# Passive and In-situ Assessment of Mental and Physical Well-being using Mobile Sensors

**Mashfiqui Rabbi***
Dept. of Information Science
Cornell University
ms2749@cornell.edu

**Tanzeem Choudhury***
Dept. of Information Science
Cornell University
tanzeem.choudhury@cornell.edu

**Shahid Ali**
Community and Family Medicine
Dartmouth Medical School
shahid.a.ali@dartmouth.edu

**Ethan Berke**
Community and Family Medicine
Dartmouth Medical School
ethan.berke@tdi.dartmouth.edu

## ABSTRACT
The idea of continuously monitoring well-being using mobile-sensing systems is gaining popularity. In-situ measurement of human behavior has the potential to overcome the short comings of gold-standard surveys that have been used for decades by the medical community. However, current sensing systems have mainly focused on tracking physical health; some have approximated aspects of mental health based on proximity measurements but have not been compared against medically accepted screening instruments. In this paper, we show the feasibility of a multi-modal mobile sensing system to simultaneously assess mental and physical health. By continuously capturing fine-grained motion and privacy-sensitive audio data, we are able to derive different metrics that reflect the results of commonly used surveys for assessing well-being by the medical community. In addition, we present a case study that highlights how errors in assessment due to the subjective nature of the responses could potentially be avoided by continuous mobile sensing.

### Author Keywords
mental health, physical health, activity inference, mobile sensing, machine learning.

### ACM Classification Keywords
H.1.2 User/Machine Systems; I.5 Pattern Recognition; J.3 Life and Medical Sciences.

### General Terms
Algorithms, Experimentation.

## INTRODUCTION
One of the pillars of population health is to improve overall quality of life by promoting cognitive, physical and mental well-being [1, 2]. Everyday behaviors are often reflective of physical and mental health and can be predictive of future health problems. The standard practice for collecting behavioral data in the health sciences relies on observational data collected in laboratory settings or through periodic recall surveys or self-reports. These proxy measures have several limitations: (i) the time and resource requirements are too high to simultaneously gather data from a large number of individuals; (ii) the measurements are prone to considerable bias and the manual and sporadic recording of information often fails to capture the finer details of behavior that may be important; and (iii) the end user effort is too high to be suitable for continuous long-term monitoring.

With continued rises in medical costs, the need for a model that screens and facilitates early diagnosis, as well as increased efforts in prevention, is an essential concern for healthcare providers and administrators. Consequently, a growing number of studies are demonstrating potential of behavior monitoring devices to assist in one or more of the three clinical applications mentioned above [3, 4, 5]. The ultimate vision is to develop a mobile sensing system that can contribute significantly to cognitive, physical and mental well-being while maintaining easy and universal applicability, security and patient privacy protection, and low cost.

## BACKGROUND
Monitoring physical activity and mental health has been extensively investigated in the past via a variety of traditional recall surveys or Ecological Momentary Assessments (EMA) [6]. Paper-based surveys like Yale Physical Activity Survey (YPAS) [7], SF-36 [8], and Center for Epidemiological Studies - Depression (CES-D) [9] are examples of commonly accepted surveys and are some of the primary metrics for assessing physical and mental well-being in medicine. These paper-based surveys utilize recall techniques to capture daily, weekly, and seasonal patterns of behavior, but may require in-person administration and are limited by recall bias, memory dependence, current mood, and their obtrusive nature [10, 11]. Furthermore, answers to the paper-based survey questions are subjective, with risk of social de-


*UbiComp'11*, September 17–21, 2011, Beijing, China.


* The work presented here was done while both the authors were in the Computer Science department at Dartmouth College.

sirability bias, and sometimes suffer from issues like "back-filling" due to non-adherence to time-sensitive protocols (i.e. completing daily surveys at end of the week [6]). Electronic survey tools such as digital diaries, smart-phone and web-based surveys allow investigators to tailor and improve survey questions dynamically, reduce recall bias by collecting responses close in time to one's experience, and avoid backfilling by digitally time-stamping the submitted surveys. However, recall surveys and EMA, both paper-based and electronic, do not capture activities continuously and rely on the subjects to be responsive, can be cumbersome, and may require periodic re-administration, thereby hindering their consistent use in the primary care setting.

More recently, a variety of sensor-based systems have emerged with great potential of circumventing recall limitation and subjective dependence of aforementioned techniques. These sensor-based systems differ in their mobility, placements, numbers, and continuity in data capture, and have their own advantages and limitations. For example, video camera and RFID equipped rooms have been used to capture physical activity and sleep patterns for users [12]. These systems usually capture a portion of user's daily life and can be prohibitively costly for mass use. In contrast, wearable single-sensor devices such as pedometers and monoaxial accelerometers have been used to measure physical activity continuously and unobtrusively while observing subjects in their natural environment [13, 14, 15]. Although useful, single-sensor based systems are limited in the types of activities they can capture. To overcome this limitation, investigators have used a combination of multiple sensors placed on different locations of the body to learn more about subject behavior [16]. These systems can capture a richer set of activities and can yield higher accuracy in recognizing activities. Another approach investigators have explored is the use of single mobile devices equipped with multiple types of sensors. Devices such as Actigraph [17], Sensewear [18], the Mobile Sensing Platform (MSP) [19] and LifeShirt system (Vivometrics, Ventura, CA) [20] provide versatility by detecting different modalities ranging from light and directional acceleration to physiological signals. These devices and new sensor-equipped smartphones have been used in a wide range of applications ranging from measurement of physical activity and energy expenditure to providing context-aware instantaneous feedback with goal of preventing future health complications. Furthermore, a network of wearable sensors communicating with a central server has demonstrated the potential benefits of monitoring and assisting in the care of the aged, a rapidly growing subset of population [21].

In contrast, the use of automated sensing techniques in assessing mental health has been very limited. Mental illness costs $30.7 billion/year [22] in the US alone. Depression, a common psychiatric disorder, has a prevalence of 5-25% [23, 24] and contributes significantly to the cost of healthcare [25, 26]. Previous research has shown that increased social activity correlates negatively with depressive behavior and can even improve depressive symptoms [27, 28]. Others have demonstrated using subjective methods that physical activity delays cognitive decline [29]. Some aspects of mental health have been approximated using proximity measurements [30]. We feel proximity based approach has three potential limitations: (i) its inability to detect reduction in physical activity and speech limits availability of useful information for clinicians to treate and manage mental dysfunction, (ii) interactions inferred from speech and colocation have been shown to be different [31], thus it is possible that behavioral indicators of social isolation may be manifesting itself but not being detected by colocation, and (iii) the system has not been compared against medically approved screening techniques validated in the medical literature and used in practice. Nonetheless, purely proximity based approaches may be a useful tool to screen for mental health risk but we believe a detailed measure of *behavior* will be more useful to practitioners for diagnosis purposes and is a worthy line of investigation. There is also a growing body of evidence supporting the use of acoustical properties of speech to detect changes in emotional health [32], variations in mood [33, 34] and periods of stress [35]. Furthermore, voice analysis can detect changes in verbal initiation (difficulty in starting sentences) and perseveration, characteristics highly associated with changes in mental health and inability to perform Instrumental Activities of Daily Living [36]. We believe these results, together with the fact that face-to-face conversations, are still perceived as a common and important medium for social communication [37], make the continuous study of human speech in people's daily life a valuable line of investigation for mental health assessment.

## CONTRIBUTIONS

In this paper we present an automated system for measuring mental well-being from behavioral indicators in natural everyday settings, in addition to measuring physical well-being. We describe the sensing and activity recognition system and its deployment in a real-world setting of older adults. We assess its reliability, its feasibility in this population, and its validity with established health instruments.

We hypothesized that continuous recording of daily audio patterns, specifically relating to the amount of human speech, would be linked to social and mental well-being. To address the challenges of survey implementation and test our hypothesis, we used a previously developed mobile multimodal sensor platform for continuous and objective evaluation of mental health in a small group of elders. Specific contributions include:

- Automatic analysis of the amount of speech occurring in natural settings while avoiding raw audio collection for privacy protection. Correlation between the total amount of human speech sensed and the well-established paper-based surveys for mental health like CES-D score, SF-36, and Friendship Scale, approached statistical significance for CES-D and showed statistically significant correlation for SF-36 Mental Health Component and Friendship Scale.

- In addition to mental health, an overall physical activity score was computed, based on a weighted count of sim-

ple physical activity categories including stationary (sitting/standing), walking on flat surface, walking up and down inclines. This physical activity score showed significant correlation with Yale Physical Activity Survey (YPAS), a survey commonly used by the medical community.

- Presentation of a case study that illustrates how automated sensing could potentially lead to improved screening of mental health disorders and can overcome some of the limitations of current paper-based surveys. Specifically, in a subject whose CES-D scores were well below those that would indicate depressive symptoms, the sensor-based measurements were in closer agreement with observational assessment by the physician and the medical trainee on the research team.

## STUDY OVERVIEW: ASSESSING MENTAL AND PHYSICAL WELL-BEING

Fine-grained sensor data were collected continuously from a group of older adults living in a continuing care retirement community. Located at the bottom of a hill, the retirement community includes a variety of facilities including a dining center, a library, group meeting rooms, and a fitness center. Apartment buildings surround the community center and are situated at a relatively higher elevation. Sloping roads connect the community center to the apartments.

Letters were sent to resident mailboxes and posted on major announcement boards across the retirement community requesting unpaid volunteers to participate in a study on automatic sensing of mental and physical well-being. Interested residents were interviewed by the team physician to establish availability for the entirety of pilot, comfort level with electronic devices, and were selected on a first come basis as a convenience sample. Eight self-selected adults were recruited: 4 singles and 2 couples; 50% male and female. The length of the pilot study was 10 consecutive days in August 2009 from 7am to 7pm; inter-individual variability was roughly $\pm 2$ hours for pickup and drop off times. One participant dropped from the study one day before the start of study due to personal reasons unrelated to the study protocol. Investigators found a replacement but lost one day of observation due to recruitment delay for that subject. All participants were informed of their right to leave the study at any time and briefed on the Institutional Review Board (IRB) approved pilot design, its goals, and the extent of use of the collected data. All participants accepted the terms and conditions by signing a written informed consent form during their face-to-face time with project investigators. Four surveys (CES-D, YPAS, SF-36, and Friendship Scale) [9, 7, 8, 38] were administered before and after the pilot and a usability survey was administered post-study only to evaluate participants' experience with the device, surveys, and to receive comments and concerns about the technology's potential and privacy concerns.

During data collection, each participant wore a 2-inch waist-mounted device equipped with multiple sensors: 3-axis accelerometer, barometer, light sensors, temperature sensor, humidity sensor, compass and microphone. Each device was equipped with a clip so the subject could wear it comfortably around the waist. Participants were trained one day prior to the start of the pilot on basic usage of the device. To tackle scenarios where the participants may feel uncomfortable to record, they were shown how to turn on and off the device and where to contact the medical trainee, who was always present on-site during the study. Also, acceptable ways of wearing the device were demonstrated. Participants were instructed to avoid placing the device in pockets and wet environments as the device is not waterproof. Participants received no special instruction on changing their daily routine, and were allowed to travel outside the facility with the device. At the start and end of the protocol (i.e., on two different days), participants completed a series of activities (short walk, walking up and down inclines) and engaged in casual conversation in different parts of the retirement community that amounted to approximately 80 minutes of total data (8-15 minutes of data for each subject). This data is later used to compare the recognition accuracy with previous published results [39, 40, 41, 42]. Devices were collected daily every evening, data was extracted, and batteries were recharged. Participants were asked to occasionally monitor the recording LED and were provided instructions to seek assistance from on-site support or contact one of the principal investigators (a computer scientist and a family medicine physician) in case there was an issue or if device stopped recording.

**Figure 1: Mobile sensing device**

Raw audio was not recorded because of privacy reasons (a detailed description of the recorded audio features are in Overview of Speech and Conversation Detection section). Although the recorded features do not allow reconstruction of audio afterwards, they enabled us to infer when human voice was present and whether there was conversation. This information is also sufficient for estimating who was speaking when, and how a specific subject was speaking (energy and pitch) [41, 43]. In this context, it is worth mentioning that during the study we learned that the privacy sensitive audio data collection was very well accepted by users, since there was no incident of a participant turning off the device or seeking assistance for turning off the device due to privacy concerns. It can be argued that we could have recorded raw audio for further analysis given such a pilot study. But we want to argue that privacy sensitive features enable us to collect data in realistic environments continuously in unobtrusive way. Furthermore, it is still possible to do interesting analysis related to health and mental well-being using non-verbal aspects of speech (e.g., pitch, speaking rate, loudness) and are supported by previous research [44, 45, 46]. In addition, the Health Insurance Portability and Ac-

countability Act (HIPAA) provides strict guidelines on protecting personal health information. Since this act prevents access and use of health information without explicit consent of all patients, recording of raw health information becomes extremely challenging, since it is almost impossible to control when unconsented individuals in the background gets recorded. Even for our experiments, the management authority of the retirement community would not have approved continuous recording of audio. Although focused recording in pre-specified rooms may have been a possibility, it defeats the goal of our study, which aims to assess mental health from passive sensing of everyday speech in naturalistic conditions.

| Participant ID | Marital Status | Age (years) |
|---|---|---|
| SU01 | Single | 90 |
| SU02 | Single | 89 |
| SU03 | Married | 83 |
| SU04 | Married | 82 |
| SU05 | Single | 93 |
| SU06 | Single | 81 |
| SU07 | Married | 84 |
| SU08 | Married | 86 |

**Table 1: Subject profile**

### Overview of Speech and Conversation Detection

Audio is processed on-the-fly to extract and record features that are informative for inferring the presence and style of speech and conversations but not enough to reconstruct the words that are spoken. We utilize the privacy-sensitive speech processing methods developed in [47, 42]. Features that are recorded include: (i) non-initial maximum auto-correlation peak, (ii) the total number of auto-correlation peaks, and (iii) relative spectral entropy – these features have been shown to be particularly useful for detecting the structure of voiced speech and are more robust than energy based methods.

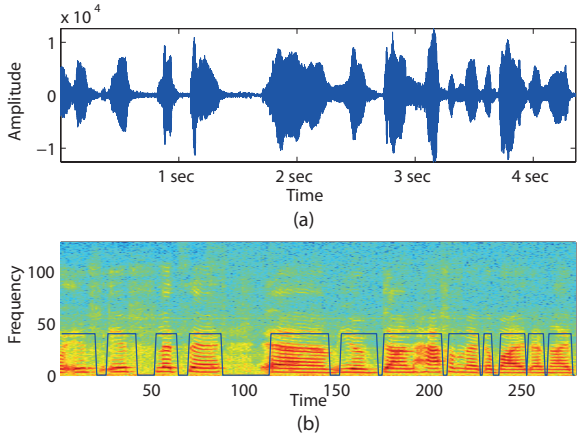


**Figure 2: (a) amplitude values for 5 second recording of audio (b) spectrogram of the recording with blue lines depicting inferred human speech**

The first step in the inference pipeline involves finding audio segments that contain human voice. A two-state hidden Markov model (HMM) is used to classify speech vs. non-speech segments using the aforementioned three recorded features. For each hidden state, the observation probability is modeled using a multi-variate Gaussian distribution (for a detailed description of the classification approach please see [47]). Upon detection of human voice, mutual information between voicing segments for each pair of microphones is used to find conversations among subjects. If two individuals are in the same physical environment and engaged in a conversation, they will take turns in speaking, which will result in high mutual information between the two sensor streams. If a third individual, who also happens to be wearing a device, also takes part in the conversation with the two, our method will also put the third person in conversation with the first two because conversation detection method will compute mutual information for each pair of users who are wearing the devices (for detailed discussion of the conversation detection technique please see [48, 42]).

### Overview of Physical Activity Detection

Physical activities such as walking on flat surface, walking up and down on inclines, and stationary (includes both sitting and standing) were detected based on features extracted from accelerometer and barometric pressure data. We segmented the data into quarter second segments and manually labeled the activity for supervised learning and testing. The features included energy, mean, variance, and suite of spectral features for the accelerometer data, and variance and signed change in pressure over various time windows for the barometric pressure data. We use simple boosted decision-stump classifiers [49] in our current experiments to train binary activity classifiers. Boosted decision stumps have been successfully used in a variety of classification [50] tasks including human activity recognition [39, 51]. For each activity $A_i$, we iteratively learn an ensemble of weak binary classifiers $C^i = {c_1}^i, {c_2}^i, {c_3}^i, ..., {c_M}^i$ and their associated weights ${\alpha_m}^i$ using a variation of the AdaBoost algorithm [52]. The final output is a weighted combination of the weak classifiers. The prediction of classifier $C_i$ is:

$$C_i = sign(\sum_m {\alpha_m}^i {c_m}^i) \quad (1)$$

The classification approach is based on the approach previously developed and validated by Lester, *et. al.*. We refer the reader to [39, 51] for further details.

### Overview of Survey Instruments

Four common surveys for measuring well-being used by healthcare practitioners were administered in a pre-post study design. Subjects completed these paper-based surveys one day prior to study commencement and repeated again one day after the study was concluded (three subjects took the post-study surveys two days after the conclusion of the study). Conducetd surveys are as follows:

**Friendship Scale** is a self-administered [38] survey that measures six dimensions of social isolation and connectedness. Each question can be scored from 0-4 points and adds

up to a total of 24 points. High scores indicate social connectedness and low scores near indicate social isolation.

**SF-36** [8] is a self-administered and commonly used survey for evaluating overall well-being. Its eight sections are weighed together to produce a mental health and a physical health score. We focused on the SF-36 Mental Health Score and compared it with sensed audio measures. As an example, we present in Table 2 the questions used to obtain the mental health score based on five out of the total eight sections.

**CES-D (Center for Epidemiological Study Depression Scale)** is one of the most frequently used surveys to screen for depressive symptoms and behaviors. Each question is scored from 0-3 points with a maximum of 60 points. Individuals with scores 16 or higher are considered to have symptoms indicative of clinical depression [9].

**YPAS (Yale Physical Activity Survey)** [7] is a survey requiring administration that recalls activities performed during a typical period in the previous month. The validated survey estimates energy expenditure (kcals/week), total time spent doing vigorous or leisure activities, and provides a total activity summary index.

| Dimension | Summary of Questions Used To Evaluate Dimensions |
|---|---|
| Social Functioning | During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups? (Not at all, slightly, Moderately, Quite a bit, Extremely)<br><br>During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)? (all the time, most of the time, a good bit of the time, some of the time, little of the time, none of the time) |
| Role Emotional | During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?<br><br>1.Cut down the amount of time you spent on work or other activities<br>2. Accomplished less than you would like<br>3. Didnt do work or other activities as carefully as usual |
| Mental Health | How much of the time during the past 4 weeks (all the time, most of the time, a good bit of the time, some of the time, little of the time, none of the time)<br><br>1. been a very nervous person?<br>2. felt so down in the dumps nothing could cheer you up?<br>3. felt calm and peaceful?<br>4. felt downhearted and blue?<br>5. been a happy person? |
| General Health | In general, would you say your health is (Excellent, Very Good, Good, Fair, Poor?)<br>- I seem to get sick a little easier than other people (definitely true, mostly true, I dont know, mostly false, definitely false)<br>- I am as health as anybody I know (definitely true, mostly true, I dont know, mostly false, definitely false)<br>- I expect my health to get worse (definitely true, mostly true, I dont know, mostly false, definitely false |
| Vitality | How much of the time during the past 4 weeks (all the time, most of the time, a good bit of the time, some of the time, little of the time, none of the time)<br><br>1. did you feel full of pep?<br>2 .did you have a lot of energy?<br>3. did you feel worn out?<br>4. did you feel tired? |

**Table 2: Questions asked in different components of the Mental Health Score of SF-36**

### Evaluation Methodology

Evaluation of data collected in everyday environment is challenging as it is often impossible to have "ground-truth" for every single data point recorded. In order to be confident in our analysis and findings, we took a multi-pronged approach to evaluation. The first step was to label some data from the subjects in the retirement community and compare the physical activity and speech classification accuracy with the accuracy numbers on labeled datasets collected from other individuals in different geographical locations and from different age groups. For speech classification, the comparison dataset included more than twenty-five individuals ranging between 18-50 years of age and about an hour of labeled speech [48, 42]. For activity classification, the comparison dataset included ten individuals ranging between 20-30 years of age and with more than ten hours of labeled data [39, 40]. The labeled data from the retirement community was much smaller in size but was collected in two different days and in different locations of the facility. The speech data was labeled at a very fine granularity to accurately label human voice that changes in the order of milliseconds. We had 16,000 labeled data points collected from the retirement community amounting to little over 4 minutes of audio. For the activity data, we had slightly over 19,000 labeled data points amounting to 80 minutes of data. Our approach for validation was to ensure consistency in classification results across data sets. Since neither the sensing device nor the algorithms changed, we believe it is sufficient to ensure that the new results are consistent with previous results and it is not necessary to collect an extensive new labeled dataset for this population, which is not feasible given the age of our subjects and their willingness to generate such a dataset.

In addition to verifying the consistency of our classification accuracy with existing results, a family care practitioner interviewed each subject during recruitment and a medical trainee was present on-site everyday throughout the day while the study was ongoing. The medical trainee was trained by the physician to administer the surveys to ensure reliability and compliance. The trainee typically had brief interactions with the subjects when he handed out the sensing devices every morning and collected them every evening. He also had occasional interactions with the elders if they chose to come by his room during the day for any questions related to the study. A few subjects used this opportunity to stop by for social interactions. This provided us with some opportunistically collected observational data and indication to whether the survey responses from the subjects were consistent with the medical trainee and the physician's observations.

## RESULTS AND DISCUSSION

In this section, we present the classification accuracy of the various activity classifiers as well as our findings related to the in-situ assessment of physical and mental well-being.

### Classification Accuracy

As mentioned previously, a small amount of labeled data is used to test whether our performance is similar to previously published results on larger datasets [48, 42] and evaluate if it is robust across diverse scenarios in the retirement com-

munity and across subjects. According to our previous research [48, 42], the typical accuracy numbers for inferring human speech ranged between 85%-95% and conversation detection was approximately 95%. For activity recognition, detection accuracy numbers ranged from 80%-95% for stationary, walking on a flat surface, walking up and walking down inclines [39, 40].

As described in the Evaluation Methodology section, four minutes of raw audio included speech both inside the meeting room and around the retirement center. These data are manually labeled for the speech and non-speech segments. Three minutes of this data is used for training and the remaining for testing. Testing performance of the human speech classifier in the meeting room is: accuracy 85%, precision 84%, recall 82%, and along the corridors of retirement center was: accuracy 83%, precision 92%, recall 84%. Overall accuracy is 83.7%, precision 90%, recall 84%.

| | SU01 | SU02 | SU03 | SU04 | SU05 | SU06 | SU07 |
|---|---|---|---|---|---|---|---|
| SU01 | | 12 | 20 | 4 | 24 | 10 | 12 |
| SU02 | 12 | | 70 | 10 | 48 | 12 | 30 |
| SU03 | 20 | 70 | | 108 | 42 | 38 | 32 |
| SU04 | 4 | 10 | 108 | | 12 | 10 | 8 |
| SU05 | 24 | 48 | 42 | 12 | | 20 | 26 |
| SU06 | 10 | 12 | 38 | 10 | 20 | | 12 |
| SU07 | 12 | 30 | 32 | 8 | 26 | 12 | |
| TOTAL | 82 | 182 | 310 | 152 | 172 | 102 | 120 |

**Table 3: Pairwise and total conversation times in minutes for different subjects over 10 day study period**

However, one concern is that the speech/non-speech classifier can overhear the sound of TV programs and classify human speech in TV programs as speech. Since in subsequent analysis, we plan to use the amount of human speech in a subject's conversational vicinity for mental health assessment, we discuss how we filter out speech sounds that occur in TV programs. In a preliminary test of a TV recording that included movie and news (including commercials), the speech inference algorithm classified 19% of the recording as speech. To filter out TV speech, we use two energy features derived from the audio signal. During speech in conversational vicinity there is a broader spread in the energy intensity in the voiced region due to non-fixed location of the speech sound (as opposed to fixed location of TV) and the energy values of voiced regions for TV are more uniform across time. Thus, the entropy of energy intensity is higher for human speech occurring in the same physical space compared to the human speech occurring on TV. And entropy of energy distribution over time will be lower for speech in the same physical space compared to speech in TV. These features are computed for every three minute windows. Thirty-three minutes of conversation data and fifty seven minutes of TV program data is used (comprising of a short movie segment and news program) to train a model. A simple Gaussian classifier, where one gaussian is trained for TV and one gaussian for conversation, is used for classification. On a separate testing dataset that included one hundred and three minutes of conversation and forty-two minutes of TV data, the classifier have a 100% classification accuracy for TV and 94% accuracy for human voice occurring in the same location of the subject. The conversation data used for training and testing in this experiment were collected from a research group meeting, casual conversation between colleagues in the department and during the orientation session at the retirement community. We apply this filtering algorithm after the speech inference step to eliminate instances of TV speech.

Table 5 shows the number of minutes inferred as human speech across different subjects for the entire 10 day data collection. During our analysis, we found that subject 08's microphone was faulty and the audio recorded by the device was unusable for speech and conversation detection. Furthermore, subject 04 stopped recording on several occasions before the afternoon and most of the data collection was heavily skewed towards the morning. We omitted subject 04's data when computing correlation with the survey scores as the amount and distribution of data collection from him was substantially different than the rest of the study group. Figure 4 shows data availability for different subjects.

Conversation is detected only between the study participants as the detection algorithm assumes that each of the conversation participant will be carrying their own microphones. The mutual information between the inferred speech streams are calculated for each pair of subjects for every two minute of data and conversation is identified using a threshold on the mutual information score that was obtained via cross-validation [48, 42]. Table 3 shows the number of minutes of conversation that took place between each pair of subjects over the 10 day period. One particular point to note is that subject 03 and 04 are a couple and they have the highest amount of conversation between them (the other couple in the experiment is not presented here because that couple includes subject 08 whose audio data, as stated above, was unusable for further analysis). In this context, conversation analysis is presented here not just for sanity checking our data but to show that it reveals important interpersonal information among the subjects of the experiment. Also conversation analysis is pivotal for speaker segmentation [42] and subsequent user specific speech analysis, like pitch, speaking rate etc., which will be used for future research.

For physical activity recognition, a total of 80 minutes of physical activity data was collected from the subjects. Out of this data, 25 minutes of data is labeled for stationary, walking on flat surface, walking up and down inclines. During labeling, it is ascertained that data from every subject is represented and all classes have roughly the same number of labeled data points. Leave one subject out cross-validation results are reported in Table 4. Here, we expect the performance numbers on the small test dataset to be higher and the actual performance will be closer to previously published results [39, 40]. We would like to reiterate that these evaluations are mainly used to check that the classification accuracy is consistent with the performance of prior systems. Furthermore, cross-validation establishes that the classification worked robustly across individuals and device placement variations.

| Actiivty | Precision | Recall | Accuracy |
|---|---|---|---|
| stationary | 92.4% | 100% | 96.05% |
| walking flat surface | 100% | 75.86% | 93.1% |
| walking up | 94.9% | 100% | 99.6% |
| walking down | 86.7% | 100% | 97.4% |

**Table 4: Performance numbers for physical activity classifier**

### Automated Measurement of Mental Well-Being

To determine the relationship between sensed speech and mental well-being, we calculated the fraction of time human speech was present within conversational proximity for each subject (as shown in table 5). Fraction of time refers to the ratio of the time human speech was present within conversational proximity and the total duration of recording. We use fraction rather than total amount of speech for the subjects because of unequal data length across different subjects. Univariate regression analysis comparing the sensed speech with SF-36 Mental Health Score revealed a positive correlation R= 0.82 (p=0.048). In order to interpret this result, it is worth delving into the SF-36 Mental Health Score. It is the average of five out of eight sections of the survey, which include Social Functioning, Role Emotional, Mental Health, and Reported Health. Table 2 contains the questions used in these five sections. The Social Functioning dimension is concerned with issues that interfere with subjective perception of General Health, and Vitality. Role Emotional is concerned with emotions that interfere with daily social and occupational activities. The Mental Health dimension captures subjective self-perception, which can be clouded by pathologic negative interpersonal interactions and subject's current mood, especially in those with mood disorders [53, 54, 55]. The General Health represents subjective perception of health compared to others. Lastly, Vitality captures the subjective perception of energy and stamina. Thus, a comprehensive look at mental health score accounts for emotional or physical stressors that interfere with socialization and seeks to quantify the effect on the emotional integrity of the participant. Directly capturing amount of human speech present in an individual's close vicinity can indicate whether an individual is socially engaged and provide valuable information about metal health. Although, the measurement of speech alone is unlikely to be a conclusive or comprehensive measurement of mental health, it can serve as a continuous indicator and early warning system.

Regression analysis comparing fraction of human speech sensed in conversational proximity and CES-D scores revealed a negative correlation R= -0.73 (p=0.096), which is slightly outside the statistical significance range in this small sample. Note that higher CES-D scores indicate increased presence of depressive symptoms. In addition to CES-D and SF-36, regression analysis comparing fraction of human speech sensed in conversation vicinity and friendship scales revealed a strong correlation of R= 0.96 (p=0.002).

Anecdotal qualitative observations provide support that sensed measurements can be used to assess mental health and are often used as one of the early screening methods for depression. Table 5 shows that subject 01 scored above 25 on CES-D and had a much lower SF-36 Mental Health score. Comparing this score to the interaction-based assessment of the participant by the medical trainee and the physician in the team (before survey results were obtained) confirmed concerns that the participant appeared socially isolated, and avoided large groups, and was dysthymic (exhibiting depressive symptoms without a formal diagnosis of depression) and unhappy. Automatically inferred speech duration for subject 01 was low compared to what was observed for the rest of the group. We also discuss another example in section below.

| Subject | Amount of Human speech (%) | SF-36 Mental Health Score | CES-D Score | Friendship Scale |
|---|---|---|---|---|
| SU01 | 2.6 | 68 | 25 | 14.5 |
| SU02 | 9.9 | 90 | 13 | 22 |
| SU03 | 11.6 | 90 | 0.5 | 21 |
| SU04 | 6.3* | 96 | 1.5 | 22 |
| SU05 | 15.3 | 90 | 5.5 | 24 |
| SU06 | 6.9 | 88 | 4.5 | 18 |
| SU07 | 12.03 | 88 | 5.5 | 21 |
| SU08 | Faulty Mic. | 96 | 2 | 24 |

**Table 5: Percentages of sensed human speech during the 10 day study and mental health scores from the surveys. (*ignored for further analysis because of lack of data throughout the day)**

| Subject | Stationary (%) | Walking (%) | Going Up (%) | Going Down (%) | Unknown (%) | Assistive Devices |
|---|---|---|---|---|---|---|
| SU01 | 63.41 | 3.98 | 10.81 | 6.24 | 15.56 | No |
| SU02 | 57.29 | 11.20 | 6.39 | 5.08 | 20.04 | Walker /Cane |
| SU03 | 62.76 | 18.43 | 1.24 | 3.17 | 14.39 | No |
| SU04 | 65.20 | 9.71 | 0.66 | 7.70 | 16.72 | Cane |
| SU05 | 54.60 | 2.72 | 14.49 | 11.42 | 16.77 | No |
| SU06 | 66.13 | 2.95 | 3.58 | 17.54 | 9.80 | No |
| SU07 | 73.39 | 4.83 | 1.82 | 2.12 | 17.84 | Wheelchair /Cane |
| SU08 | 62.85 | 3.25 | 7.42 | 11.62 | 14.85 | No |

**Table 6: Percentages of different physical activities for subjects during the 10 day study (Going up and going down refers to walking up and walking down inclines respectively)**

### Automated Measurement of Physical Well-Being

Table 6 shows the breakdown of classified activities, namely walking on flat surface, walking up and down inclines, and being stationary (sitting and standing). The *unknown* category contained all segments of the data that were not associated with the four classified types of activity. Comparing the percentage of time spent doing a specific activity with daily patterns and physical limitations of individual subjects shows good qualitative correlation. Subject 05 reported walking and participating in other physical activities multiple hours daily and had the highest amount of climbing up, down and walking and the lowest percentage of stationary data in the group. Lastly, subject 07 had a neurological disease that limited the person's ability to walk and take part in other physical activities. This is reflected in the amount of

stationary activity that is inferred–the highest in the group–and also by the lower values for all other activities. Furthermore, it is worth mentioning that the retirement community is situated on a hilly area with sloping pathways. A few of the subjects took long daily walks along those path and as a result have high percentages of walking up and walking down inclines.

To demonstrate the potential to partly capture the information gathered by surveys using automatically sensed behavior, we compute a weighted sum of physical activity measure inferred using Equation 2 and compare it to the YPAS Activity Index. In the equation, $a_i$ (in percentage) represents different types of activities (stationary, walking, walking up inclines, walking down inclines) and $w_i$ corresponds to the weights of different activities in the overall inferred activity score $A$.

$$A = \sum_i w_i a_i \quad (2)$$

A multi-dimensional linear regression is utilized to determine relations between YPAS scores and the measures computed from the sensed data. Coefficients (or weights) generated by this regression are as follows: 0.6 for stationary, 3.0 for walking, 5.0 for walking up inclines, 1.0 for walking down inclines, and -1.5 for unknown. These weights for different activities make intuitive sense and reflect the increasing intensity of activity and calories burnt as one goes from stationary, walking down, walking on flat surface to walking up. Furthermore, possible explanation for the negative weight for *unknown* is two fold: (i) it may represent activities that do not correspond to any physical activity; (ii) a high relative proportion of *unknown* makes the values for other physical activities lower than usual (i.e., it has a negative effect on other physical activities).

The total minutes for different activities inferred from sensor measurements can be found in Table 6 and the weighted sum can be computed based on Equation 2 to obtain the activity score. The *inferred activity score* demonstrates a correlation of R=0.88 with p=0.01 with the YPAS activity index. Of note, subject 04's data is not included in the correlation analysis because of lack of data throughout the day.

An unexpected finding was observed for subject 02 regarding the relationship between inferred activity score ($A$) (from equation 2) and the YPAS activity score. Direct observation of the person's daily activities indicated that the subject spent most of the day being active as a volunteer and also spent time cooking/cleaning her home. The inferred activity level was comparable to the rest of the group and this paralleled observational assessment by the medical trainee who was always at the facility during data collection. However, YPAS reported a much lower estimate of their physical activity level. Despite being administered by the medical trainee, inaccurate recall of daily activities, human error in calculating amount of time spent doing a specific activity, and limitation of the survey in capturing the time spent participating in the specific volunteer task the person engaged in may have contributed to this discrepancy.

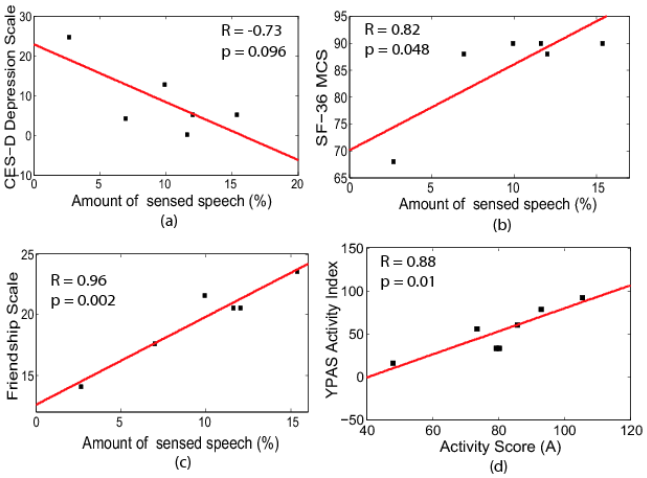


**Figure 3: Correlation between (a) sensed human speech and CES-D (b) sensed human speech and SF-36 Mental Health Score (c) sensed human speech and friendship scale (d) physical activity score $A$ and YPAS**

### Case Study: Limitations of Survey-based Assessment

Our primary goal was to validate sensor-based measurement of in-situ behavior with paper-surveys, the current gold standards in health industry for assessing mental and physical health. We also wanted to explore the potential limitations of self-reported measurement of behavior and how continuous sensing of behavior using easily wearble mobile sensors can supplement and improve currently available tools and methods in health behavioral sciences. Analysis of total amount of speech for subject 06 provides a case example illustrating the limitations of the standard methods for measuring mental health and suggests that automated measurements could potentially overcome these limitations. A comparison of subject 06's CES-D and the Mental Health component of the SF-36 scores and the inferred activities indicate that: (i) there are no major mental health concerns detected by CES-D and SF-36 and (ii) the surveys disagree with the measurement made using the mobile sensors (6.9% of human speech sensed in conversation vicinity, which is second lowest after subject 01), and raised concerns of social isolation, a major risk factor for emotional well-being [56]. The sensor based measurement was in agreement with direct observations made by the medical trainee. This subjective measurement of mental health could have been inappropriately influenced by one or more factors including skewed self-perception, misinterpretation of questionnaires, and purposeful misrepresentation. Thus, continuos sensing could be utilized in preventing exclusive dependence on subject for accurate recall, interpretation and response to questions.

## LIMITATIONS AND SHORTFALLS

Our study has several of limitations. The sample size was small and it is difficult to draw definitive conclusions and should be viewed as a promising pilot study. Similarly, this study was focused on older adults. Additional investigation needs to be performed on other populations to see if findings are generalizable. While we did collect a small amount of observational data, we did not perform continuous direct observation. Although this strategy would have made compar-

ison between subjective behavior and actual behavior easier, it would have likely biased observations by disrupting the natural setting of participants.

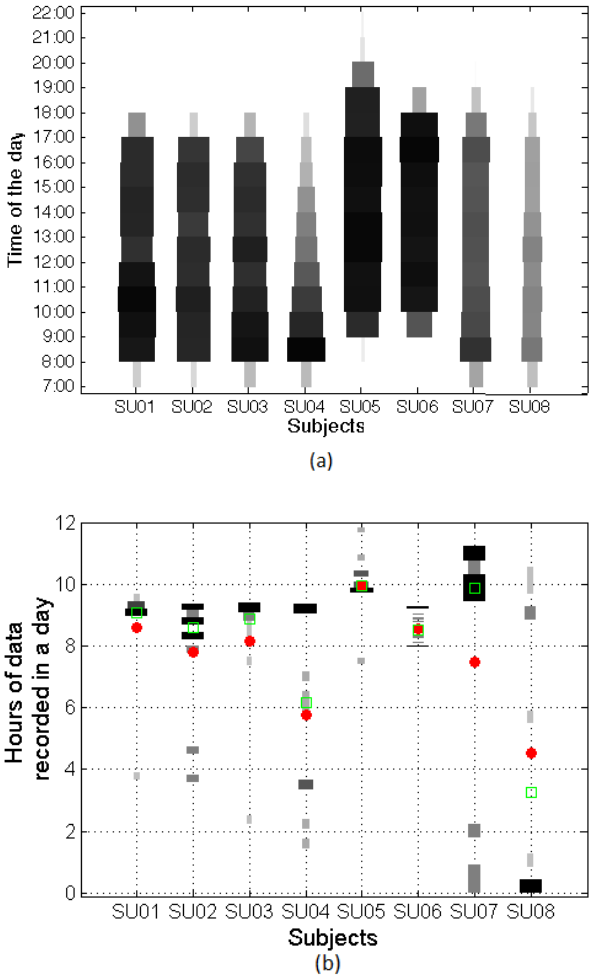


**Figure 4: Amount of data recorded for different subjects through out the 10 day pilot study. (a) shows the amount of recorded data at different parts of a day. Wider and darker horizontal bars represent more recording during that part of the day. (b) shows distribution of the hours of data recorded on each day for a given subject. For example, subject 02 recorded close to 9 hours of data on most days but has a couple of days where she records between 3.5-4.5 hours of data. The red dot represent the mean and the green square represents the median number of recorded hours per day for each subject.**

Figure 4 shows the amount of data recorded at different parts of the day during the 10 day study. It highlights the fact that subject 08 provided the least amount of data because he choose not to wear the device during voluntary duties outside the retirement facility. Subject 04 provided data mostly in the morning and his data distribution is heavily skewed towards the morning as can also be seen in Figure 4. To increase reliability and decrease user burden, participants dropped off the sensing devices at the end of the day and picked them up the following morning. This enabled offloading of data and recharging of the batteries overnight. Due to picking up and returning the device at different times, unequal amount of data was collected for different subjects. Finally, significantly less data were collected on day 8 due to insufficient charging of the devices during the previous night. Despite these limitations, this pilot work demonstrates the power and potential of utilizing commonly available sensors with sophisticated processing techniques to improve the detection of specific physical and behavioral activities. As more people are carrying sensors as part of everyday mobile devices, the potential to detect health problems and monitor treatment could become more efficient and effective.

## CONCLUSION

In this paper, we demonstrated that daily human behaviors, inferred from mobile sensors, correlate highly with well-established survey metrics, including measures of depressive symptoms, in older adults. In an informal usability survey, most study participants stated that they found the device easy to use, comfortable to wear, and all participants thought the sensor based approach was preferable to surveys. Though the strong quantitative results combined with qualitative acceptance is encouraging, there are scopes for future improvements in scalability, usability, robustness and in detecting finer aspects of mental health. We are currently focusing on implementing our sensing and inference system on smartphones. We believe that the recent proliferation and pervasiveness of smartphones can enable us to truly scale the passive objective well-being system for the masses. These smartphones will not only improve usability but also enable us to collect large amount of data in diverse environments and create models that are robust across scenarios. In addition to data collection, we plan to focus on different aspects of mental health, for example aspects that change rapidly overtime (e.g., non-chronic stress, mood) and that are less transient (e.g., personality). We believe this work highlights a first step towards our goal by showing that passive and continuous sensing of activity and behavior is feasible and comparable to traditional, more cumbersome methods of assessment for an increasing population like older adults. With continued advancement of these technologies, there is great potential to improve early detection of changes in well-being and overall quality of life.

## ACKNOWLEDGMENTS

This work was supported by NSF IIS-0845683, NSF CNS-0910842 and National Institute on Aging 1K23AG036934.

## REFERENCES

1. R. Camicioli, MM Moore, G. Sexton, DB Howieson, and JA Kaye. Age-related brain changes associated with motor function in healthy older people. *Journal of the American Geriatrics Society*, 47(3):330, 1999.
2. P.L. Sheridan, J. Solomont, N. Kowall, and J.M. Hausdorff. Influence of executive function on locomotor function: divided attention increases gait variability in Alzheimer's disease. *Journal of the American Geriatrics Society*, 51(11):1633–1637, 2003.
3. Daniel Olguin Olguin, Benjamin N Waber, Taemie Kim, Akshay Mohan, Koji Ara, and Alex Pentland. Sensible organizations: technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(1):43–55, 2009.
4. Megan P Rothney, Emily V Schaefer, Megan M Neumann, Leena Choi, and Kong Y Chen. Validity of physical activity intensity predictions by actigraph, actical, and rt3 accelerometers. *Obesity (Silver Spring)*, 16(8):1946–52, 2008.
5. S. Consolvo, D.W. McDonald, T. Toscos, M.Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1797–1806, 2008.

6. Susan J Wenze and Ivan W Miller. Use of ecological momentary assessment in mood disorders research. *Clinical psychology review*, 2010.
7. L. Dipietro, C.J. Caspersen, A.M. Ostfeld, and E.R. Nadel. A survey for assessing physical activity among older adults. *Medicine and science in sports and exercise*, 25:628–628, 1993.
8. Sf-36, http://www.sf–36.org, 2008.
9. E M Andresen, J A Malmgren, W B Carter, and D L Patrick. Screening for depression in well older adults: evaluation of a short form of the ces-d (center for epidemiologic studies depression scale). *American Journal of Preventive Medicine*, 10(2):77–84, 1994.
10. A A Stone, J E Schwartz, J M Neale, S Shiffman, C A Marco, M Hickcox, J Paty, L S Porter, and L J Cruise. A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology*, 74(6):1670–80, 1998.
11. J E Schwartz, J Neale, C Marco, S S Shiffman, and A A Stone. Does trait coping exist? a momentary assessment approach to the evaluation of traits. *Journal of personality and social psychology*, 77(2):360–9, 1999.
12. Sunlee Bang, Minho Kim, Sa-Kwang Song, and Soo-Jun Park. Toward real time detection of the basic living activity in home using a wearable sensor and smart home sensors. *Proc IEEE Engineering in Medicine and Biology Society*, pages 5200–3, 2008.
13. DM Bravata, C Smith-Spangler, V Sundaram, AL Gienger, N Lin, R Lewis, CD Stave, I Olkin, and JR Sirard. Using pedometers to increase physical activity and improve health: a systematic review. *JAMA*, 298(19):2296, 2007.
14. C. Tudor-Locke, S.B. Sisson, T. Collova, S.M. Lee, and P.D. Swan. Pedometer-determined step count guidelines for classifying walking intensity in a young ostensibly healthy population. *Applied Physiology, Nutrition, and Metabolism*, 30(6):666–676, 2005.
15. R Bergman and D Bassett Jr. Validity of 2 devices for measuring steps taken by older adults in assisted-living facilities. *Journal of physical activity & health*, 5:S166, 2008.
16. L. Bao and S.S. Intille. Activity recognition from user-annotated acceleration data. *Pervasive Computing*, pages 1–17, 2004.
17. Actigraph, http://www.theactigraph.com/.
18. Sensewear, http://sensewear.com/.
19. T Choudhury, G Borriello, S Consolvo, D Haehnel, B Harrison, B Hemingway, and J Hightower. The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing*, pages 32–41, 2008.
20. Ansgar Conrad, Frank H Wilhelm, Walton T Roth, David Spiegel, and C Barr Taylor. Circadian affective, cardiopulmonary, and cortisol variability in depressed and nondepressed individuals at risk for cardiovascular disease. *Journal of psychiatric research*, 42(9):769–77, 2008.
21. S. Reder, G. Ambler, M. Philipose, and S. Hedrick. Technology and Long-term Care (TLC): A pilot evaluation of remote monitoring of elders. *Gerontechnology*, 9(1):18–31, 2010.
22. Revenues and expenditures reports from 2006.
23. L J Kirmayer, J M Robbins, M Dworkind, and M J Yaffe. Somatization and the recognition of depression and anxiety in primary care. *American Journal of Psychiatry*, 150(5):734–41, 1993.
24. M Olfson, B Fireman, M M Weissman, A C Leon, D V Sheehan, R G Kathol, C Hoven, and L Farber. Mental disorders and disability among patients in a primary care group practice. *American Journal of Psychiatry*, 154(12):1734–40, 1997.
25. G Simon, J Ormel, M VonKorff, and W Barlow. Health care costs associated with depressive and anxiety disorders in primary care. *American Journal of Psychiatry*, 152(3):352–7, 1995.
26. H J Henk, D J Katzelnick, K A Kobak, J H Greist, and J W Jefferson. Medical costs attributed to depression among patients with a history of high medical expenses in a health maintenance organization. *Archives of General Psychiatry*, 53(10):899–904, 1996.
27. Vivian Isaac, Robert Stewart, Sylvaine Artero, Marie-Laure Ancelin, and Karen Ritchie. Social activity and improvement in depressive symptoms in older people: a prospective community cohort study. *American Journal of Geriatric Psychiatry*, 17(8):688–96, 2009.
28. Hayden B Bosworth, Judith C Hays, Linda K George, and David C Steffens. Psychosocial and clinical predictors of unipolar depression outcome in older adults. *International journal of geriatric psychiatry*, 17(3):238–46, 2002.
29. W.W. Walk. A prospective study of physical activity and cognitive decline in elderly women. *Arch Intern Med*, 161:1703–1708, 2001.
30. Anmol Madan, Manuel Cebrian, David Lazer, and Alex Pentland. Social sensing for epidemiological behavior change. In *Ubicomp '10: Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 291–300, New York, NY, USA, 2010. ACM.
31. Danny Wyatt, Tanzeem Choudhury, James Kitts, , and Jeff Bilmes. Inferring Colocation and Conversation Networks from Privacy-sensitive Audio with Implications for Computational Social Science. *To appear in ACM Transactions on Intelligent Systems and Technology*, 2010.
32. D.J. France, R.G. Shiavi, S. Silverman, M. Silverman, and M. Wilkes. Acoustical properties of speech as indicators of depression and suicidal risk. *IEEE Transactions on Biomedical Engineering*, 47(7):829–837, 2002.
33. R. Cowie and E. Douglas-Cowie. Automatic statistical analysis of the signal and prosodic signs of emotion in speech. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 3, pages 1989–1992, 2002.
34. R.W. Frick. Communicating Emotion:: The Role of Prosodic Features. *Psychological Bulletin*, 97(3):412–429, 1985.
35. Elliot Moore, Mark A Clements, John W Peifer, and Lydia Weisser. Critical analysis of the impact of glottal features in the classification of clinical depression in speech. *IEEE Transactions on Biomedical Engineering*, 55(1):96–107, 2008.
36. D.N. Kiosses, S. Klimstra, C. Murphy, and G.S. Alexopoulos. Executive dysfunction and disability in elderly patients with major depression. *American Journal of Geriatric Psych*, 9(3):269, 2001.
37. B.L. Kirkman, B. Rosen, P.E. Tesluk, and C.B. Gibson. The impact of team empowerment on virtual team performance: The moderating role of face-to-face interaction. *The Academy of Management Journal*, 47(2):175–192, 2004.
38. G Hawthorne. Measuring social isolation in older adults: Development and initial validation of the friendship scale. *Social Indicators Research*, pages 1–28, 2006.
39. Jonathan Lester, Tanzeem Choudhury, Nicky Kern, Gaetano Borriello, and Blake Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *In Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 766–772, 2005.
40. Jonathan Lester, Tanzeem Choudhury, Gaetano Borriello, and Blake Hannaford. A practical approach to recognizing physical activities. In *Proc. of Pervasive*, 2006.
41. D. Wyatt, T. Choudhury, J. Bilmes, and J.A. Kitts. Inferring colocation and conversation networks from privacy-sensitive audio with implications for computational social science. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):7, 2011.
42. D. Wyatt, T. Choudhury, and J. Bilmes. Conversation detection and speaker segmentation in privacy sensitive situated speech data. In *Proceedings of Interspeech*, pages 586–589, 2007.
43. D. Wyatt, J. Bilmes, T. Choudhury, and J.A. Kitts. Towards the automated social analysis of situated speech data. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 168–171. ACM, 2008.
44. F. Dellaert, T. Polzin, and A. Waibel. Recognizing emotion in speech. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 3, pages 1970–1973. IEEE, 1996.
45. D. Gatica-Perez, I. McCowan, D. Zhang, and S. Bengio. Detecting group interest-level in meetings. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. Citeseer, 2005.
46. R.T. Hurlburt, M. Koch, and C.L. Heavey. Descriptive experience sampling demonstrates the connection of thinking to externally observable behavior. *Cognitive Therapy and Research*, 26(1):117–134, 2002.
47. S. Basu. A linked-HMM model for robust voicing and speech detection. In *Proceedings of International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, 2003.
48. Tanzeem Choudhury and Sumit Basu. Modeling conversational dynamics as a mixed-memory markov process. In *Proc. of NIPS*, 2004.
49. Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28:2000, 1998.
50. Antonio Torralba and Kevin P. Murphy. Sharing visual features for multiclass and multiview object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):854–869, 2007.
51. Ulf Blanke and Bernt Schiele. Daily routine recognition through activity spotting. In *LoCA '09: Proceedings of the 4th International Symposium on Location and Context Awareness*, pages 192–206, Berlin, Heidelberg, 2009. Springer-Verlag.
52. P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2002.
53. J.B. Nezlek, R.M. Kowalski, M.R. Leary, T. Blevins, and S. Holgate. Personality moderators of reactions to interpersonal rejection: Depression and trait self-esteem. *Personality and Social Psychology Bulletin*, 23(12):1235, 1997.
54. T L Simoneau, D J Miklowitz, and R Saleem. Expressed emotion and interactional patterns in the families of bipolar patients. *Journal of Abnormal Psychology*, 107(3):497–507, 1998.
55. Saul Shiffman, Arthur A Stone, and Michael R Hufford. Ecological momentary assessment. *Annual review of clinical psychology*, 4:1–32, 2008.
56. T E Seeman. Social ties and health: the benefits of social integration. *Annals of Epidemiology*, 6(5):442–51, 1996.