CogWatch – Cognitive Rehabilitation of Apraxia and Action Disorganisation Syndrome

D3.2.1 Report on data analysis for action recognition 1

Deliverable No.		DN.3.1	
Workpackage No.	**WP3**	Workpackage Title	**Activity Recognition & Prediction**
Task No.	**T3.2**	Activity Title	**Action Recognition**
Authors (per company, if more than one company provide it together)		**Charmayne Hughes, Joachim Hermsdörfer, Marta Bienkiewicz (TUM)** **José M. Cogollor, Javier Rojo, Sandra Campo (UPM)** **Martin Russell (UoB)**	
Status (F: final; D: draft; RD: revised draft):		**F**	
File Name:		**CogWatch_M11_D3 2 1_Final15Oct2012.doc**	
Project start date and duration		**01 November 2011, 36 Months**	

EXECUTIVE SUMMARY

This report discusses the issues that are relevant to the collection of data for action recognition and prediction. The main purpose of the report is to explain the technologies that a) are currently used in the development of the first prototype, and b) will be used in the future to develop the action prediction models for the second prototype, in month 33 of the project.

The report is presented in six main sections, which address the collection and analysis of data from the clinical screening phase. Section two begins with a review of kinematic-based action recognition, and details previous research that has utilized kinematic technology in the context of performance quality in activities of daily living (ADL). Section two concludes by presenting primarily data on a Kinematic-based Action Recognition Algorithm (KARA) that is being developed by TUM. The utility of such an algorithm is that it can predict actions during the execution of a movement, rather than having to wait until the patient has made an error. Although this work is still in the preliminary stages, it is expected that future versions of the CogWatch prototype will implement KARA.

Section 3 discusses the use of a low-cost, easy-to-implement motion capture system (Microsoft Kinect™) for action recognition. The Kinect™ was compared to a commercially available ultrasonic motion capture system (Zebris) during a tea making task. The results were promising, and suggested that the Kinect™ may be a viable addition to the CogWatch system as a motion capture technology. Present research aims to investigate whether KARA is able to distinguish between motions at the lower sampling frequency used by the Kinect™.

In section 4, we detail information regarding sensor-based action recognition. The EECE group at UoB has conducted preliminary experiments to assess the ability of CogWatch instrumented coasters (CIC) attached to the base of a kettle to detect sub-goals during four contexts (pour, toy, pour-toy, rest). Their results suggest that activity recognition performance is poor when the raw CIC data is used, but that simple thresholding of the force sensitive resistor (FSR) outputs removes significant variability and improves recognition accuracy. UoB-EECE is currently working on improving data analysis methods from CIC outputs, which will be the primary basis of action recognition in prototype 1.

Section 5 discusses the data and results obtained from video recordings in apraxia patients and healthy controls during a two tea-making task, which is the ADL scenario chosen for the clinical screening phase. The data obtained from these studies has been analysed, and we report results on error types, error frequencies, and action sequencing in ADL performance. This data has been modelled using Bayesian Logic Networks (BLN) (see D3.1 report on action recognition techniques), and will be integrated into the task models used in prototype 2.

Last, in section 6 we provide an outlook for data collection and analysis in the upcoming 12 months, and introduce the tasks that will be used to evaluate prototype 2.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

REVISION HISTORY

Revision no.	Date of Issue	Author(s)	Brief Description of Change
V0	24/09/2012	**Charmayne Hughes, Joachim Hermsdörfer, Marta Bienkiewicz (TUM)** **José M. Cogollor, Javier Rojo, Sandra Campo (UPM)**	Draft for Peer Review
V1	25/09/2012	Laura Pastor, María Teresa Arredondo (UPM)	Formatted draft
V2	09/10/2012	**Charmayne Hughes, Joachim Hermsdörfer (TUM)** **Martin Russell (UoB)**	Revised Version
V3	15/10/2012	**Maria Teresa Arredondo, Matteo Pastorino (UPM)**	Typographical and content checks
Final	15/10/2012	Alan Wing	Typos checked, page numbering format change, minor pagination change.

LIST OF ABBREVIATIONS AND DEFINITIONS

Abbreviation	Abbreviation
AADS	Apraxia and Action Disorganisation Syndrome
ADA	American Disability Association
ADL	Activity of Daily Living
AS	Action Segment
BLN	Bayesian Logic Network
CIC	CogWatch Instrumented Coasters
CVA	Cerebrovascular Accident
EECE	Electronic, Electrical and Computer Engineering
FSR	Force Sensitive Resistor
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
HTA	Hierarchical Tree Analysis
IAS	Intelligent Autonomous Group
IRF	Infinite Response Filter
KARA	Kinematic Action Recognition Algorithm
MSE	Mean Square Error
PDF	Probability Density Function
SDK	Software Development Kit
TUM	Technische Universität München
UoB	University of Birmingham

This page intentionally blank

1. INTRODUCTION

This is the 11th month report from CogWatch Work Package 3 “Activity Recognition and Prediction”. WP3 involves two of the CogWatch partners, the University of Birmingham (UoB) and the Technische Universität München (TUM). In addition, researchers from the Universidad Politécnica de Madrid (UPM) have worked with TUM to develop and test technologies that are of potential use to the CogWatch system.

As stated in the Annex, the objectives of WP3 are to:

- Explore psychological and pattern-based models to construct *action recognition* techniques;
- Apply these techniques to the data collected from the patient studies to develop an *action recognition system;*
- Review psychological models and apply advance statistical methods to provide a reliable *action prediction* model that will be able to identify patients’ *intentions*, *predict* actions and assess the *progress* of a task.

There are a variety of technologies that are used to recognize and predict actions in ADL, and include video-based, kinematic-based, and non-marker kinematic-based action recognition. In order to evaluate error production and action sequencing in apraxia patients, video-based technologies will be used. This data will then be manually coded to ascertain the types and frequency of errors, as well as grasp selection errors during ADL performance. This information is then used to develop the Bayesian Logic Network (BLN) models that will be used in the second CogWatch prototype (T3.4).

Kinematic-based technologies will be used to detect errors as they are unfolding, which can be used to provide cueing information to participants about errors before they happen. This technology is also used to develop a kinematic action recognition algorithm (KARA), which can segment an ADL into action segments using a kinematic-based criterion. Although this work is in the preliminary stages, we are confident that KARA can be implemented into the second CogWatch prototype.

Non-marker kinematic-based technologies have also been exploited as a suitable, low-cost and easy to implement motion capture system for kinematics analysis in ADL. We have recently conducted an experiment that compares the data collected from the Microsoft Kinect™ sensor to the data collected from an ultrasonic motion capture system. The data indicate a moderate to strong correlation between signals, demonstrating the potential uses of the Kinect™ system in cognitive rehabilitation.

1.1 Proposed timetable

- First draft September 14th
- Completion September 23rd
- Reviewed via UoB-Psychology Quality management 26th September
- Copy editing end 28th September
- Submitted to EU October 1st (well within 45 days of nominal deadline).

2. KINEMATIC-BASED ACTION RECOGNITION

One of the corollary aims of WP3 is to analyze and characterize movement production quality in patients with apraxia, relative to healthy controls. The study of human movement often focuses on kinematics, describing the movements of the body through space and time, but without reference to the forces involved (Whittle, 2002; Shumway-Cook & Woollacott, 2001). Kinematic movement analysis provides insights into the effects of maturation and development of motor learning (Thelen, 1995), skill development (Newell & van Emmerik, 1989; Vereijken, van Emmerik, Whiting, & Newell, 1992), and the effects of peripheral or central nervous system injury during ADL (Hermsdörfer, Hentze, & Goldenberg, 2006). In addition, kinematic measurements can elucidate the motor strategies in goal-oriented tasks, provide essential information of person's motor capabilities, as well as evaluate upper-extremity therapies (McCrea, Eng, & Hodgson, 2002). For these reason, the study of human movements is applied to cerebrovascular diseases which impact the quality of life for individuals who suffer from diseases such as Apraxia or Action Disorganization Syndrome (AADS) (Bickerton et al., 2011).

2.1 Previous findings

The examination of movement production quality in patients with apraxia and healthy young and elderly controls is examined using high frequency video cameras, marker-based motion capture systems (Ascension, Zebris [TUM], Qualisys [UoB]), and a non-marker-based system (Kinect™). The information can also be used to detect gradual deviations from prototypical object use and provide instructions in order to achieve more adequate movements (see T3.3.2). The following section will provide an overview on the technologies applied in the CogWatch project.

2.2 Marker-based systems

Motion analysis is typically recorded using marker based techniques. There are various systems that utilize different technologies, including cinematographic, electromagnetic (Ascension¸ Liberty Polhemus, Flock of Birds), optoelectronic (OptoTrak, CODA, Selspot), infrared marker systems (Peak Motus, Vicon), and ultrasonic systems (Zebris). Generally speaking, these systems continuously record the 3-dimensional coordinates of small markers fixed to predefined body parts. The Technical University Munich currently utilizes ultrasonic (Zebris) as well as electromagnetic (Ascension, Polhemus) technology to characterize kinematic performance during tool use actions and ADL.

Electromagnetic systems do not suffer from marker dropouts, provide real time 6-DOF data, and are highly accurate. However, they also require that markers must be attached to body segments with cables, and the quality of data is influenced by interference from metallic objects or other magnetic fields. Optoelectronic systems employ active light emitting diodes (LED’s) markers (Ferrigno & Pedotti, 1985; Ladin, 1995), which are triggered and pulsed sequentially using a computer. As with electromagnetic systems, the LED’s must be attached to body segments via cables.

In general, the choice of each system depends on the requirements of the user but, although marker-based systems are popular and offer precision and the advantage of automatic tracking, which eliminates problems with marker merging or misidentification, marker based methods have several limitations: (i) markers attached to the subject can influence the subject’s movement; (ii) a controlled environment is required to acquire high-quality data; (iii) the time required for marker placement can be excessive; and (iv) the markers on the skin can move relative to the underlying bone, leading to what is commonly called skin artefacts (Cappozzo, Della Croce, Leardini & Chiari, 2004; Chiari, Della Croce, Leardini, & Cappozzo, 2004; Rothi & Heilman, 1997). Marker-based systems also relatively expensive and as such are not an appropriate choice for the CogWatch system. That said, they will be able to provide invaluable information about the quality of movement production, which will be integrated into the models that describe actions on a higher level (see T3.3). Furthermore, marker-based systems can serve as the gold standard, against which other kinematic systems can be compared (e.g., Kinect™).

2.3 Kinematic-based action recognition algorithm

In order to recognize actions and errors as they are being executed, we are currently developing a kinematic-based action recognition algorithm (KARA), which is capable of determining action segments during an ADL.

2.3.1 Task and Procedure

In this study, we collected data from a single person (female, 32 years of age) during the performance of the tea making task. The instructions specified the sequence in which the task should be performed: 1) reach and grasp cup, 2) transport cup to target position, 3) reach and grasp tea bag, 4) place tea bag into cup, 5) reach and grasp kettle, 6), transport kettle to position over cup, 7) pour water from kettle into cup, 8) place kettle back on table, 9) reach for tea bag, 10) dip tea bag into cup. Further, the participant was informed that they should use the left hand to perform the task and to rest the right hand on the table. The instructions also emphasized that the task be performed at a naturalistic pace. The participant performed a total of 20 tea-making trials. The entire session took 30 minutes.

The apparatus used in the experiment consisted of a water kettle, a cup, a tea bag, and a tea bag box, which were placed on a table top (90 cm in height). The objects were arranged so that the cup was located 10 cm from the front edge of the table and 40 cm to the right of the participant's body midline. The kettle was located 10 cm from the front edge of the table and 40 cm to the left of the participant's body midline. The tea bag box (with the tea bag inside) was located 20 cm from the front edge of the table and 30 cm to the left of the body midline.

Movement data was recorded using an ultrasonic three-dimensional motion capture system (Zebris® CMS 30, Medizintechnik GmbH, Tübingen) as the reference system. Specifications for the Zebris system are listed in Table 1. This system features three sonic microphones which receive packets of ultrasound sent from emitters that can be placed on relevant body segments. The system calculates the distance between the receiver and each emitter by measuring the time delay between when the ultrasound packet was sent, and when it reaches the receivers. After calculating the distance from the three emitters, the coordinates of the emitters can be triangulated with an absolute accuracy < 1.0 mm (Overhoff, Lazovic, Liebing, & Macher, 2001).

The Zebris system uses a right-handed Cartesian coordinate system, with the transverse plane (front-back) defined as the x axis, the sagittal plane (left-right) defined as the y axis, and the coronal plane defined as the z axis. To collect kinematic data from the Zebris system, a receiver (1 cm diameter) was fixed to the dorsum just proximal to the space between 1st and 2nd metacarpophalangeal joint of the left hand. The spatial coordinates of the markers were sampled at 120 Hz. The data were processed by the software WinData 2.19.14 (Zebris, Medizintechnik GmbH, Tübingen).

Table 1. System CMS30 Zebris features

System Dimensions	95 x 160 x 235(W x H x D)
System Weight	1.1 kg
Sensor Dimensions	360x 320 x 30 (W x H x D)
Sensor Weight	0.7 kg
Max number of marker channels	15
Buffer memory	60 Hz
Interface to PC	USB/Parallel Port
Measurement rate	Max. 300 Hz /number of selected markers
	until 160 Hz- 2.0 m max
	200 Hz – 1.6 m
Resolution	1/10 mm 1/100 mm (30cm)

2.3.2 Data Analysis

The tea making task was comprised of ten subtasks, which were broken down into four action segments (ASs). AS1 was composed of subtasks 1 and 2 (reach and grasp cup, transport cup to target position). AS2 was composed of subtasks 3 and 4 (reach and grasp tea bag, place tea bag into cup). AS3 was comprised of subtasks 5-8 (reach and grasp kettle, transport kettle to position over cup, pour water from kettle into cup, and place kettle back on table). AS4 was comprised of subtasks 9 and 10 (reach for and dip tea bag into cup).

The process of AS identification was achieved by manually segmenting the time series into the four ASs, by first recognizing the most characteristic kinematic feature of a given AS, and then determining the start and end of that segment based on secondary kinematic characteristics.

AS1 was defined as the time period between when the hand left the start position to the time the cup was placed on the target. AS1 was always the first action segment in the time series, and was identified by calculating the first two peaks in the resultant velocity profile with a value greater than 5000 mm/s. AS1 onset was determined as the time of the sample in which the resultant velocity of the hand first exceeded 5% of peak velocity. Movement offset was determined as the time of the sample in which the second resultant velocity peak dropped and stayed below 5%.

AS2 was defined as the time period between when the hand reached for the tea bag to the time the tea bag was placed in the cup. Analysis revealed that the kinematic features of AS2 were not as pronounced as the other ASs. As such, AS2 onset and offset were determined after all other ASs had been defined. AS2 onset was defined as the time of the sample in which the velocity of the hand first exceeded 5% of peak velocity for the resultant velocity peak that immediately succeeded AS1 offset. AS2 offset was defined by first determining the resultant velocity peak that immediately preceded AS3 onset, and then finding the time of the sample in which the second resultant velocity peak dropped and stayed below 5%.

AS3 was defined as the time period between when the hand reached for the kettle to the time the kettle was placed back on the table top. The most recognizable kinematic characteristic of AS3 was the kettle pour subtask. Kinematic analysis revealed that this subtask was distinguishable by a lack of velocity peaks within a 1300 - 3000 ms time period. The first step in classifying action segment three was to determine all peaks in the resultant velocity profile with a value greater than 5000 mm/s, and then calculate the time period between each peak throughout the entire time series. When the time period between two neighbouring peaks exceeded 1300 ms, and 12% of total movement time, this segment was classified as the kettle pour subtask. The peak in velocity that immediately preceded and followed the kettle pour subtask that had a peak velocity value greater than 6000 mm/s was then determined. The start of AS3 was classified as the time of the sample in which the velocity of the hand first exceeded 5% of peak velocity for the peak that immediately preceded the start of the kettle pour subtask. The end of AS4 was classified as the time of the sample in which the resultant velocity dropped and stayed below 5% of peak velocity for the peak that immediately followed the end of the kettle pour subtask.

AS4 was defined as the time period between when the hand reached for the tea bag located in the cup to the time the hand released the tea bag. The most recognizable kinematic characteristic of AS4 was the presence of multiple peaks of similar velocities. To this end, the first step to classify AS4 was to find three or more peaks in the z velocity profile that featured peak velocity values between 3000 – and 10000 mm/s, and that occurred within a 2500 ms time frame. To classify the end of AS4, we calculated the time of the sample in which the resultant velocity dropped and stayed below 5% of peak velocity for the last identifiable peak in the range. The start of action segment was first determined by calculating the 4th preceding peak in velocity, and then calculating the time of the sample in which the velocity of the hand first exceeded 5% of peak velocity.

2.3.3 Results and outlook

The following variables were calculated for each AS: movement time (ms), percentage total movement time (%), z dimension peak velocity (mm/s), resultant peak velocity (mm/s), and time to peak resultant velocity (% AS). Average z dimension peak velocity (mm/s) and average peak resultant velocity (mm/s) were also calculated for AS4. These are displayed in Table 2.

The preliminary results of KARA indicate that an ADL such as tea making can be divided into different subtasks. These subtasks, or action segments, have distinct characteristics, which can be classified on the basis of their kinematic traits. The data was collected at a sampling rate of 120 Hz, which afforded the possibility to detect critical features in each AS. However, it remains to be seen whether KARA is able to accurately detect ASs when a lower sampling rate is used (i.e., a 30 Hz sampling rate used by the Microsoft Kinect™ sensor). Furthermore, the ability of KARA to segment actions into subcomponents is highly influenced by various factors, and as such the present evaluation was performed by one individual, and was highly constrained in task context.

Table 2. Mean kinematic and standard deviation (in parentheses) values during each of the four action segments

	AS1	AS2	AS3	AS4
Movement time (ms)	1148 (115)	4140 (364)	6084 (675)	3847 (509)
Movement time (%)	11.8 (1.3)	26.1 (3.3)	38.19 (2.37)	24.3 (1.7)
z dimension peak velocity (mm/s)	1437 (386)	6341 (1145)	4341 (1908)	3047 (1235)
Peak resultant velocity (mm/s)	7585 (888)	8591 (701)	4767 (4403)	9392 (1808)
Time to peak resultant velocity (% phase)	39.1 (8.2)	10.0 (1.6)	8.9 (7.4)	25.5 (15.3)
Average z dimension peak velocity (mm/s)	N/A	N/A	N/A	2377 (404)
Average resultant peak velocity (mm/s)	N/A	N/A	N/A	4329 (854)

In months 11 to month 24, staff from the Research Group “Movement Science” at TUM will work to refine KARA so that it is able to automatically segment an ADL with high trial-to-trial kinematic variability, different spatial arrangement of manipulated objects, and feature interleaving of two subtasks. Furthermore, data will be collected from healthy young adults, healthy elderly controls, and apraxic individuals, in order to ascertain whether KARA can automatically segment ADL scenarios in populations with different kinematic profiles. Lastly, it is hoped that researchers are able to automate the process of kinematic action, not only in a tea-making task, but in all ADL scenarios described in section 5.3.

3. NON-MARKER KINEMATIC-BASED ACTION RECOGNITION

A critical element of the CogWatch system is to provide personalized, long-term and continuous cognitive rehabilitation of ADL for stroke AADS patients in a home environment. It is being designed to be personalised to suit the needs of individual patients at the same time as being practical and affordable for home installation so that rehabilitation takes place in familiar environments performing familiar tasks. As such, it is imperative that the home-based action recognition system is affordable and easy to implement. Fortunately, emerging technologies have led to the rapid development of low-cost and easy to use marker-less motion capture systems which offer an attractive solution to the problems associated with marker based methods. The most popular of these is the Kinect™ system developed by Microsoft.

In 2011, Microsoft released the Software Development Kit (SDK) focused on Kinect™, which has allowed developers and authors to develop and create different applications depending on the objective pursued with Kinect™**.** Developers have designed non-gaming applications for areas such as interior design (NConnex), tracking consumer behaviour (Kimetric), and for educational purposes (Kinect™ Math UWB B). Regarding the present work, the SDK has supported the corresponding functions in order to develop the correct software and interface to ensure the communication with the device and the performance of the whole system.

Considering rehabilitation applications, Kinect™ has been used before to assist in the recovery from physical and cognitive deficits after stroke (Alankus, Proffitt, Kelleher, & Engsberg, 2010; Chang, Chou, Wang, & Chen, 2012; Chang, Chen & Huang, 2011). Moreover, GestSure Technologies has developed an application to help doctors navigate MRIs and CAT scans during surgery; Jintronix has developed a software application that allows patients with recovering from stroke to perform physical therapy exercises from within their own home; and the Johns Hopkins University uses Kinect™ and gesture based tele-surgery to help in fine and precise manipulation of surgical tools while conducting surgeries.

TUM and UPM-ROMIN have recently conducted a study that analyzes kinematic data obtained from the Kinect during the performance of an ADL (i.e. making a cup of tea), and compared it to data collected at a higher sampling rate from an ultrasonic motion capture system. In the following sections we provide more specific information about the components, software architecture, and general data processing steps required by the Kinect™.

3.1 Kinect™ components

Kinect™ (Figure 1) provides two RGB cameras, a depth sensor, a multi-array microphone and an infrared projector which allow:

- Full-body 3D motion capture
- Facial recognition
- Voice recognition

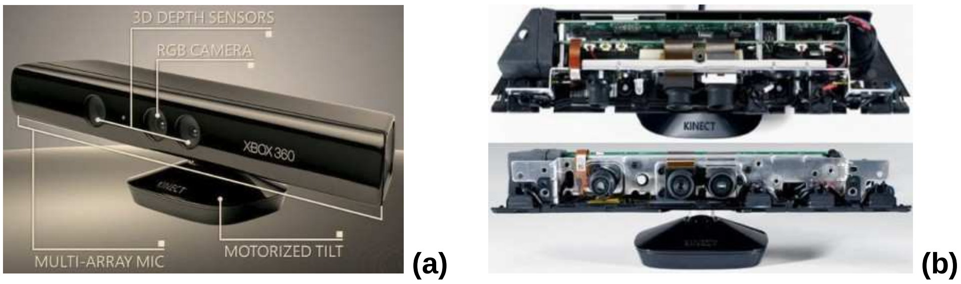
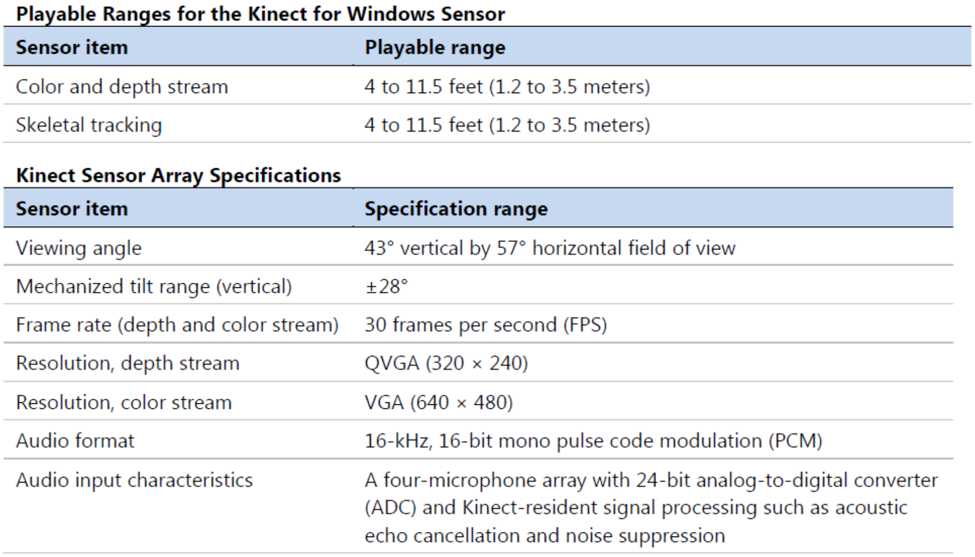


Figure 1. (a) A photo showing all the different parts of the sensor, external view of the 3D depth sensors, RGB camera, multi – array microphone, and a motorized tilt taken at Microsoft's E3 2010. (b) Internal view of the device

The RGB cameras are used for getting a colour image of the workplace and the infrared to obtain information about the depth of the different elements involved in the task. The method of determining 3D position for a given object or hand in the scene is described by the inventors as a triangulation process (Freedman, Shpunt, Machline, & Arieli, 2010). Essentially, a single infrared beam is split by refraction after exiting a carefully developed lens. This refraction creates a point cloud on an object that is then transmitted back to a receiver on the assembly.

Using complex built-in firmware, Kinect™ can determine the three-dimensional position of objects and hands in its line-of-sight by this process. The main advantage of this assembly is that it allows 3D registration without a complex set-up of multiple cameras and at a much lower cost than traditional motion labs and robotic vision apparatuses. Figure 2 provides the relevant information about Kinect™ features:



Playable Ranges for the Kinect for Windows Sensor

Sensor item	Playable range
Color and depth stream	4 to 11.5 feet (1.2 to 3.5 meters)
Skeletal tracking	4 to 11.5 feet (1.2 to 3.5 meters)

Kinect Sensor Array Specifications

Sensor item	Specification range
Viewing angle	43° vertical by 57° horizontal field of view
Mechanized tilt range (vertical)	±28°
Frame rate (depth and color stream)	30 frames per second (FPS)
Resolution, depth stream	QVGA (320 × 240)
Resolution, color stream	VGA (640 × 480)
Audio format	16-kHz, 16-bit mono pulse code modulation (PCM)
Audio input characteristics	A four-microphone array with 24-bit analog-to-digital converter (ADC) and Kinect-resident signal processing such as acoustic echo cancellation and noise suppression

Figure 2. Kinect^{TM} principal features for capturing image video and motion

3.2 Kinect™ software architecture

Figure 3 shows a block diagram representing the whole process from the data acquisition from Kinect™ to the plot and analysis of the signal:

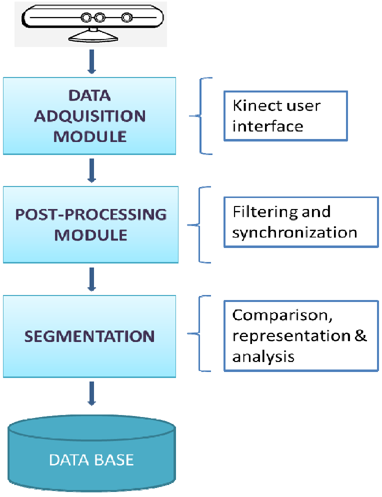


Figure 3. Modular methodology used for data analysis

First of all, a user interface (Figure 4) has been designed in order to interact with Kinect™ and make the acquisition of the data easier. This interface lets the user choose between different options related to the mode in which Kinect™ records the information from the cameras. One of the advantages of selecting the mode is that the user can customize the video image. Instead of recording the whole body, Kinect™ can record only the data from the upper half of the body. This feature is pertinent to the CogWatch system, which requires information obtained strictly from hand movements during ADL performance.

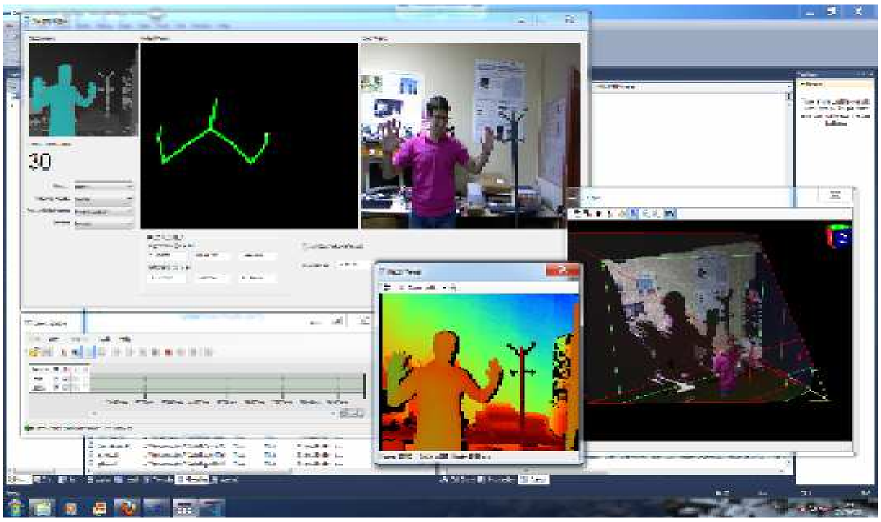


Figure 4. Kinect™ user interface

Besides the different images from the whole scenario, another functionality provided by the interface is the possibility of visualizing in real time the x, y, z positions of the hand (a fixed point in the wrist). The interface, which has been programmed using C++ in Microsoft Visual Studio, has implemented a final option of saving all the position data in a file with a specific format selected by the user. The most common file formats used during the experiments are .xls and .csv.

Secondly, once all the data is saved, all the data is post-processed by loading the files saved before using Matlab in order to plot the signals and study if any filtering is needed. As shown in the following section, filtering has been focused on Butterworth. A Butterworth filter is applied to obtain maximum flatness in the band pass. For this article, it is applied in its digital version that is an infinite response filter (IRF). The Butterworth filter provides the best Taylor Series approximation to the ideal low-pass filter response at analog frequencies $f_r = 0$ and $f_r = \infty$. In this case, no stability problems are found in the domain of the input data.

After filtering the signal, Kinect™ data is compared with the data obtained from Zebris, and used to determine whether the Kinect™ can recognize handmade movements, and if it is a suitable motion capture system for cognitive rehabilitation.

3.3 Validation of Kinect™

Given that traditional marker-based kinematic systems are relatively expensive, they are not an appropriate choice for the CogWatch system. TUM and UPM are involved in a collaborative project to determine whether the Kinect™ can recognize handmade movements, and if it is a suitable motion capture system for cognitive rehabilitation.

3.3.1 Task and Procedure

In order to test the suitability of the Kinect^{TM} device as a motion capture system we collected data from a neurologically healthy 32 year-old female, and a 47 year-old female with apraxia. Both participants were right handed. The apraxia patient had left-brain damage which resulted in hemiparesis of the dominant right hand. As such the apraxia patient had the use of the non-dominant left hand only. Given that the apraxia patient was hemiparetic, the control participant was also required to use only the left hand to perform the task. The instructions emphasized that the task must be performed at a normal pace. The participant performed a total of 20 tea-making trials. The entire session took 30 minutes.

The sequence of actions and apparatus is identical to that used to develop KARA (section 2.2.1). Kinematic data was collected by a Kinect sensor with a 30 Hz sampling frequency, and an ultrasonic motion capture system (Zebris) at 120 Hz.

3.3.2 Data Analysis

In the first step of data analysis, the kinematic data was loaded into a custom written MatLab program (The MathWorks®, Version R2010a), and the 3D coordinates were low-pass filtered at a 5 Hz cut- off, using a first order Butterworth filter. The data was also filtered using an IRF, which was selected because it reduced the amount of delay in the signal, and it is known to be more computational efficiency than a finite response filter. This low-pass filter adequately smoothed the signal, rejected movement artefacts and high frequency phenomena (e.g., aliasing), and eliminated power line harmonic interferences. Figure 5 shows the signal before (blue signal) and after (red signal) filtering Kinect data:

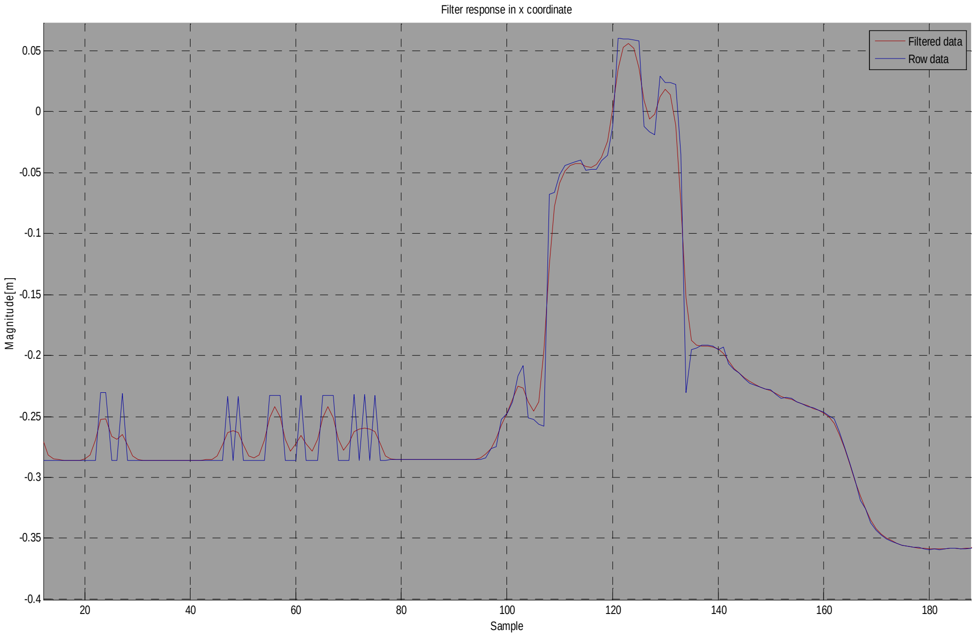


Figure 5. Filtering of Kinect™ data

In the next step, the Zebris data was then down-sampled to 30 Hz and then normalized by computing standardized z-scores using the mean and standard deviation of the vector for each axes (transverse, sagittal, coronal). This procedure allowed the comparison of the obtained data from the two motion capture systems to be done directly. The difference between Kinect^{TM} and Zebris data were quantified utilizing Mean Square Error (MSE) and cross-correlation measures separately for each axes.

MSE was computed using the matlab *mean squared normalized error performance* (mse) function, which measures the expected squared distance between an estimator (Zebris) and the observations (Kinect^{TM}). Because the signals from the Kinect^{TM} and Zebris have different units, MSE analysis was conducted on the normalized data (mean = 0, SD = 1). As such, the reported MSE values in Figure 6 are dimensionless. That said, MSE values can be interpreted in a qualitative fashion, such that a MSE of 0 means the estimator (Zebris) predicts observations (Kinect^{TM}) with perfect precision. In contrast, larger MSE values indicate that the observation values (Kinect^{TM}) differed from the estimator (Zebris).

Cross-correlations were computed by using the Matlab corrcoef function, and that function was used to estimate the dependence of the values obtained from the two motion capture systems. The correlation coefficient ranges from -1 to 1. A value of 1 implies that a linear equation describes the relationship between the Zebris and Kinect™ perfectly. Small values or a value of 0 implies that there is no linear correlation between the variables. Negative values indicate an inverse relationship between the variables (i.e., the values of one of the variables increase, the values of the second variable decrease).

3.3.3 Results

Hand position for each axes for the control participant and apraxic individual are displayed in Figure 6. It was observed that Kinect^{TM} (blue signal) was able to adequate track 3D hand positions, and was similar to that collected by a marker-based ultrasonic motion capture system Zebris (green signal). The average MSE and cross-correlation values for each axis are displayed in Table 3.

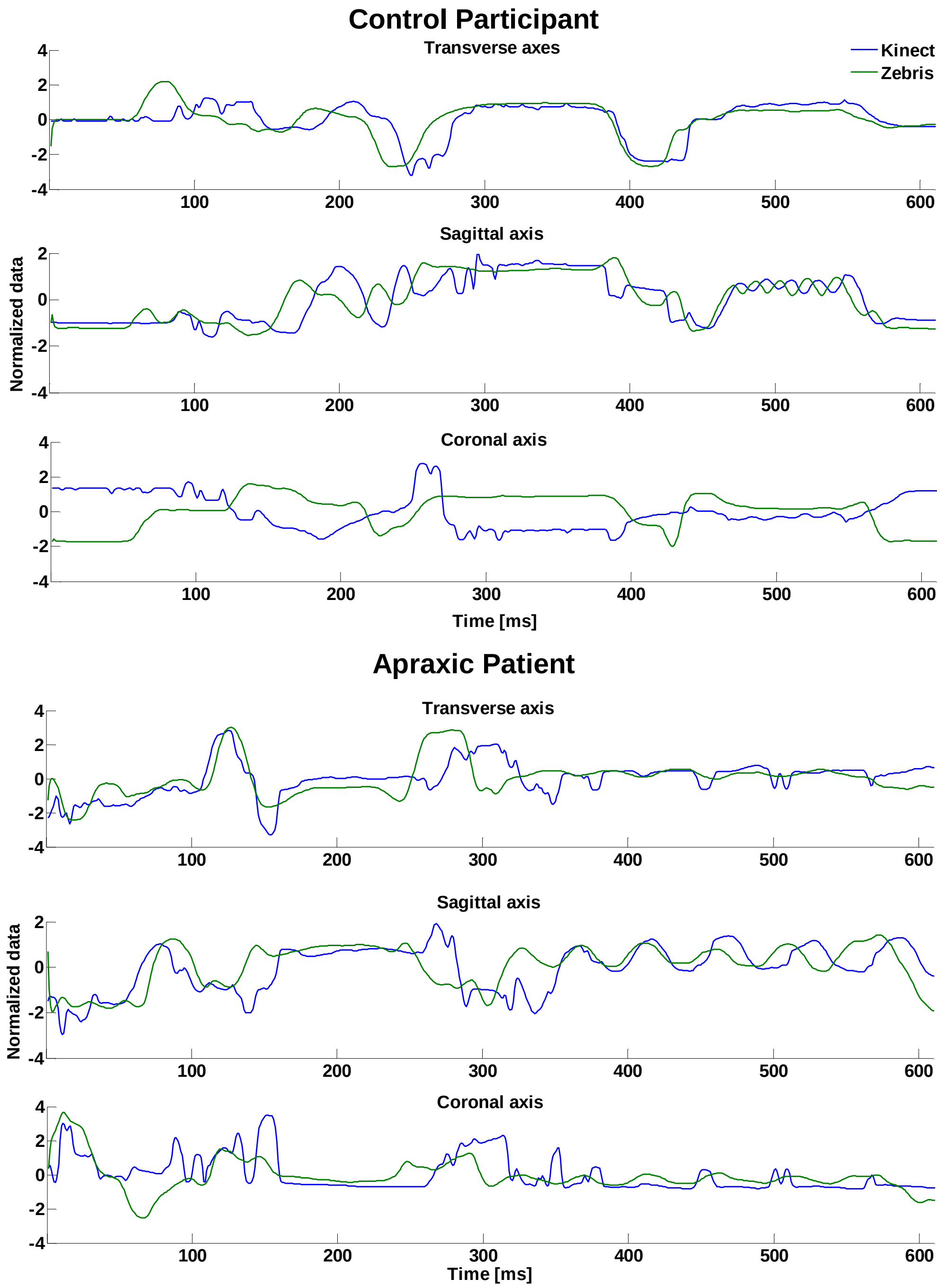


Figure 6. Comparison of position data for the Kinect™ (blue lines) and the Zebris system (green lines) from a single tea making task trial for the control participant (top graph) and the apraxia patient (bottom graph)

In general, average MSE values for the control participant ranged from 0.938 to 1.143. The MSE values for the apraxic individual were slightly higher, ranging from 1.265 to 1.322. Statistical analysis indicated that MSE values for the control participant and apraxic patient were significantly different from one another, F(1,19) = 10.017, p = 0.005. This difference was primarily driven by the MSE values in the sagittal axes, which were lower for the control participant than the apraxic patient, F(1,19) = 4.478, p = 0.018.

These results indicate that the overall observation (Kinect^{TM}) and estimator values (Zebris) were similar to one another ($p > 0.05$) regardless of axes, and that the obtained data from the Kinect^{TM} and Zebris motion capture systems were similar regardless of the neurological status of the participant for the transverse and coronal axes, but were significantly different from another in the sagittal axes.

Table 3. Mean Square Error (MSE) and cross-correlation values for the transverse, sagittal, and coronal axes

	Mean square error (MSE)			**Cross – correlation**		
	Transverse	**Sagittal**	**Coronal**	**Transverse**	**Sagittal**	**Coronal**
Control	1.143 (0.454)	0.938 (0.417)	1.074 (0.453)	0.427 (0.114)	0.489 (0.137)	0.525 (0.075)
Apraxic	1.265 (0.343)	1.322 (0.224)	1.292 (0.306)	0.367 (0.172)	0.338 (0.112)	0.353 (0.153)

Cross-correlation analysis indicated that the relationship between the Kinect^{TM} and Zebris systems ranged from 0.427 to 0.525 for the control participant. Based on the Cohen scale(Cohen, 1988), these results indicate a medium correlation between motion capture systems for the transverse and sagittal axes, and a strong correlation between motion capture systems for the coronal axis. The cross-correlations for the apraxic patient ranged between 0.338 and 0.367, indicating a moderate correlation between the tested motion capture systems for all three axes. Statistical analysis indicated that the cross-correlations were similar between the control participant and the apraxic patient [F(1,19) = 0.923, p = 0.349] and across all three axes [F(1,19) = 2.089, p = 0.138].

3.4 Summary and outlook

Overall, the results indicated a moderate to strong correlation between signals in the control participant, and moderate correlations between signals in the apraxia patient. Furthermore, although the sagittal axes MSE values differed between the control participant and apraxia patient, the transverse and coronal axes MSE values were similar for both participants. Taken together, the research presented indicates that the Kinect^{TM} device is able to adequately track hand movement during an ADL, regardless of the neurological status of the individual. These findings are indeed promising given that the sampling frame of the Kinect^{TM} device is lower than the Zebris system (30 Hz compared to 120 Hz), the position of the hand is determined using RGB cameras and a depth camera, and that the Kinect^{TM} sensor costs significantly less than its marker-based counterparts.

As future work, more detailed experiments will involve multiple ADL tasks (e.g., making toast, putting on a shirt, etc.) and various AADS patient populations (e.g., left-brain damage, aphasic). Overall, while more researches could be done into the capabilities and limitations of the device in physical therapy, this work determined that Kinect™ has a high potential for use in stroke rehabilitation as a tool for both clinicians and stroke survivors.

4. SENSOR-BASED ACTION RECOGNITION

The EECE group at UoB have been utilizing sensor technologies in order to detect the actions at the sub-goal level. The types of sensors include radio frequency identification (RFID) tags, accelerometers, force sensitive resistors (FSR), and force sensitive handles. The rationale for the choice of sensors was presented in detail in CogWatch deliverable D2.1.

In preparation for prototype 1, researchers at UoB have focused their efforts on detecting sub-goals using CogWatch instrumented coasters (CIC) which can be attached to the base of a jug or cup. The CIC communicates with a host computer via Bluetooth, where its outputs are sampled at 200Hz and stored in a file. A CIC data file comprises a sequence of six dimensional feature vectors, comprising *x*, *y* and *z* accelerometer values and the outputs of three force sensitive resistors (FSRs), produced 200 times per second.

4.1 Tasks

Naive participants performed four types of activity with the instrumented kettle:

- “pour” – participants were asked to pour liquid from the kettle into a mug. They were told that the kettle should be at rest on the desk surface at the start and end of the activity.
- “toy” - participants were asked to move the kettle arbitrarily, whilst avoiding the “pour” activity. Again they were told that the kettle should be at rest on the desk surface at the start and end of the activity.
- “pour-toy” - participants were asked to pour liquid from the kettle into a mug and then to toy with the kettle. They were told that the kettle should be at rest on the desk surface at the start and end of the activity. This data was used only for testing the detector.
- “rest” – participants were asked to leave the kettle at rest on the desk top.

A total of 96 files were recorded. These were partitioned into a training set of 63 files and a test set of 33 files. For each data file a label file was created indicating the sequence of actions represented by the data. No timing information was included in the label file.

4.2 Visualization of the data

Figure 7 shows CIC data for an example of the “pour” activity lasting approximately 9.75s. The times at which the jug is raised and replaced on the surface are evident from the FSR graphs in the figure. Before the lift the three FSRs have different values, due, presumably, to the characteristics of the individual FSRs and the precise orientation of the mug. However, after the lift all of the FSRs give a consistent value of approximately 4100 until the jug is put down again after approximately 1200 samples. The FSRs then take between 200 and 800 samples to return, approximately, to their original values.

The accelerometer values are more difficult to interpret and appear to be noisier than the FSR values when the jug is at rest. The third accelerometer reading (acc 3) corresponds to the vertical axis (assuming that the cup is upright), therefore it is measuring the effect of gravity as 1g accelerating upwards. When the mug is tipped most of the structure in the accelerometer graphs is the gravity component moving across the axes. The gravity component across the three axes always adds up to 1g but it is quite hard to distinguish between the effects of gravity and more general motion when the item is rotated.

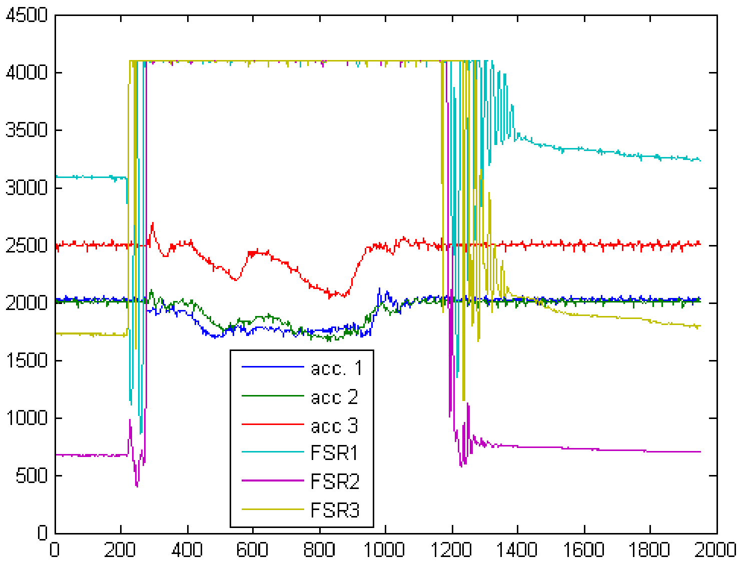


Figure 7. CIC outputs for an example of the "pour" activity lasting approximately 9.75 seconds

4.3 Data Analysis

All experiments were conducted using the Cambridge University Engineering Department's Hidden Markov Model (HMM) Toolkit, HTK (Young et al., 2006), which is the *de facto* standard for offline experiments in HMM based automatic speech recognition.

4.3.1 HMM parameter estimation

Basic HMM parameters, such as the number of states, the number of components in the Gaussian Mixture Model (GMM) state output probability density functions (PDFs), and the structure of the initial state probability vector and state transition probability matrix, are chosen manually. Once these have been set, the remaining parameters are estimated from data.

The "pour" activity was modelled as a 9 state left-right HMM. This was chosen to try to capture the sequential structure of the "pour" action. Intuitively, the 'segments' of the "pour" activity are (1) the jug at rest, (2) the jug is raised, (3) the jug is moved, (4) the jug is tipped, (5) the jug is moved, (6) the jug is lowered, and (7) the jug is at rest. Two additional states were included to obtain a better piecewise constant approximation to the complex tipping movement. Only transitions from a state to itself or the next state were permitted, to reinforce the sequential structure of the activity. Initially, each state of the "pour" HMM was associated with a single Gaussian PDF.

The "toy" and "rest" actions were modelled as single state HMMs with multiple- component GMM state output PDF. For example, it was envisaged that different GMM components would capture different positions of the object during toying.

4.3.2 HMM model optimization

Initialisation was implemented using the HInit tool in HTK (Young et al., 2006). All states are first associated with single Gaussian state output PDFs. Each training file is segmented uniformly into *N* segments, where *N* is the number of states in the corresponding model (thus *N=9* for "pour" and *N=1* for "toy" and "rest"). The (vector) means and variances of these segments were chosen as the initial mean and variance of the multivariate Gaussian PDFs associated with the corresponding state. These initial HMM parameters were refined using 10 iterations of the Baum-Welch maximum likelihood (ML) optimisation algorithm (using the HTK tool HERest).

Two approaches to increasing the number of GMM components in the models were tested. These are described in full in D3.3.1. The most successful approach was as follows: For the "pour" model, each state output PDF was split into two PDFs along its direction of maximum variation using the HTK HHed tool, and a further 5 iterations of Baum-Welch training were applied. For the "toy" and "rest" models the same splitting technique was used separately to produce single state HMMs with 2, 4, 8, 16, 32 and 64 component GMM state output PDFs. In all cases a further 5 iterations of Baum-Welch parameter optimisation were applied.

4.3.1 Activity detection

The 'recognition grammar' (

8) is designed to set up a competition between the "pour", "toy" and "rest" models, in terms of which one can provide the best explanation of the data at any point. During recognition, each of the solid square boxes is replaced by the corresponding HMM. Given a sequence of test vectors o corresponding to a particular movement of the jug, the recognition algorithm (known as the Viterbi decoder) finds the sequence of states s such that the joint probability of o and s is maximised. From this state sequence the optimal explanation of the data as a sequence of "pour", "toy" and "rest" activities can be recovered.

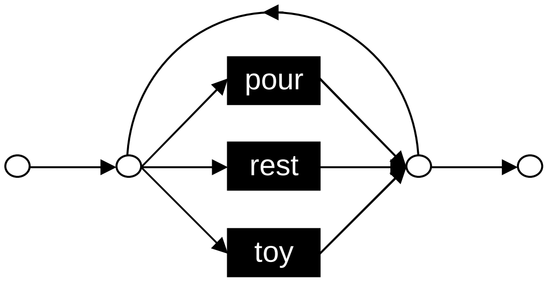


Figure 8. Recognition network for "pour" detection

4.4 Results

4.4.1 Activity recognition using the “raw” CIC data

Activity recognition accuracy using the raw data from the CIC is very poor. The prime cause is variability in the outputs of the FSRs in the CIC. It has already been observed that the values from FSRs are consistent (and approximately equal to 4096) when the jug is not resting on the surface, but variable when it is. This causes most instances of “rest” to be recognized as “toy” and causes confusion between “pour” and “toy”. Although misrecognition of “rest” as “toy” is not practically significant, it indicates serious problems with the recognition algorithm that need to be addressed.

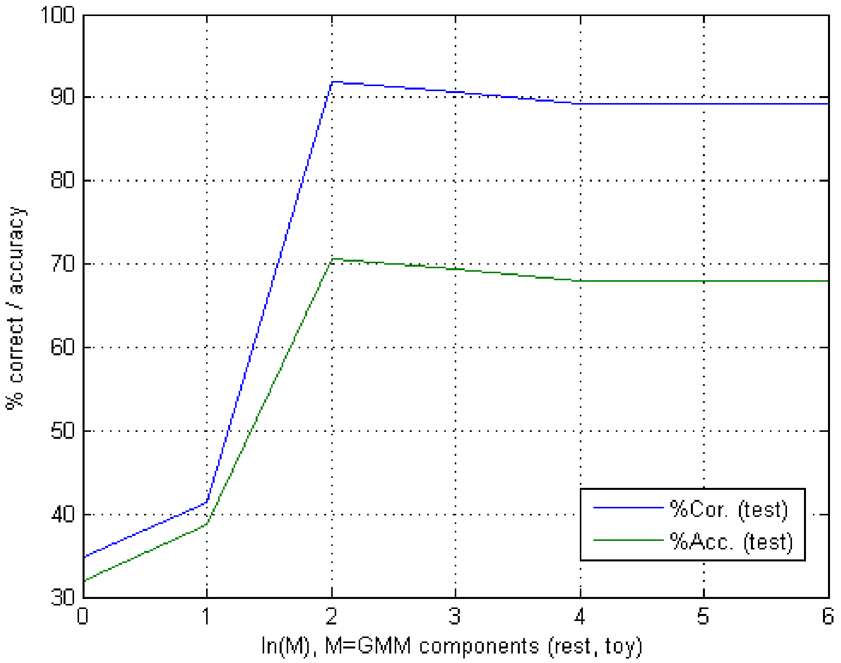


Figure 9. Activity detection results using thresholded FSR data

4.4.2 Activity recognition using thresholded FSR data

A simple solution to the variability in the outputs of the FSRs is to threshold their outputs so that values above 4000 are assigned the value 1, and those below 4000 are assigned 0. This is referred to as “thresholded FSR data”. The results of experiments with thresholded FSR data are presented in 9. The figure shows % accuracy and % correct as a function of the number of GMM components in the “toy” and “rest” HMMs. These are defined as:

$$\%Corr. = \frac{N-S-D}{N} \times 100, \ \%Acc. = \frac{N-S-D-I}{N} \times 100,$$

where *N*, *S*, *D* and *I* are the numbers of test samples, substitution errors, deletion errors and insertion errors, respectively. These are computed from an alignment of the recognizer output and the true transcription, using the HTK tool HResults. The figure of 70% accuracy for 8 component GMM state output PDFs in the "toy" and "rest" models are the best results achieved to date.

4.5 Discussion

Although simple thresholding of the FSR outputs removes significant variability and improves recognition accuracy, it is not a good solution. For example, an important additional cue for detecting "pouring" is the increase in weight of the mug that has milk poured into it. Calibration of the FSRs (deliverable D.2.3.1) has shown that the FSRs can detect weight changes of the order of 5 grams, which is less than the weight of milk that is typically added to a cup in tea making. However, this change in FSR output is within the variability that is seen in the raw FSR outputs, and would be thresholded to the 0 value corresponding to the mug resting on a surface.

A more satisfactory solution would be to use the derivatives of the FSR outputs. These should be equal to 0 when the mug is at rest (either on the surface or lifted above the surface) but adding milk to the mug should result in positive derivatives. However, because the FSR outputs are noisy, either the derivatives need to be estimated over a long interval or the FSR data needs to be low-pass filtered before the derivatives are computed. The latter is a more satisfactory solution.

Another issue with the FSR data is whether a more stable output could be produced by averaging the values from the three FSRs. This is not entirely straightforward because the relationship between load and FSR output is non-linear. However, once the FSRs have been calibrated (D.2.3.1) linearization of their outputs becomes possible.

In addition, there is currently no pre-processing of the accelerometer outputs. While the *z* axis accelerometer data is important in detecting raising and lowering of the jug, the direction information in the *x* and *y* axis accelerometer is not currently being used explicitly and may be noise if the relative positions of the jug and mug vary.

Moving away from the CIC data, another issue is that the "toy" data may represent more "extreme toying" than one would expect to observe in practice, especially if movements of the jug are restricted because it contains liquid. It is possible that some of the "toy" / "pour" confusions arise because the "toy" data includes activity that is very close to "pouring".

4.6 Summary and outlook

In summary, the result of the experiment to detect the "pour" activity presented here are encouraging. However, potential improvements to the CIC outputs and the difficulty of the current version of the task suggest that better performance may be achievable. The EECE group at UoB will continue to ameliorate these issues, as the activity recognition in prototype 1 will be based primarily on the CIC data. More detailed information of the experiments, as they pertain to predictive models, will be presented in the forthcoming D.3.3.1 deliverable.

5. VIDEO-BASED ACTION RECOGNITION

To analyze error production and action sequencing in apraxia patients, the clinical screening and laboratory experiments are recorded by several video cameras. The videos are then evaluated with respect to error production (e.g., using the wrong object for a given action, failing to turn on the kettle), and action sequencing. This information is then used to develop the Bayesian Logic Network (BLN) models that will be used in the second CogWatch prototype (T3.4).

In addition, the video data will provide information about errors related to grasp selection. For example, neurologically healthy individuals will often grasp objects with an initially uncomfortable grasp posture if this affords control over the object during later stages of the movement (Rosenbaum et al. 1990). Extending this work, Randerath and colleagues (Randerath, Li, Goldenberg, & Hermsdörfer, 2009) found that the selection of grip type in apraxic individuals is impaired, and is influenced by task context. In general, nearly all patients (except for a few LBD patients) grasped the objects with grips that ensured end-state control when asked to demonstrate how the tool is typically used. However, when asked to transport the tool to another location, participants selected end-state control compliant grasp postures in approximately 50% of trials. The importance of grasping objects should not be underestimated. Taking hold of a kettle with heated water with a grip that does not lead to end-state control might result in hot water being spilled over the table top and onto the patient. The CogWatch system should account for such issues, and provide appropriate feedback to the patient so that they can correct their movements.

5.1 Action Sequencing and Error Recognition

Given the deficits in action sequencing and the errors in the movement quality or space (see D1.2) in apraxic populations a model needs to be able to describe both low-level motor defects, (e.g., grasping an object with an inappropriate grip), and high-level errors (e.g., performing a task in a wrong sequence). Furthermore, there exists a great deal of freedom in how an ADL task can be performed, such that the same goal can be reached by significantly different action sequences. In these tasks, subsequent actions depend not only on the previous one, but on all actions that have already been performed, since they determine which other actions are still needed to complete the task at hand.

The Bayesian Logic Network model developed by the Intelligent Autonomous Group (IAS) at the Technical University of Munich (TUM) is able to handle the high degree of variation often observed in ADL tasks (Tenorth, 2011). In general, BLNs are statistical relational models that combine the expressiveness of first-order logics, necessary to describe the complex interactions between actions and the parameters associated with these actions, with the representation of probability in a probabilistic logical language.

From training data, partially-ordered models can learn which actions are relevant and which ordering relations are important, such that actions that occur in all observations of a task are considered more relevant than those that are only rarely observed, and ordering relations that consistently hold are also more likely to be important. Thus, the advantage of this approach is that the system learns a model that is able to describe complex tasks including their partial order from observed data.

5.2 Initial phase: Clinical screening

Although a number of different tasks will be considered during the CogWatch project, the ADL scenario chosen for the clinical screening phase is a two tea-making task. This task requires that the patient make one cup of tea with milk and two sweeteners, and another cup of tea with a lemon slice and one sugar cube. The results of the clinical screening study will be compared with data obtained from young healthy adults and elderly control participants. The data was analyzed with respect to error production and action sequencing, which will be used to develop the Bayesian Logic Network (BLN) models that will be used in the second CogWatch prototype (T3.4).

5.2.1 ADL Scenario: The two tea-making task

The main ADL scenario for the clinical screening phase is a two tea-making task. This task was chosen because it is highly relevant to everyday life, should be familiar to the majority of participants, and is sufficiently complex to ensure the inclusion of a substantial number of apraxic patients, and also enables analysis about the selectivity of the effects of apraxia. Tea making has also been thoroughly studied in the literature, and thus provide a basis from which we can compare our results. Figure 10 shows a hierarchical tree based description of the task.

Six patients (age = 55.83 y, SD = 14.17, 4 men, 2 women) with lesions following a single cerebrovascular accident (CVA) participated in the study. There were 2 left-handed and 4 right-handed patients. Six healthy participants served as the control group (age = 35.50 y, SD = 9.18, 2 men, 4 women). None of the control participants had any history of neurological disorders or any constraints of upper limb movements. Five control participants were right-handed, and one control participant was left-handed.

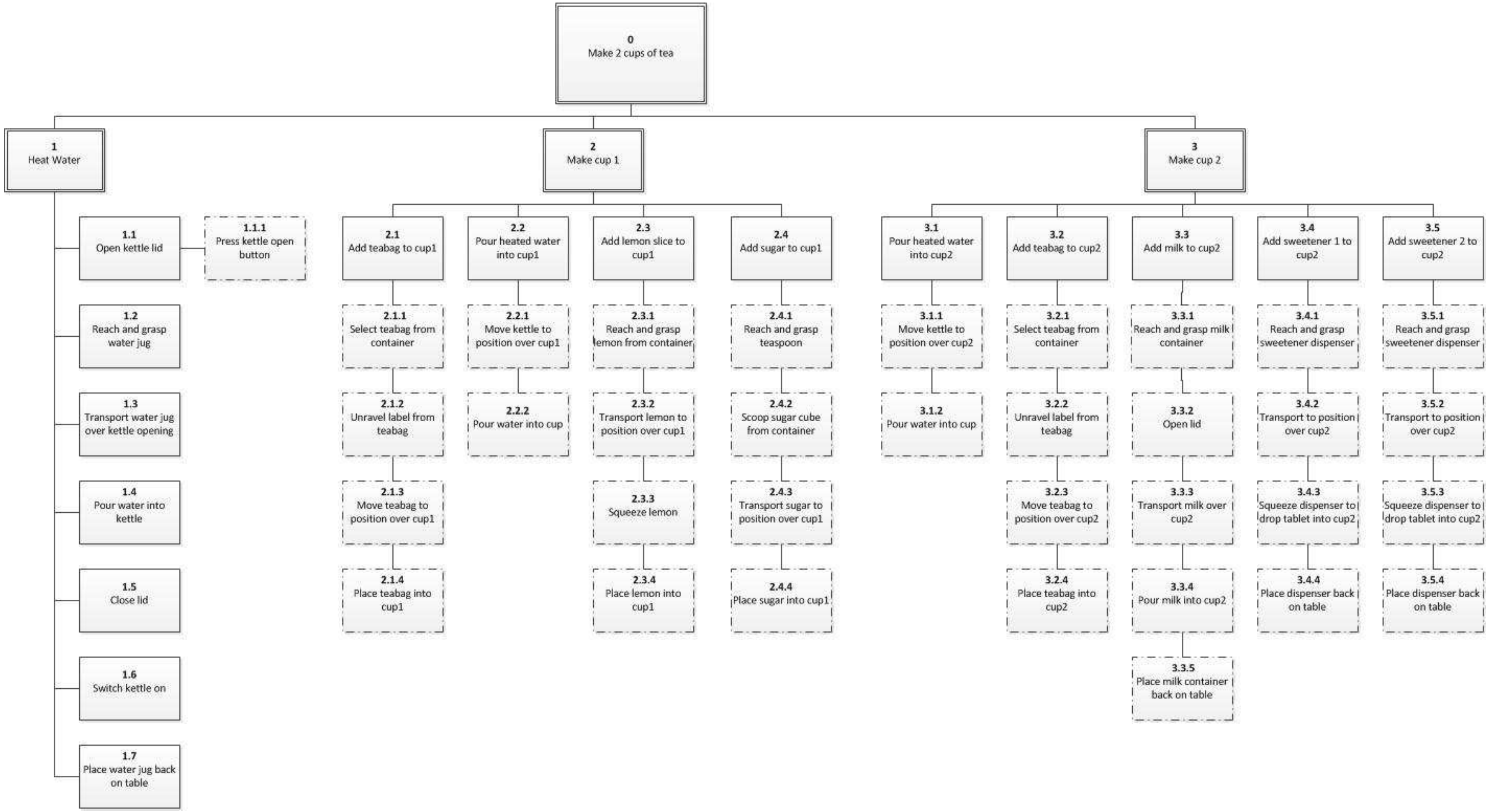
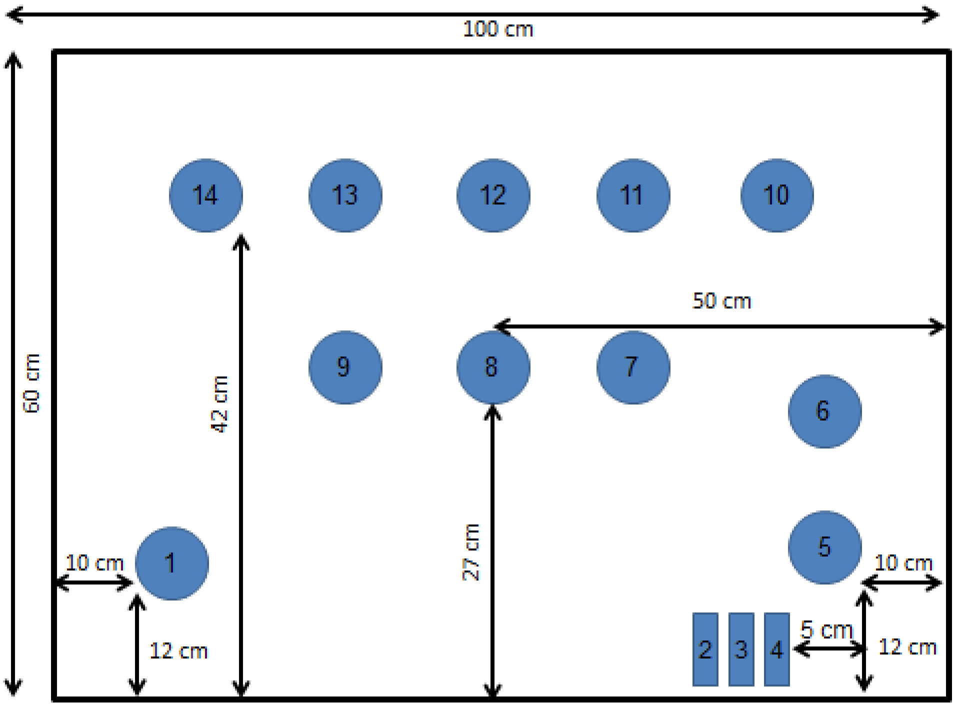

0
Make 2 cups of tea
1
Heat Water
1.1
Open kettle lid
1.1.1
Press kettle open button
1.2
Reach and grasp water jug
1.3
Transport water jug over kettle opening
1.4
Pour water into kettle
1.5
Close lid
1.6
Switch kettle on
1.7
Place water jug back on table
2
Make cup 1
2.1
Add teabag to cup1
2.1.1
Select teabag from container
2.1.2
Unravel label from teabag
2.1.3
Move teabag to position over cup1
2.1.4
Place teabag into cup1
2.2
Pour heated water into cup1
2.2.1
Move kettle to position over cup1
2.2.2
Pour water into cup
2.3
Add lemon slice to cup1
2.3.1
Reach and grasp lemon from container
2.3.2
Transport lemon to position over cup1
2.3.3
Squeeze lemon
2.3.4
Place lemon into cup1
2.4
Add sugar to cup1
2.4.1
Reach and grasp teaspoon
2.4.2
Scoop sugar cube from container
2.4.3
Transport sugar to position over cup1
2.4.4
Place sugar into cup1
3
Make cup 2
3.1
Pour heated water into cup2
3.1.1
Move kettle to position over cup2
3.1.2
Pour water into cup
3.2
Add teabag to cup2
3.2.1
Select teabag from container
3.2.2
Unravel label from teabag
3.2.3
Move teabag to position over cup2
3.2.4
Place teabag into cup2
3.3
Add milk to cup2
3.3.1
Reach and grasp milk container
3.3.2
Open lid
3.3.3
Transport milk over cup2
3.3.4
Pour milk into cup2
3.3.5
Place milk container back on table
3.4
Add sweetener 1 to cup2
3.4.1
Reach and grasp sweetener dispenser
3.4.2
Transport to position over cup2
3.4.3
Squeeze dispenser to drop tablet into cup2
3.4.4
Place dispenser back on table
3.5
Add sweetener 2 to cup2
3.5.1
Reach and grasp sweetener dispenser
3.5.2
Transport to position over cup2
3.5.3
Squeeze dispenser to drop tablet into cup2
3.5.4
Place dispenser back on table

Figure 10. Hierarchical tree representation of the two tea-making task

Participants sat at a table with a dimension of 100 cm x 60 cm. The spatial arrangement of the objects on the table is shown in Figure 11, with a total of 14 objects located on the work surface. Each participant was asked to perform a 2 cup tea-making task, in which one cup of tea required milk and two sweeteners, and the other cup of tea required lemon and one sugar cube. Participants were informed that all the things required to make the tea are on the table, and that they were to inform the experimenter if they required help stabilizing an object. Two trials were performed. Actions were recorded by a video camera (Panasonic HDC-SD909) located 45° to the right side of the table.



1. Electric kettle
2. Teaspoon
3. Dessert spoon
4. Fork
5. Water jug
6. Jug of milk
7. Cup2
8. Saucer for used teabags
9. Cup1
10. Slices of lemon
11. Jar of teabags
12. Sweetener
13. Jar of coffee
14. Jar of sugar cubes

Figure 11. Spatial layout of objects for the two tea-making task

5.2.2 Results

5.2.2.1 Error Types and their frequency

Control participants successfully completed the task in 75% of trials (total 3 errors made in three trials). All three errors were considered to be substitution errors, where the participant added two sugar cubes to cup2 and one sweetener to cup1 (33% of errors), or added two sugar cubes of sugar to cup2 and one sugar cube to cup1 (67% of errors).

Figure 12 shows the proportions of errors during the tea making task for apraxic patients. Apraxia patients committed errors in 46% of trials, with a total of 35 errors recorded. Patients also committed an error in at least one trial; with the number of errors per trials ranging from 0 - 5 (mean = 2.0, SD = 1.5). The most frequently occurring error was that of omission (51% of errors) with patients failing to pour water from the jug into the kettle, put tea bags into one or both cups, or adding sweetener to the cup that required it.

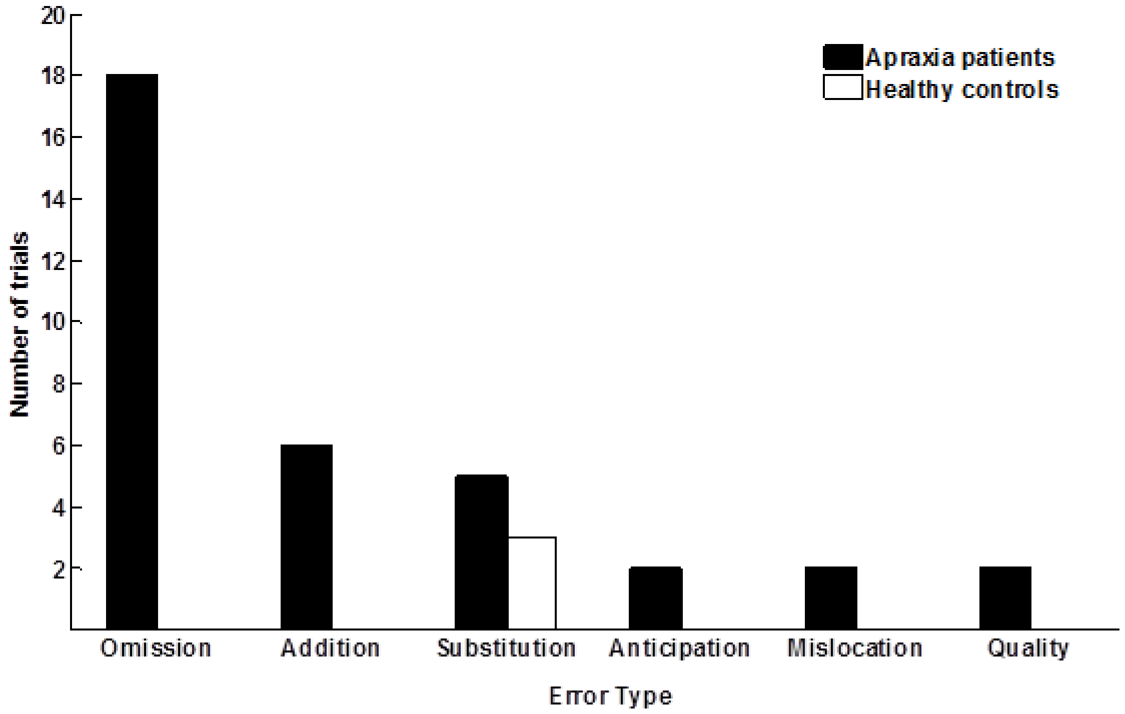


Figure 12. The distribution of errors by error type during the two tea-making task

Patients produced addition and substitution errors in 17% and 14% of trials, respectively. Examples of addition errors include adding coffee to a cup of tea, or putting sugar or lemon into the cup that did not require it. There were also a small number of trials in which patients committed mislocation, anticipation, and quality errors (6% each). In these trials, apraxia patients failed to open the packet of sweetener before pouring the contents into the cup (mislocation errors), added coffee into a cup instead of a tea bag (substitution errors), or failed to pour enough water into the kettle to fill two cups of tea (quality errors).

Lastly, the presence of distractor objects on the workspace (i.e., jar of coffee, dessert spoon, fork) did not impact the quality of performance. Out of the six apraxia patients tested, only two patients committed an error involving a distractor object (i.e., coffee into a cup instead of a tea bag). This finding is counter to previous research (Moores et al., 2003; Schwartz et al. 1998) suggesting that semantically related distractors compete for selection with appropriate target objects for action. More detailed analysis on a larger group of apraxia patients is needed to ascertain how distractors influence action sequencing and error production during ADL performance.

In sum, the error production results are consistent with previous research (Buxbaum et al 1998; Schwartz et al 1998) demonstrating that omission errors are the most commonly committed type of error during ADL. In addition, errors did not appear to be related to the laterality of lesion, hemiparesis, or aphasia type. Future research into error production in apraxic populations will continue to examine this issue in order to ascertain the variables that correlate to error production. Further, the data obtained from the clinical studies will be used to develop the TMs for prototype 2.

5.2.2.2 Modeling action sequencing and error production

Figure 13 depicts the conditional probabilities inside the precedes-node of the BLN. The visualizations contain all nodes that have a probability of at least 0.002. The ellipse dimensions are proportional to the node's probability, the thickness of the edges is proportional to the conditional probability of the target given the origin times the probabilities of the target and origin nodes. In order to improve clarity, the redundant relations between actions have been pruned. For example, in instances in which action A always precedes action B (probability = 1) the edge A – B was not drawn.

As can be seen, the algorithm is able to successfully recover the partial-order structure from the data obtained from both healthy and patient populations. The results of the BLN approach indicate that the relevance of the nodes (i.e., actions) is different between the two groups, indicated by the different sizes. There are more nodes in the patient group, which were caused by addition or substitution errors or by alternative ways of solving the task using a different set of tools. Apraxic patients were more consistent in some relational orderings, typically pouring the water into both cups before adding the ingredients (sugar, sweetener, lemon, milk), which can be seen by the very bold arrows between some of the actions. In comparison, control participants added the ingredients before pouring water into the cups (indicative of a strong ordering relation), but the order in which the ingredients were added was not consistent (i.e., weak ordering relation). This can be seen from the comparatively strong ordering relations between the action groups (e.g. from adding the tea bag to adding sugar or sweetener), compared to much weaker relations between actions (e.g., adding sugar or sweetener).

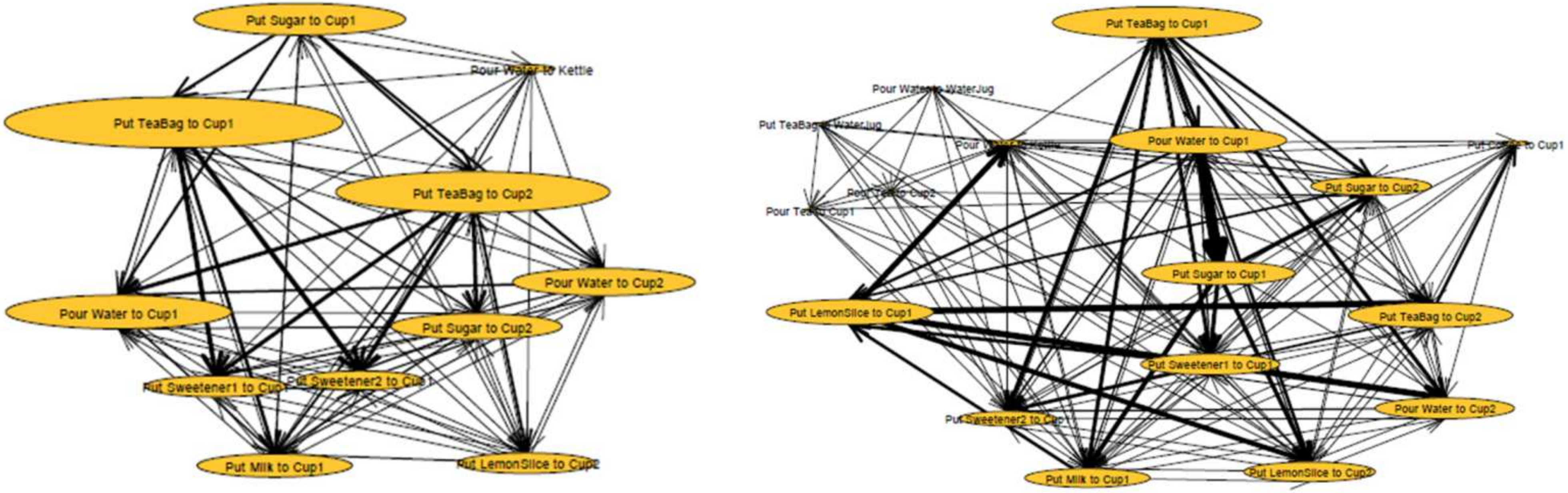


Figure 13. Learned dependencies in six healthy controls (left) and eleven apraxia patients (right)

6. ACTION RECOGNITION DATA ANALYSIS OUTLOOK

In the laboratory study phase, more detailed analysis of ADL performance will be obtained from healthy young adults, healthy elderly individuals, and AADS patients. In this section, we outline how potential participants are identified and recruited, the new TUM CogWatch Laboratory, and the four ADL tasks used to develop the action recognition and prediction models for prototype 2.

6.1 Laboratory Facilities

The studies will take place in the CogWatch laboratory of the Lehrstuhl für Bewegungswissenschaft, Fakultät für Sport- und Gesundheitswissenschaft at TUM (see Figure 14 for a layout of the CogWatch laboratory). The aim of the CogWatch Laboratory is to provide an environment that closely replicates a familiar environment of the AADS patients and control participants. With this in mind, the data collection station is located behind the kitchen set up, and all equipment is out of plain sight during the experiment.

The TUM CogWatch Laboratory is outfitted with an adjustable kitchen environment. The kitchen set up is designed so that it replicates a standard kitchen, but is complies with the American Disability Association (ADA) standards for the needs of individuals who are bound to wheel chairs. The Kitchen set up features a fully functioning basin and sink, two height-adjustable wall shelves, and a height-adjustable kitchen preparation area.

In order to capture the necessary data, a metal cage is attached in the room (not shown) that spans the outside area of the kitchen set up. The cage is used to mount the five high speed Flea cameras (Flea3 FL3-U3, Point Grey Research Inc, Richmond, BC. Canada) and the Zebris motion capture system. In addition, the Kinect™ motion capture sensor is located at the front of the kitchen set up, and provides an unobstructed view of the food and drink preparation area. If required, a second Kinect™ system can be attached to the cage in order to collect data from the sink and shelf area.

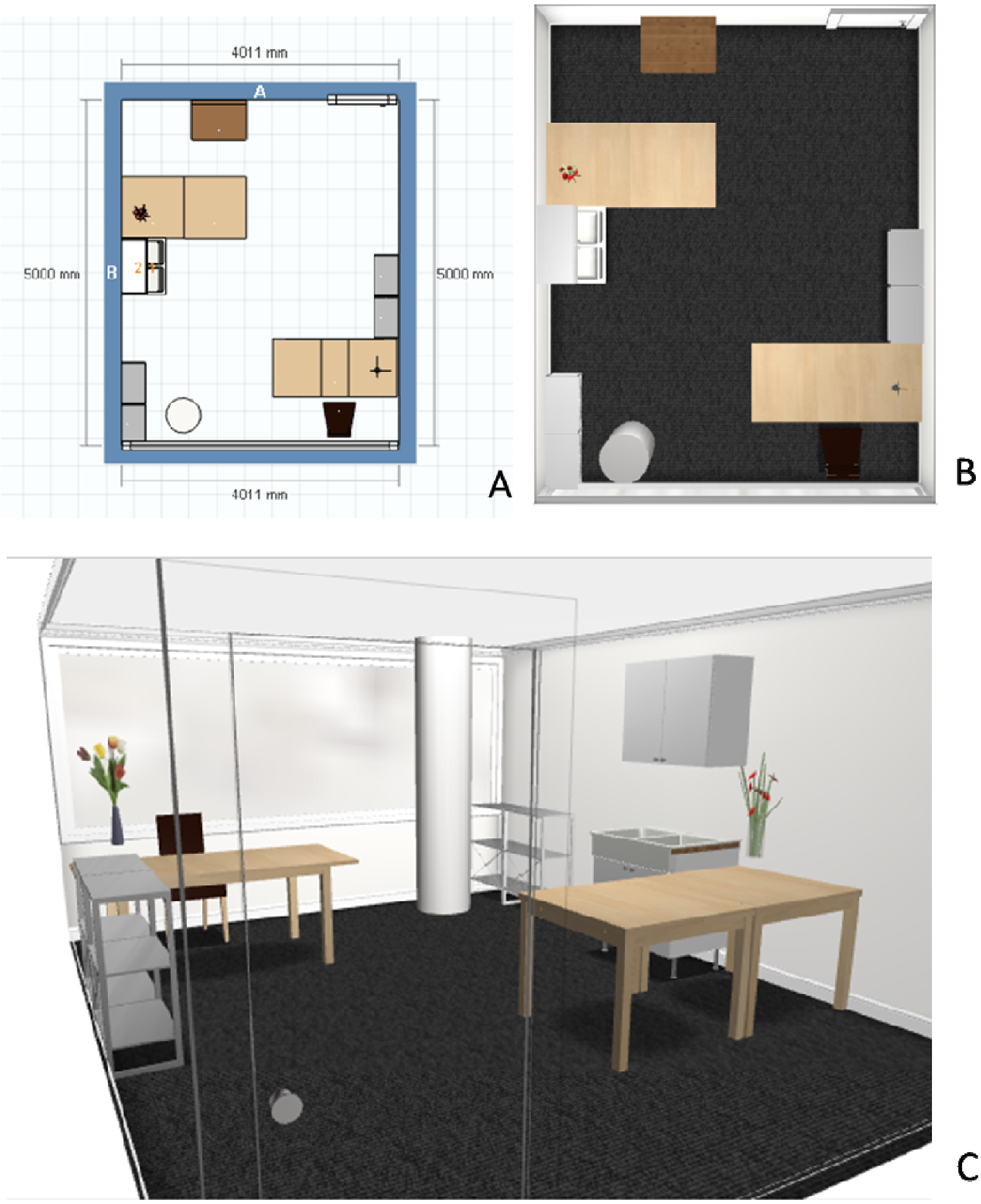


Figure 14. Illustration of the TUM CogWatch Laboratory used to test the four ADL scenarios. A-B) Birds eye view of the lab plan. C) 3D perspective depicting the kitchen set up (right side) and the experimenter work station (back left side)

6.2 ADL Scenarios

There are four different scenarios where the patient will execute the four ADL tasks preparing a cup of tea, preparing a jam on toast, teeth brushing, and dressing. These four tasks were chosen because they comprise basic care ADL, and they feature a number of actions steps directed at a distinct end goal. Furthermore, they have been studied by previous researchers, which provide a basis with which the results can be compared.

These tasks are divided into component subtasks using the hierarchical task analysis approach employed in human factors. The following sections summarize the manipulated objects, distracter objects (when required), and the location of the objects. Each section ends with a description of the instructions and experimental procedure.

6.2.1 The tea-making task

The first task is focused on the preparation of tea, with additional possibilities of adding sugar and milk that can be specified by the patient prior to movement initiation. As introduced in the first deliverable (D.1.1), the task consists of several sub-tasks and basic actions. The tea-making task is the chosen scenario for the first CogWatch prototype 1. Figure 15 shows a hierarchical tree based description of the task.

6.2.1.1 Assumptions

Pilot testing has indicated a substantial number of individual differences in tea preparation. For this reason, the task is to be performed in a controlled environment, with the following assumptions having been made:

- Water is inside a water jug
- The kettle is already plugged in
- Five tea bags are located in a specific box, placed on the table
- All objects are located within reaching distance on the table top, and the start location of the objects is standard across participants
- The tea bag must be removed from the cup in order for the task to be completed.

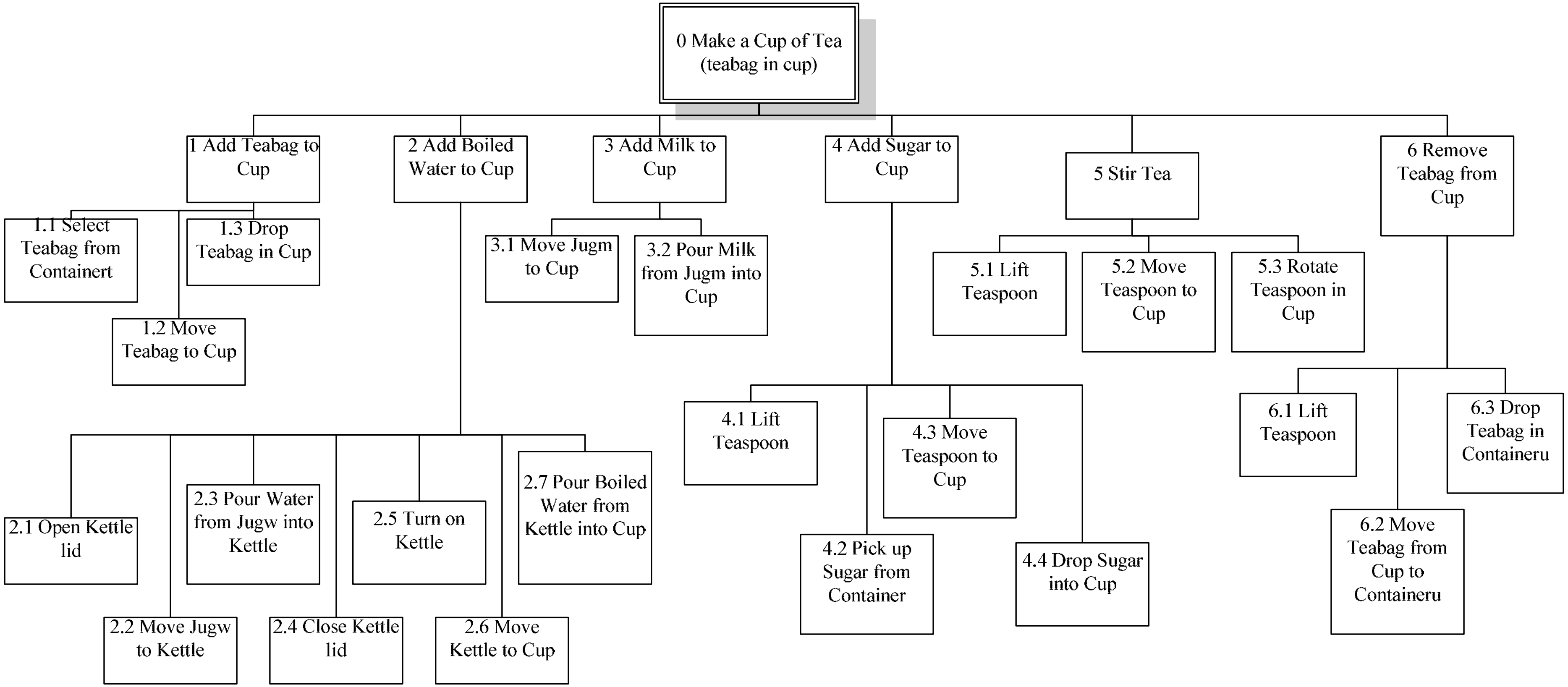


Figure 15. Task tree for tea preparation. Shown is the task tree for tea preparation without milk or sugar. For tasks that involve milk and/or sugar, the trees "add milk" and/or "add sugar" will be components of the task tree

6.2.1.2 Objects involved in the task

The kitchen environment consists of many objects that may or may not be necessary for the task at hand. For the tea-making task, we have classified the manipulated objects into those that are *necessary*, and those that are *distractors*.

The necessary objects for successful task completion are:

- A jug that contains water
- A kettle
- Two cups
- A sugar jar
- A milk jug
- Two teaspoons
- A tea bag container
- Five tea bags.

Distractor objects include:

- A knife
- A fork
- A dessert spoon
- A cereal bowl.

6.2.2 Toast making task

Meal preparation is a basic ADL that individuals perform at least once per day. Successful meal preparation performance is not only necessary for fundamental functioning, but also allows an individual to be more independent in their environment. The meal preparation ADL scenario chosen for further testing is a toast making task (see D1.1). The task description in tree form is shown in Figure 16.

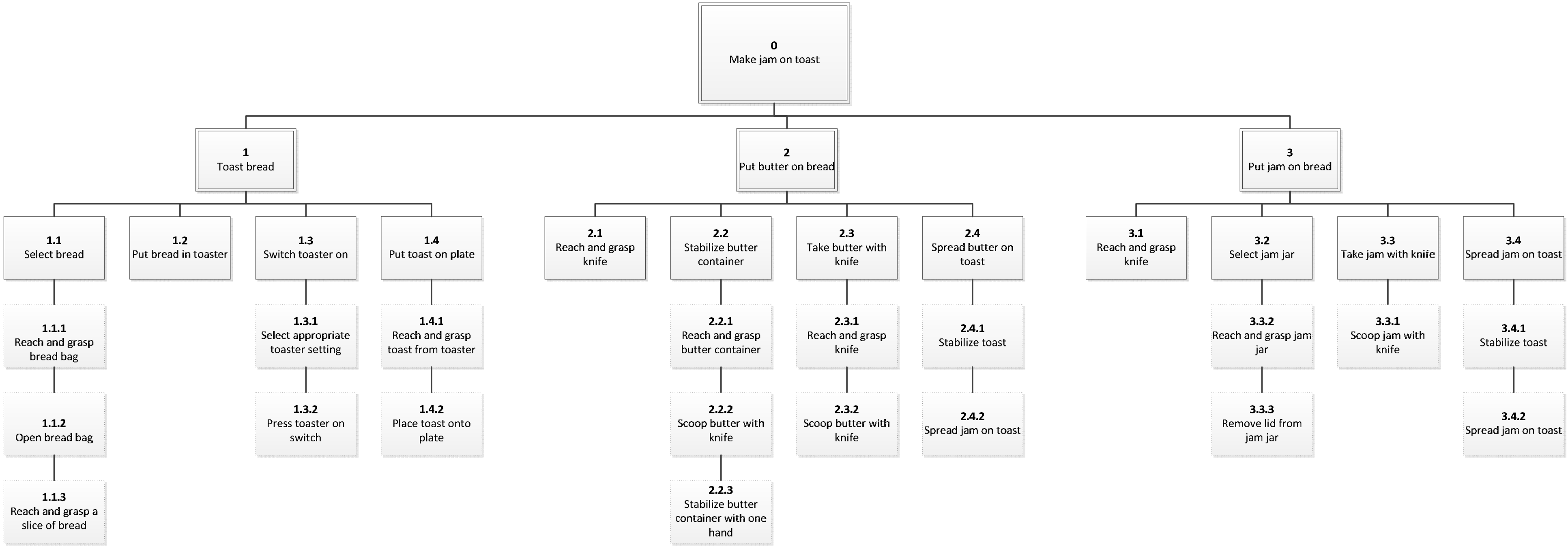


Figure 16. Task tree for toast preparation

6.2.2.1 Assumptions

- Prior to the experiment, participants will indicate whether they would use white or whole wheat bread when eating jam. This bread preference will be used in the actual experiment.
- If participants do not eat jam on toast, or are unfamiliar with the task, the spread ingredient can be substituted to another. For example, Nutella is a common spread ingredient in Germany, whereas Marmite is a common spread ingredient in the U.K., New Zealand and Australia.
- For participants with right hand hemiparesis, the toaster will be located on the right side of the set up. The set up will be reversed for those who do not have hemiparesis, or have left brain damage.
- Participants are informed that the toaster is set to a medium toast level, but that they are allowed to change the toaster settings if they prefer.
- The jam lid will be closed at the start of each trial.
- The bread will be in a closed bread bag at the start of each trial.

6.2.2.2 Objects involved in the task

The necessary objects for successful task completion are:

- A toaster
- A loaf of bread
- A jar of jam
- A plate
- A knife
- A spoon.

Distractor objects include

- A cup
- A sugar jar
- A milk jug.

6.2.3 Brushing teeth

One of the aims of the health practitioner and therapist is to improve the ability of apraxic individuals to successfully perform tasks of personal hygiene. ADLs in this category include bathing, hair brushing, teeth brushing and shaving. Due to practical and safety issues, we have chosen a teeth brushing task to assess ADL personal hygiene and grooming. The task description in tree form is shown in Figure 17.

6.2.3.1 Assumptions

- Various tooth paste flavours (fresh mint, pepper mint, cinnamon) and sensitivity relief tooth pastes will be provided. Participant will select the tooth paste they prefer prior to the experiment
- Tooth paste will have a screw cap lid
- Participants will have completed a survey indicating their teeth brushing (step 4) regiment
- The hand towel will be located on a railing next to the wash basin
- In the situation where participants wear dentures, participants will be asked to perform a denture care routine. As such, appropriate denture products will be available to the participant

6.2.3.2 Objects involved in the task

The necessary objects for successful task completion are:

- A toothbrush
- Toothpaste
- A sink and faucet
- A water glass
- A mirror
- A facecloth/hand towel

Distractor objects include:

- A comb
- A bar of soap
- A box of tissue (Kleenex)
- A hairbrush

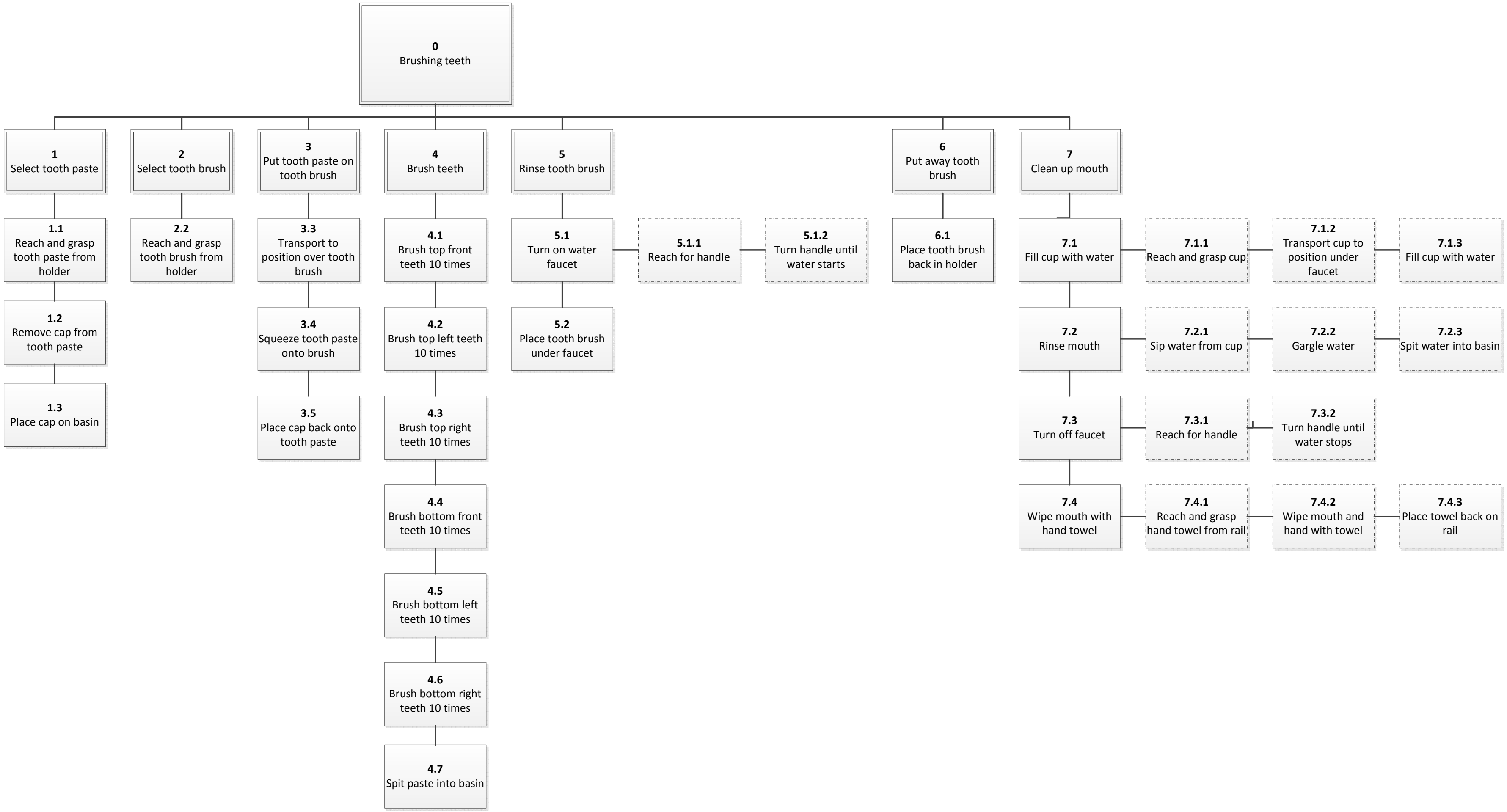

0
Brushing teeth
1
Select tooth paste
1.1
Reach and grasp tooth paste from holder
1.2
Remove cap from tooth paste
1.3
Place cap on basin
2
Select tooth brush
2.2
Reach and grasp tooth brush from holder
3
Put tooth paste on tooth brush
3.3
Transport to position over tooth brush
3.4
Squeeze tooth paste onto brush
3.5
Place cap back onto tooth paste
4
Brush teeth
4.1
Brush top front teeth 10 times
4.2
Brush top left teeth 10 times
4.3
Brush top right teeth 10 times
4.4
Brush bottom front teeth 10 times
4.5
Brush bottom left teeth 10 times
4.6
Brush bottom right teeth 10 times
4.7
Spit paste into basin
5
Rinse tooth brush
5.1
Turn on water faucet
5.1.1
Reach for handle
5.1.2
Turn handle until water starts
5.2
Place tooth brush under faucet
6
Put away tooth brush
6.1
Place tooth brush back in holder
7
Clean up mouth
7.1
Fill cup with water
7.1.1
Reach and grasp cup
7.1.2
Transport cup to position under faucet
7.1.3
Fill cup with water
7.2
Rinse mouth
7.2.1
Sip water from cup
7.2.2
Gargle water
7.2.3
Spit water into basin
7.3
Turn off faucet
7.3.1
Reach for handle
7.3.2
Turn handle until water stops
7.4
Wipe mouth with hand towel
7.4.1
Reach and grasp hand towel from rail
7.4.2
Wipe mouth and hand with towel
7.4.3
Place towel back on rail

Figure 17. Task tree for teeth brushing

6.2.4 Dressing

Dressing is one of the more complex skills which requires many sequential movements and involves both gross and fine motor control, eye-hand coordination, and balance. It is apparent that even if the task involves putting on a loose-fitting pullover sweat shirt, there are a great number of sequential actions that are required. Part of the CogWatch system consists of wearable devices (e.g., textiles, motion trackers) that acquire multi-parametric behavioural (e.g. grip force, hand configuration, position and movement, body posture, position and movement) and physiological (e.g., heart rate, blood pressure) data. As such, one of the ADL scenarios will involve a simple dressing task that utilizes the technology developed by RGB. The task description in tree form is shown in Figure 18.

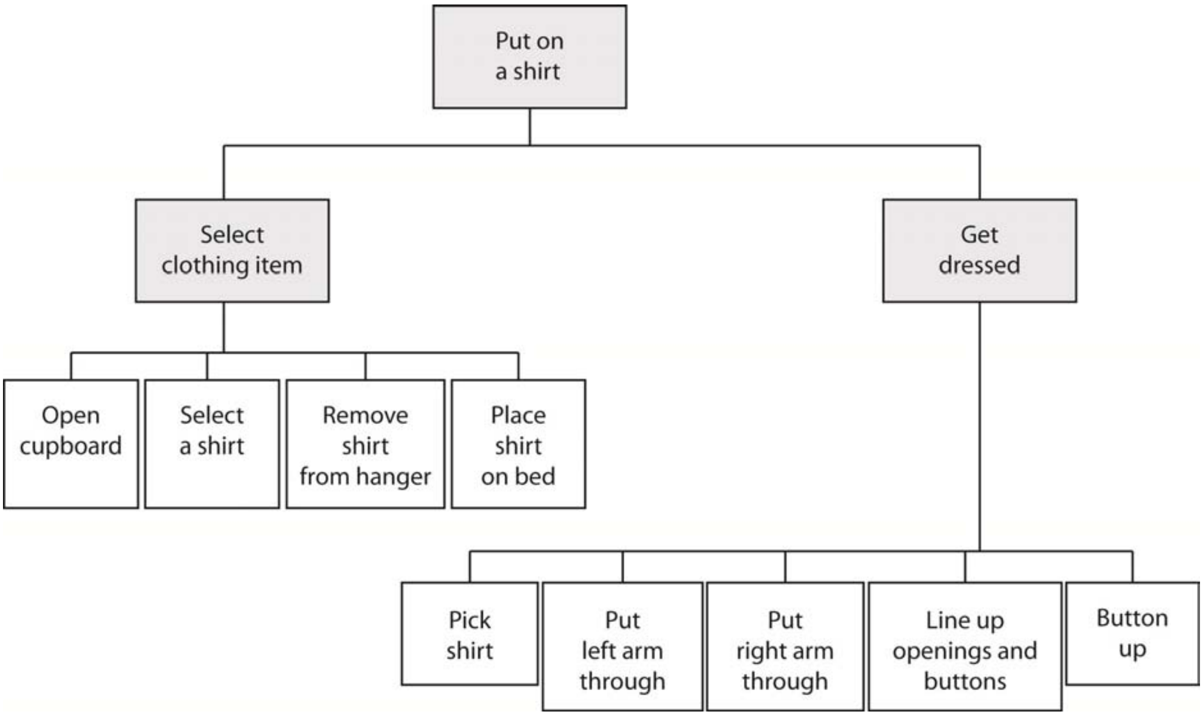


Figure 18. Task tree for dressing. Shown is an example of putting on a Velcro shirt

6.2.4.1 Assumptions

Due to compromised fine motor skills, the clothes chosen for the dressing task should be designed with ease of dressing in mind. Clothing choices should consider patients with lowered mobility, hemiparesis, and wheelchair dependent individuals who have difficulty or are unable to weight-bear. As such, the following assumptions have been considered:

- Velcro brand fasteners or easy touch closures will be used instead of buttons or zippers
- Elasticized waistbands will be used, as these will make it easy to pull on garments when challenged by hemiparesis or lowered hand dexterity

- Adaptive clothing should utilize front or side opening styles to facilitate assisted dressing

6.2.4.2 Objects involved in the task

The necessary objects for successful task completion include:

- A wardrobe
- Clothes on hangers
- Empty clothes hangers

Distractor objects include:

- A tie
- A pair of pants
- A pair of shoes
- An iron

6.3 Data collection

The data collection procedures are identical to those used in the Initial Phase: Clinical Screening experiments. Specifically, the experiments will be recorded by several video cameras in order to reconstruct the surface of the human as well as the joint angles. These measures allow for detailed, fine-grained analysis of human motions and for a comparison with previous executions of similar actions. These descriptions also provide temporal and spatial information about task performance (e.g., different action sequences or wrong objects used for a given action). The sequencing of actions will be used to develop the HTA models and BLN models used in prototype 1 and 2, respectively.

ADL performance will be segmented into subtasks (or action segments), and the movement time for each AS will be evaluated. This analysis might prove useful in classifying the degree of ADL performance impairment, as well as improvements over time.

In addition, future experiments will use motion analyses to examine arm and object movements throughout each trial. Such measurements provide the basis for feeding back information about correct or incorrect use of a tool to the acting patient. The information can also be used to detect gradual deviations from prototypical tool use and provide instructions in order to achieve more adequate movements. Kinematic measurements will also be used to further develop and evaluate KARA in a number of ADL scenarios.

7. CONCLUSIONS

This deliverable has outlined the technologies exploited to develop the CogWatch algorithms for action recognition and prediction. General information about each action recognition technology has been provided, and progress on the development and evaluation of each is detailed.

The report is presented in four main sections, separated into the different exploited technologies. Section two begins with a review of Kinematic-based action recognition, and details previous research that has utilized kinematic technology in the context of performance quality in ADL. Section two concludes by presenting primarily data on a Kinematic-based Action Recognition Algorithm (KARA) developed by TUM, which will be able to predict action and errors from the motion capture data. This technology will be implements into CogWatch prototype 2.

Section 3 discusses the use of the Microsoft Kinect™ sensor as a low-cost, easy-to-implement action recognition system. When evaluated against a commercially available motion capture system (Zebris), the Kinect™ performed quite well, with the results indicating a moderate to strong correlation between signals in the control participant, and moderate correlations between signals in the apraxia patient. Furthermore, although the sagittal axes MSE values differed between the control participant and apraxia patient, the transverse and coronal axes MSE values were similar for both participants. In summary, these results suggest that the Kinect™ is a viable addition to the CogWatch system as a motion capture technology. Future work will evaluate the ability of the Kinect™ to be a useful motion capture system for KARA.

Section 4 outlines the data collection, analysis, and results obtained from sensors located on the base of a tea kettle (CIC). This data has been modelled as a HMM, and suggests that the variability and low recognition accuracy issues apparent when modelling the raw CIC data can be resolved by simple thresholding of the FSR outputs. The EECE group at UoB will continue to ameliorate these issues, as the activity recognition in prototype 1 will be based primarily on the CIC data.

Section 5 provided information regarding video-based experiments in apraxia patients and healthy controls during a two tea-making task. The data from 14 Apraxia patients has been analyzed, and we report results on error types, error frequencies, and action sequencing in ADL performance. This data has been modelled using Bayesian Logic Networks (see D3.1 report on action recognition techniques), and will be integrated into the task models used in prototype 2.

In section 5 we provide an outlook for data collection and analysis in the upcoming 12 months, and introduce the tasks that will be used to evaluate prototype 2. The results of this work will be integrated into the Hierarchical Tree Analysis (HTA) and BLN modelling techniques used in the algorithms for the task models.

REFERENCES

Alankus, G., Proffitt, R., Kelleher, C., & Engsberg, J. (2010). Stroke therapy through motion-based games: A case study. *Proc. of ACM ASSETS'10*.

Bickerton, W. L., Samson, D., Williamson, J., & Humphreys, G. W. (2011). Separating forms of neglect using the Apples Test: validation and functional prediction in chronic and acute stroke. *Neuropsychology, 25,* 567-580.

Buxbaum, L.J., Schwartz, M.F., & Montgomery, M.W. (1998). Ideational apraxia and naturalistic action. *Cognitive Neuropsychology, 15,* 617-643.

Cappozzo, A., Della Croce, U., Leardini, A., & Chiari, L. (2004). Human movement analysis using stereophotogrammetry Part 1: theoretical background. *Gait and Posture, 21,* 186-196.

Chang, Y.J., Chou, L.D., Wang, T.Y., & Chen, S.F. (2012). A Kinect-based Vocational Task Prompting System for Individuals with Cognitive Impairments. *Personal and Ubiquitous Computing*, DOI: 10.1007/s00779-011-0498-6.

Chang, Y.J., Chen, S.F., & Huang, J.D. (2011). A Kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities, *Research in Developmental Disabilities, 32*, 2566-2570.

Chiari, L., Della Croce, U., Leardini, A., & Cappozz, A. (2004). Human movement analysis using stereophotogrammetry Part 2: Instrumental errors. *Gait and Posture, 21*, 197-211.

Ferrigno, G., & Pedotti, A. (1985). Elite: A digital dedicated hardware system for movement analysis via real-time TV signal processing. *IEEE Transactions on Biomedical Engineering, 32*(11), 943-950.

Freedman B., Shpunt A., Machline M., & Arieli Y. (2010). Depth Mapping Using Projected Patterns. U.S. Patent 2010/0118123.

Hermsdörfer, J., Hentze, S., & Goldenberg, G. (2006) Spatial and kinematic features of apraxic movement depend on the mode of execution. *Neuropsychologia, 44,* 1642-1652.

Ladin, Z. (1995). Three-dimensional instrumentation. In P. Allard, I. A. F. Stokes & J. P. Blanchi (Eds.), *Three-dimensional analysis of human movement* (pp. 3-17). Champaigne, IL: Human Kinetics.

McCrea, P.H., Eng, J.J., & Hodgson, A.J. (2002). Biomechanics of reaching: clinical implications for individuals with acquired brain injury. *Disability and Rehabilitation, 24*(10), 534-541.

Moores, E., Laiti., L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, *2*, 182-189.

Newell, K.M. & van Emmerik, R.E.A. (1989). The acquisition of coordination: Preliminary analysis of learning to write, *Human Movement Science, 8*, 17-32.

Overhoff, H. M., Lazovic, D., Liebing, M., & Macher, C. (2001). Total knee arthroplasty: coordinate system definition and planning based on 3-D ultrasound image volumes. *Proc. 15th Int. Congr. Exhibit. Computer Assisted Radiology and Surgery (CARS), 1230*, 292-299.

Randerath,J., Li, Y., Goldenberg, G., & Hermsdörfer, J. (2009). Grasping tools: Effects of task and apraxia. *Neuropsychologia, 47*, 497-505.

Rosenbaum, D.A., Marchak, F., Barnes, H.J., Vaughan, J., Slotta, J.D., & Jorgensen, M.J. (1990). *Constraints for action selection: overhand versus underhand grips*. In: Jeannerod M (ed) Attention and performance XIII. Motor representation and control. Lawrence Erlbaum, Hillsdale, pp 211–265.

Rothi, L.J.G., & Heilman, K.M. (1997). *Apraxia: The Neuropsychology of Action*. East Sussex (UK): Psychology Press.

Schwartz, M. F., Buxbaum, L. J., Montgomery, M. W., Fitzpatrick-DeSalme, E., Hart, T., Ferraro, M., Lee, S. S., & Branch-Coslett, H. (1998). Naturalistic action production following right hemisphere stroke. *Neuropsychologia, 37*(1), 51-66.

Shumway-Cook A., Woollacott M. (2001) *Motor Control: Theory and Practical Applications*, Second Edition. Baltimore, MD: Lippincott Williams & Wilkins.

Tenorth, M. (2011). Knowledge processing for automous robots. Unpublished doctoral dissertation, Technische Universität München, München, Deutschland.

Thelen, E. (1995). Motor development: A new synthesis. *American Psychologist, 50*(2), 79-95.

Vereijken, B., van Emmerik, R.E.A., Whiting, H.T.A., & Newell, K.M. (1992). Free(z)ing degrees of freedom in skill acquisition", *Journal of Motor Behavior, 24*, 133-142.

Whittle, M. W., (2002). Gait analysis: an introduction. 3rd edition. Oxford , Butterworth-Heinemann.

Young, S., Evermann, G., Gales, M.J.F., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., & Woodland, P. (2006). The HTK book, version 3.4, Cambridge University Engineering Department.