Empath: A Continuous Remote Emotional Health Monitoring System for Depressive Illness

Robert F. Dickerson
Dept. of Computer Science
UVA Center for Wireless
Health University of Virginia
Charlottesville, VA
rfdickerson@cs.virginia.edu

Eugenia I. Gorlin
Dept. of Psychology
University of Virginia
Charlottesville, VA
eig7qr@virginia.edu

John A. Stankovic
Dept. of Computer Science
UVA Center for Wireless
Health University of Virginia
Charlottesville, VA
stankovic@cs.virginia.edu

ABSTRACT

Depression is a major health issue affecting over 21 million American adults that often goes untreated, and even when undergoing treatment it is hard to monitor the effectiveness of the treatment. To address these issues, we have created a real-time depression monitoring system for the home. This system runs 24/7 and can potentially detect the early signs of a depression episode, as well track progress managing a depressive illness. A cohesive set of integrated wireless sensors, a touch screen station, mobile device, and associated software deliver the above capabilities. The data collected are multi-modal, spanning a number of different behavioral domains including sleep, weight, activities of daily living, and speech prosody. The reports generated by this aggregated data across multiple behavioral domains are aimed to provide caregivers with more accurate and thorough information about the client's current functioning, thus helping in their diagnostic assessment and therapeutic treatment planning as well for patients in the management and tracking of their symptoms. We present data of a case study showing the value of the system, deployed over a period of two weeks in a home during a depressive episode. Larger scale studies are planned for the future.

Categories and Subject Descriptors

H.3.4 [**Systems and Software**]: User Profiles and Alert Services

General Terms

Algorithms, Experimentation

Keywords

Emotional Health Monitoring, Clinical Depression

Wireless Health '11, October 10-13, 2011, San Diego, USA

1. INTRODUCTION

Depression is a major health issue that affects over 21 million American men and women each year. Depression often goes unrecognized and untreated, and even once treatment begins it is often difficult to monitor its effectiveness. This poses particular challenges for the diagnosis and treatment of depression, particularly for those who avoid visiting a doctor or therapist due to social stigmas or a lack of energy. Currently, depression diagnosis is often based on subjective screening questionnaires or structured clinical interviews that rely on timely in-person visits as well as accurate recollections by the patient. This makes early detection of depression symptoms exceedingly difficult among this population. Yet early detection and treatment of this debilitating disorder has been shown to improve patient outcomes considerably. Along with depression's detrimental affect on mood, it can lead to other associated problems because of reduced social interactions, decrease in personal hygiene, increased alcohol use, and neglect of medications for current medical conditions. Assessment and treatment are often hampered by a lack objective data to corroborate patients' retroactive self-reports about their current functioning; hence an objective symptom-monitoring tool could complement subject self-report measurement and enhance diagnostic accuracy.

Depression has several behavioral and psychosomatic manifestations [24, 22]. Independently, each has been studied and is well-documented in clinical research as well as in the widely used Diagnostic and Statistical Manual of Mental Disorders (DSM IV) [2]. For example, severe forms of depression have been shown to affect individuals' vocal prosody. Frequently, depressive episodes affect sleeping patterns, leading to increased or decreased sleep duration as well as diminished sleep quality (with frequent bouts of waking in the night, more restlessness during sleep, etc.) Depressive episodes are also commonly characterized by lack of social interaction and signs of anhedonia, i.e. the lack of pleasure in doing things one previously enjoyed and the withdrawal from one's usual activities of daily living. Appetite changes and resulting weight gain and loss are another commonly observed symptom and a DSM criterion for depression. Behavioral changes associated with depression onset also include reduction in gross motor activity and slowing of gait. Each of these components, on their own, do not give caretakers a complete picture of an individual's condition, since depression is syndromatic. Observing the combination of several behavioral markers can aid in the correct classification of the overall phenomenon as depression and in the prevention of false positives. We believe that by monitoring several factors together, and taking advantage of systematic temporal patterns of change across different behavioral domains, we can help clinicians to predict and reliably detect the onset of depression. To address this aim, we propose a 24/7 depression-monitoring product for in-home sensing, ideal for use in single-person homes. This product can aid in detecting the early signs of depression and can provide information about the effectiveness of any treatment. The end result could be improved quality of life and possible improvement of other medical conditions and problems caused by or related to depression. Additional goals are to minimize deployment cost and to make the system as passive and unintrusive as possible, to enable greater user adoption.

The contributions of this paper include: 1) presentation of an emotional health monitoring system installed in a home that can recognize danger signs for a depressive episode by combining objective measures such as activities and motion, speech prosody, sleep quality, and weight monitoring with subjective measures; 2) an integrated a set of user interfaces for patients and caregivers; 3) system installed and real data collected from an apartment and the results are presented to show the value of the system.

The remainder of the paper is organized as follows: Section 2 summarizes existing in-home health monitoring solutions. In Section 3 we present the Empath system and discuss the implementation of each component. In Section 4 we present results from a two week deployment collecting real data and perform controlled experiments on the sleep and speech to test specific problems. We conclude in Section 5.

2. RELATED WORK

We have seen an emergence of research into wireless sensor networks and smart environments for remote-monitoring for health-care [15] applications. At-home and mobile aging applications have been proposed to detect the cognitive, physical, and social changes that occur in the elderly that challenge their health [28]. Wireless networked sensors embedded in people's living spaces or carried on a person can collect objective information about behavioral patterns in real-time [27, 26]. Systems have been introduced to deal with quality of patient care, in particular for the impending worldwide "silver tsunami" where the aging population could overload the capacity of current hospitals. It is economically and socially advantageous to reduce burden of hospitals by enhancing prevention and early detection so people can stay at home for as long as possible. A few systems have been developed for this purpose, one example is AlarmNet [28], and assisted living and residential monitoring network for pervasive adaptive health-care using an extensible and heterogeneous architecture. Intel Research Seattle and University of Washington have built a system to infer activities of daily living (ADLs) using sensor tags placed on everyday objects such as toothbrushes and coffee cups. Their goal is to create an unobtrusive system to help manage ADLs for the senior population [19]. University of Rochester has built a five-room house outfitted with infrared sensors, computers, bio-sensors, and video cameras as they test concepts and prototype products. Georgia Tech built an Aware Home [14] as a prototype of an "intelligent space" combining context-aware and ubiquitous sensing, computer vision-based monitoring, and acoustic tracking for ubiquitous computing of everyday activities. MIT is working on their PlaceLab [12] initiative, which is a part of the House_n project, a one-bedroom condominium with hundreds of sensors installed in nearly every part of the house. Oregon State presented a technique for monitoring motor activity as a means of predicting cognitive changes in the elderly. It is able to detect both acute and gradual changes that may indicate the need for medical intervention [8]. There has been much focused research to improve ADL detection accuracy [26, 13].

The *Quantified Self* is a guiding principle that posits that a person should be an active participant in managing their own health and lifestyle through self-experimentation. Several sensor devices have been used for this purpose that collect data about a person's exercise, diet, and vital signs (blood pressure, resting heart rate) to give the, valuable feedback about their efforts to maintain a positive lifestyle. A few systems target emotional wellness. For instance, the Optimism App [16] is an application for both the desktop and mobile platforms that logs self-reported mood as well as medication use, exercise, and sleep quality. These mood charts have been recommended by psychiatrists and therapists as tools for their clients to use in monitoring their own mental health. A group from Digital Ecosystems and Business Intelligence Institute is working to integrate different kinds of patient data such as daily activities, bodily functions and emotions, as well as mental-health data reported by therapists, all of which is collected and collectively mined to reveal interesting patterns [6]. Researchers at the Rhode Island Hospital have developed a telemedicine-based depression protocol using simple display in-home healthcare, with pilot studies showing that it could improve geriatric depression [23]. The subjects in the study were favorable to the technology, reporting that the frequent checks from the display were reassuring and helped them to better understand their condition.

A project most related to depression monitoring was done by MIT and Mass General Hospital using their LiveNet system [25]. Subjects wore mobile physiologic sensing technology to track depression symptoms over time and to measure objective measures of depression. The patient pool came from psych wards and the technology was used to validate whether electro-convulsive therapy (ECT) was having positive effects on patients' depressive state. The measured data included skin conductance response, heart rate variability, movements, and vocal characteristics.

To our knowledge, there has been no system yet that has been implemented to provide continuous emotional monitoring in the home by combining objective symptomatic behavioral factors with the subjective factors. Continuous and daily self-report instruments such as Optimism go a long way toward mitigating self-reflection errors but do not incorporate enough potential factors that could be useful for episode monitoring such as speech, sleep, weight, and movement. Although LiveNet is able to collect similar behavioral features, it requires costly and cumbersome mobile equipment, and not designed to be deployed in the home. In addition, it did not track bodyweight and subjective measures which are important factors for depression.

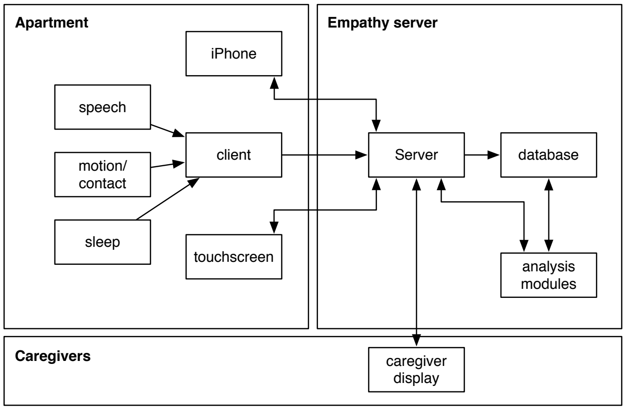

Figure 1: The communications architecture provides a reliable system for collecting data from the patient's apartment and uploading to a webserver for processing. Behavioral measures can be delivered to caregivers or patients on a variety of platforms.

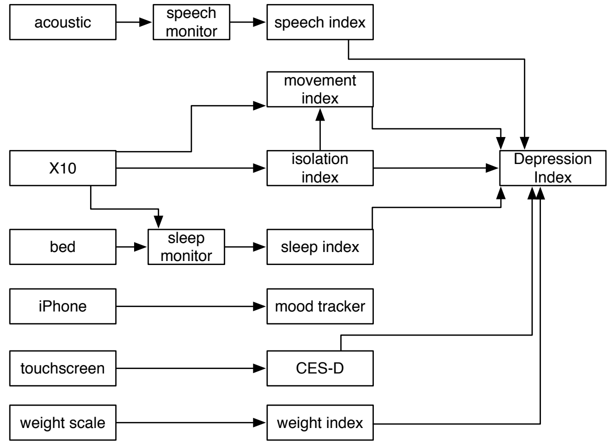

Figure 2: Each module is responsible for processing lower level data to high-level knowledge at defined intervals. The behavioral factors can be combined to arrive at a global depression risk index.

3. EMPATH PLATFORM

We implemented an integrated system of sensors and analysis code called *Empath*, an abbreviation for Emotional Monitoring for PATHology. Each module of the system addresses the factors listed among the DSM-IV criteria for depression as well as other factors identified in the depression literature.

3.1 System Architecture

When designing Empath, we considered the potential of scaling the system for multiple home deployments with various stakeholders such as caregivers (therapists, physicians, family) and patients needing access to the data, therefore a federated architecture with a single webserver and database was chosen. The communication infrastructure is presented in Figure 1. Each of the data collection modules (speech, activity, and sleep) archives its data locally. The synchronization client is responsible for connecting to the server and resolving disconnections with periodic retries. If any new data has been generated since disconnection, the new data are bundled and sent to the server. The webserver acts as a mediating layer between the sensing and user interfaces to the backend MySQL database.

The behavior analysis routines run on the server. Each module (sleep, weight, movement, etc) are programmed to activate at various intervals such as daily, weekly, or biweekly. All processing occurs on data stored in the database tables, which is processed into statistics or factor scores which then stored into different tables. Figure 2 shows the details of this process.

3.2 Sleep Monitoring

A number of clinical studies have found that depression results in disruptions in sleep patterns. Three sleep pattern abnormalities have been well documented in depressed patients [7]: sleep continuity problems such as difficulty falling asleep or staying asleep or waking up early, decreased slow-wave delta sleep, and alterations in the nature and timing of Rapid Eye Movement (REM) sleep. These abnormalities are present in about 80% of people with major depression, and hence shows the importance of monitoring sleep.

To measure the quality of sleep, there are self-report questionnaires such as the Pittsburgh Sleep Quality Index (PSQI) [4]. However, studies have shown that subjects with insomnia are not accurate in their subjective report of variables such as sleep latency, sleep duration, and the number of disturbances. Therefore, there is a need for objective instruments that can measure sleep quality where the subjective ratings fail. However to date, there are few low-cost and unobtrusive sleep monitoring systems. The most accurate instruments are polysomnography devices, but there are major drawbacks to using them since they need to be worn, require monitoring in a sleep lab, and need specialists to analyze the data. Therefore, they are expensive to use and not feasible in home environments. The actigraph is an accelerometer device that can be attached to any of the limbs (e.g. wrist, legs) to provide data on movement, however they still need to be worn. The Zeo is a headband that measures electrical signals on the skin of the scalp to estimate the stage in sleep, however they need to be worn and their accuracy has not been thoroughly evaluated. There are new research into various types of unobtrusive sleep monitoring solutions [20]. In our previous work [11], we have shown how WISP tags, active RFID devices with accelerometers, can be used to detect motions in the bed for measuring restlessness and potentially the quality of sleep. But to date, the RFID receiver device (which is placed under the bed) is too costly ($600 each), so we decided not to use our WISP solution.

To detect sleep cheaply and non-invasively, we built a custom solution using the Synapse SNAPpy RF motes for wirelessly transmitting data. We attached three independent triaxis accelerometers to the mote as shown in Figure 3 and they are sampled at 1 Hz. Data is processed on the client PC which determines the amount of deflection since the last sample. Since the accelerometers roughly indicate the force due to gravity as a vector, we use the dot product of the last sampled vector with the new to determine the amount of deflection since the last sampling. If that deflection exceeds a threshold (we used 3° based on controlled experiments for this particular mattress) a movement event has occurred. The advantage of this approach is that the true orientation of the accelerometer does not have to be established to mea-

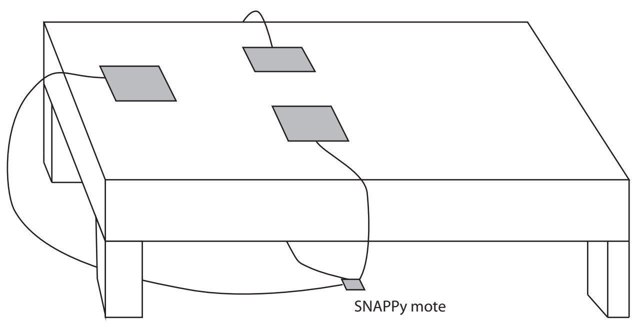

Figure 3: The sleeping monitoring setup. Three tri-axis accelerometers are taped onto the mattress and connected to a wireless Synapse mote.

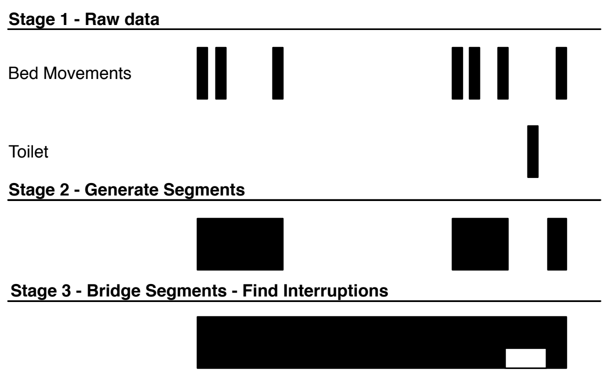

Figure 4: The discrete bed movements are processed by a multi-stage filter to discover the duration of sleeping and the number of interruptions. When other sensors fire (such as a toilet), an interruption has been created inside the sleeping segment.

sure movements, which allows us to do detection without calibration and it allows continued operation even when the sensors may have been knocked out of place. Since we only store events and not raw measurements, we can eliminate noise and lower the amount of storage required.

The next stage of the sleep analysis is to convert the discrete bed movements to sleeping activity segments. The process is illustrated in Figure 4. The first stage of the algorithm performs segmentation on the movement data using a rule that a sleeping segment will have at least ten movements and lasts for at least 20 minutes. Although the thresholds are contrived, this method performs well for eliminating noise such as if the user touches or lays something on the bed. The next stage joins possible segments together if no other sensors in the apartment fired such as a refrigerator, toilet, etc. The assumption, at least for single person homes, is that if no other sensors fire, you can assume the patient is still on the bed. Restlessness can also be recorded by the number of discrete movement events during a sleeping activity segment. More studies need to be done to correlate bed movements with sleeping quality, but should come correlation exist, it could be important for depression.

We created the following scoring strategy for calculating the *sleeping factor*. The parameters were chosen to match similar components used on the PSQI exam [4]. The following are determined and summed to get a score between 0 and 9:

1. *Terminal insomnia* - getting up early in the morning (>7AM=0 pts;6-7AM=1 pt;5-6AM=2;<5AM=3 pts).
2. *Interruptions* - if time taking an interruption is 1 hr or over, +3pts.
3. *Sleep Duration* - (>7 hrs=0 pts;6-7 hrs=1 pt;5-6 hrs=2 pts;<5 hours=3 pts)

3.3 Weight Monitoring

Weight measurements are taken by the Withings WIFI bodyweight scale. When a reading is taken, the data is uploaded using WIFI to the Withings webserver. When our weight module needs to evaluate changes in bodyweight, it polls the Withings server for new data using the WS-API web interface. The weight monitoring module uses the past two weeks of historical weight data to detect any significant weight gain or loss. Guided by the DSM criterion, if the patient's bodyweight has changed at least 5% in the past two weeks, it could be a sign of appetite changes and depression. If no new measurements have been taken for a week, an alert appears on the patient's touch-screen device. The following scoring system is used for the *weight factor*: 0 pts - within 5% of body weight, 1 pt - 5% gained or lost, 2 pts - over 10% of body weight gained or lost.

3.4 Speech Analysis

Acoustic features of voice such as pitch, utterance duration, and speech pause time have been shown in previous studies to detect the severity of depression [10, 18, 1, 5]. However, these studies were done in controlled environments under the oversight of speech pathology experts analyzing the patient's voice at a fixed distance from a microphone. The challenge is to incorporate speech monitoring to work at real-time in natural home settings. Our solution, uses a microphone attached to the patient's touchscreen device. A prompt appears on the screen telling the patient to give a free response to the question "How was your day today?". Speech segments are recorded at 44.1K sampling rate mono channel and only taken when the input exceeds the silence threshold and simple filters are used to remove noise. Pitch contours are generated from the signal using a pitch detection algorithm (PDA) [17] implemented in the Edinburgh Speech Tools (EST) Library. It is important to note that the fundamental frequency in speech cannot determined as simply as taking a DFT since the pitch requires estimating the missing fundamental. Human perception of pitch is more determined by the ratio of the ascending harmonics. The fundamental frequency for male voices fall within 60-200 Hz and females 120-400 Hz. The standard deviation on the pitch contour is used to infer the amount of vocal inflection. The speech pause time, the silent interval between phonations, were estimated by the duration of silence between successive pitch contours. Large gaps (greater than 1s) were not used in the calculation. Once the statistics are computed, they are uploaded from the client to the server. The original file can be kept on the client PC or deleted if privacy is a concern. There is a challenge to compute a *speech factor* score since the relationship between mood and speech characteristics seem to be dependent on a particular person. We use multi-variate linear regression model to fit the self report mood to the speech features. Once those parameters can be found, then the speech monitoring can be done automatically.

3.5 Activity Detection

We used X10 PIR motion detectors and door/window contact reed switches (DS10A) for basic activity detection. On the client PC, a W800RF32A antenna receives the X10 packets, and sends it by a serial port to the PC. We built our own driver for parsing the X10 packets. X10 devices do not guarantee reliable communication, instead, they will send five duplicate packets in series when a sensor fires. Each packet contains byte-compliment pairs. If at least one packet gets through to the receiver, then a packet delivery ratio can be computed. We found that poor packet delivery ratio can be a result of poor antenna range and obstructions such as walls. Parity check errors occur from two devices sending messages at the same time. The solution to this problem is to move the sensors farther apart.

We use activity detection to detect symptoms that are related to loss of energy or anhedonia and social isolation. In particular, Empath examines two factors that are linked to depression 1) home occupancy and 2) movement levels.

3.5.1 Home Occupancy and Movement Level

Depression can express itself in anhedonia or by social isolation. In this case, patients will leave their homes less often than normal. We measure the percentage of time spent in the home versus away, and monitor for anomalies in this pattern. Many things can contribute to this factor changing and it is sensitive to false positives such as 1) going on a vacation 2) medical problems (cold, broken leg) and 3) weekend and weekday work schedules.

We use a simple algorithm to predict the time the patient was in their home. The basic principle is that segmentation of periods where people are inside or outside of their homes should occur when a front door sensor fires. In our deployment, this is the only portal through which someone could enter or leave the apartment. However, each time the door opened the patient did not necessarily enter or leave their home since they can open the door to let a breeze in or more light in. Next, each segment is labelled as occupied or not occupied by using the other X10 devices in the home (kitchen sensors, bathroom, etc). If the sum of events exceeds a threshold, the segment is labelled as *home occupied*. This simple approach works sufficiently well, however it has a single point of failure- the front door sensor. If the sensor malfunctions, the system cannot define crisp boundaries for home occupancy, and instead would have to rely on clustering of other activities to estimate occupancy. We suggest having double redundancy on front door sensors to improve reliability.

A score from 0 to 3 is generated for the social isolation score, and one point is given for each increase in time spent at home by one standard deviation. Next, the *movement factor* is computed from the number of sensor firings that go off in a day. We scale the activity level to the amount of time spent in the home.

3.5.2 Activities of Daily Living

Activities of daily living are logged for future reference, but not used yet for any calculations of factors since misclassification levels are too high and the challenge of determining anomalies has not been solved yet. Activities that can be monitored include 1) cooking 2) hygiene and 3) cleaning. We differentiate two types of cooking: preparing light meals and snacks, to preparing more complex meals. Contact switches are placed on the microwave, the oven, and the cabinets (spice and sauces), the refrigerator, and the freezer. Our simple recognition algorithm detects the opening of the spice cabinet or the use of the oven and stove as being a *complex meal*. Using the microwave or the freezer without the previous mentioned sensors, it is considered a *light meal*. Detecting whether someone is eating out, or not eating at all is challenging, and cannot be easily determined using our activity recognition system. This is why augmenting the ADL data with a bodyweight scale is important. For hygiene, we detect showering, using the bathroom sink, opening cleaning closets, and opening the trash lid. A motion sensor was placed in the shower unit and over the sink, and contact sensors are placed between the trash can and the lid and storage cabinets.

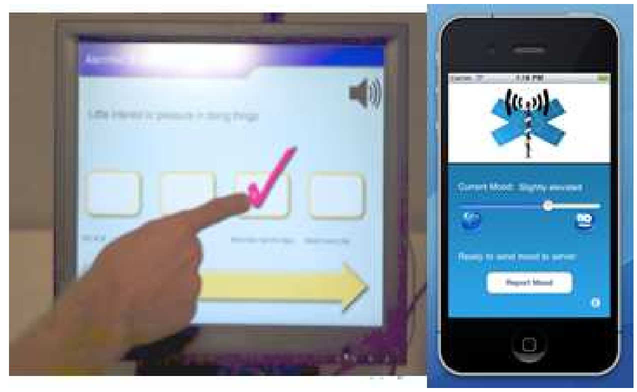

Figure 5: Two user interfaces were created for the patients. The touchscreen administers the CES-D exam and the iPhone runs an App that does frequent experiential sampling of mood.

3.6 Patient Display and Subjective Scores

The patient interface shown in Figure 5 runs on a touchscreen inside the patient's apartment. Its primary purpose is to receive continuous subjective scores from the items in the CES-D [21] exam. The test is available at all times, but encouraged to be taken once a week. The exam consists of 20-items, where each item is scored on a scale from 0-3 points. The sum of these items are used to predict the severity of the episode. A score of 15-21 might suggest mild to moderate depression and over 21 a possibility of major depression. The implementation is was built with Adobe AIR 2, which connects to the server and transmits the scores through an XML protocol. Notifications and alerts can be sent from the server to the patient that appear on the screen. Some examples of alerts include: reminders to check body weight on the scale and to complete a late CES-D exam. We plan on expanding the touchscreen's capabilities to serve as a mood coach, social planner, and mood journal. Personal behavioral factors similar to the caregiver is presented to patient for positive feedback.

We created an iPhone application (shown on the right in Figure 5)using the iOS 4.3 SDK that serves as an input device for instantaneous mood measurements. The patient to

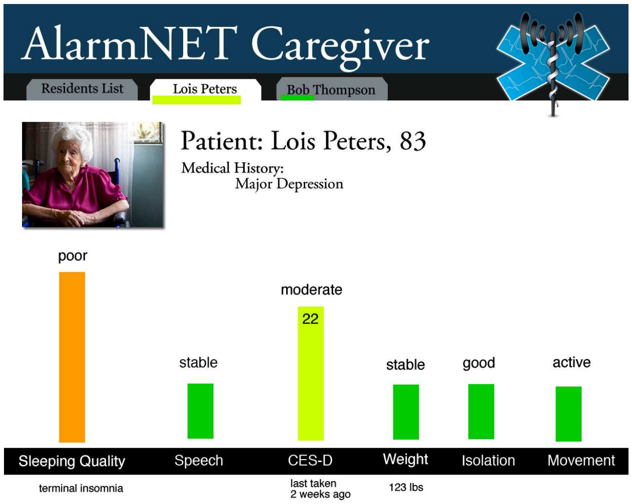

Figure 6: The caregiver display shows an overview of attending patients and a breakdown of their objective and subjective measurements. More data can be "drilled down" by selecting a factor, which then a time plot or table is shown.

input instantaneous mood on a 1-10 ladder on the continuum of extreme depressed to extreme elevated mood. We created this mobile interface so that the patient does not necessarily need to be in his or her home to input data into the system. This experiential sampling approach [9] is useful for collecting instantaneous measurements. The iPhone's local notifications can be enabled to alert the patient when a new measurement is recommended. The application's capabilities can be expanded in the future to record other types of emotions, such as levels of anxiety or irritability which are also typically experienced during depression.

3.7 Caregiver Display

We developed a user interface especially for caregivers such as therapists, nurses, or doctors. The caregiver's screen is shown in Figure 6. The caregiver's list of attending patients are presented with an overview of their depression risk factor. When a patient is selected, a summary of the current behavioral factors: sleeping quality, social isolation, CES-D score, weight, movement levels, and speech analysis are presented as a bar graph. Each factor is represented on a scale (from green to red) representing the risk for a particular factor. When the caregiver selects the factor, a new view appears either with a time-series plot or table showing detailed information. For instance, when the CES-D is selected, historical tests and items can be individually evaluated. For sleep, detailed statistics can be shown such as bed time, number of interruptions, and sleep durations. To put each patient's history in context and to see if a patient is improving, annotations can be added to the display indicating when a patient started new therapy or medication. This system does not perform diagnosis, rather it exposes all the factors in a presentable way to improve diagnosis.

4. DEPRESSION CASE STUDY

In this section, we present a case study of the Empath system deployed in a real apartment over a period of 14 days.

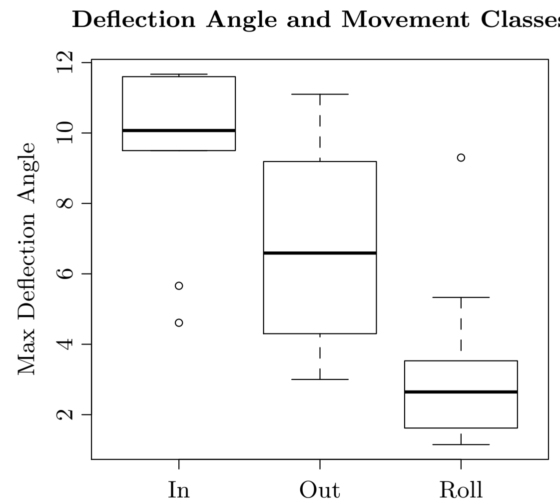

Figure 7: The maximum deflection angles for the actions: getting in the bed, out of the bed, or rolling from side to side.

The following results are not meant to make any scientific claims or prove any medical hypotheses, however it shows an example of the system in operation and how it is able to collect data about a depressive episode continuously in the home. The subject who volunteered for this study has had a history of depressive illness and during the period of data collection, scored a 30 and 37 on the CES-D indicating moderate to severe depressive symptoms. During the case study, the subject was not undergoing any medications or undergoing any form of therapy.

It took less than one hour to install Empath in the subject's home. X10 devices were attached to the stove, freezer, refrigerator, kitchen sink, microwave, spice cabinet, plate cabinet, glasses and cups cabinet, front door, cleaning closet, medicinal closet, bathroom sink, trash can, wardrobe closet, and shower. The weight scale was placed on the floor of the bathroom. A PC with the client software was placed in the living room. The total cost of the system excluding the laptop and phone is less than $500.

The subject used the iPhone App to record his mood twice daily. Due to the diurnal variation of symptoms during depression, these measurements can vary greatly. We took measurements more frequently than would be typically needed by a patient using Empath. But the high-granularity of data is useful for comparing against the objective factors.

4.1 Sleep Analysis

4.1.1 Controlled Experiment

From initial testing, we discovered that the placement of the accelerometers on the bed must carefully be considered. We tried different options, starting with placing them on the rim of sides of the mattress. This worked well for detecting when the user entered and left the bed, it could not capture rolling around. The final configuration chosen was to place two accelerometers on the sides of the mattress inset about one foot. The third accelerometer was tested at

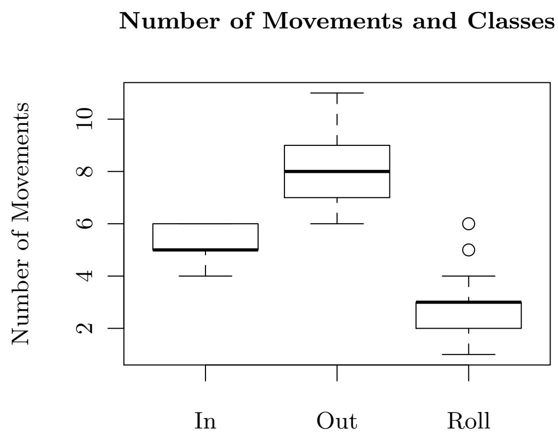

Figure 8: The number of movements needed for each action were compared across groups.

various places, and we decided to place it under the pillow. Detecting feet moving would be useful, however the weight from the feet were not heavy enough to depress the mattress.

We performed the following experiment to test the bed sensors, and to investigate whether the accelerometers on their own can recognize different types of movements produced relating to sleep. The mattress tested was a Serta posturepedic twin-size mattress and the sheets were placed on top of the accelerometers. The subject remarked that she could not feel the accelerometers when lying on the bed.

The subject performed the following rolling movements five times each: rolling right, to the center, rolling left, and back to the center. Additionally, we tested getting in and out of the bed. Five times each, the subject entered the bed to the left, then to the right, and five times each again the subject got out of the bed on the left and the right. When measuring the maximum deflection angle of the series of movements, we see in Figure 7 that getting into the bed produced the greatest deflection angle. As expected, rolling on the bed produced a lower deflection angle. We ran a one-sided T-test that the mean deflection angle of rolling in bed is less than getting out of the bed with a resulting value of $p<0.001$. From the following data, it suggests that threshold boundaries such as one standard deviation from the mean can be set between class types. We decided to investigate whether the number of movements can be used to tell the difference between the same class types. We found that one roll in the bed produced fewer movements than getting into the bed and getting out of the bed as shown in Figure 8. Getting out of the bed produced the most amount of movements since this action requires several steps such as pushing oneself into a seated position, then placing the legs on the floor and finally exiting the bed. The problem with using the number of movements as a feature is that multiple rolls could be confused with getting into and out of the bed. Since our deployment has multiple sensors, we decided to use the algorithm described earlier for determining the sleep periods

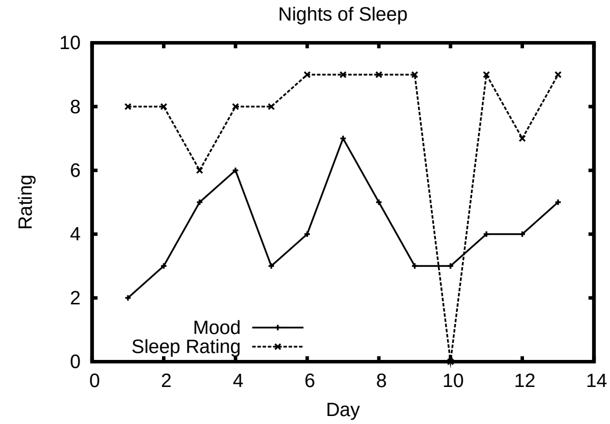

Figure 9: For 13 days, the duration of sleep, number of interruptions, and the time spent for each interruption were recorded. The sleep rating score was computed and compared against the mood.

	Estimate	Std. Error	t value	Pr(>|t|)
intercept	-0.26398	0.78950	-0.33	0.7439
$\mu(F_0)$	-0.01203	0.00917	-1.31	0.2141
$\sigma(F_0)$	0.02561	0.01528	1.68	0.1196
$\mu(SPT)$	-0.33934	0.09636	-3.52	0.0042
$\sigma(SPT)$	0.36225	0.27552	1.31	0.2131

Table 1: A linear model of speech features were fit to the subjective self-report mood using multi-variate linear regression. For this subject, the mean speech pause time was the strongest indicator variable for the mood.

and interruptions. From this study, we show that using only accelerometer data from the bed motion we can get similar accuracy than using an entirely instrumented apartment.

4.1.2 Case Study

Each morning, the subject reported his subjective rating of the previous night's rest as being good or poor. Figure 9 shows the sleep quality rating and mood for each night. We inverted the sleep score by taking the difference from 9, since we wanted to present in the graph poor sleep quality with a lower number. The nights where the subject responded that his sleep was poor were on days 2, 3, and 10, which appears to correlate with our sleep quality index. The graph suggests that for this subject, the previous day's mood highly affects the sleeping quality that night.

These results show how Empath's sleep monitoring solution can approximate sleep quality with some degree of accuracy. However, one challenging problem we aim to solve, is determining the *sleep efficiency*, the amount of time spent in bed attempting to sleep rather than actually sleeping. We plan to run studies showing the relationship between actual sleep times and bed motion.

4.2 Speech Analysis

4.2.1 Deployment Analysis

For each day, two speech samples were taken once in the

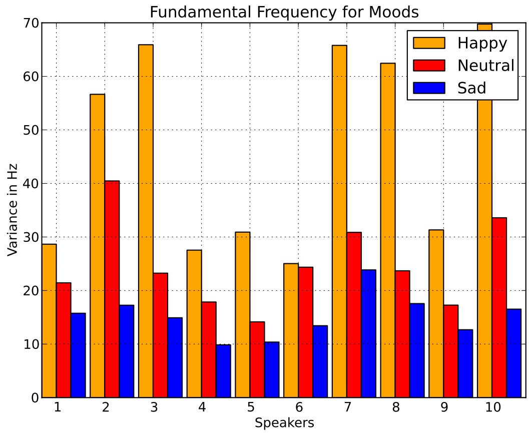

Figure 10: The vocal modulation of speech (standard deviation) using the pitch detection algorithm for each of the ten speakers in the EmoDb dataset.

morning and once at night roughly when the self-report scores were collected on the iPhone App. The following prompts were used to guide the subject's response: for the morning sample, the subject discussed what he had planned for the day, and during the evening the subject discussed what he had done that day. Extracting speech samples in this manner have several advantages: 1) the microphone was only inches from the subject gave high quality recordings and no voice discrimination was necessary. 2) the free responses were generally 1-2 minutes long which gave us a long sample to do the analysis more accurately.

We use the iPhone self-report measure as our dependent variable and fit the following linear function:

$$mood = \beta_0 + \beta_1\mu(F_0) + \beta_2\sigma(F_o) + \beta_3\mu(SPT) + \beta_4\sigma(SPT)$$

The parameters of the speech samples considering fundamental frequency (F_0) and speech pause time (SPT) with a subjective mood are shown in Table 1. The model fit the data extremely well with a residual error of 0.0916 on 12 degrees of freedom (p<0.011).

4.2.2 Dataset Analysis

Next, we decided to test if this approach would work with a larger set of speakers. We tested the speech component against a known public dataset, the Berlin Database of Emotional Speech (EmoDB) [3]. This database contains emotional utterances that were spoken by actors and each sample were evaluated using perception tests by others to determine their naturalness. There were five male and five female speakers, and each said the same ten different utterances with varied emotions. We used the data labelled happy, sad, and neutral.

Each of the ten utterances in each mood group were concatenated together to form a long running instance. We run the feature extraction algorithm from Empath on these waveforms to determine the frequency curves. The modulation of fundamental frequencies for neutral, sad, and happy data for each of the ten speakers are plotted in Figure 10. Each of the speakers showed a decrease in the level of modulation in their voice as the affect went from happy to sad.

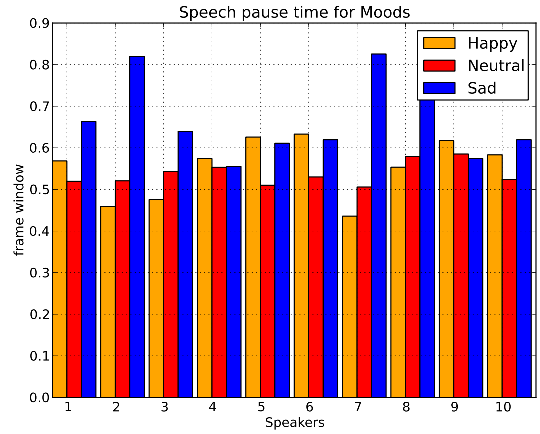

Figure 11: The speech pause time for each of the ten speakers in the EmoDb dataset.

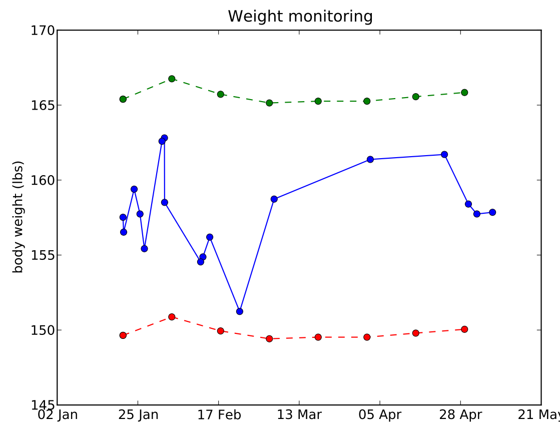

Figure 12: Body weight measurements were taken over a period of three months, and the staying within healthy limits (within upper and lower bounds)

Those with the higher variation across groups were the females (speakers 2, 3, 7, 8, and 10). Speech pause time however did not always yield significant difference between the classification types. It can be a discerning factor for some speakers, and not for others. It is is clear that both variables are important to help predict affect in the voice.

4.3 Weight Monitoring

For three months the subject took his weight at various intervals as shown in Figure 12. Every two weeks the evaluator created new upper and lower limits for the weight using the historical information from the last evaluation time. For the period of evaluation, the subject remained within 5% of his body weight in each of the two week intervals, and thus the weight gain/loss risk factor was zero across the experiment time. The weight monitoring solution is not yet resilient to planned changes in body weight such as starting a new diet or weight training.

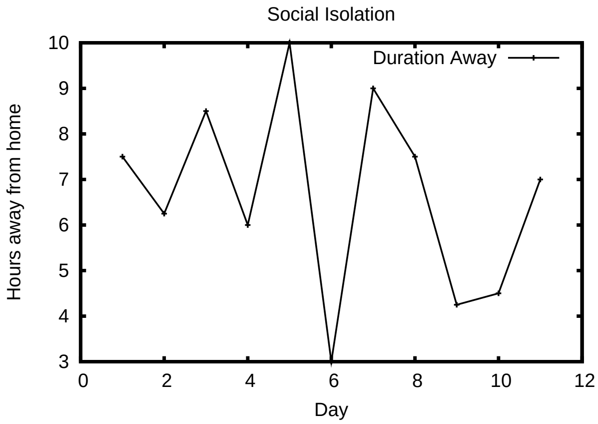

Figure 13: For 11 days, the number of hours spent away from home were recorded.

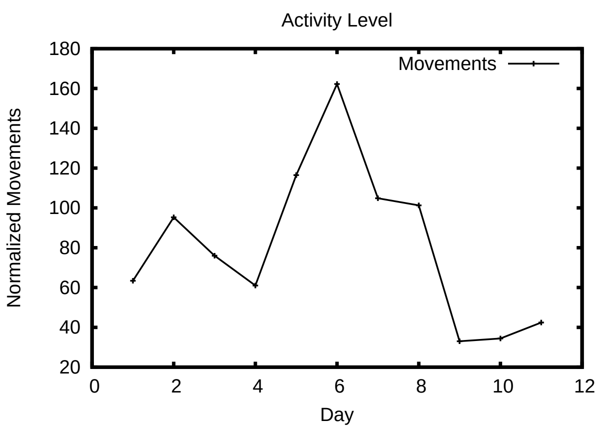

Figure 14: For 11 days, the number of movements detected in the home were normalized to the time spent in the home.

4.4 Apartment Occupancy

The occupancy detection algorithm was used on 11 days worth of data. Figure 13 shows for each of the 11 days the amount of time spent away (vacant) each day. For this particular dataset, we found no relationship between mood and time spent away from the home, by running an ANOVA on the linear relationship between mood and duration. The assumption here is that higher levels of vacancy correlate to less social isolation. There are complications to this measure as if the subject stays at home, but receives visitors, the factor will be lower than it should be. In addition, times spent on vacation can produces errors in this estimation. We see that this is where other factors are important in this measure.

4.5 Motion Levels

For each day, we recorded the number of sensor firings in the home to give us a gross estimate of the amount of motion and activities occurring in the home. Those who stay still, and therefore do not interact with many devices, and will receive a lower movement factor. We realized that the number of firings do not give us an fair measure of the activity level, since a person who scurries about their apartment for a few hours would receive a lower score than someone who spent the entire day in the apartment but spending most of the time on the couch. So we normalized the score based on the apartment occupancy times. Figure 14 shows the results of producing this factor against the reported mood. The sixth day was the most active for the participant, since day was spent cleaning the apartment. We ran an ANOVA on the linear model again to find a relationship between the movement factor and mood levels, but no significance were found. This method gives us an approximation of energy levels, that may correlated heavily to psychomotor retardation that depressed individuals experience.

4.6 Depression Index and Integration

We use various factors together for arriving at a depression risk index. People exhibit depression in different ways, so relying on single measures is not accurate. In this case study, for instance, the speech factor and sleeping factors were most highly indicative of depression, while the weight, movement, and isolation were within healthy limits. The depression index is a weighted sum of various subcomponents, however should any of the subfactors be in extremely high risk, the depression index should be elevated to a high level. Studies on the population need to be run first to make decisions on the weighting factors.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we presented Empath, a continuous monitoring system for the home that can collect continuous objective and self-report measurements concerning mood and factors linked to depression.

We have shown through controlled studies that the sleep monitoring solution can accurately measure sleeping data, and could potentially be an objective tool that can match the objective measurements in the PSQI exam. Based on accepted datasets, the speech analysis solution seems to scale to other speakers as well. The week long deployment shows all the factors Empath can generate presenting its usefulness to both to the patient and to caregivers looking to get more data about depression conditions.

5.1 Future Plan for Technology Updates

For this work, we focused on sleep, occupancy, speech, and bodyweight. However, there are many other factors that can give patients and caregivers better knowledge about the condition. In particular, the activity detection module can be refined to report specific events that are occurring in the day. Some important activities include: 1) taking medication 2) social interaction 3) exercise 4) alcohol use and 5) recreation and play. One challenge is to detect activities that do not occur inside the home such as social interaction and exercising. Hence, we are investigating ways to expand the activity recognition onto mobile devices.

5.2 Future Plan for Medical Research

Because Empath is designed for people living by themselves, a target group that could benefit from such a system are the elderly living in by themselves or in assisted living. It is estimated that over 15% of people over the age of 65 have depressive symptoms. As technology enables the elderly to age in their homes, they will be in less contact daily with caregivers. Beyond monitoring for clinical depression, Empath could be used by soldiers returning from war and monitoring Post-traumatic Stress Disorder (PTSD). PTSD shares several symptoms of depression, such as isolation, loss of interest in activities and life in general, difficulty falling or staying asleep, and difficulty concentrating.

6. ACKNOWLEDGMENTS

We would like to thank Bethany Teachman, Professor of Clinical Psychology and Seki Balogun, M.D. Professor of Geriatrics for their valuable discussions and feedback. Also Enamul Hoque and Timothy Hnat assisted in the sleep monitoring solution. This work was supported, in part, by NSF grants ECCS-0901686 and IIS-0931972.

7. REFERENCES

[1] M. Alpert, E. Pouget, and R. R. Silva. Reflections of depression in acoustic measures of the patient's speech. In *Journal of Affective Disorders*, 2001.

[2] APA. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. APA, 4 edition, 1994.

[3] F. Burkhardt, A. Paeschke, M. Rolfes, and W. Sendlmeier. A database of german emotional speech. *Interspeech*, 2005.

[4] D. Buysse, C. Reynolds, and T. Monk. The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. In *Journal of Psychiatry Research*, 1988.

[5] A. Flint, S. Black, and I. Campbell-Taylor. Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression. In *Journal of Psychiatric Research*, 1993.

[6] M. Hadzic, F. Hadzic, and T. Dillon. Mining a patient data: towards better treatment strategies for depression. In *Functional Informatics and Personalized Medicine (IJFIPM)*, volume 3, 2010.

[7] C. Hammen. *Depression*. Psychology Press Ltd., 1997.

[8] T. Hayes, MPavel, and J. Kaye. An Unobtrusive In-home Monitoring System for Detection of Key Motor Changes Preceding Cognitive Decline. *IEEE EMBS*, pages 1–4, jun 2004.

[9] J. Hicks, N. Ramanathan, D. Kim, M. Monibi, J. Selsky, M. Hansen, and D. Estrin. Andwellness: an open mobile system for activity and experience sampling. In *Wireless Health 2010*, 2010.

[10] G. Hoffmann, J. Gonze, and J. Mendlewicz. Speech pause time as a method for the evaluation of psychomotor retardation in depressive illness. In *British Journal of Psychiatry*, 1985.

[11] E. Hoque, R. Dickerson, and J. Stankovic. Monitoring body positions and movements during sleep using wisps. In *Wireless Health 2010*, pages 1–10, Aug. 2010.

[12] S. Intille, K. Larson, E. Tapia, J. Beaudin, P. Kaushik, J. Nawyn, and R. Rockinson. Using a live-in laboratory for ubiquitous computing research. In *Pervasive*, 2006.

[13] T. Kasteren, A. Noulas, G. Englebienne, and B. Krose. Accurate activity recognition in a home setting. In *UbiComp*, 2008.

[14] J. A. Kientz, S. N. Patel, B. Jones, E. Price, E. D. Mynatt, and G. D. Abowd. The georgia tech aware home. In *CHI '08 extended abstracts on Human factors in computing systems*, 2008.

[15] J. Ko, C. Lu, , M. Srivastava, J. Stankovic, A. Terzis, and M. Welsh. Wireless sensor networks for healthcare. In *Proceedings of the IEEE*, 2010.

[16] O. A. P. Ltd. Optimism apps, 2011.

[17] Y. Medan, E. Yair, and D. Chazan. Super resolution pitch determination of speech signals. In *IEEE Transactions on Signal Processing*, volume 39, pages 40 –48, jan 1991.

[18] A. Nilsonne, J. Sundberg, S. Ternstrom, and A. Askenfelt. Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression. In *Journal of the Acoustic Society of America*, 1988.

[19] D. Patterson, D. Fox, H. Kautz, and M. Philipose. Expressive, tractable and scalable techniques for modeling activities of daily living. In *UbiHealth 2003: Ubiquitous Computing for Pervasive Healthcare Applications*, 2003.

[20] Y.-T. Peng, C.-Y. Lin, M.-T. Sun, and C. A. Landis. Multimodality Sensor System for Long-Term Sleep Quality Monitoring. *IEEE Transactions on Biomedical Circuits and Systems*, 1(3):217–227, 2007.

[21] L. Radloff. The CES-D scale: A self-report depression scale for research in the general. *Applied psychological measurement*, 1977.

[22] D. Schrijvers, W. Hulstijn, and B. Sabbe. Psychomotor symptoms in depression: a diagnostic, pathophysiological and therapeutic tool. *Journal of Affective Disorders*, jan 2008.

[23] T. Sheeran. Viability and impact of telehealth-based depression care in home care. Technical report, Rhode Island Hospital, 2010.

[24] C. Sobin and H. Sackeim. Psychomotor symptoms of depression. *American Journal of Psychiatry*, jan 1997.

[25] M. Sung, C. D. Marci, and A. Pentland. Objective physiological and behavioral measures for identifying and tracking depression state in clinically depressed patients. Technical report, MIT, 2005.

[26] E. M. Tapia, S. S. Intille, and K. Larson. Activity recognition in the home using simple and ubiquitous sensors. *Pervasive*, 2004.

[27] G. Virone, N. Noury, and J. Demongeot. A system for automatic measurement of circadian activity deviations in telemedicine. *IEEE Transactions on Biomedical Engineering*, 49(12):1463–1469, dec 2002.

[28] A. Wood, G. Virone, T. Doan, Q. Cao, and L. Selavo. ALARM-NET: Wireless sensor networks for assisted-living and residential monitoring. Technical report, University of Virginia, Jan 2006.