divide the set of training samples $S_i$ at node $i$ into left and right subsets for each split candidate/threshold pair $\phi$ according to the depth feature, where

$$S_{i,L}(\phi) = \{I_t, x\} \ | \ f_\phi(I_t, x) < \tau; S_{i,R}(\phi) = S_i \setminus S_{i,L}(\phi)$$

2. Find $\phi_{max}$ which maximizes the gain according to

$$g(\phi) = H(S_i) - \frac{|S_{i,L}|}{|S_i|} S_{i,L}(\phi) - \frac{|S_{i,R}|}{|S_i|} S_{i,R}(\phi)$$
where $H(S_i)$ is the entropy of the normalized PDF of $S_i$.

3. If $g(\theta_{\text{max}})$ is greater than a minimum gain $g_{\text{min}}$ and the tree depth is less than $D_{\text{max}}$, recursively repeat steps 1 to 3 for the left $S_{i,1}(\theta_{\text{max}})$ and right $S_{i,2}(\theta_{\text{max}})$ subsets.

4. If $g(\theta_{\text{max}})$ is less than a minimum gain $g_{\text{min}}$ or the tree depth equals $D_{\text{max}}$, store the probability density function $P_t(p|I_i,x)$ across all parts.

Decision trees tend to overfit data, particularly when training sets are small [59]. Accordingly, the decision trees were used in an ensemble of $T$ trees to form a decision forest. For a given pixel $x$, each tree was traversed until a leaf node was reached. The final PDF was given as the average of the individual PDFs provided by each tree.

We measured the impact of six training parameters on the multiclass classification, or per-pixel, performance of the intermediate classification: 1) the number of training images; 2) the maximum tree depth $D_{\text{max}}$; 3) minimum cutoff gain $g_{\text{min}}$; 4) maximum magnitude of the offset vectors $\theta_{\text{max}}$; 5) the number of decision trees ($T$); and 6) the number of samples per image $N$. A forest was trained by varying each parameter while holding all other parameters constant. A confusion matrix was generated for each forest between the most likely predicted body part and the ground-truth part using a set of labeled holdback images. Similar work by [59] and recommendations for multiclass classifier performance [125] measurement suggest the use of the Overall Success Rate (OSR), or simply the sum of the diagonal of the confusion matrix. This measure works well for balanced data sets, but our data sets are proportionally imbalanced. Accordingly, the intermediate part classification performance was calculated as the unweighted average recall (UAR) [126] of each class in the confusion matrix to account for the unbalanced classes.

4.4.3 Aggregate Part Proposals

For each image in the holdback set ($I_h$), the random decision forest provided an intermediate classification of the set of $N$ pixels, where each pixel was assigned a PDF across all body parts. The PDF of each sampled pixel was taken as a representative point of an underlying density function or feature space. Final body part position proposals provided the most likely position of each part by searching for clusters present in the feature space. The coordinates of each pixel $x$ were transformed from projective space into pixel $\hat{x}$ in world space creating a 3D density
function for each body part. Initial estimates for each part \( \hat{x}_p^0 \) were identified as any pixel with a part probability above a threshold \( \phi_p \)

\[
\hat{x}_p^0 = \{ l, \hat{x} \mid P(p \mid l, x) > \phi_p \} \tag{4}
\]

The maxima \( \hat{x}_p \) of each density function were then determined for each part using a mean shift mode seeking algorithm [127] with a weighted Gaussian kernel from [59]

\[
\hat{x}_p^{t+1} = \frac{\sum_{i=1}^{N} P(p \mid l, \hat{x}) \cdot d_i(\hat{x}) \cdot e^{-\frac{|x_p - x_i|^2}{h_p^2}}}{\sum_{i=1}^{N} e^{-\frac{|x_p - x_i|^2}{h_p^2}}} \tag{5}
\]

where \( h_p \) is a per-part bandwidth that influences the convergence rate and final number of part proposals. The confidence of each mode was given as the sum of the contributions of each pixel to the final mode

\[
\alpha_p = \sum_{i=1}^{N} P(p \mid l, \hat{x}) \cdot d_i(\hat{x}) \cdot e^{-\frac{|x_p - x_i|^2}{h_p^2}} \tag{6}
\]

The mode with the highest confidence was used as the final part position proposal.

We evaluated the aggregate part proposal performance of the mode-finding algorithm over the set of holdback images given a starting threshold \( \phi_p \). The distance between all proposed modes for each part was compared in world-space to the ground-truth part centers. For each part, a confusion matrix was constructed between the ground-truth part position and the proposed position. A true positive (TP) was scored for the first mode within a minimum per-part distance \( \Delta_p \) of the ground-truth part center. Any other mode within \( \Delta_p \) or any mode outside \( \Delta_p \) was scored as false positive (FP). A true negative (TN) was scored if the part did not exist in the image and the classifier did not propose any modes. A false negative (FN) was scored for all modes proposed in an image that did not contain a corresponding part.

Precision-recall (PR) curves were generated for each body part as a function of mode find starting threshold \( \phi_p \) over the holdback image set. The average precision (AP) was calculated for each part. The aggregate part proposal performance was calculated as the mean average precision (mAP) across all parts. The optimal starting threshold \( \hat{\phi}_p \) was determined for each part using the Equal Error Rate (EER) – the point where precision and recall were equal.
4.4.4 Task-based Hand Positions

Modes were provided by the classifier for each part for each image in the validation image set using the final decision forest and starting thresholds ($\Phi_p$). The mode with the highest confidence ($\alpha_p$) was used as the final part proposal. A confusion matrix was constructed between the ground-truth part position and the proposed position in world space for each part. For each image, a true positive (TP) was scored if the proposal was within a minimum per-part distance $\Delta_p$ of the ground-truth part center. If the proposal was outside $\Delta_p$ it was scored as false positive (FP). A true negative (TN) was scored if the part did not exist in the image and the classifier did not propose any modes. A false negative (FN) was scored if a position was proposed for the part in an image that did not contain a corresponding ground-truth part. An $F_\beta$ measure was used to evaluate the classifier performance, where $\beta$ is a factor weighting the importance of precision and recall.

$$F_\beta = \frac{(1+\beta^2) \cdot \text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$ (7)

An $F_{0.5}$ measure was calculated for each part $p$ in each validation category $\text{action}$ the validation data set. The scaling factor of $\beta = 0.5$ was selected to prioritize the correctness of any classified parts. The final $F_{0.5}$ measure per part was calculated as the mean measure over all trials.

For each frame in each validation trial, the left and right hand positions (determined by the mode with the highest classification confidence) were converted to one of the six participant task activities defined by $\text{activity}$. Each $\text{activity}$ was characterized by a pre-defined spheroid in world space. When a hand entered a region defined by a spheroid and persisted for three or more consecutive frames the associated $\text{activity}$ was considered active for use by the POMDP decision policy.

A confusion matrix was created for each validation trial to measure the COACH’s ability to track each hand washing step = {turn on water, get soap, rinse hands, turn off water, dry hands} [11]. A step was considered complete if the associated $\text{activity}$ occurred in sequence. A true positive (TP) was scored when a user completed a task activity and the COACH correctly identified the activity as complete. A true negative (TN) was scored if a task activity was not completed and the COACH did not indicate the activity was completed. A false positive (FP) was scored if a
task activity was not completed but the COACH identified the activity as completed. A false negative (FN) was scored if a task activity was completed but the COACH did not indicate it as correctly completed. An $F_1$ measure from (7) was calculated for each trial, equally weighting precision and recall. The overall performance of the system was measured as the average $F_1$ measure over all trials.

To provide a performance comparison, the left and right hand positions, as determined using the colour-based flocking tracker, were then converted to one of the six participant activities defined by activity. The predefined activity spheroids were flattened to a two-dimensional circle in projective space. Again, if a hand entered a region defining an activity and persisted for three or more consecutive frames the associated activity was considered active. The completion of a task step and associated scoring was then repeated as with the depth tracking trials, resulting in an overall performance $F_1$ measure for the colour-based tracker.

## 4.5 Results

### 4.5.1 Data Collection

Depth images were captured over fifty-one (51) trials where thirty (30) unique participants performed a real hand washing task. Trials were conducted in a fully functional washroom located in Toronto Rehabilitation Institute’s HomeLab [61]. Participants were associated with the research team and were included in the study on a voluntary basis. A total of 34,999 frames test data were collected. After removing all frames that did not contain a foreground image, a total of 28,763 frames of useable data remained. The training data set ($I_t$) was comprised of eight (8) trials and a total of 4,971 frames of depth data. The holdout data set ($I_h$) included two (2) trials and a total of 854 images. The validation data set ($I_v$) included the remaining forty-one (41) trials and a total of 22,938 images.

### 4.5.2 Intermediate Multiclass Classification

Initial decision trees were trained using the following default parameters: 100 thresholds ($\tau$), 3000 offset vectors ($\theta$), maximum offset ($\theta_{\text{max}}$) 500, minimum gain ($g_{\text{min}}$) = 0.05, 4000 samples per image ($N$), maximum tree depth ($D_{\text{max}}$) = 20 (root at depth 0, PDF at depth 19), three trees ($T$), and 4971 training images ($I_t$) split across the number of trees. Final decision tree parameters and the number of trees in the forest were identified by varying a single parameter and setting all
others to the default value. Figure 4.2 shows the results of the parameter optimization training. Investigating each chart suggests that the forest achieves a maximum accuracy at $D_{\text{max}} = 12$, $I_i = 4971$, $\theta_{\text{min}} = 0$, $\theta_{\text{max}} = 250$, $N = 3000$ and $T = 3$. Note that while increasing the number of trees yields increased forest classification accuracy, increasing the number of images in each tree resulted in a larger increase.

![Figure 4.2: Unweighted average recall (UAR) versus training parameter.](image)

4.5.3 Aggregate Part Proposals

The holdout image set ($I_h$) was used to generate precision-recall (PR) curves for each part over the range of starting threshold probabilities $0 \leq \varphi_p \leq 1$. Table 4.1 shows the average precision (AP) and optimal starting threshold ($\bar{\varphi}_p$) for each part, and a mean average precision $\text{mAP} = 0.846$. Notable is the high starting threshold value $\bar{\varphi}_p = 0.95$, suggesting that the intermediate classification of the “head” is reliable.

| TABLE 4.1: AGGREGATE PART PROPOSAL PERFORMANCE PARAMETERS AND OUTCOMES |
|---------------------------------|-----------------|-----------------|-----------------|
| 25%                            | 35%             | 45%             | 55%             |
| 2814 14 20                     | 67%             | 68%             | 69%             |
| 1000 3000 5000                 | 70%             | 71%             | 72%             |
| 1723 4971 Images               | 73%             | 74%             | 75%             |
| 0 0.1 0.2                      | 76%             | 77%             | 78%             |
| 0 250 500                      | 79%             | 80%             | 81%             |
| 123                            | 82%             | 83%             | 84%             |
| 2500 5000                      | 85%             | 86%             | 87%             |
| 3500                            | 88%             | 89%             | 90%             |
| (a) Maximum tree depth ($D_{\text{max}}$) | (b) Number of training images | (c) Minimum cutoff gain (gmin) | (d) Maximum offset ($\theta_{\text{max}}$) | (e) Samples per image ($N$) | (f) Trees in forest ($T$) |
4.5.4 Task-based Hand Positions

Table 4.2 shows the mean $F_{0.5}$ measure for each body part $p$ within each action category, along with the number of frames within each action category from the set of validation and training data sets. The mean $F_{0.5}$ measure was calculated as the average $F_{0.5}$ score over all validation trials for each body part using the EER starting thresholds ($\hat{\phi}_p$). Over all validation frames, the $F_{0.5}$ measure for the left hand, right hand, and head were 0.713, 0.789 and 0.890 respectively. In all cases classification of the head outperforms classification of the left and right hands except in the walk action where the head is often not in the scene. Additionally, classification of all body parts substantially underperforms during the “walk” and “turn” actions compared to the “wash” and “towel” actions.

A disproportionately large number of training images were in the “wash” action, represented by 697 (81.6%) of all training images. Conversely, only 32 (3.8%) and 29 (3.4%) of all training images were in the “turn” and “towel” action categories respectively. In comparison, a total of 15,613 (68.1%) validation images were in the “wash” action, while 1,505 (6.6%) and 2,968 (12.9%) were in the “turn” and “towel” actions respectively.

<table>
<thead>
<tr>
<th>Action</th>
<th>Body Part (p)</th>
<th>Mean $F_{0.5}$ Measure</th>
<th>Validation Frames (%)</th>
<th>Training Frames (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walk</td>
<td>Left Hand</td>
<td>0.219</td>
<td>2852 (12.43%)</td>
<td>96 (11.24%)</td>
</tr>
<tr>
<td></td>
<td>Right Hand</td>
<td>0.315</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Head</td>
<td>0.305</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wash</td>
<td>Left Hand</td>
<td>0.879</td>
<td>15613 (68.07%)</td>
<td>697 (81.62%)</td>
</tr>
<tr>
<td></td>
<td>Right Hand</td>
<td>0.908</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Head</td>
<td>0.997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turn</td>
<td>Left Hand</td>
<td>0.082</td>
<td>1505 (6.56%)</td>
<td>32 (3.75%)</td>
</tr>
<tr>
<td></td>
<td>Right Hand</td>
<td>0.199</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Head</td>
<td>0.497</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Towel</td>
<td>Left Hand</td>
<td>0.355</td>
<td>2968 (12.94%)</td>
<td>29 (3.4%)</td>
</tr>
<tr>
<td></td>
<td>Right Hand</td>
<td>0.699</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Head</td>
<td>0.994</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>Left Hand</td>
<td>0.713</td>
<td>22938 (100%)</td>
<td>854 (100%)</td>
</tr>
<tr>
<td></td>
<td>Right Hand</td>
<td>0.789</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Head</td>
<td>0.890</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The performance of the COACH depth tracker is compared to the tracking performance in the unsupervised trials [12] and using the colour-based tracker over the validation image set in Table 3. Based on the data presented in Table 3, the precision of the depth-based tracking system is 0.994 and the recall is 0.938. The resulting F1 measure score is 0.965.

Table 4.3: The performance of the COACH while monitoring the task-based activities of the participants over the set of validation trials.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Current Study</th>
<th>Previous Trials with Colour Tracker[12]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Depth Tracker</td>
<td>Colour Tracker</td>
</tr>
<tr>
<td>True Positive</td>
<td>180</td>
<td>158</td>
</tr>
<tr>
<td>True Negative</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>False Positive</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>False Negative</td>
<td>12</td>
<td>34</td>
</tr>
<tr>
<td>Precision</td>
<td>0.994</td>
<td>0.981</td>
</tr>
<tr>
<td>Recall</td>
<td>0.938</td>
<td>0.822</td>
</tr>
<tr>
<td>F1 measure</td>
<td>0.965</td>
<td>0.894</td>
</tr>
</tbody>
</table>

4.6 Discussion and Conclusion

The purpose of this study was to develop and integrate a novel hand tracker into the COACH prompting system using depth images captured from an overhead perspective. Existing depth-based body part, pose and posture tracking methodologies were not suitable for this application. The work of Shotton et al [59] was adaptable to our application, however using the approach required a substantial amount of training data in order to develop a functional classifier.

We captured a large set of depth and RGB images, totaling 28,763 usable frames of video. However, manually creating fully labeled training and holdback data sets was time consuming, taking three to five days for each trial. To reduce manual labeling time, we employed a partial labeling approach where we uniquely labeled body parts with application to the COACH prompting system, generically labeling all other pixels in the foreground image as body. To further reduce the data preparation time, we simply labeled the part centers for the validation image set.

Using unbalanced data to train decision trees in a random decision forest provided comparable results to the foundational studies we used to develop our methodology [e.g., 59, 113, 114]. The intermediate multiclass classification accuracies we achieved were comparable to the results
presented in those studies. This is likely due to the fact that the Shannon entropy used to determine the optimal split criteria is robust to proportionally unbalanced probability density functions. These results suggest that fully labeled training data are not necessary when training a random decision forest. Furthermore, when evaluating the performance of a decision forest, the unweighted average recall [126] is an appropriate and effective measure for an unbalanced data set. Determining the effect of the unbalanced training data quantitatively against a more balanced set would further substantiate this claim.

Varying the values of training parameters and evaluating the performance of the resulting intermediate classifier clearly showed the optimal values for those parameters or the trend caused by varying the parameters. In the case of the maximum tree depth ($D_{\text{max}}$), a reduction in classification accuracy was visible in Figure 4.2(a) with depths beyond the optimal value. This is likely a reflection of the tendency decision trees have to overfit to the data. A similar trend is visible when considering the maximum offset ($\theta_{\text{max}}$) in Figure 4.2(d). Shotton et al. [59] propose that this is likely due to overfitting to an increase spatial context. We also propose that increasing the maximum magnitude of the offset vectors increases the likelihood that an offset vector will generate an offset probe outside the bounds of the image. All offset probes landing outside the image result in the same probe value. Increasing the number of vectors consistently reaching outside the image functionally reduces the total number of unique offset vectors usable by the training routine. Increasing the number of training samples ($N$), shown in Figure 4.2(e), beyond the optimal value also leads to a decrease in the classification accuracy of the system. This is largely due to overfitting the decision trees to the training data. A clear trend is visible in 4.2(b), indicating that an increase in the number of training images increases the classifier accuracy, which is in agreement with similar studies [59]. Finally, setting a minimum cutoff gain ($g_{\text{min}}$) exponentially reduces the accuracy of the classifier (Figure 4.2(c)). The advantage of increasing the minimum cutoff gain is that the memory penalty is reduced. Larger gains result in less complex decision trees. In the case of our forest the size of each decision tree is reduced by a factor of approximately $2^n$, where $g(1 \ldots n) = (0, 0.05, 0.1 \ldots n)$. However, the overall size of our trees, trained with four body parts, is relatively small. If the complexity of the trees is increased by classifying more parts or by adding additional training images, sacrificing accuracy to reduce the memory penalty may be worthy of consideration.
Aggregating the underlying PDFs from the intermediate classifier effectively provided final part position proposals. In particular, proposals for the position of the head were given with a mAP of 0.931. Interestingly, the optimal threshold to identify starting points for the mode find algorithm was 0.95 for the “head”, suggesting a high confidence in the intermediate classification for that part. In contrast, the starting thresholds for the “left hand” and “right hand” were 0.65 and 0.60 respectively, indicating a lower confidence in the intermediate classification. Regardless, the aggregate left and right hand position proposals still resulted in an average precision of 0.802 and 0.805 respectively. This is not surprising as the “head” is much less variable in terms of position, shape and size than the “left hand” and “right hand”. The depth tracker was able to track all parts with a mAP of 0.846 over the holdout image set. These results are comparable to the tracking accuracy of 0.879 reported by [114] over 500 real images, 0.731 by [111] over 8800 real and 900k synthetic images and 0.754 by [124] over 9000 real images. Furthermore, the depth tracker outperformed the colour tracker which achieved an accuracy of 0.771 over the set of holdout images.

The performance of the final hand tracker over the set of validation trials yielded highly variable results (Table 4.2). Overall, the system performed well based on high F0.5 measures for each body part. However, evaluating the F0.5 measures for each action of the participants shows that the classifier didn’t perform as well in some cases. In particular, the actions “walk” and “turn” performed poorly for all body parts. Additionally, classification of the “left hand” was not overly accurate during the “towel” action. This is explainable by investigating the proportion training and validation image frames in each action. A disproportionately large number of training image frames (81.62%) were in the “wash” action. Additionally, only 11.24% of all training image frames were in the “walk” action, with the remaining images split evenly between the “turn” and “towel” action. This disproportionate distribution among action categories resulted in a decision forest biased toward classifying body parts in the “wash” action. Indeed, the relatively decent performance of the classifier during the “towel” action is likely a coincidental result of the similarity between human motions during the “wash” and “towel” actions. The other “walk” and “turn” actions, in comparison are much more unique and as a result provided poorer classification performance. To improve the performance of the system during these under-represented actions, future work will look to increase the training image sets in these actions, as well as incorporating improved regression forest approaches such as in [128] and [129].
The COACH system performed exceptionally well at correctly identifying the participants’ activities over the 22,938 validation images. Over the forty-one (41) trials, participants could have potentially completed 205 task steps. Of the 192 completed steps the system correctly identified 180 as complete. Of the remaining thirteen (13) steps that were not completed, the system correctly identified twelve (12) as not completed. Considering the thirteen (13) incorrectly classified activities: seven (7) occurred because the classifier failed to correctly identify the locations of the hands for the duration of the activity; three (3) were the result of the hands completing the action outside the defined region; two (2) occurred when the hands were fully obscured from the camera’s view for the duration of the activity (e.g., user leaning over the sink); and the last was a system issue where several frames were not recorded properly preventing proper classification of the hands. A critical point to address when considering the success of the hand tracker is the paucity of False Positive results. Depth tracking was easily able to disambiguate the completion of certain task actions compared to the hands simply passing over action activity regions. False Positives triggered by the hands moving above activity regions was a significant contributor to poor performance in our previous colour-based tracker [12].

Despite measurable gains in intermediate classification through individual parameter optimization, our study did not investigate the potentially complex interactions between all the training parameters. Future work will look to further improve intermediate classification by concurrently optimizing training parameters. In particular, the concurrent optimization of the number of training images per tree, maximum tree depth and number of trees in the decision forest will likely have the largest impact on overall classification performance. Furthermore, the most significant gains are likely achievable by increasing the training data set – particularly by including more images in task actions that underperformed during the validation trials. The value of a larger training set is perhaps most supported when considering the results achieved by Shotton et al. [59], who used an enormous training set of 900,000 real and synthetic images. From a more practical perspective, in some instances a participant’s hands were occluded by the head (e.g., when a participant leaned over the sink) as a result of the overhead perspective. Consideration of different mounting locations (e.g., offset from center may yield more favorable results system performance results.
Notwithstanding the limitations of this study, our results strongly support the furthered development of a depth-based hand tracker for the COACH prompting system. To this end, we now readdress our initial research questions:

1. *What classification performance can we achieve using a random decision forests classifier trained with unbalanced training data compared to the previous colour-based tracker?* Unbalanced training data was shown to be an effective source of training data for a random decision forest. Similarly, the performance evaluation of decision forests was shown to be possible using the unweighted average recall metric.

2. *What training parameters impact the overall performance of the classifier?* We evaluated the impact of the maximum tree depth \(D_{\text{max}}\), the number of training images, the minimum cutoff gain \(g_{\text{min}}\), the maximum probe offset \(\theta_{\text{max}}\) and the number of training samples per image \(N\). Each of these parameters impacts the performance of the classifier. Optimal values were discovered for the maximum tree depth \(D_{\text{max}}\), the minimum cutoff gain \(g_{\text{min}}\), the maximum probe offset \(\theta_{\text{max}}\) and the number of training samples per image \(N\). When investigating the impact of the number of training images on classifier performance results suggested that an optimal value had not yet been, implying additional training images would improve performance.

3. *What is the mean average precision of classifier when proposing hand positions in three dimensions compared to ground-truth hand positions?* The final decision forest, trained using optimized training parameters, achieved a mean average precision of 0.846 over the *left hand, right hand, and head*, body parts.

4. *How accurately can the COACH track the completion of task-based activities of users through the task of hand washing in a real washroom?* The COACH tracker classified the actions of the participants with a resulting \(F_1\) measure score of 0.965 out of a maximum value of 1.

Future work will now look to address the limitations of the current study. A key component of this process will be to further test the efficacy of the new depth-based hand tracker through applications to different daily tasks. Implicit to this extension to additional tasks will be the validation of the tracker in different home environments (i.e., outside the washroom used to
evaluate hand washing performance) while simultaneously improving the classification ability by incorporating additional, diverse training data. Ultimately, the full integration of the new hand tracker into the COACH system will provide the foundation for additional, unsupervised, real-world efficacy studies of the COACH prompting system as an aid to older adults with dementia.

4.6.1 Acknowledgements

The study was funded by grants from the Alzheimer Association (ETAC), Canadian Institute of Health Research (Operating Grant), and the NSERC CREATE CARE program.
Chapter 5

5 Conclusion

An aging global population and an associated increase in the prevalence of dementia [2, 3] has created an opportunity for the development of a variety IAT that may help support the needs of this population [8]. The COACH [11, 37-47] is a prototype IAT designed to promote the independence of older adults with dementia while completing ADL (e.g., hand washing) in their homes. At the conception of this dissertation the COACH was a product of three iterations of development, each geared toward supporting the task of hand washing. Each version of COACH was developed to address technical limitations and deficiencies encountered with the previous prototype. Additionally, each supervised clinical trial [11, 37, 41], showed improvements in the number of task steps older adults with dementia could independently complete (i.e., without requiring the help of a human caregiver) while washing their hands. In some cases, the COACH was even shown to reduce the number of interventions required by a human caregiver [11].

This dissertation looked to extend the COACH system. The primary objective (PO) was to continue development of the COACH toward an unsupervised, real-world in-home deployment. The secondary objectives were: (SO1) to identify the needs older adults with dementia and their family caregivers have for IAT designed to support the completion of ADL in a home environment; and (SO2) to identify technical factors limiting the performance of the COACH in an unsupervised, real-world deployment in a clinical setting. Following my introductory chapter, Chapters 2, 3 and 4 presented three independent journal papers (two published and one in review) reporting on my work completed to satisfy my primary and secondary objectives.

5.1 Dissertation Contributions

5.1.1 Research Contributions

The three primary research contributions of this dissertation are:

1. A holistic user needs assessment (Chapter 2, SO1) [see 28];

2. An evaluation of the COACH in an unsupervised, real-world clinical deployment (Chapter 3, SO2) [see 12]; and
3. The development of an improved COACH system to simultaneously overcome the limitations identified in the unsupervised, real-world clinical deployment while addressing the needs of the systems’ users (Chapter 4, PO) [see 130].

I briefly review the contributions here and throughout I identify how each of my research questions have been answered in light of the objectives of the dissertation, and where relevant I identify how my findings are interconnected and bridge existing gaps in the literature.

5.1.1.1 Secondary Objective 1 (SO1): to identify the needs older adults with dementia and their family caregivers have for IAT designed to support the completion of ADL in a home environment

Incorporating the needs of older adults with dementia and their family caregivers into the process of designing and testing IATs can increase the likelihood of their acceptance, adoption and effectiveness [8, 27]. Unfortunately, the needs of users of IATs designed to support older adults with dementia are relatively unknown. The small number of published needs assessment studies typically focus on task-based needs, and yield results that are divided preventing generalization and in some cases leading to inconclusive data [9, 27]. In addition, the majority of people living with dementia reside in a home environment in Canada [2] and in almost every country worldwide [1]. Yet, little is known about the needs people with dementia have for an in-home IAT, particularly when considering the features and functions of such devices.

The first study (Chapter 2) adds to the literature through a 94-item assessment questionnaire designed to elicit the needs of older adults with dementia and their family caregiver through caregiver report, framed within the concept of IATs in a home environment. The questionnaire was constructed by adapting existing scales relevant to the study [e.g., 52, 53] and drawing prevalent themes from the literature [33, 55, 56]. A follow-up focus group with six (6) family caregivers of people with dementia living in the Greater Toronto Area, Toronto, Canada provided content validity to the questionnaire beyond the statistical methods used. A total of 106 participants voluntarily completed the questionnaire.

Respondents reported that older adults with dementia are able to at least partially complete fundamental ADL (e.g., eating, drinking, using the bathroom) with some level of independence (i.e., without the help of a human caregiver). However, respondents also indicated that the majority of older adults with dementia cannot participate in higher-level ADL (e.g., preparing
complex meals, cleaning the house). These findings support the development of prompting and cuing systems for fundamental ADL that leverage the remaining abilities of older adults with dementia. Respondents also suggested that support with daily tasks that involve an invasion of privacy (e.g., getting dressed, bathing, showering) or are typically completed independently (e.g., washing hands or brushing teeth) are most in need of automated support. Finally, respondents reported that any IATs integrated into the home environment must be familiar and unobtrusive, autonomous and simple to use. The results of this study answers the first two research questions:

**R1:** What activities of daily living are most challenging for older adults with dementia and their family caregivers in a home environment? The findings suggest that older adults with dementia generally retain the ability to, at least, complete fundamental ADL (e.g., eating, drinking, using the bathroom) with some level of independence. The results also indicate that the majority of older adults with dementia cannot participate in higher-level ADL (e.g., preparing complex meals, cleaning the house). Family caregivers reported that providing support for tasks that involve an invasion of privacy (e.g., getting dressed, bathing, showering) or tasks that are typically completed independently (e.g., washing hands or brushing teeth) are most desirable for IATs.

**R2:** What features and functions are required for the successful integration of an intelligent assistive technology that is intended to support the completion of daily activities of older adults with dementia and their family caregivers in a home environment? Respondents reported two key findings with respect to the successful integration of an IAT into a user’s home. First, an IAT must be familiar and unobtrusive in a home environment. Secondly, an IAT must be autonomous and simple to use.

The first of the two secondary objectives (SO1) is met by consolidating the major results of this study. Older adults with dementia and their family caregivers believe that autonomous IATs developed to unobtrusively support the completion of private or personal tasks by empowering users to leverage their remaining abilities with minimal interactions between the system and the caregiver are useful in a home environment. These findings also contribute to the literature by presenting generalizable results on both the task-based needs of older adults with dementia and their family caregivers, and the role IATs can play in supporting those needs in a home environment. For example, these findings complement the work of Rialle, Ollivet, Guigui, and
Herve [131] who explored the general acceptance of technologies in a home environment through a questionnaire. They found a defined split in perceptions, with respondents either holding a strong sense of confidence in the helpfulness of these technologies or rejecting the value of such interventions. Our findings also agree with the work of Demeris et al. [132] who found that users of in-home technologies were concerned with ease of use, device autonomy (or conversely, requirements for user training), and human-machine interfaces. Our findings also agree with other studies investigating in-home ADL specifically [e.g., 55], but do not generalize well to studies exploring task support in a more general sense (i.e., out of the home) [56, 132], suggesting that more work is still required to understand the comprehensive needs of IAT users.

Identifying the needs of older adults with dementia and their family caregivers for IATs in a home environment is a critical first step in developing a technological intervention intended to meet these needs. However, the COACH is an existing IAT that has undergone three substantive stages of development for over a decade. Certainly, the overarching goal of developing the COACH for use in a home environment has remained constant. However, through each stage of development of the COACH, the local objectives have necessarily changed for various reasons (e.g., technical limitations, ethical restrictions, and practical challenges). Furthermore, to date all performance and efficacy studies involving the COACH have been supervised (i.e., a researcher has been present, controlling the trial), and in many cases have occurred in either a lab setting or controlled clinical environment (e.g., in a washroom specifically intended for use by participants of the study). For these reason, a thorough understanding of the current capabilities of the COACH in an unsupervised, real-world deployment is also required before further development of the system is possible. I address this requirement in the next section of this chapter.

5.1.1.2 Secondary Objective 2 (SO2): to identify factors limiting the technical performance of the COACH in an unsupervised, real-world deployment in a clinical setting

The COACH has been proven effective at increasing the number of hand washing task steps an older adult can independently complete (i.e., without the help of a human caregiver) in a supervised, clinical or lab setting [11, 37, 41]. The second study (Chapter 3) sought to extend the literature by contributing a testing methodology and resulting performance data on one of the first, and certainly the largest, unsupervised, real-world deployments of an IAT designed to support older adults with dementia. Furthermore, this study was designed to empirically
evaluate the performance of the COACH and each of its three main components (hand tracking, policy and prompting module) to identify limitations preventing an unsupervised real-world, in-home deployment.

A total of forty-one (41) hand washing trials were completed by 27 consenting participants recruited by convenience sample over a four month period at Toronto Memory Program (TMP) [97]. The COACH was installed in a washroom that was accessible by all patients of the TMP as well as staff, visitors and in some cases to the public. In comparison to previous performance trials with the COACH, in this deployment the COACH was completely unsupervised (i.e., members of the research team did not interact with the system over the entire trial). The system was configured to run continuously over the four month period, only interacting with users who were included in the study (i.e., when user-system interaction was enabled by manually pressing the study inclusion button). This forced the COACH to determine when a user began and ended the hand washing task, rather than relying on a human to demarcate these points in time. The result was a more accurate, real-world evaluation of the system’s ability to monitor and assist with the hand washing task.

The performance results of this study revealed that the hand tracker was able to track the unfiltered, frame-by-frame positions (i.e., x, y coordinates) of both the left and right hands with a failure rate of 27.32% and 26.11% respectively. These hand positions enabled the system to correctly infer whether a participant completed or failed to complete a task step with an accuracy of 54.9%. Using the task step inferences, the policy module was able to track the users’ task progression with an accuracy of 46.9% when using the real trial data. Utilizing inaccurate tracking data, the prompter was only able to provide appropriately timed and relevant prompts for ten (10) out of a total of 108 prompts issued to the system’s users. To evaluate the performance of the policy independent of the hand tracker, the trials were simulated with manually labeled ground-truth hand tracking data. The simulations resulted in a task progression accuracy of 97.6% from the policy, indicating that the hand tracker was restricting the performance of the policy.

According to the results of this study the colour-based hand tracker could not accurately identify the position of the users’ hands, severely reducing the overall performance of the COACH system in an unsupervised, real-world environment. Failure modes resulting in poor hand
tracking performance were predominantly attributable to the reliance on colour (e.g., mistaking skin coloured clothing as a hand, disambiguating the head of a bald person from the hands), and the inability of a two-dimensional colour-based tracker to fully identify user-object interactions (e.g., hands passing over the faucet versus actually using the water). Isolating the COACH policy and prompting modules by creating ground-truth hand tracker data and re-simulating the trials yielded very strong system performance results, further implicating the hand tracker as the weakest component. Combining these performance results overwhelmingly supported that inaccurate user hand tracking severely impacted the performance of the COACH in an unsupervised, real-world environment. Furthermore the performance results obtained by evaluating the COACH in this study allow me to address my third and fourth research questions:

**R3:** How effectively does the COACH infer the progression of older adults with dementia through the task of hand washing in an unsupervised, real-world clinical deployment? The COACH was able to correctly identify 46.9% of the hand washing task steps completed by older adults with dementia in an unsupervised, real-world clinical deployment.

**R4:** What technical factors explain occurrences where COACH failed to infer the progression of older adults with dementia through a hand washing task in an unsupervised, real-world clinical deployment? Two-dimensional, colour-based hand tracking was predominantly responsible for occurrences where COACH failed to track the progression of older adults with dementia through the hand washing task in an unsupervised, real-world clinical deployment. Specifically, the two-dimensional colour-based hand tracker could not locate users’ hands in images that had many other skin-coloured pixels (e.g., bald users, users wearing sleeveless garments, skin-coloured clothing), and could not correctly identify some user-object interactions using only two-dimensional hand position locations (e.g., hands over the faucet versus hands in the water).

Notably, for this study participants were recruited by convenience sample, and participants contributed an average of 1.52 trials with two participants completing six trials. Multiple trials from the same individual may result in similar movements, actions or tracking error types. However, the failure modes of the colour-based hand tracker identified in this study did not indicate a bias presented by repeated trials from the same user.

The results of this study allow me to achieve the second of my secondary objectives (SO2). Namely, the colour-based, two-dimensional hand tracker must be improved in order to deploy
the COACH in an unsupervised, real-world environment. A method of overcoming the reliance on a colour model through the development of a tracking approach that is colour, texture and lighting invariant is paramount to overcoming the limitations of the current tracker. Additionally, the introduction of a third dimension representing the distance of objects in the camera scene from the camera (i.e., depth) would provide the location of each user’s hands in a three-dimensional space, further reducing occurrences of failed hand tracking.

Nevertheless, neither the needs of the users of the COACH nor the technical limitations of the system in a real-world setting alone will provide appropriate technical design criteria for the future development of the COACH. The needs of the COACH users must be combined together with the capabilities of the COACH system to provide a set of design criteria for the next COACH prototype. The process of identifying design criteria, redesigning the COACH to satisfy the criteria and testing the new prototype COACH system are the topic of the next section.

5.1.1.3 Primary Objective (PO): to continue development of the COACH toward an unsupervised, real-world in-home deployment

The House of Quality (HOQ) is an industry driven approach used to improve an existing technology or product [57, 58]. The main function of the HOQ is to translate the needs of customers into technical measures or design criteria. These design criteria can subsequently be used to continue development of an existing product in a way that maximizes customer satisfaction within the current and potential capabilities of the product [57, 58]. The primary objective of my dissertation can be framed within the context of a HOQ approach. The COACH is an existing product; an intervention designed to help older adults independently wash their hands. The customers of the COACH are the users of the system. Using a HOQ approach allows the synthesis of the needs of the system’s users (Chapter 2) with the capabilities of the system (Chapter 3) into a list of technical design criteria ranked in importance (at satisfying user needs within the capabilities of the system) relative to each other. This list of design criteria can then be used to select the most appropriate or relevant criteria which can subsequently be used to direct future development of the COACH system.

Two HOQ matrices were constructed. The first (see Appendix 1) was designed to determine required changes for the three major COACH modules: (1) hand tracking, (2) policy, and (3) prompting. The results of the HOQ analysis provided a relative ranking of the importance of
subcomponents relative to the other subcomponents. The ranking identified the specific module and subcomponents where development resources could most significantly contribute to satisfying user needs. Four of the five highest ranked subcomponents were within the hand tracking module: depth tracking, kinematic skeleton modeling, insensitivity to lighting variations, and improved user-object interactions. The second HOQ matrix (see Appendix 2) was designed to identify features and functions deemed essential to the successful installation and use of COACH in a home environment. Required changes were grouped into three categories: (1) user interface, (2) physical appearance, and (3) cost. The results of the HOQ analysis ranked the subcomponents in importance relative to the other subcomponents. The top five (of a total 11) subcomponents represented 86% of the total importance with almost an even distribution. In order of importance, the subcomponents representing the design criteria were: caregiver interface appearance, camera and monitor appearance, speaker and display appearance, device autonomy, and initial hardware cost.

The HOQ approach indicated that the overall performance of the COACH in an unsupervised, real-world environment would be improved by developing a new hand tracker that is lighting and colour invariant and incorporates three-dimensional tracking to develop a three-dimensional model of the users, improving detection of user-object interactions. To fulfill these design criteria I developed and tested a depth-based tracking algorithm that is lighting, shape, texture and colour invariant. The new part tracker was integrated into the COACH and was validated on forty-one (41) real-world hand washing trials, correctly identifying 192 of 205 hand washing steps. The successful implementation of a HOQ approach and successful development of a new hand tracker for the COACH enables me to address the final two research questions:

**R5:** What technical design criteria, used to develop an improved COACH prototype, can be identified by synthesizing the needs of older adults with dementia and their family caregivers for assistive technologies in a home environment with the capabilities of the COACH in an unsupervised, real-world deployment? Using a House of Quality approach to synthesize the needs older adults with dementia and their family caregivers have for an IAT in a home environment with the capabilities of the COACH in an unsupervised, real-world deployment provides a set of design criteria for future development of the COACH. Specifically, development must focus on new hand tracker that is lighting and colour invariant and incorporates three-dimensional tracking to develop a three-dimensional model of the users
**R6:** How effectively do(es) the COACH module(s) developed to satisfy the most relevant technical design criteria identified in research question 5 perform in a real-world setting compared to the module(s) implemented in the current prototype? The new depth-based hand tracker developed to satisfy the design criteria obtained using a HOQ approach was able to track the COACH users’ hands with a specificity of 0.938, sensitivity of 0.923 and accuracy of 0.937 over forty-one (41) trials with thirty (30) participants. The existing colour-based hand tracker was able to track the users’ hands with a specificity of 0.822 and 0.466, sensitivity of 0.769 and 0.975, and accuracy of 0.819 and 0.549 with the colour-based tracker using the data from the current study and in the real-world trials respectively.

**TABLE 5.1: COMPARISON OF STEP TRACKING PERFORMANCE BETWEEN DEPTH-BASED HAND TRACKER AND COLOUR-BASED HAND TRACKERS.**

<table>
<thead>
<tr>
<th></th>
<th>Current Study</th>
<th>Previous Trials with Colour Tracker [12]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Depth Tracker</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Positive (TP)</td>
<td>180</td>
<td>96</td>
</tr>
<tr>
<td>(Step completed by user and identified as complete by COACH)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True Negative (TN)</td>
<td>12</td>
<td>39</td>
</tr>
<tr>
<td>(Step not completed by user and not identified complete by COACH)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Positive (FP)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(Step not completed by user and identified complete by COACH)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Negative (FN)</td>
<td>12</td>
<td>110</td>
</tr>
<tr>
<td>(Step completed by user and not identified complete by COACH)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.938</td>
<td>0.822</td>
</tr>
<tr>
<td>(TP/(TP + FN))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>0.923</td>
<td>0.769</td>
</tr>
<tr>
<td>(TN/(TN + FP))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.937</td>
<td>0.819</td>
</tr>
<tr>
<td>((TP + TN)/(Total Outcomes))</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The primary objective (PO) of my dissertation is to employ a knowledge-driven approach to continue development of the COACH toward an unsupervised, real-world deployment in the homes of older adults with dementia. In the first unsupervised, real-world trial ever conducted with the COACH system, the colour-based hand tracker was identified responsible for the majority of failures experienced by the system. Ultimately, the colour-based tracker correctly identified 46.9% of the hand washing task steps completed by older adults with dementia in this environment. I successfully accomplished my PO by developing a new depth-based hand tracker that identified 93.7% of task steps completed in forty-one (41) trials – a 46.8% improvement in performance. With the new depth-based hand tracker, the performance of the COACH system can now be evaluated again in a real-world context.
5.1.2 Engineering Contributions

The substantial engineering contributions of this work are listed here:

1. The development of a robust tracker:
   a. Colour, lighting, texture and shape invariant;
   b. First depth-based tracker using an overhead perspective;
   c. Classifier trained using partially labeled, unbalanced data;
   d. Approximately 5,000 training images with the hands and head fully labeled;
   e. Approximately 24,000 validation images with the hands and head part centres labeled;
   f. Generic training algorithm that can be used to train any image classifier.

2. The creation of a user-configurable application to:
   a. Capture depth and/or colour images from a depth camera;
   b. Label depth images for random decision forest classifier training;
   c. Train random decision forests given a set of training images;
   d. Annotate and label validation images;
   e. Compute performance metrics of random decision forest classifiers.

3. The implementation of a House of Quality approach to IAT:
   a. Industrial tool to translate the needs of customers into technical measures or design criteria.
   b. Allowed the integration of the current capabilities of the COACH along with design themes prevalent in the literature.

5.2 Overall Limitations

Unfortunately, this study is not without limitations. The determination of the IAT needs of the COACH users were identified using an exploratory questionnaire. The exploratory nature of the study was two-fold: (i) the questionnaire itself was constructed and validated through the study, and (ii) the research itself was exploratory, probing areas of inquiry that do not have associated areas of scholarly reference (criterion validity). Accordingly, a larger and more diverse sample taken from across Canada would provide a source of external validity and perhaps shed insight into social or regional factors that may impact the needs expressed by respondents. Furthermore,
a larger sample would allow further statistical testing to ensure construct validity and proper factor extraction from the questionnaire items.

The observational methodology I used to evaluate the performance of the COACH in a real-world setting was also exploratory in nature. At the time of this study, little research was available specifying methodologies or evaluation frameworks that were appropriate for the evaluation of IATs, specifically when considering vulnerable populations in a home environment. Accordingly, the technical performance evaluation methodology I used to evaluate the COACH was developed by extending the methodologies used in previous supervised COACH trials and incorporating relevant approaches from the literature. Caution must be taken in interpreting the results of the performance evaluation which are restricted to the technical performance of the system. Little inference can be made to the effectiveness of the system as an automated hand washing prompting aid. In particular, the metrics used to evaluate the prompts provided by the system were reported for the sake of completeness. However, subsequent analysis of the prompting (e.g., prompt timing and appropriateness) was not conducted beyond the metrics provided in acknowledgement of the poor performance of the system’s decision making module caused by ineffective hand tracking. Furthermore, constraints impose by our ethics review board during Study 2 prevented the collection of participant data other than MMSE scores. Accordingly, little was reported on the nature of the participants, and analysis of the system performance was not considered in light of other participant characteristics (e.g., age, gender).

Lastly, the final study of my dissertation centered on the technical development of a new hand tracker. The first step of this process sought to identify technical design criteria through the application of a House of Quality approach. This approach utilized a potentially limited set of data representing the user needs, acquired through the needs assessment of Study 1. Accordingly, this approach may have been biased by the relatively small sample and may not have been exhaustive affecting the reproducibility of the study. Furthermore, the results of the performance evaluation of this study focused on the technical efficacy of the new hand tracker rather than the actual performance of the COACH as an automated prompting system. Additionally, the hand tracker was developed and tested in a single washroom. Although the hand tracker is theoretically insensitive to changes in location, the system is sensitive to changes in camera perspective warranting further study in diverse locations before generalizing the performance
results. Finally, random decision forests are known to overfit training data which can introduce classification error in the validation image set or during operation of the system using the classifier. In the final study of my dissertation three decision trees were used in an ensemble as a random decision forest to decrease the tendency of the classifier to overfit. However, a measure of the significance of the overfitting was not empirically determined and methods of reducing the overfitting (e.g., pruning) were not explore in favour of evaluating the performance of the classifier in a practical, hand washing application following optimization of key forest parameters.

5.3 Future Work

Understanding the needs of the users of IATs is critical to the successful integration of these systems into the homes of older adults with dementia and their family caregivers. Certainly, continued efforts directed at identifying these needs are warranted. Specifically, situating our findings within other similar studies will help expand our understanding of user needs to include potential privacy concerns, types of care (e.g., reminder-based, task support, physical assistance) and the overall suitability of technology at satisfying these needs. However, identifying the needs of the system users as well as the more practical features and functions required of an in-home technology will not necessarily translate into useable and useful IAT. These needs must be presented in such a way as to provide design guidelines or specifications for IAT developers to use, for example within a design framework. Such a design framework would allow a formal and explicit pairing of user needs with the capabilities of an IAT. This framework would allow potential users of an IAT to find an appropriate intervention, and allow developers of IATs to find appropriate target populations. Toward this goal, I provided the intellectual base and am named a collaborator on a research grant that was awarded by the Alzheimer Society of Canada in 2012. The overall objective of this project is to develop design guidelines (based on user needs) within a generalizable framework that can be used by IAT developers that support activities of daily living completion by older adults with AD.

With respect to the COACH specifically, the needs of the users are not restricted to the task of hand washing alone. Accordingly, the COACH system must be expanded to facilitate concurrent support for additional daily tasks in addition to hand washing. To accomplish this, improvements to the hand tracker, policy and prompting module must be made. Firstly the hand tracker must be
modified to detect interactions between the users of the COACH and additional environmental objects (as required by each specific task). Although in many cases this modification will be trivial, in some cases systematic or methodological issues may arise (e.g., the overhead perspective used for hand washing may prevent the detection of some user-object or user-environment interactions in different tasks) requiring more substantive changes to the hand tracking. Secondly, the COACH policy must be expanded to concurrently monitor multiple ADL. The inclusion of additional POMDP policies presents two challenges. Firstly, a hierarchical POMDP structure must be implemented, with a high-level POMDP monitoring which task(s) is(are) currently being completed, while multiple lower-level POMDPs must be implemented modeling the individual tasks (e.g., hand washing, tooth brushing). Secondly, the specification of a POMDP takes a substantial amount of time and expert knowledge, particularly with respect to validation of the proper functioning of the decision process. We have achieved some success at rapidly specifying and validating POMDP policies for specific ADL (e.g., hand washing, tooth brushing, tea making) using a Syndetic Assistance Process (SNAP) [110]. However, this approach has not been used to develop hierarchical POMDPs; an area of future development. Finally, the COACH prompting module must be expanded to facilitate provision of additional prompts, cues and task step reenactments (videos) for any additional tasks.

The COACH prompting module represents one of the two methods of human-computer interaction (HCI), occurring between the COACH and older adults with dementia engaging in the hand washing task. The second method of HCI occurs when caregivers interact with the COACH (e.g., initial configuration, maintenance, acquisition of statistics). User interface (UI) design is critical in both of these methods of interaction. When considering interactions with older adults with dementia completing the ADL, user interface considerations include the content and context of the prompts, cues and videos, as well as the mode of delivery. Some work has been done explicitly looking at communication strategies with COACH supporting a hand washing task [47, 133]. However, expansion of the COACH to additional ADL will require additional understanding of the communication requirements of those ADL, as well as the potential communication challenges experienced by concurrently supporting multiple ADL. Little work has been done to understand interactions between the COACH and the caregivers, restricted to pilot work investigating high-level design criteria [134-136]. Future work must consider several aspects of caregiver-COACH interactions including factors identified through
Study 3 (Chapter 4) such as appearance, usability and autonomy, and cost as well as the presentation of relevant information, safety and protection of personal data. Additionally, the COACH must be able to adapt to the changing characteristics and increasing responsibilities of the family caregiver [31]. For example, IATs must accommodate the possibility of remote care provision (e.g., primary family caregivers not living in the same city, primary family caregivers taking vacation while still providing care). Advances in other forms of dementia care technology such as telehealth systems [137, 138] have created an opportunity for IATs to adapt to this changing care context [139], but approaches to combining these remote support technologies must be explored.

While working to further understand user needs for an in-home IAT, extend the COACH to additional ADL and improve HCI between the COACH and its users, efficacy studies of the COACH in an unsupervised, real-world deployment must continue. These efficacy studies must investigate the capabilities of the system in a real-world environment with the new depth-based hand tracker as well as continue to validate the measures and metrics used to evaluate the performance of the COACH. These measures and metrics will necessarily include longitudinal studies observing various desirable clinical outcomes such as the ability of the COACH to change routines over time, and to improve task performance over time. Furthermore, longitudinal studies may also reveal undesirable outcomes such as participant desensitization to prompts over time. The results of these efficacy studies must be iteratively integrated into future designs of the COACH along with stronger understandings of user needs yielding new prototypes. Only then will these prototypes, along with the UI be suitable for real-world effectiveness trials in the homes of older adults with dementia and their caregivers. However, the challenge of conducting effectiveness trials with IAT in the homes of older adults with dementia and their caregivers is not strictly limited to a scarcity of IAT ready for such evaluation. Introducing IAT into the homes of a vulnerable or compromised population is wrought with debate and uncertainty including topics such as the legal, privacy and ethical issues surrounding the deployment of IAT in the homes of older adults with dementia. To ensure the success of future IAT effectiveness studies in a home environment, a detailed understanding of these factors (e.g., the legal doctrines of capacity and consent, the right to privacy and human rights legislation in Canadian law) must be obtained.
5.4 Final Remarks

The aging of the world’s population is resulting in an increased prevalence of neurocognitive disorders associated with age, such as Alzheimer’s disease and other forms of dementia [140][2]. The loss of independence associated with dementia affects both the person afflicted with the disorder as well as people involved in providing support [4]. The demands associated with support are increasingly being met by informal caregivers (e.g., family members, friends) resulting in a significant increase in caregiver burden, negatively impacting the lives of these caregivers [4]. IATs have emerged in recent years with the potential to alleviate some of the burden experienced by family caregivers, while simultaneously supporting the independence of both the care recipient and care provider [e.g., 11, 12-16]. However, despite the potential benefits these devices offer, few IATs are currently available as commercial products and adoption and acceptance of these devices remains minimal [8]. Two significant factors have been attributed to the limited development and acceptance of these devices. First, IATs generally remain in the early stages of development, with few IATs reaching the stages of clinical or real-world trials [8], resulting in little data regarding their efficacy or effectiveness. Secondly, the needs of the end users of these IAT are generally unexplored, and often limited to task-based needs obtained through case studies or technical trials [9, 27]. Little is known about how these IATs can integrate into the lives of their users, specifically within the context of an in-home IAT.

Combining these two factors limiting the penetration of IATs into the marketplace results in a development environment where little is known about the needs of IAT users and even less is known about the capabilities of existing IATs at satisfying these needs. This is hardly an environment conducive to effective development of IATs to support such a vulnerable population.

Through the work presented throughout this dissertation, I attempt to address these limiting factors. First I conducted a holistic assessment including both the task-based needs and practical features and functions older adults with dementia and their family caregivers have for an in-home IAT. Secondly, I evaluated the performance of the COACH in an unsupervised, real-world deployment. Finally, I integrated the needs of the users with the performance results of the COACH to identify technical design limitations reducing the efficacy of the COACH in an unsupervised, real-world environment. Using this approach, I developed an overall methodology that effectively targeted the development of the COACH toward satisfying the needs of the
system’s end users. Furthermore through the process I contribute a holistic and generalizable needs assessment investigating the role of IATs in the homes of older adults with dementia and their family caregivers. I also conduct one of the first, and largest, real-world trials of an IAT designed to encourage independent ADL completion by older adults with dementia. Finally, I developed and validated a lighting, shape, colour and texture invariant overhead body part tracker for the COACH using a large training and validation set. Overall, my findings suggest that the COACH is once again ready to be evaluated in an unsupervised, real-world clinical environment toward the ultimate goal of a deployment in the homes of older adults with dementia.
6 References


[38] A. Mihailidis, "The development of an intelligent cognitive orthosis to facilitate handwashing for persons with moderate-to-severe dementia," PhD, Bioengineering Unit, University of Strathclyde, Glasgow, Scotland, 2001.


Appendix 1: House of Quality used to determine required changes for the three major COACH modules: (1) hand tracking, (2) policy, and (3) prompting

<table>
<thead>
<tr>
<th>Technical Measure (H/W)</th>
<th>Tracking</th>
<th>Promoting</th>
<th>Decision Making</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>Target</td>
<td>Priority</td>
<td>Overall</td>
</tr>
<tr>
<td>Usability</td>
<td>4.7</td>
<td>3.1</td>
<td>0.13</td>
</tr>
<tr>
<td>Ease of use</td>
<td>5.3</td>
<td>5.0</td>
<td>0.05</td>
</tr>
<tr>
<td>Speed</td>
<td>3.5</td>
<td>5.0</td>
<td>0.20</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>2.5</td>
<td>3.0</td>
<td>0.20</td>
</tr>
<tr>
<td>Customer tracking</td>
<td>3.3</td>
<td>3.1</td>
<td>0.13</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>3.4</td>
<td>3.3</td>
<td>0.12</td>
</tr>
<tr>
<td>Usability</td>
<td>4.7</td>
<td>3.3</td>
<td>0.20</td>
</tr>
<tr>
<td>Ease of use</td>
<td>5.6</td>
<td>3.4</td>
<td>0.05</td>
</tr>
<tr>
<td>Speed</td>
<td>4.4</td>
<td>5.0</td>
<td>0.20</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>3.0</td>
<td>5.0</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Customer Requirement (H/W): Technical Measure (R/W)

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Technical Measure</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Promoting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Making</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer</td>
<td>Importance</td>
<td>Overall</td>
</tr>
<tr>
<td>tracking</td>
<td></td>
<td>3.2</td>
</tr>
<tr>
<td>satisfaction</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>effectiveness</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>usability</td>
<td></td>
<td>3.1</td>
</tr>
<tr>
<td>ease of use</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td>2.9</td>
</tr>
</tbody>
</table>
Appendix 2: House of Quality used to identify features and functions deemed essential to the successful installation and use of COACH in a home environment

<table>
<thead>
<tr>
<th>Customer Requirements (WHA Ts)</th>
<th>Mean Score (10)</th>
<th>Current Importance (50)</th>
<th>Target Importance (50)</th>
<th>Overall Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Supporting COACH with Technology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiarity &amp; Autonomy</td>
<td>7.1</td>
<td>4</td>
<td>4</td>
<td>3.33</td>
</tr>
<tr>
<td>Use house to make things more simple</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3.33</td>
</tr>
<tr>
<td>Its important that I don't need any training</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.00</td>
</tr>
<tr>
<td>I can't have time to play with it</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3.33</td>
</tr>
<tr>
<td>Can't use computer</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.00</td>
</tr>
<tr>
<td>Can't use computer</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.00</td>
</tr>
<tr>
<td>Use it to learn new things</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3.33</td>
</tr>
<tr>
<td>Use it to extend enough</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
<td>2.50</td>
</tr>
<tr>
<td>Use it with different places on the floor</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.00</td>
</tr>
<tr>
<td>Use it to control</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.00</td>
</tr>
<tr>
<td>Easy to do something meaningful</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3.00</td>
</tr>
<tr>
<td>Professional</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.00</td>
</tr>
<tr>
<td>Time to play with setting specifications</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3.00</td>
</tr>
<tr>
<td>Professional</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.00</td>
</tr>
<tr>
<td>Think the display should be white</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
<td>2.50</td>
</tr>
<tr>
<td>Condition some higher tech</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.00</td>
</tr>
<tr>
<td>Cannot learn to use technology</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.00</td>
</tr>
<tr>
<td>Cannot learn to use technology</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.00</td>
</tr>
<tr>
<td>It's not worth the money</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.00</td>
</tr>
<tr>
<td>It's not worth the money</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4.00</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Planning (WHA Ts)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical Measures (HOWS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User Interface</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appearance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical Measures (HOWS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User Interface</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appearance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relative Importance of HOWS:

- User Interface: 21.4
- Appearance: 30.6
- Cost: 18.0

Percentage Importance of HOWS:

- User Interface: 21.4%
- Appearance: 30.6%
- Cost: 18.0%

Technical Measures (HOWS):

- User Interface
- Appearance
- Cost

Percentage Importance of HOWS categories:

- User Interface: 21.4%
- Appearance: 30.6%
- Cost: 18.0%
Appendix 3: Doctoral Publication List

Chapters in Books

   My contribution (40% of the overall work) includes: Conception and outlining, writing multiple sections, editing.

   My contribution (20% of the overall work) includes: Conception and outlining, writing multiple sections, editing.

Peer-Reviewed Journal Articles

   My contribution (30% of the overall work) includes: Data collection, data analysis, writing multiple sections, revisions, and editing.

   My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.

   My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.
   
   My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.

**Peer-Reviewed Conference Proceedings/Articles**

   
   My contribution (20% of the overall work) includes: Conception and outlining, writing multiple sections, editing.

   
   My contribution (30% of the overall work) includes: Data collection, data analysis, writing multiple sections, revisions, and editing.

**Non-Peer-Reviewed Articles/Conference Proceedings**

   
   My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.

   
   My contribution (35% of the overall work) includes: Idea conception, writing multiple sections, editing.

   
   My contribution was 100% of the overall work.

My contribution (15% of the overall work) includes: Writing multiple sections.


My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.


My contribution (10% of the overall work) includes: Revisions, and editing.


My contribution (15% of the overall work) includes: Idea conception, revisions, and editing.


My contribution was 100% of the overall work.


My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.


My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.


My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.
   My contribution (25% of the overall work) includes: Data collection, data analysis, writing multiple sections, revisions, and editing.

   My contribution (5% of the overall work) includes: Writing a section.

**Peer-Reviewed Abstracts**

   My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.

   My contribution (25% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, outlining draft, revisions, and editing.

   My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.

   My contribution (45% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, outlining draft, revisions, and editing.
My contribution (25% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, outlining draft, revisions, and editing.

My contribution (45% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, outlining draft, revisions, and editing.

My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.

My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.

My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.

My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.

My contribution (95% of the overall work) includes: Study conception, design and implementation, data collection, data analysis, writing complete draft, revisions, and editing.

Invited Presentations/Lectures/Keynotes

1. **Czarnuch, S.** (2013). *Activity tracking from an overhead depth camera: From joint proposals to a skeleton model*. Speaker series lecture conducted at the Department of Computer Science. Memorial University. St. John's, NL.
   My contribution was 100% of the overall work.

   My contribution (95% of the overall work) includes: Presentation conception, design, review, editing and delivery.

   My contribution (50% of the overall work) includes: Presentation conception, concept design, review, editing and lecture.

   My contribution was 100% of the overall work.

   My contribution was 100% of the overall work.

   My contribution was 100% of the overall work.

   My contribution was 100% of the overall work.

8. **Czarnuch, S.** (2009, December). Exploring interdisciplinary research methodologies. Guest lecture in the Faculty of Arts, department of Sociology at Trent University, Peterborough, Ontario.
My contribution was 100% of the overall work.


   My contribution was 100% of the overall work.

**Dissertation**


   My contribution was 100% of the overall work.
Copyright Acknowledgements

Chapter 2: Open Access

Chapter 3: © 2013 IOS Press and the authors

Chapter 4: This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible

All other chapters © Stephen Czarnuch