Automatic Assessment of Cognitive Impairment Through Electronic Observation of Object Usage

Mark R. Hodges[1], Ned L. Kirsch[2], Mark W. Newman[3], and
Martha E. Pollack[1,3]

[1] Computer Science and Engineering
[2] University of Michigan Medical School
[3] School of Information
University of Michigan, Ann Arbor, MI, USA
{hodgesm, nlkirsch, mwnewman, pollackm}@umich.edu

Abstract. Indications of cognitive impairments such as dementia and traumatic brain injury (TBI) are often subtle and may be frequently missed by primary care physicians. We describe an experiment where we unobtrusively collected sensor data as individuals with TBI performed a routine daily task (making coffee). We computed a series of four features of the sensor data that were increasingly representative of the task, and that we hypothesized might correlate with severity of cognitive impairment. Our main result is a significant correlation between the most representational feature and an apparent indicator of general neuropsychological integrity, namely, the first principal component of a standard suite of neuropsychological assessments. We also found suggestive but preliminary evidence of correlations between the computed features and a number of the individual tests in the assessment suite; this evidence can be used as the basis of larger-scale studies to validate significance.

1 Introduction

Cognitive impairments such as dementia and traumatic brain injury (TBI) can be difficult to detect and assess—one study showed that up to 75% of cases of dementia or probable dementia go undiagnosed by primary care physicians [1]. Additionally, cognitive ability may vary from day to day and, since therapists cannot observe patients on a daily basis, they are often forced to rely on questioning patients about their activities and to "play detective" using the answers given [2]. This paper describes work to automatically assess cognitive impairments caused by traumatic brain injury by using wireless sensors to observe individuals performing an everyday task, and then extracting features that are correlated to important neuropsychological assessments. The use of simple wireless sensors can potentially enable unobtrusive assessment on a daily basis in a naturalistic setting.

We asked individuals with TBI to make a pot of coffee and electronically observed them by placing RFID tags on relevant objects and having the subjects wear a bracelet with an RFID reader. We chose coffee making as a common

The original publication is available at `www.springerlink.com`

functional task that serves as a proxy for everyday activity performance. We analyzed the collected sensor data and compared it to the subjects' scores on a standard suite of neuropsychological assessments. Our analysis considered a series of four features computed from the data, which were increasingly representative of the task. We found a significant correlation between the most representational feature—edit distance from a correct task plan—and the first principal component of the assessment suite, which appears to serve as an indicator of general neuropsychological integrity. Because of the large number of tests in the neuropsychological suite, we were unable to collect sufficient data to demonstrate statistical significance between our computed features and individual tests, but we did find suggestive yet preliminary evidence of correlations, which can be used to structure larger scale investigations.

2 Motivation

The study we report on in this paper involved subjects who were being treated for traumatic brain injury (TBI). TBI is not uncommon: approximately 0.46% of Americans are hospitalized for brain injury each year and individuals aged 15-24 are far more likely than any other age group with over 0.9% hospitalized each year for brain injury [3]. Regrettably, TBI is frequently seen in wounded veterans returning from the Iraq War. Improved body armor has helped soldiers survive explosions that they might not have survived before, but many soldiers are suffering brain damage as a result of the blasts. The increase in Traumatic Brain Injury has been so dramatic that it has been called the "signature wound" of the Iraq War [4]—in one study of servicemembers arriving at Walter Reed Army Medical Center with injuries caused by explosions, 59% of the soldiers were found to have TBI and 56% of those were considered moderate or severe [5]. The existence and severity of TBI can be difficult to assess, in part because it cannot always be detected with imaging tests [6].

While our study was restricted to subjects with TBI, we anticipate that the approach we are using can generalize to other causes of cognitive impairment, notably including dementia. This is important, because the world's population is aging. In 2000, 12.4% of the U.S. population was aged 65 and older, and it is predicted to increase to 19.6% by 2030 and 20.6% by 2050. The oldest subgroup, that of individuals aged 80 and older, is expected to rise even more dramatically, more than doubling from 3.3% of the population in 2000 to 5.4% in 2030 and 8.0% in 2050 [7]. Trends worldwide are similar [8]. Dementia becomes much more common with age, affecting fewer than 1% of individuals in North America aged 60-64, but nearly 13% aged 80-84 and more than 30% of those over age 85 [9]. Without scientific advances to lower the incidence rates or the progression of Alzheimer's—the most common form of dementia—it is estimated that between 7.98 and 12.95 million people in the United States will have Alzheimer's Disease in 2050, four times the number that did in 2002 [10]. While our work overall is motivated by the challenges in assessing a wide range of cognitive impairments, it is important to keep in mind that we have so far only conducted tests using patients with traumatic brain injury.

By using wireless sensors placed in a home environment, passive and ongoing observation and re-evaluation may be possible without disruption of the individual's life or schedule. This would allow the observation of subjects over an extended period of time to provide information about both short-term and long-term changes in impairment. Short-term changes, for example improvement caused by successful medication or treatments, or sudden degradation resulting from a side-effect of medication, could be quickly detected and acted upon. Simplifying long-term observation means subtle changes could be more easily detected and that day-to-day variation could be distinguished from long-term changes.

3 Background

3.1 Automated Detection of Cognitive Impairments

Researchers at the Oregon Health & Science University are developing several techniques for the automatic detection of cognitive impairments, including automatically observing users play a modified version of the game FreeCell. One study focuses on mouse movement during the game [11] while others focus on the subjects' performance over time, comparing it to the performance of an automated solver. Using the results, it was possible to differentiate the three mildly cognitively impaired subjects from the six others [12]. Work with several other computer games, specially created to perform assessments of cognitive impairments is underway with some promising early results [13]. They have also studied automatically monitoring mobility because slowed mobility may be a predictor of future cognitive decline. The time to answer a phone call was used to measure mobility [14], as were passive infrared detectors and several models to infer the mobility of subjects more directly as they move about a residence [15].

Research by Glascock and Kutzik used various sensors to observe activities of daily living (ADLs) in a subject's home. The output from the sensors, however, was analyzed manually [16]. Other research by Barger, Brown, and Alwan observed subjects using motion detectors to detect behavioral patterns. Although basic analysis of behavioral patterns was performed, this analysis was only loosely tied to performed activities [17]. Finally, Hoey, et al use an estimate of the subject's functionality in their system that assists a user with dementia during handwashing. This estimate is updated over time as the user completes the handwashing task. While detection of cognitive impairments was not the focus of that research and accuracy results are not given for their system, this approach could potentially be expanded to those goals [18].

Similar ideas have been used to address other conditions such as automatically observing autism using accelerometers placed on the wrists and chest [19]. Preliminary studies have also examined using toys with sensors to support assessment of a child's development [20].

3.2 Activity Recognition

Activity recognition is an active field of research that uses various sensors to monitor individuals, applying interpretation algorithms to recognize the activities they are performing. While we do not perform activity recognition in the current study—we instead assess the performance of known activities—there are clearly connections between the two tasks. Different applications in activity recognition focus on a wide range of activities. Recognizing whether a subject is moving in ways such as jumping or walking [21], identifying a user's common destinations in a city [22], and differentiating whether a user is taking medication, making cereal, or eating cereal [23] are all examples of tasks distinguished by activity recognition systems.

Several types of sensors can be used to observe interactions with objects, such as RFID readers, motion detectors and accelerometers designed to detect object-use interactions, as well as electric current and water flow detectors [24]. In each case, the sensors measure approximations of object usage: with RFID readers, for example, proximity of a hand and an object is used as a proxy for object interaction; with accelerometers, movement of the object serves as a proxy. Several techniques have been used to analyze this data, including probabilistic methods and decision trees [24, 25].

There are also many approaches to activity recognition that are not based on the analysis of interactions with objects, including the use of GPS [22], small wearable sensing platforms that have several sensors including accelerometers [26], and data-rich sensors such as video cameras or microphones [21].

4 Methodology

Our study involved subjects performing an everyday activity that could be monitored using wireless sensors. We hypothesized that patterns of errors made in the performance of such activities are associated with the severity and type of a patient's cognitive impairment and further, that we could use wireless sensors to accurately detect those errors. We were concerned both with predicting overall neuropsychological integrity, and with identifying more specific neuropsychological profiles, such as isolated difficulties with memory, attention, or executive reasoning.

4.1 Selection of a Task

We chose to observe subjects preparing a pot of coffee using a drip coffee maker common in North America. This task was selected because it is performed by many people on a regular basis, so individuals could be assessed as they perform their daily routine. Variation in the ways that individuals make coffee is somewhat limited so patterns can be analyzed more easily, but there is still opportunity for mistakes or inefficiencies when the subjects perform the task. The same task was used successfully in our previous study of behavioral patterns [27]. In the current study, subjects did not fully prepare a cup of coffee but only started the coffee pot brewing so that they did not handle hot liquids, discussed more in section 4.3. While the particular task we used, coffee making, was selected for reasons just given, in the end it is simply a common functional task that serves as a proxy for everyday activity performance. We expect that studying a range of similar tasks would be necessary before our methodology would be put into place in individuals' households, so that an individual who was, say, a tea- rather than coffee-drinker could still benefit from the approach.

4.2 Selection of Technology

Radio Frequency Identification (RFID) technology has been used successfully to study object-use interactions in several activity recognition projects (as discussed in section 3.2). RFID uses tags placed throughout an environment, along with readers that detect nearby tags. An important advantage of this technology is that one can use passive RFID tags which require no power source. Other advantages to using RFID are that tags are available in a small form factor (approximately the size of a postage stamp) and that they are inexpensive (less than $0.20).

Following earlier work in automated activity recognition, we had each subject wear an RFID reader on the wrist, and we thereby recognized the objects with which the subject was interacting. Specifically, our subjects wore the Intel iBracelet, with a range of about 10cm, that is depicted in Figure 1 [28].

Fig. 1. The Intel iBracelet RFID Reader

This design is beneficial for privacy concerns as well: a subject may take off the bracelet to avoid observation. Likewise, other individuals will not confuse the system as long as they don't wear the bracelet. Disadvantages include the fact that the bracelet is somewhat bulky and has a relatively short battery life (about three hours). If our methodology were used in the home environment, however, the subject would only need to wear the bracelet when making coffee, and could remove it for the rest of the day.

4.3 Experimental Setup

Sixteen subjects with traumatic brain injuries and full neuropsychological evaluations were recruited to participate in the study. For each trial, the subject was to start a pot of coffee—putting in water, a filter and ground coffee and turning the coffee maker's power on. Subjects were each asked to perform five trials on five different days (13 completed at least three trials and 10 completed all five trials).

The subjects performed the trials in a kitchen at the medical center where they were receiving care for their cognitive impairment. The coffee pot and all supplies were placed on a counter in the kitchen, next to a sink for water. Subjects were asked if they knew how to make coffee and given basic instructions if they did not. No physical demonstrations were given. If subjects asked how much material to put in, they were told to use enough for six cups of coffee (about half the capacity of the coffee pot).

The material that was set out included the coffee pot and carafe, a canister of ground coffee, a bag of filters, a mug and a spoon. Twelve tags were used: four on the coffee pot, four on the canister of ground coffee, and one each on the other objects. Multiple tags are needed for some objects to reliably detect interaction due to the range of the iBracelet (the shorter range is desirable to avoid a higher rate of false positives).

The experimental setup was influenced by the fact that the subjects had cognitive impairments and were performing the task within the clinic. We placed the supplies on the counter, rather than away in cabinets, to make the task easier for the subjects to complete in order to avoid causing frustration by having subjects searching in an unfamiliar kitchen if they forgot where a supply was located. This should not be necessary when observing subjects in a home environment.[4]

5 Automatic Assessments

The sensor data collected in each trial consists of a series of time-stamped interactions with RFID tags, a sample of which is shown in Figure 2. From the collected sensor data, we computed four features that we hypothesized might correlate with the subjects' cognitive impairments. The features are increasingly representative of the task, ranging from very simple—how long does it take the subject to complete the trial—to much more detailed—how "far off" is the subject's behavior from a correct instance of task performance.

5.1 Trial Duration

Our first hypothesis was that a more severely impaired individual might take longer to prepare the pot of coffee than a less impaired individual, as a result of confusion, mistakes, or simply performing steps more slowly. Therefore, the first feature we computed is the duration of the trial: how long it takes the subject to complete the task.

Given a trial with n detected interactions, we define this feature using the following formula: $TrialDuration(t) = EndTime_n - StartTime_1$ where $StartTime_i$

[4] Out of an abundance of caution and on the advice of the clinic staff, we also did not have subjects pour out a cup of coffee once the pot had brewed. This was to ensure that the individuals would not be handling hot liquids and decrease the potential of injuring patients at the medical facility. This should not be a barrier to using a similar system in-home since we expect that many cognitively impaired individuals regularly make coffee and, anecdotally, several participants in our study noted that they regularly made coffee at home (the percentage of participants who make coffee regularly is unknown since that was not part of the formal interview).

Time Stamp	Tag Detected
1200503935	Carafe
1200503935	Filters
1200503936	Filters
1200503939	Ground Coffee 3
1200503956	Ground Coffee 1
1200503989	Coffee Maker Lid 3

Fig. 2. Stream of time-stamped interactions from a portion of a trial. When multiple tags are placed on one object, a number is given indicating which tag has been detected.

and $EndTime_i$ indicate the start and end times of the i^{th} action in the temporal sequence of trial t. That is, the feature is simply measured as the time between the first interaction that is detected and the last.

5.2 Action Gaps

Note that the trial duration feature is extremely simple and has very limited representational power: it would not distinguish between two people who are behaving in very different ways, provided only that the total amount of time for each trial was the same. We next moved to a somewhat more representational feature, which is based on the hypothesis that more severely impaired individuals might have more periods during which they were not interacting with any objects, on the assumption that during those periods they are considering what step to take next. The second feature measures these periods of inactivity during the trial which we call Action Gaps. We define the number of Action Gaps with length g of trial t:

$$ActionGaps_g(t) = \sum_{i=1}^{n-1} \begin{cases} 1, & \text{if } StartTime_{i+1} - EndTime_i > g \\ 0, & \text{otherwise} \end{cases}$$

We examine the number of brief action gaps using $g = 3$ seconds and the number of longer action gaps using $g = 10$ seconds.

5.3 Object Misuse

We next moved to a feature that accounts for the specific objects used in task performance. One way of determining whether someone is being effective in carrying out a task is to examine the number of times he or she interacts with each object used in the task. We thus hypothesized that failure to interact with a required object—e.g. to "touch" the coffee filters—indicates a problem, as does an excessive number of interactions. For the simple task of making coffee, we manually determined a reasonable range of interactions with each object, shown in Table 1. The filters, for example, may be used once or twice—once to open the bag of filters and perhaps again if the user closes the bag in a separate interaction (remember that the tag is on the bag of filters, not individual filters themselves).

Note that we do not state a maximum number of accepted interactions with the Ground Coffee or the Mug or Carafe (to get water) because these are difficult to define—unlike closing the lid which is one distinct interaction, putting ground coffee in the coffee pot may involve touching the ground coffee multiple times to get several scoops and filling the water may require using the mug multiple times to fill the coffee pot. The Spoon is not included in this feature because it was rarely detected—it would also be difficult to use since it is not required but, like the Grounds, may be used multiple times.

For each trial, we then counted the number of times the subject interacted with each object b ($touch_b$) and determined whether that number was outside the accepted range and, if so, how far outside the range it was.[5]

$$ObjectMisuse(t) = \sum_{b \in Objects} \begin{cases} 0, & min_b \leq touch_b \leq max_b \\ min_b - touch_b, & touch_b < min_b \\ touch_b - max_b, & touch_b > max_b \end{cases}$$

Object b	min_b	max_b
Lid	2	2
Ground Coffee	1	∞
Filters	1	2
Mug or Carafe (Getting Water)	1	∞
Power Switch	1	1

Table 1. Number of Accepted Interactions for each Object

5.4 Edit Distance

Our final approach to automatically measuring performance moves even further in the direction of matching the subject's performance to an explicit model of correct performance. With this approach, we begin with a representation of how to make coffee—a "plan" for the task. The plan we used in our analysis is a partial order over object interaction, depicted in Figure 3, with "Water" indicating using the carafe, mug, or both to get water from the sink and put it into the coffee maker. Note that the use of the partial order allows us to score as "correct" alternative task executions that are reasonable: we score as correct both executions in which water is added before the filter and ground coffee and

[5] We also investigated a few variations of the Object Misuse metric, to address the possibility that touching an object too many times could have a disproportionately large impact compared with touching too few times. These variations were approximately as successful as the basic metric here; because the variations and results did not appear to be interesting, they are not presented in this paper.

those in which those actions are reversed. However, we judge to be incorrect executions in which the power switch is turned on before the filters are used.

We then further constrain our plan for correct executions to those in which object interactions are not interleaved and using filters is followed directly with using ground coffee. These two criteria are added for the same reason: for a basic task like making coffee, we hypothesize that it is more likely that a mistake occurred than that the individual chose to interleave actions (like getting ground coffee, then water, and then ground coffee again). Using filters and ground coffee are kept together because we view them as really one general action: putting ground coffee in the coffee maker.[6]

Fig. 3. Partially Ordered Plan of Object Interaction for Making Coffee

Although we manually created the plan to represent making coffee, other research on activity recognition has addressed the question of automatically constructing plans for everyday activities by mining the web for descriptions of these activities [29]. Such an approach could be adopted to extend our work.

Once we have a plan that models correct task executions in terms of object interactions, we next have to define a measure of deviations from that plan. For that, we adopted the notion of edit distance, which is frequently used in the literature on natural language processing [30], but which has also been used in prior work on activity recognition [25]. More specifically, we make use of the Levenshtein distance which allows the insertion, deletion, or substitution of a character [31]. We compute the distance between the sequence of observed object interactions and each of the legal executions of the plan for the task.

Note that to compute the edit distance, we merge consecutive interactions with the same object (for example, multiple usages of the ground coffee are just shown once as long as no other objects are used in between). We then define $EditDistance(t)$:

$$EditDistance(t) = min_{e \in LegalExecutions}(Distance(t, e))$$

With our very simple plan, there are only two legal executions: the one in which placement of the filters and the ground coffee precedes the filling of the water canister, and the one in which these occur in the reverse order. Hence Edit Distance is easy to compute, involving determination of just two distances.

[6] The assumptions we have made in our model may be too constraining—perhaps many unimpaired individuals do interleave using filters and grounds with getting water, for example. This suggests a further elaboration, in which the plans are probabilistic—with the probabilities representing the plausibility of certain sequences being performed. This elaboration, however, is outside the scope of this project.

The edit distance is intended to provide a fairly fine-grained measure of the relationship between the "correct" task performance, at least as modeled in our plan, and the subject's actual performance.

6 Neuropsychological Assessment

Neuropsychological impairments are assessed with a battery of tests that sample a broad range of cognitive domains. Many of these tests assess general functioning, such as intellectual ability. Others are very specific, having been chosen because they are known to be associated with functioning that is mediated by a specific brain locus (e.g., left versus right hemisphere, anterior or lateral frontal lobe, specific sub-regions of the areas that mediate expressive or receptive language), or because they provide critical information about a cognitive domain that is central to performance of everyday activities (e.g., attentional shifting). The measures employed for this type of assessment are meticulously normed, often in the context of multiple samples, such that statements can often be made about a patient's performance relative to the population at large, to specific cohorts (e.g., those of same gender and similar age or education), or relative to specific clinical comparison groups (e.g., is the profile most consistent with a cerebro-vascular accident, dementia, or depression) [32–34]. The neuropsychological assessments we used are given in Table 4 in the appendix.

We obtained the results of neuropsychological tests from the patient records of our 16 subjects to use as ground truth. We then computed the correlations of our computed features with the individual neuropsychological assessments listed, using an individual's average value over the five trials for each computed feature and applying one-tailed non-parametric evaluation. In addition to the individual neuropsychological assessments, we applied principal component analysis (PCA) to the complete set of neuropsychological assessment data for the subjects in order to examine how well our computed features correlate with larger trends in the assessment data. PCA is a standard statistical technique that finds linearly independent components that explain as much variance in the data as possible. Each component is a linear combination of the assessments where the sum of the squares of the component coefficients is one. The first principal component is the linear combination that has the largest possible variance; the second principal component is the linear combination that has the largest possible variance and is uncorrelated with the first principal component; the third is uncorrelated with either of the first two components, and so on. To perform the PCA, some of the summary assessments in Table 4 were replaced with their component scores, a standard statistical practice.

The first principal component of the neuropsychological data we used accounts for 26.4% of the total variance in the data and the top five principal components together account for 72.0% of the total variance. Table 5 in the appendix shows the first five principal components and the variance explained by each component. Table 4 indicates which factors, if any, each assessment (or any subtest of that assessment) has a loading of 0.6 or higher.

After computing the principal components, the domain expert on our team (Kirsch) interpreted them. The first principal component includes a diverse set of measures of general intelligence. It appears to be a good proxy for general neuropsychological integrity, including measures of intellectual functioning, verbal and nonverbal reasoning, memory, and complex attention. The interpretation of the lower-order components is less clear, although the second could be seen as a measure of general motor integrity; the third as representing verbal memory and concept formation; the fourth, the ability to retain verbal information over time; and the fifth, strategy formation and modification.

7 Results

7.1 Assessing Neuropsychological Integrity

Recall that the main question we ask in this study is whether we can assess a patient's cognitive status by observing performance of an everyday activity using wireless sensor networks. Our main result is quite promising: we found a statistically significant correlation ($p < 0.01$) between the Edit Distance feature and the first principal component of the neuropsychological assessments, which, as just described, can serve as a proxy for overall generalized neuropsychological integrity. Importantly, we did not find such a correlation with any of the simpler features (Trial Duration, Action Gaps, or Object Misuse). The ability to predict neuropsychological integrity, at least within the scope of this experiment and in particular for the population of TBI patients involved, is a strong indication that it is possible to conduct the types of automatic assessments that motivate this work. Figure 4 shows the plot of Edit Distance and the first principal component, with the regression line.

Fig. 4. Plot and Regression Line of Edit Distance with the First Principal Component of the Neuropsychological Assessments

7.2 Assessing Other Metrics of Impairment

Although general neuropsychological integrity is a very important metric, it is also interesting to see how our features assess other metrics of cognitive impairments. The reason for doing this is based on domain practice—in addition to a concern with overall neuropsychological integrity, it is often important for a rehabilitation team to understand different aspects of a patient's impairment: does it involve memory impairment? Problems with focus of attention? Decreased motor coordination? There is a potential for these to be blurred in a single measure of overall integrity, important though that summary measure is. For instance, an individual whose impairment involves decreased motor coordination or processing speed may have unimpaired executive function, and thus still be able to follow a "plan" for making coffee successfully, but perform the task more slowly. An increased Total Duration might help tease out the types of cognitive difficulties facing this patient. To address this, we also look at the next four (the second through fifth) principal components as well as the twenty-nine individual assessments.

Additional statistical analysis such as a Bonferroni correction is required to state that a correlation between two variables exists with statistical significance. However, because of the large number of assessments and features, achieving statistical significance at this strict level would require the collection of data from a huge number of subjects—many more than were in the scope of this project. Nonetheless, our results on the individual tests in this section are important as exploratory data analysis and as a foundation for further research in the area. Although we recognize that correlations at the variable level are of questionable significance because of the number of analyses performed, we are nevertheless presenting these findings because the coherence of the relationships and the number of associations we found provide important direction for future research. Additionally, while only suggestive, the variable level correlations provide tentative guidance in regard to further refinement of "markers" that clinicians can use when attempting to make a determination of the mechanisms for a patient's failure. Analysis of mechanisms (e.g., decreased processing speed vs. executive functioning) may then lead, in turn, to choices on the clinician's part about interventions that are specifically targeted to the underlying impaired cognitive mechanism.

Beyond this, it is noteworthy that we saw more correlations than would have been expected by chance (22 actual correlations for the individual assessments versus 14.5 expected by chance), especially when looking at what would be strict p-values if the Bonferroni correction were not required ($p < .01$: 8 actual correlations versus 2.9 by chance). Moreover, many of the identified correlations "make sense" from a neuropsychological standpoint, in a manner similar to the example in the previous paragraph. The results from correlations with principal components will also be presented, although the number of correlations with the second through fifth principal components (2) is what was expected by chance.[7]

Edit Distance The Edit Distance feature achieved the best results with the individual evaluations as well, having a suggestive correlation with the fourth principal component as well as with 7 of the 29 (24%) neuropsychological assessments.[8] Recall that the fourth principal component appeared to represent the ability to retain verbal information over time. The correlations with individual evaluations are predominantly and compellingly with memory features; we spec-

[7] These numbers include the results from the additional five variations of Object Misuse noted in Section 5.3 although those results are not presented here.

[8] We identify a suggestive correlation whenever there would be a statistically significant correlation if the Bonferroni correction were not needed. Because it is necessary, these correlations are not significant but are still of interest for their value in guiding future studies.

ulate that they could also be said to measure the integrity of the left-cerebral hemisphere and the capacity to engage in sequential and logical thinking. Generally the assessments that have suggestive correlations with Edit Distance are also factors with a high loading in the principal components with which Edit Distance is correlated but the slightly weaker interpretation is likely due to the less efficient analysis of individual assessments. Table 2 summarizes the Edit Distance results and compares them to the other features. Table 3 shows more detail, giving the assessments with which Edit Distance had a suggestive correlation.

Feature	Correlations with Principal Components	# Suggestive Correlations with Individual Features
Edit Distance	1st ($p < 0.01$) Suggestive: 4th	7
Total Duration	-	6
Action Gaps ($\geq$ 3s)	-	5
Object Misuse	-	3

Table 2. Summary of Results from each Feature.

	Computed Feature			
Assessment	Edit Distance	Total Duration	Action Gaps $\geq$ 3s	Object Misuse
WAIS III Processing Speed		* #		
WMS-R Visual Reproduction II	*	*		*
CVLT II Total	*			
CVLT II Long Delay Free Recall	*	*		*
CVLT II Discriminability	*			*
Trails B			* #	
Animals	*	*	*	
WRAT 4 Reading			* #	
TPT Memory	*			
Finger Tapping - Dominant	*			
Finger Tapping - Non-Dominant		* #	* #	
GPB - Non-Dominant		* #	* #	

* indicates a suggestive correlation

notes additional coverage: a metric not also correlated with Edit Distance

Table 3. Suggestive Correlations between Neuropsychological Assessments and Computed Features (Assessments with no suggestive correlations are not shown)

Total Duration and Action Gaps Total Duration and Action Gaps also proved to be promising features. Though neither had suggestive correlations with any of the principal components, they did with a number of neuropsychological assessments. Total Duration had a suggestive correlation with 6 of the 29 (21%) neuropsychological assessments. Similarly, Action Gaps of 3 seconds or greater suggestively correlated with 5 (17%) of the neuropsychological assessments. These results are less coherent from a neuropsychological perspective than the Edit Distance results but the correlation between processing speed and Total Duration is very logical. And while the results are not as good as the Edit Distance results, they are still valuable: between the three features presented thus far, there are suggestive correlations with over 12 of the 29 neuropsychological tests (40%), including 5 that did not have suggestive correlations with Edit Distance. We also tested Action Gaps of 10 seconds or greater but this only had a suggestive correlation with one metric (GPB - Non-Dominant which also correlated with two other features); we hypothesize that the poor result for this feature is due to the low frequency of gaps that long.

Object Misuse The results from the Object Misuse feature were the least successful—as shown in Table 3 the feature had fewer suggestive correlations than Edit Distance, Total Duration or Action Gaps of 3 Seconds, and none with assessments that were not also correlated with Edit Distance.

8 Discussion and Conclusion

We have presented an approach to using RFID-based sensing of individuals as they perform a simple task, with the aim of assessing their level of cognitive impairment. We presented four features, with increasingly representational power, that can be computed from the collected sensor data, and evaluated them using the results of the subjects' performance on standard neuropsychological assessments as well as with the principal components of those assessments. The most knowledge-rich feature we computed, Edit Distance, had a statistically significant correlation with the meaningful first principal component, a measure of general neuropsychological integrity. We also presented the results of exploratory analysis of the correlations between the four types of features and the individual assessments; these results are helpful to guide future research into other metrics of impairment without the need for a massive amount of data collection.

There are many practical concerns for the in-home implementation of a system that could automatically assess impairments. Compliance with the system is important since the user must wear the bracelet and complete the task to be assessed; individuals at risk for or developing an impairment may be particularly forgetful about doing this. Other sensor modalities, such as accelerometers placed on the objects, motion detectors, or current or water-flow sensors might be considered which do not have this drawback. On the other hand, the privacy implications of observing individuals in a home environment are important to address and we feel these may be somewhat alleviated by using a system which can clearly be prevented from observing an individual's behavior (by taking the bracelet off).

A great deal of future work remains, including collecting additional data and performing further analysis to investigate the suggestive individual correlations identified in this study. Additionally, further study is needed to examine whether these or similar techniques can successfully differentiate impaired from unimpaired subjects. Observation of other kinds of impairments (particularly dementia) and longitudinal studies of individuals at risk for cognitive impairments are necessary to understand the ability of these techniques to detect the onset of impairment and potentially to develop new techniques to observe change in an individual's performance over time. There are a number of ways in which the scope of the research can be expanded, particularly applying these assessment techniques to other activities beyond coffee making and using them in a home environment. Lastly, studying other types of clinical assessments (such as speech and occupational therapy) as well as development of other computed features, particularly those that might correlate with different assessments from the computed features presented here, are also areas for future work.

Acknowledgments. We thank Intel Research Seattle for the iBracelet RFID reader used in this study. We also thank the staff of MedRehab at the University of Michigan, particularly Erin Spirl, Michelle Shenton and Pippin Gilbert, for their help with subject recruitment and data collection.

References

1. Holsinger, T., Deveau, J., Boustani, M., John W. Williams, J.: Does this patient have dementia? JAMA **297** (June 2007) 2391–2404
2. Wilson, D., Consolvo, S., Fishkin, K., Philipose, M.: In-home assessment of the activities of daily living of the elderly. Extended Abstracts of CHI 2005: Workshops - HCI Challenges in Health Assessment (April 2005) 2130
3. Sosin, D.M., Sniezek, J.E., Thurman, D.J.: Incidence of mild and moderate brain injury in the united states, 1991. Brain Injury **10** (1996) 47–54
4. Zoroya, G.: Scientists: Brain injuries from war worse than thought. USA Today (September 2007)
5. Okie, S.: Traumatic brain injury in the war zone. New England Journal of Medicine **352** (May 2005) 2043–2047
6. Bagley, L.J., McGowan, J.C., Grossman, R.I., Sinson, G., Kotapka, M., Lexa, F.J., Berlin, J.A., McIntosh, T.K.: Magnetization transfer imaging of traumatic brain injury. Journal of Magnetic Resonance Imaging **11** (January 2000) 1–8
7. United States Census Bureau: International data base. (July 2007)
8. United Nations Population Division: World population prospects. (July 2007)
9. Ferri, C.P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., Hall, K., Hasegawa, K., Hendrie, H., Huang, Y., Jorm, A., Mathers, C., Menezes, P.R., Rimmer, E., Scazufca, M., for Alzheimer's Disease International: Global prevalence of dementia: a delphi consensus study. The Lancet **366** (2006) 2112–2117

10. Sloane, P.D., Zimmerman, S., Suchindran, C., Reed, P., Wang, L., Boustani, M., Sudha, S.: The public health impact of alzheimer's disease, 2000-2050: Potential implication of treatment advances. Annual Review of Public Health **23** (2002) 213–231
11. Jimison, H.B., Pavel, M., McKanna, J.: Unobtrusive computer monitoring of sensory-motor function. Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference (September 2005) 5431–5434
12. Jimison, H.B., Pavel, M., McKanna, J., Pavel, J.: Unobtrusive monitoring of computer interactions to detect cognitive status in elders. IEEE Transactions on Information Technology in Biomedicine **8**(3) (2004) 248–252
13. Jimison, H.B., Pavel, M., Le, T.: Home-based cognitive monitoring using embedded measures of verbal fluency in a computer word game. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2008)
14. Pavel, M., Adami, A., Morris, M., Lundell, J., Hayes, T.L., Jimison, H., Kaye, J.A.: Mobility assessment using event-related responses. 1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare (2006) 71–74
15. Pavel, M., Hayes, T.L., Adami, A., Jimison, H., Kaye, J.: Unobtrusive assessment of mobility. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006 (2006) 6277–6280
16. Glascock, A., Kutzik, D.: Behavioral telemedicine: A new approach to the continuous nonintrusive monitoring of activities of daily living. Telemedicine Journal **6** (2000)
17. Barger, T.S., Brown, D.E., Alwan, M.: Health-status monitoring through analysis of behavioral patterns. IEEE Transactions on Systems, Man and Cybernetics, Part A **35** (January 2005) 22–27
18. Hoey, J., von Bertoldi, A., Poupart, P., Mihailidis, A.: Assisting persons with dementia during handwashing using a partially observable markov decision process. Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007) (2007)
19. Albinali, F., Goodwin, M., Intille, S.: Recognizing stereotypical motor movements in the laboratory and classroom: A case study with children on the autism spectrum. Ubicomp (2009)
20. Westeyn, T.L., Kientz, J.A., Starner, T.E., Abowd, G.D.: Designing toys with automatic play characterization for supporting the assessment of a childs development. Workshop on "Designing for Children with Special Needs" at the Seventh Conference on Interaction Design for Children (IDC) (2008)
21. Ben-Arie, J., Wang, Z., Pandit, P., Rajaram, S.: Human activity recognition using multidimensional indexing. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(8) (2002) 1091–1104
22. Liao, L., Fox, D., Kautz, H.: Location-based activity recognition. In Weiss, Y., Schölkopf, B., Platt, J., eds.: Advances in Neural Information Processing Systems 18. MIT Press, Cambridge, MA (2006) 787–794
23. Pentney, W., Popescu, A.M., Wang, S., Kautz, H.A., Philipose, M.: Sensor-based understanding of daily life via large-scale use of common sense. AAAI (2006)
24. Logan, B., Healey, J., Philipose, M., Tapia, E.M., Intille, S.: A long-term evaluation of sensing modalities for activity recognition. In: Proceedings of Ubicomp 2007, Innsbruck, Austria (September 2007)
25. Patterson, D.J., Fox, D., Kautz, H., Philipose, M.: Fine-grained activity recognition by aggregating abstract object usage. In: ISWC '05: Proceedings of the Ninth IEEE International Symposium on Wearable Computers, Washington, DC, USA, IEEE Computer Society (2005) 44–51

26. Lester, J., Choudhury, T., Kern, N., Borriello, G., Hannaford, B.: A hybrid discriminative/generative approach for modeling human activities. Proceedings of International Joint Conference on Artificial Intelligence (July 2005)
27. Hodges, M.R., Pollack, M.E.: An 'object-use fingerprint': The use of electronic sensors for human identification. In: Proceedings of Ubicomp. (2007) 289–303
28. Smith, J.R., Fishkin, K.P., Jiang, B., Mamishev, A., Philipose, M., Rea, A.D., Roy, S., Sundara-Rajan, K.: Rfid-based techniques for human-activity detection. Commun. ACM **48**(9) (2005) 39–44
29. Wyatt, D., Philipose, M., Choudhury, T.: Unsupervised activity recognition using automatically mined common sense. Proceedings of AAAI 2005 (July 2005)
30. Kukich, K.: Techniques for automatically correcting words in text. ACM Computing Surveys **24**(4) (1992) 377–439
31. Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady **10** (February 1966) 707–710
32. Smith, G.E., Ivnik, R.J., Lucas, J.: Assessment techniques: Tests, test batteries, norms and methodological approaches. In Ricker, J.M..J., ed.: Textbook of Clinical Neuropsychology, New York, Taylor & Francis (2008) 38–58
33. Grant, I., Adams, K.M.: Neuropsychological Assessment of Neuropsychiatric Disorders. Oxford University Press, Oxford (2008)
34. Lezak, M.D.: Neuropsychological Assessment, Fourth Edition. Oxford University Press, Oxford (2004)

Appendix: Neuropsychological Assessments Tables

Table 4 lists the individual neuropsychological assessments that we use as ground truth to measure the severity of a subject's impairment, with any principal components for which the assessment had a loading of 0.6 or higher. Table 5 gives the percentage of variation attributed to each principal component.

Wechsler Adult Intelligence Scale (WAIS) III Verbal Comprehension (1)
WAIS III Perceptual Reasoning (1)
WAIS III Working Memory (1)
WAIS III Processing Speed (1,3)
Wechsley Memory Scale-Revised (WMS-R) Logical Memory I (3)
WMS-R Logical Memory II (3)
WMS-R Visual Reproduction I
WMS-R Visual Reproduction II
California Verbal Learning Test II (CVLT II) Total (1)
CVLT II Long Delay Free Recall (4)
CVLT II Recall Discriminability (4)
Trails A
Trails B (2,5)
Booklet Category Test (BCT) Error Total
Wisconsin Card Sorting Test (WCST) Concepts (3)
WCST Perseverative Errors (5)
Controlled Oral Word Association Test (COWAT-FAS) Total (1)
Animals (1)
Wide Range Achievement Test (WRAT) 4 Reading (1)
WRAT 4 Mathematics (5)
Peabody Individual Achievement Test-Revised (PIAT-R) Reading Comprehension
Peabody Picture Vocabulary Test-Revised (PPVT-R) (1)
Tactual Performance Test (TPT) Total (2)
TPT Memory (1)
TPT Location (2)
Finger Tapping Test - Dominant
Finger Tapping Test - Non-Dominant
Grooved Pegboard (GPB) - Dominant (2)
GPB - Non-Dominant (2)

Table 4. List of Standard Neuropsychological Assessments Used. Parentheses Indicate Principal Components in Which the Assessment has a Loading of 0.6 or More.

Component	% of Variance	Cumulative %
1	26.4	26.4
2	15.0	41.5
3	12.8	54.2
4	9.0	63.3
5	8.7	72.0

Table 5. Principal Component Analysis of Neuropsychological Assessments