System for the Collection, Storage, Analysis, and Reporting of Objective Behavioral Measures

Ph.D. Dissertation Proposal

Robert F. Dickerson
rfdickerson@cs.virginia.edu
Department of Computer Science
University of Virginia
Charlottesville, VA, USA

Thesis

> This thesis investigates how a continuous, multi-modal monitoring system can be built that is functional, feasible, and effective at recognizing depressive symptoms in many homes.

1 Introduction

Depression is a major health issue that affects over 19 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, 26]. 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 its own, does not give caretakers a complete picture of an individual's condition, since depression is marked by a combination of symptoms. To address this aim, we propose a 24/7 depression-monitoring system for in-home sensing, ideal for use in single-person homes. This system can aid in detecting the early signs of depression or can provide information about improvement of symptoms. The end result could result in improved quality of life for many suffers from this disease.

2 Proposed Research

Before such a system can be realized, there are many technical challenges that first need to be considered and solved. These include the aggregation and inference architecture and various subcomponents such as user interfaces, speech affect monitoring, sleep monitoring. Finally, scalability will be considered for porting this system to large populations over long periods of time such as months and years.

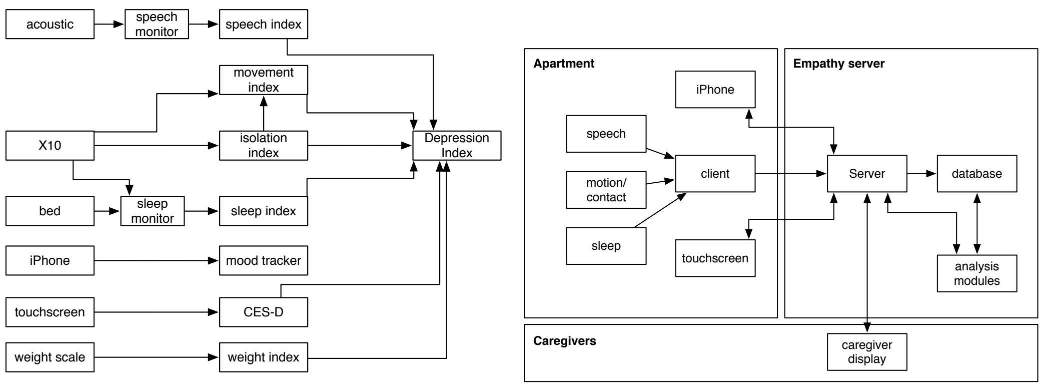
2.1 Disease Inference

Most medical applications using emplaced sensor networks or body sensor networks collect vital signs independently, and present the signals (such as the ECG) to doctors and other caregivers. Depression, however is a complex disease that cannot be isolated to abnormalities in an individual factor, rather it is based on a combination of several factors. Therefore, we need a wide range of sensing modalities to increase diagnostic accuracy. In addition, there must be comprehensive techniques to combined the data into a useful form. Creating such as a system is a challenging problem and requires the following:

- **Heterogeneity**: The system should be able to fuse heterogeneous and both subjective and objective data in a systematic way. Subjective and objective data used together can help get more accurate diagnosis accuracy and treatment prescription.
- **Uncertainty**: The system should have a way to handle the inherent uncertainty of each type of information. Some sensing modalities are more accurate than others, and the inference chain should be aware of this.
- **Extensibility**: Devices will be added, removed, or upgraded in the the homes. The system should be able to handle changing inputs. The system will be able to interchange components and be able to operate with some or all of the components.
- **Domain expert knowledge**: Easily integrate the domain experts prior knowledge. Rule-bases will be written in a declarative fashion by domain experts to input information about various symptoms.
- **Automatic knowledge discovery**: Domain experts can miss many rules that can infer depression. Therefore, automatic discovery of anomalies and patterns across other patients can be used to determine alarms.

2.1.1 State-of-the-Art

We have seen an emergence of wireless sensor networks and smart environments for use in health-care [15] monitoring. 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 [30]. Wireless networked sensors embedded in people's living spaces or carried on a person can collect objective information about behavioral patterns in real-time [28,29]. Systems have been introduced to deal with quality of patient care, in particular for the coming worldwide silver tsunami where a large number of retiring elders 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. Many systems have been developed. One example of a WSN is AlarmNet [30], and assisted living and residential monitoring network for pervasive adaptive healthcare using 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 [21]. 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 [11] initiative, which is a part of the House_n project. It is 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 [9]. Much research has been done to improve ADL detection accuracy [13,28].



(a) Inference chain (b) Communication network

Fig. 1. Design of the Empath system

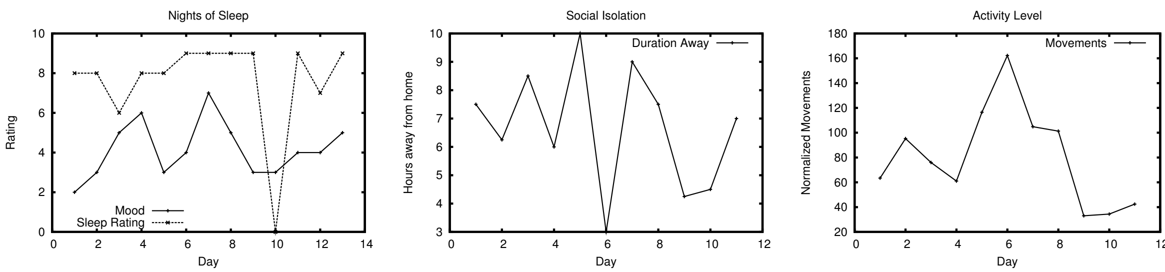
Collecting behavioral symptoms of clinical depression from sensor devices is not an entirely new idea. Along with isolated studies targeting specific symptoms done in the past few decades, a project lead by MIT and Mass General Hospital in Boston has created the LiveNet system [27]. They used mobile physiologic sensing technology on patients in psychiatric wards to track severe depression symptoms over time to determine whether electro-convulsive therapy (ECT) were having positive effects on their on depressive state. The measured data includes skin conductance response, heart rate variability, movements, and vocal characteristics. LiveNet is able to collect similar behavioral features as we are proposing, however the system requires wearing costly and cumbersome mobile equipment, it is not intended to be placed in homes. In addition, it does not track sleep, weight, and self reported daily mood which are important factors for depression. Preliminary data from their experiments have shown differences between non-depressed and depressed states, however the researchers have not focused on the presentation of the data into useful forms for caregivers.

2.1.2 Preliminary Results

We created an initial system that addresses this problem, with an integrated set of sensors and analysis code called *Empath* (Emotional Monitoring for PATHology). Each module of the system follows closely the factors listed in the Diagnostic and Statistic Manual for Disorders (DSM-IV) [2] criteria for depression as well as other factors identified in depression clinical studies. The data collection and networking architecture is designed to accommodate multiple home installations with various stake-holders such as caregivers (therapists, nurses, physicians, family) and patients sharing access to the data (Fig. 1(b)). Various sensors and data processing modules run on a laptop in the home and data is stored locally. Intermittedly, a synchronization protocol connects to a webserver and uploads data to the backend. In the backend, the data is stored onto a MySQL database. At various intervals, symptoms and behavioral trends are assessed, and the summary is stored back onto the database for inspection later. Figure 1(a) shows the details of the current evaluation routines. Currently, subfactor scores are generated using a primitive rule-based heuristics such as sleep should greater than 7 hours with fewer than 1 interruption per night, and weight should not deviate 10% from normal.

The system was deployed in a real apartment for over 14 days. Although the results from the pilot are not meant to make any scientific claims or prove any medical hypotheses, it shows potential that such a system could be used to monitor a depressive episode continuously in the home. It took less than one hour to install Empath in the participant's home. X10 devices were attached to the stove, freezer, refrigerator, kitchen sink, microwave, cabinets, bathroom sink, trash can, wardrobe closet, and shower. A wireless weight scale was placed on the floor of the bathroom. A PC with the wireless receiver, software, and database was in the living room. The total cost of the system excluding the PC is less than $500. Each day, the participant recorded his subjective mood on an iPhone application.

Each morning, the subject reported his subjective rating of the previous night's rest as being good or poor. Figure 2(a) shows for each night, the sleep quality rating and mood for each night. We inverted the sleep score by taking the difference from 9, since we wanted to illustrate 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, it seems to roughly correlate with our sleep quality index. The graph suggests that for this subject, the previous night's mood highly effects



(a) Sleep rating (b) Home vacancy (c) Movement Levels

Fig. 2. Preliminary data from a case study

the sleeping 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 compared with actually sleeping. We plan to run studies showing the relationship between actual sleep times and bed motion.

The occupancy detection algorithm was used on 11 days worth of data. Figure 2(b) 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.

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 2(c) 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.

2.1.3 Proposed Approach

To address the *heterogeneity* problem, we will create various modules in the inference chain that will process the data creating higher levels of inference. For instance, binary switch sensors in the kitchen will infer cooking events, which will be combined with the body weight scale to infer changes in appetite. Also, the system will be opportunistic, leveraging both objective and subjective data. When certain subjective data is unavailable, the system will use corresponding objective data and vice versa. Currently, systems do not use both types of data and new ways of aggregating the data dynamically to improve understanding and account for missing data.

In addition, since the system is dynamic with data coming from various sources, we will have comprehensive metadata labels on each data stream describing device model or streams that produced the data, the sampling rate, report rate, and the context under which the data was collected. When new devices are installed, a new stream is generated, metadata and capabilities are added into the system dynamically. We consider that some devices will be able to improve diagnostic accuracy than others (such as wearing a Zeo headband vs. accelerometer on the bed). Our system will support making alarms from deviations greater when coming from sources that are more accurate. New algorithms and standardized interfaces between evaluation modules will be created to solve the *extensibility* problem.

The inference chain will be run in parallel using two techniques, a rule-based strategy and a data-mining strategy. Under the rule-based approach, the domain expert writes a series of rules that are translated into evaluation routines. For example, medical experts will define scoring metrics for proper behaviors and create a point-based system for computing the corresponding factor. For instance, weight change factor is related to a percentage change of weight, this is defined a-priori and corresponds similarly to how current psychological instruments are created (HAM-D, CES-D, PSQI, etc). These rules can be population-general, or relative to a patient's particular lifestyle. For instance, many people, especially the elderly require only 6 hours of sleep a night, so the rules are defined based on the normal behavior.

The second technique using data mining to find patterns and anomalies that could characterize depression this addresses the inherent *uncertainty* problem with *automatic knowledge discovery*. We will explore the effectiveness of this strategy using an existing public corpus of data gathered from thousands of depressed patients. The STAR*D dataset, one of the largest and longest studies of depression, was used to assess the effectiveness of depression treatments in patients diagnosed with major depressive disorder, in both primary and specialty care settings. Over a seven-year period, the study enrolled 4,041 outpatients, ages 18-75 years, from 41 clinical sites around the country, which included both specialty care settings and primary medical care settings. We will evaluate the ability to infer depression from using the association rules learned from the dataset. There are several advantages to this technique. 1) rule-bases are hard to create by hand, and will often miss important information that can be trained automatically 2) large amounts of training data can consider options where the availability of data is scarce such as if the patient does not have a sleep monitoring solution, but instead has an activity monitor.

2.1.4 Evaluation Plan

We will evaluate the inference scheme using real data from the STAR*D dataset. The results from the instruments will be translated into data that the home monitoring solution would produce. Since the participants of the study were did not have instrumented homes, we do not have ground truth however from the Quality of Life Survey and other surveys we will make the best effort to simulate the data. The following experiments will be run using the STAR*D dataset, separating training data from evaluation data where appropriate (case 3).

1. **Single modality**: We will consider only using one modality to infer depression. For instance, we will only use the sleep quality index to infer the depression state, and not weight or activity levels. This will be the base case for current objective data collection.

2. **Rule-based strategy**: We will use only the evaluation modules that use the rule-based strategy to infer depression.

3. **Rule-based and data mining**: We will use the rule-based system with the automatically mined patterns to infer depression.

We expect to see more examples of depression recognized with a more accurate assessment of the severity (as dictated from the HAM-D) scores in the case 3 using the rule-bases and the association rules.

2.2 User interfaces

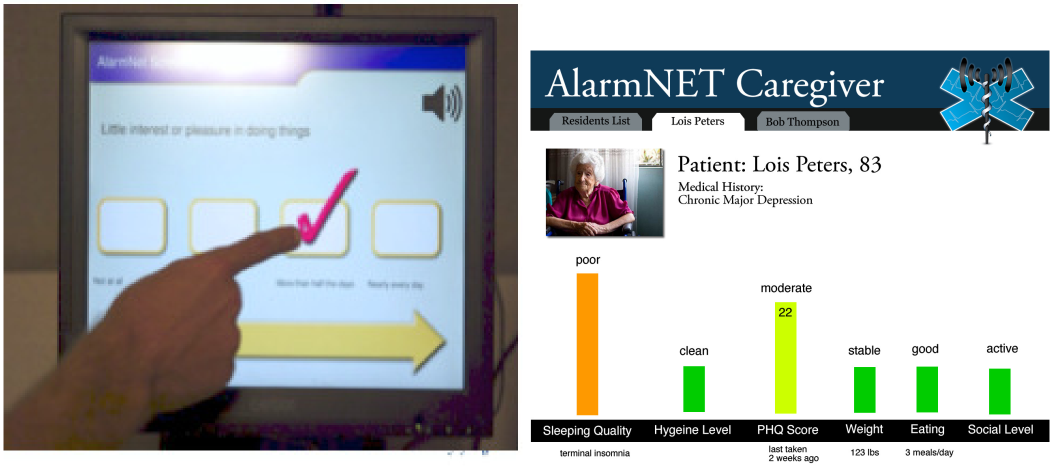
In order to properly track the effectiveness of a treatment to make diagnosis, a caregiver needs tools that can visualize important subsets of the data. There are challenges for creating these tools since the visualization needs to be appropriate for the type of stakeholder: clinicians such as therapists, assisted living care staff, and the patient can use.

- **Relevant**: the visualization of the data should be relevant to caregivers and improve their diagnosis. The presentation of data could be different for the type of caregiver such as nurses, resident staff, and doctors.

- **Adaptable**: the presentation of data should be rendered regardless of the modality. The type of data should define whether it is presented as a time-series plot, bar plot, or something specific. Anomalies and high risk parts of the data should be rendered so that experts will immediately recognize it.

- **Easy to Use**: The interface should provide the caregiver with intuitive and efficient ways to access the data.

- **Scalable**: The interface should allow a caregiver to manage large number of patients and/or search for patterns or symptoms under long periods of time.



(a) Touchscreen (b) Caregiver Display

Fig. 3. User interfaces created

2.2.1 State-of-the-Art

There has been several interfaces designed for patients concerned with their own health. The *Quantified Self* is a guiding principle for a variety of research in this domain, and posits that a person should be an active participant in managing their own health and lifestyle through self-experimentation and sensor devices. Adopters of this philosophy use devices to collect data about their exercise habits, diet, and vital signs such as blood pressure and resting heart rate to give the participant valuable feedback about their efforts to maintain a positive lifestyle. A few systems target emotional wellness: for instance, the Optimism App [17] is an application for both the desktop and mobile platforms that tracks instantaneous 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 [7]. Researchers at the Rhode Island Hospital have developed a telemedicine-based depression protocol using simple telemonitor in-home healthcare, with pilot studies showing that it could improve geriatric depression [25]. The subjects in the study were favorable to the technology, reporting that the frequent checks from the monitor were reassuring and helped them to better understand their condition.

For interfaces to present data to the caregiver, limited work has been done. A few companies such as WellAWARE system provide web interfaces into reports displayed on a web browser concerning sleep quality, activity levels, bathroom visits, and other physiological information.

2.2.2 Preliminary Development

We have created preliminary versions of both the patient interface and the caregiver interface. The current version of the patient interface shown in Figure 3(a) runs on a touchscreen and/or tablet placed in an accessible room inside the patients home. The personal behavioral factors are shown to the patient, providing continuous objective measurements for positive feedback. The touchscreen provides an interface for subjective data collection, for instance the self-report questionnaires CES-D and the PHQ-9 exam. In addition, the speech monitoring is elicited by free response questions from the device, and collected by the microphone.

The current caregiver's screen is shown in Figure 3(b). The caregiver's list of attending patients is presented with an overview of their depression risk factor. When a patient is selected, a summary of the current behavior factors: sleeping quality, social isolation, CES-D score, weight, movement levels, and speech analysis are presented as a bargraph. 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 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, annotations can be added to the display indicating when a patient started new therapy or medication so that the caregiver can monitor if the patient is improving. This system does not perform diagnosis, so exposing all the factors to the caregiver is useful for the caregiver.

2.2.3 Approach

We will use rigorous human factors principles to guide the design of the user interface. The principle investigator has had experience in research with human computer interaction and running user studies [12, 16, 22, 23].

We propose to enhance this interface by expanding the capabilities of the device to become a multi-purpose tool serving as a mental health trainer, social planner, and mood journal. Additional questionnaires such as the Pittsburgh Sleeping Quality Index (PSQI) can be administered after a detected night of insomnia. Additional user feedback will be investigated. A social planner will be integrated that coordinates a record of activities that are occurring in the patients assisted living care center, senior center, or other organization, and handles RSVPs and attendance records for the patient and also monitors the same for friends. Thus, it becomes a valuable input for measuring social involvement.

To guide the development of the next version of the caregiver display, and to ensure showing *relevant* data to caregivers, we have identified several medical, mental health, and caregiving professional with different backgrounds and roles to serve as experts elicit requirements and evaluate the interface:

- Seki Balogun, M.D. - Professor of Geriatrics, College of Medicine
- Bethany Teachman, Ph.D. - Professor of Clinical Psychology
- Karen Rose, Ph.D, R.N. - Professor of Nursing
- Sue Liberman, Director of Branchlands Assisted Living Care

Requirements elicitation and a series of iterations. Once a final version is implemented, a usability study will be performed on the interface from a series of tasks. Examples of these tasks include investigating the sleep quality, improvements after a treatment began, and others.

2.2.4 Evaluation

We will evaluate whether the caregiver interface 1) contains the correct data appropriate for depression diagnosis, 2) presents the data in a useful way for medical professionals doing diagnosis or tracking patient health, and 3) can scale to large number of modalities and patients.

Evaluation will consist of a series of user studies:

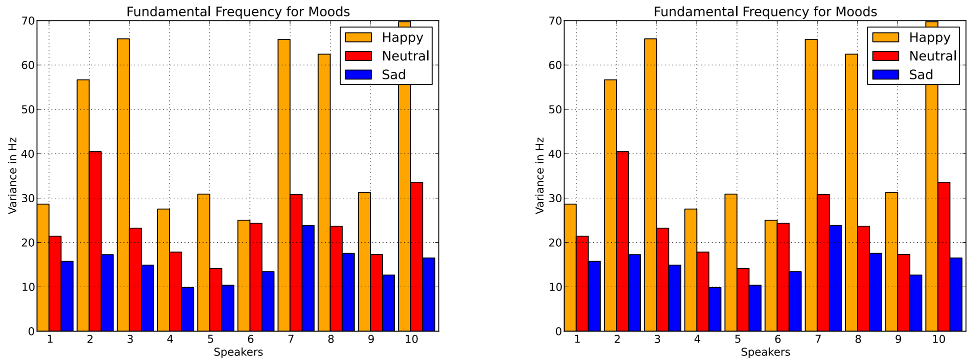
1. Wireless health community - we will demonstrate of the caregiver display to participants at the Wireless Health 2011 conference in October. A survey will be administered eliciting the background of the participant (medical or technical) and giving feedback on the usefulness of the tool.
2. Expert evaluation - we will have the above mentioned experts evaluate the interface for *ease-of-use* and scalability generating simulated data from the STAR*D dataset, and having them recognize instances of high risk of depression.

2.3 Emplaced speech affect monitoring

Acoustic features of voice such as pitch, utterance duration, and pause-time have been shown in previous studies to detect the severity of depression [1,5,10,20]. 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. Our challenge will be to implement these solutions to work at real-time in natural home settings.

Emplaced acoustic processing is a challenging problem which requires research in the following:

- **Vocal discrimination**: Patients will often be speaking only in conversation, and therefore segmenting the patient from others.
- **Large Coverage**: Full coverage of a living space will require multiple microphones, we will investigate the how to network the devices and preprocessing the data enough to extract necessary features.



(a) Speech Pause Time (b) Fundamental Frequency

Fig. 4. Speech features analysis from the EmoDb database

- **Ambient noise canceling**: In order for feature extraction, there must be preprocessing to cancel out background noise especially from appliances such as TVs and music.
- **Unsupervised**: Correlating affect with speech features often require training. We will investigate ways to automate this process using other modalities, both subjective and objective.

2.3.1 State-of-the-Art

Closely related to speech emotion prediction, there has been work in environmental compensation for speech recognition [19] in particular for mobile devices. Two techniques are used. 1) Cepstral Mean Normalization (CNN) is the simplest in that it computes a long-term mean value of the feature vectors and subtracts this mean value from each of the vectors. The effectiveness of CNN is limited when the environment in not adequately modeled by a linear channel. 2) Data-driven compensation methods are based on the fact that the effect on the environment on cepstra and log spectra of clean speech feature vectors can be modeled by additive correction factors. The correction factors are computed by examples of how clean speech vectors are affected by the environment.

...

2.3.2 Previous work

In the current implementation, the client PC with the touch-screen is also connected to a microphone. We use the SoX toolkit for initiating audio recordings and doing post-processing. 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. Pitch contours are generated from the raw data using a pitch detection algorithm [18] implemented in the Edinburgh Speech Tools (EST) Library. It is important to note that determining the fundamental frequency from speech is not as simple as taking a FFT, since determining the pitch requires estimating the missing fundamental. Human perception of pitch is more determined by the ratio of the ascending harmonics. Male voices fall within 60-200 Hz and 120-400 Hz for females. The standard deviation on the contour is used to infer the amount of vocal inflection. The speech pause time, the silent interval between phonations, were determined by the duration of the silence between successive pitch contours. Large gaps (greater than 1 s) were not used in the calculation. Once the statistics statistics are computed they are sent from the client to the server. The original file can be kept on the client PC or deleted if privacy is a concern.

We evaluated preliminary results of using these two features. 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 to determine their naturalness. There were five male and five female speakers, and each said the same ten different utterances with varied emotions. We consider the data labelled as happiness, sadness, 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.

	Estimate	Std. Error	t value	$\Pr(>\lvert t\rvert)$
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

Tab. 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.

The modulation of fundamental frequencies for neutral, sad, and happy data for each of the ten speakers are plotted in Figure 4(b). Each of the speakers showed a decrease in the level of modulation in their voice as the affect went from happy to sad. 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 in 4(a). 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.

There is a challenge to getting a *speech factor* score since the relationship between mood and speech seem to be dependent on the particular person. Our first implementation used multi-variate linear regression to fit a function of a patient's self reported mood to the speech features. For two weeks data on one person, there is preliminary evidence shown in Table 1 that there is linear relationship between the variables and mood that can be trained.

2.3.3 Approach

We will develop new or extend existing *vocal discrimination* and segmentation algorithms. Since many homes are mixed gender (husband and wife), we feel that certain assumptions if validated, could make the segmentation problem easier. We will use many speech features such as fundamental frequency, mel-cepstrum coefficients, and characteristics from certain phonemes (glottals) that are typically unique across speakers.

Ambient noise will be filtered using traditional signal processing approaches such as bandpass filters and FIR filters. In addition, we will use plug sensors to recognize when a TV is on.

We will to extend this work using a new *unsupervised* approach to automatically infer days of sadness and happiness, so that the patient does not need to perform training when the system is installed. In particular, accurate training samples is hard to obtain artificially from a speaker. We will cluster the historical data across time periods. Next, we recognize isolated clusters, and infer labels onto the cluster. Next we will automatically label incoming data by pairing it to the closest related cluster. More algorithms and approaches are expected to be developed.

2.3.4 Evaluation

We will run a series of experiments using data from the ground truth data from the EmoDb dataset. We will modify the data to simulate using the system in a real home.

1. Unsupervised approach: We will compare the unsupervised approach to the supervised approaches (linear regression, naive bayesian, SVMs) and present the results.

2. Ambient noise: We will add to the signal noise from an apartment (television, music) and evaluate how it impacts the classification task. Next, we will use the ambient noise filtering and present the results.

3. Multiple speakers: We will evaluate how the system can handle multiple people in the environment talking (one at a time). The segmentation algorithm will be evaluated by its accuracy at determining the speaker of interest.

4. Device-limited devices: We will down-sample the signal into compressible forms, and evaluate the classification accuracy.

2.4 Sleep monitoring

A number of studies have found that depression results in disruptions in sleep patterns both behaviorally and using electroencephalogram (EEG) recordings. Very prevalent are sleep disturbances such as sleep onset insomnia, mid-nocturnal insomnia, early morning insomnia, or even hypersomnia. Three sleep pattern abnormalities have been well documented in depressed patients [8]. They include 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 show the importance of the sleep modality.

Many challenges exist in particular, how to ascertain sleep quality from accelerometer data:

- **Restlessness**: Relationship between movements and sleep quality
- **Detecting anomalies**: sleep duration, number of interruptions, and other features are specific to the people.
- **Cost**: Creating a monitoring device that is less than $100 for the bed since alternatives such as Actigraph, Zeo, and others are both costly and invasive.

2.4.1 State-of-the-Art

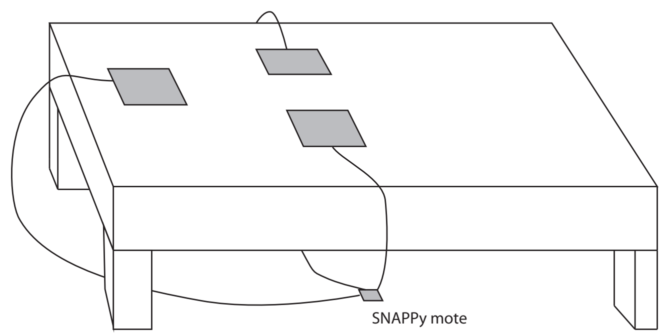


Fig. 5. The sleeping monitoring setup. Three tri-axis accelerometers are taped onto the mattress and connected to a wireless Synapse mote.

Self-report questionnaires such as the Pittsburgh Sleep Quality Index (PSQI) [4], are answered by patients every two weeks, however, studies have shown that subjects with insomnia are not accurate in their subjective report of variables such as sleep latency, sleep duration and number of disturbances and overall tend to underestimate their ability to sleep. There is a need for objective instruments that can measure sleep quality where the subjective ratings fail. Because of its importance, many sleep-monitoring systems have been developed. These systems attempt to recognize sleeping disorders by providing healthcare providers with quantitative data about irregularity in sleeping periods and durations or the amount of agitation and restlessness experienced during the night. These solutions vary in cost, comfort, and accuracy. However to date, there are few low-cost and unobtrusive sleep monitoring systems. The most accurate are polysomnography devices, but there are major drawbacks to using them since they need to be worn, and require professional monitoring and thus expensive to use. The actigraph is an accelerometer device that can be attached to any of the limbs (e.g. wrist) to provide data on movement, however they still need to be worn. Pressure pads such as the bladder can be used, but they are uncomfortable and moderately expensive. The Zeo is a headband that measures electrical signals on the scalp to estimate the stage in sleep, however they need to be worn and their accuracy has not been evaluated.

2.4.2 Approach

Solutions for providing sleep quality are still invasive or costly. The WISP receiver costs over $600 each. We built a custom solution using a Synapse SNAPpy RF motes for wirelessly transmitting data. We attached three independent tri-axis accelerometers to the mote and they are sampled at 1 Hz as shown in Figure 5. Data is processed on the client PC which determines the amount of deflection since the last sample the side of the mattress when a person's weight is applied. Since the accelerometers indicate the direction of the force due to gravity as a vector, we take the dot product of the last sampled vector and new to determine the amount of deflection since the last sampling. If that deflection exceeds a threshold, 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 measure 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, also since we only store events and not raw measurements, we eliminate noise and lower the storage requirements and increased scalability.

The next stage of the sleep analysis is to convert the discrete bed movements to sleeping segments. The process is shown in Figure 6. 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 statically defined, 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 (refrigerator, bathroom, etc). The assumption that if no other sensors in other locations fire, you can assume the patient is still on the bed. We created a scoring system for calculating the sleeping factor. The parameters were chosen to match similar components used on the PSQI exam [4].

To make judgements on sleep irregularity, historical information about the patient's typical sleeping behaviors will be considered. For instance, if the patient typically sleeps 5 hours each night, and no other symptoms are expressed, the system gives no alert to poor sleeping quality. But if the typical sleeping duration is 8 hours, and reduces down to 5 hours a poor night's rest is alerted.

2.4.3 Evaluation Plan

We will evaluate our sleep monitoring solution by running a user study with ten people with known normal and abnormal behaviors. Each will sleep on the bed for a period of a week or more. We will have the participant complete subjective measurement of the restfulness after each night, to investigate the correlation between movement levels, interruptions, and the subject's perception of a good nights' rest.

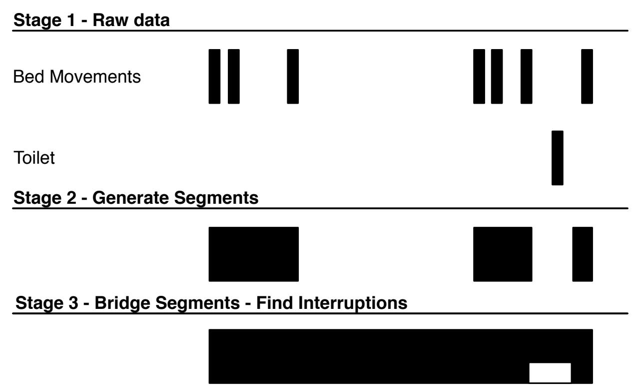


Fig. 6. Process of extracting sleeping periods and interruptions from the movement data on the mattress.

2.5 Scalability

Smart-homes generate large amount of data on a daily basis, as future sensing devices are integrated into the home, we see that there are scalability problems both in storing the data, and processing the data so that it is conditioned for knowledge-mining, and that the presentation of the data to the caregiver is appropriate given lots of information. Storage of data from multiple-deployments might exist on a cloud, lowered cost of data warehousing.

Some challenges include:

- **Access-time**: Optimizing access time and indexing into the database, giving priority for more at-risk cases, and predicting types of queries needed by the caregiver or the inference logic.
- **Heterogeneity**: the data formats vary based on the sensing modality: high-data rate periodic to low-data rate aperiodic.

2.5.1 State-of-the-Art

Handling scale of data is handled in two manners. First is by managing the amount of data that the backend database needs to store by filtering. For sensor networks, the VANGO system [6] was created to handle capturing and transmitting raw high-frequency data such as human voices in real time.

On the database, different strategies are used. Typically for medical applications, since scaling has not been considered the data streams are placed in a relational database table with a timestamp, value, and other corresponding metadata. Time series databases are typically implemented by storing the values as a binary large object (BLOBs) in a relational database. Other databases such as Hadoop is made to support petabytes of data across thousands of nodes. Scaling and performance is improved by the assumption that strict data coherence model is not necessary, since typical operations a write once, read many times.

2.5.2 Previous Work

The infrastructure for transporting data from clients to servers for high data-rate sensor stream data on the World Wide Web is challenging. RFIDs, cell phones, and sensor nodes produces streams of sensor data that help computers monitor, react to, and affect the changing status of the physical world. We presented a new web primitive called stream feeds that extend traditional XML feeds such as blogs and Podcasts to accommodate the large size, high frequency, and real-time nature of sensor streams. We demonstrated that our extensions improve the scalability and efficiency over the traditional model for Web feeds such as blogs and Podcasts, particularly when they are used for in-network data fusion. The main contribution was that high data rate, aperiodic, real-time data streams could become first-class citizens on the web.

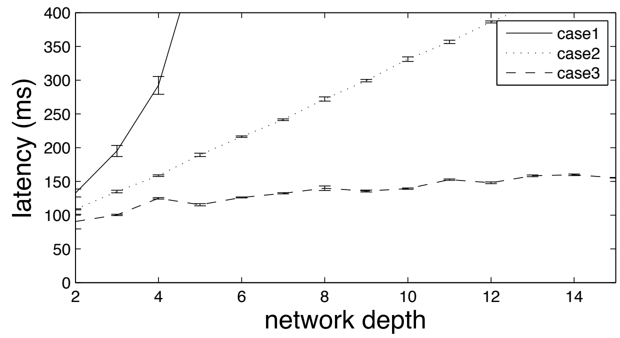


Fig. 7. Evaluating the performance of pulling only (case 1), pushing and pulling (case 2), and pushing with filtering (case 3). As the network gets deeper, filtering with the ability to push outperforms the alternatives.

Stream feeds provide the abstraction of a dynamic content object containing a stream of sensor data. The data object can be located with a URL, can be linked to other data objects, and can be accessed over the HTTP protocol. When the client requests a URL, the client opens a TCP socket to the server and sends a HTTP GET request, as usual. The server responds immediately with all the data that is already contained in the sensor stream, followed by updates about the sensor stream in real-time as new data objects are added to the stream. The user can also filter the sensor stream by passing filter parameters for lower bounds and upper bounds. This type of query restricts the response to only contain the values in the data stream that have a value attribute in this range. The values that can be filtered depending on the XML schema that the sensor stream is using, so the client must be familiar with this schema. The values in the sensor stream are then filtered using syntactic filtering of the data points based on the XML tag names and values. This can be performed automatically for any XML schema. The URL-based stream feed interface conforms to REST (Representational State Transfer) principles describing how resources are defined and addressed. The advantage to using the RESTful interface is that there is an inherent standardization of the operations that can be applied to the resources, without needing to explicitly define descriptions of the methods that can be applied to them using the web services description language (WSDL). The URL contains all the information needed to return to a particular state of a web service. By including the name of the sensor, the temporal scope, and the filter parameters all in the URL, stream feeds make it easy to link to stream feeds, or even to link to partial stream feeds.

We evaluated the performance of the streamfeeds with pushing and pulling support as well as filtering through large-scale simulation. The results are shown in Figure 7. From the experiment, we show the advantages of using filtering on the server, being able to use a publish-subscribe push system for inference chains. We will use this insight in the scaling solution of this thesis. Each component has an exposed web service API, and the control knobs can be set from SOAP protocols various methods and capabilities exposed through the WSDL interface. Each platform has different capabilities such as compression methods and archival support, and these information can be made available to scalability controllers on different levels.

2.5.3 Approach

We will address the scalability of the system in terms of bandwidth and storage by a 2-way cross-layer semantics driven filtering and compression scheme. The novelty is that the filtering and compression mechanisms are dynamically driven, and decisions can be made at runtime. The the operating parameters of the system based on following context such as whether the patient sick or not sick, activities (sleeping, exercising, eating), irregular patterns, the type of data being received, and network congestion.

The key to the solution for cross-layer filtering and compression, is that each module in the system have the following generic "control knobs":

- Binary (turn on/off)
- Only on event or event threshold
- Adjust rate

- Drop data (e.g., all normal readings)
- Use patterns and send codes
- Aggregate: average, repetition
- Compression (loss-less or lossy)

One of the problems with assessing the performance of the filtering and compression is whether it is done correctly. Usually compression techniques consider perfect signal reconstruction, often measured by PRD. This is not sufficient since strategies that remove noise or unimportant information should not be penalized for producing a result that is sufficient for the application that conforms to the Quality of Service (QoS) policy.

2.5.4 Experiments

We will evaluate the system with the scalability through simulation of data outputs from thousands of homes. This will require creating a sensor data generator which is similar to a simulator the principle investigator presented at the cyberphysical systems retreat in 2009. The metrics will be judged in terms of the compression-ratio (storage), fidelity (error-loss), and execution-time (for the compression and reconstruction time).

2.6 Contributions

This proposed research and thesis will investigate the above challenges and result in the following contributions:

- An extensible multi-modal largely passive depression monitoring system useful to caregivers (primary care physicians, psychologists, etc.) to monitor their patients' behavior (in therapy, free homes, etc.) and thereby, track their response to treatment, therapies and their general well being.
- New interfaces validated by user studies for presenting to the caregiver useful data, easily, and can scale to large number of people.
- Advancements in emplaced speech emotion monitoring, such as speaker discrimination, ambient noise subtraction, and new features and classifiers to improve emotion prediction in real environments.
- A passive and cheap sleep monitoring solution that can collect useful features such as insomnia, restlessness, and sleeping patterns that can be used to infer sleep quality.
- Solutions for handling the storage and bandwidth needs of multiple deployments over long periods of time, including dynamic cross-layer filtering and compression to provide for fast query times using optimal space.

To our knowledge, there has been no system yet that has been implemented to provide continuous emotional monitoring in the home by combining multimodal inputs from sleep, body weight, activities of daily living, and subjective factors. For various subcomponents of the system, such as voice and sleep monitoring, many advancements to the core technology must be made to make them appropriate in real deployments. Although there are daily subjective 'mood diaries' that run are on the web or smartphones are effective at mitigating self-reflection errors, they do not incorporate enough objective behavioral factors that could be useful for caregivers and patients monitoring their own health.

3 Research Plan

We have identified a number of papers to be produced and published as part of this proposed dissertation. Papers currently published are as follows:

1. R. Dickerson, J. Lu, J. Lu, and K. Whitehouse. Stream Feeds - An Abstraction for the World Wide Sensor Web. Conference on the Internet of Things (IOT'08). Zurich, Switzerland, March 2008. (25
2. R. Dickerson, J. Lu, J. Li, B. Chantree, J. Lu, J. Stankovic, K. Whitehouse. MetroNet: Case Study for Collaborative Data Sharing on the World Wide Web. Demo The Seventh International Conference on Information Processing in Sensor Networks, 2007 (IPSN '07).

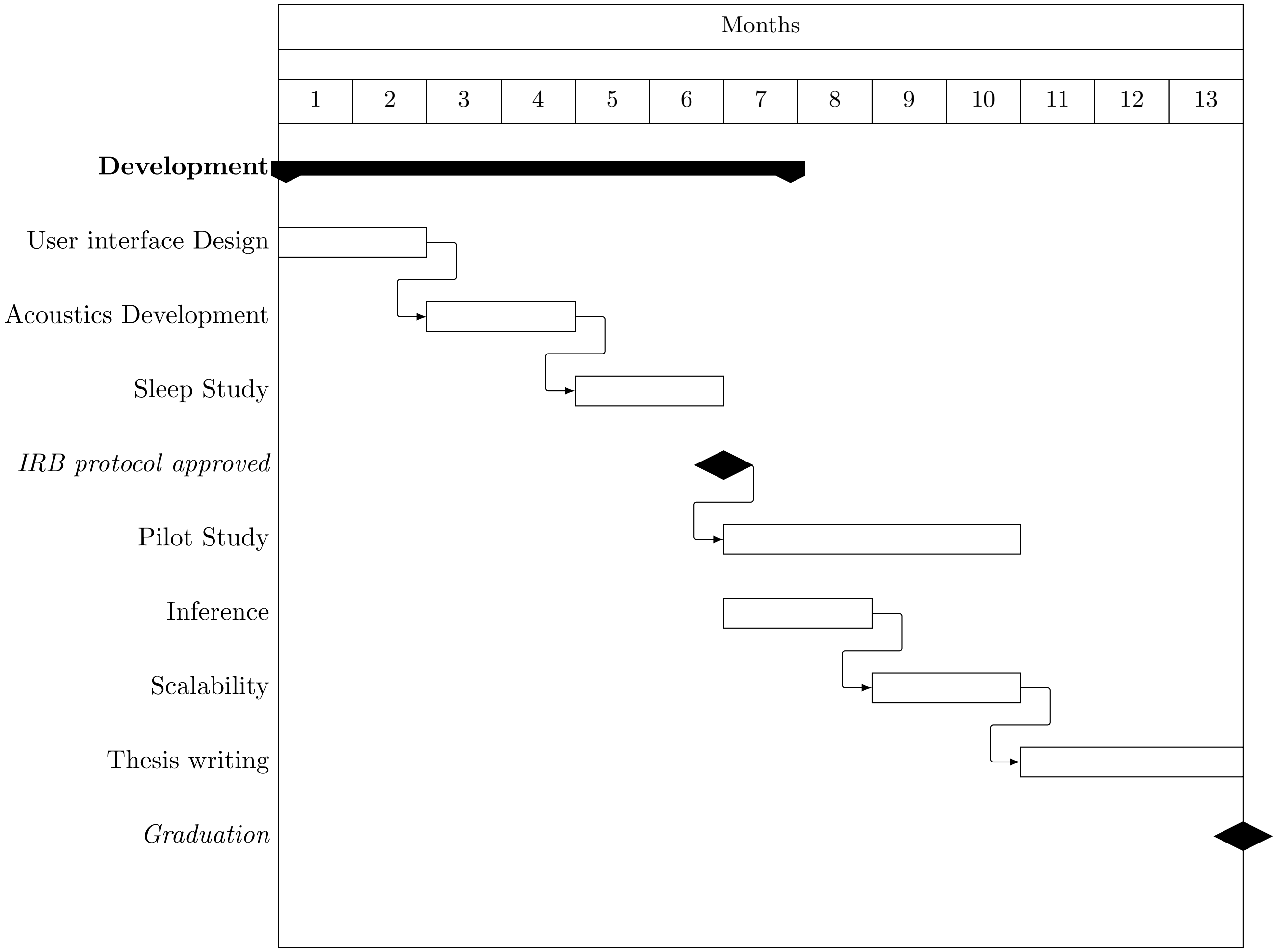


Fig. 8. Thesis research and development plan prior to graduation

3. E. Hoque, R. Dickerson, and J. Stankovic. Monitoring Body Positions and Movements During Sleep Using WISPs. In Wireless Health 2010, pages 1-10, August 2010.

4. R. Dickerson, G. Gorlin, and J. Stankovic. Empath: A Continuous Emotional Health Monitoring System for Depressive Illness. In Wireless Health 2011, October 2011.

5. E. Hoque, R. Dickerson, and J. Stankovic Monitoring Sleep with WISP Tags. In Security and Trends in Wireless Identification and Sensing Platform Tags: Advancements in RFID

3.1 Timeline

The expected timeline is shown in Figure 8. Development and evaluation of the subcomponents are first, while the IRB protocol is reviewed. Upon approval and enlistment of 3-4 patients with a history of depression, a pilot study will be run with replicated versions of the system. Finally, work at a system-level will be completed such as the inference and scalability. The last three months of the summer 2012 will be used for thesis writing, with an intended graduation date of August 2012.

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] B. Greenstein, C. Mar, A. Pesterev, S. Farshchi, E. Kohler, J. Judy, and D. Estrin. Capturing high-frequency phenomena using a bandwidth-limited sensor network. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, SenSys '06, pages 279–292, New York, NY, USA, 2006. ACM.

[7] 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.

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

[9] 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.

[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] 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.

[12] K. Johnsen, R. Dickerson, A. Raij, B. Lok, J. Jackson, M. Shin, J. Hernandez, A. Stevens, and D. S. Lind. Experiences in using immersive virtual characters to educate medical communication skills. In *Proceedings of the 2005 IEEE Conference 2005 on Virtual Reality*, VR '05, pages 179–186, 324, Washington, DC, USA, 2005. IEEE Computer Society.

[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] B. Lok, R. E. Ferdig, A. Raij, K. Johnsen, R. Dickerson, J. Coutts, A. Stevens, and D. S. Lind. Applying virtual reality in medical communication education: current findings and potential teaching and learning benefits of immersive virtual patients. *Virtual Real.*, 10:185–195, November 2006.

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

[18] 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.

[19] P. Moreno. *Speech Recognition in Noisy Environments*. PhD thesis, Carnegie Mellon University, 1996.

[20] 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.

[21] 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.

[22] A. Raij, K. Johnsen, R. Dickerson, B. Lok, M. Cohen, T. Bernard, C. Oxendine, P. Wagner, and D. S. Lind. Interpersonal scenarios: Virtual real? In *Proceedings of the IEEE conference on Virtual Reality*, VR '06, pages 59–66, Washington, DC, USA, 2006. IEEE Computer Society.

[23] A. B. Raij, K. Johnsen, R. F. Dickerson, B. C. Lok, M. S. Cohen, M. Duerson, R. R. Pauly, A. O. Stevens, P. Wagner, and D. S. Lind. Comparing interpersonal interactions with a virtual human to those with a real human. *IEEE Transactions on Visualization and Computer Graphics*, 13:443–457, May 2007.

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

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

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

[27] 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.

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

[29] 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.

[30] 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.