

Multi-Modal Mobile Sensing of Social Interactions

Aleksandar Matic, Venet Osmani, Alban Maxhuni, Oscar Mayora

{name.surname@create-net.org}

Ubiquitous Interaction Group, CREATE-NET, via alla Cascata 56/D, 38123 Trento, Italy

Abstract— The level of participation in social interactions has been shown to have an impact on various health outcomes, while it also reflects the overall wellbeing status. In health sciences the standard practice for measuring the amount of social activity relies on periodical self-reports that suffer from memory dependence, recall bias and the current mood. In this regard, the use of sensor-based detection of social interactions has the potential to overcome the limitations of self-reporting methods that have been used for decades in health related sciences. However, the current systems have mainly relied on external infrastructures, which are confined within specific location or on specialized devices typically not-available off the shelf. On the other hand, mobile phone based solutions are often limited in accuracy or in capturing social interactions that occur on small time and spatial scales. The work presented in this paper relies on widely available mobile sensing technologies, namely smart phones utilized for recognizing spatial settings between subjects and the accelerometer used for speech activity identification. We evaluate the two sensing modalities both separately and in fusion, demonstrating high accuracy in detecting social interactions on small spatio-temporal scale.

Keywords-social interaction detection; nonverbal cues; mobile phone sensing; accelerometers; speech activity detection; wearable computing.

I. INTRODUCTION

The association between social interactions and health of individuals has been established both on theoretical basis and through empirical evidences [1]. Subjects with a low quantity of social relationships are typically less healthy, psychologically and physically, while manifesting higher risks for a wide range of potential implications – from tuberculosis and psychiatric disorders to accidents and even mortality [1]. In contrast, individuals who maintain a certain level of social engagements are shown to be more successful in coping with stress, and in case of the elderly, they are highly functional and independent [2]. Several studies have demonstrated that the amount of social activity is negatively correlated with depressive behavior while socialization can improve depressive symptoms [3] [4]. In addition to depression, participation in social interactions was analyzed in bipolar disorder patients, for identifying the occurrence of both manic and depressive episodes. Therefore, monitoring social interaction constitutes an important instrument for coaching and diagnosis of various disorders, while for healthy subjects the level of social activity represents a wellbeing index that can be used to encourage a healthier life style [5] [2].

The standard methods for monitoring social interactions in the health sciences rely on self-reports and recall surveys that suffer from several limitations: 1) difficulties in recalling activities that occurred in the past, 2) self-reports are subjective and may be affected by the current mood, and 3) a high effort

for continuous long-term monitoring [5]. In addition, it was demonstrated that self-reports correspond poorly to the communication behavior recorded by independent observers [6], while engaging human observers to record communications in groups is inefficient particularly if the size of the group is large [7]. In this work, we refer to social interaction as co-located, face-to-face interaction, excluding electronically mediated interactions such as chat, social network activity and other kinds of electronic communication.

The use of automated sensing techniques for recognizing face-to-face social interactions has been explored in the past decade, aiming to overcome the limitations of self-reporting methods. Recognizing the occurrence of social interaction in an automatic and mobile way is typically based on sensing proximity of subjects and/or on detecting speech activity. One of a few solutions which detects social encounters by sensing both spatial settings and the speech activity status is the Sociometric Badge [7], a pendant-like hardware worn at chest-level. In the healthcare setting, Sociometric Badge was applied to monitor nurses in the Boston hospital and to measure their group performance in order to improve the management of healthcare staff. However, mobile solutions for gathering interaction data typically rely either on proximity or on audio analysis. Sociopatterns sensing platform [8], for instance, detects social interactions by examining spatial settings among subjects using a wearable active Radio Frequency Identification (RFID) tags. The device was applied in the healthcare setting to provide a picture of the contact patterns between patients, healthcare staff, visitors and tutors [9]. A frequently applied approach for inferring social activity through the detection of proximity relies on the use of Bluetooth embedded in a widely adopted device – the mobile phone [10], [11]. Bluetooth scans indicate the presence of nearby devices in a radius of 10m, which does not provide sufficient information to detect an ongoing social interaction. Such approaches rely on statistical modeling of human behavior at large scale in order to quantify social dynamics, but failing to detect the social encounters that happen on a small spatio-temporal scale. The use of Wi-Fi is another approach to detect interpersonal distances and to infer social interactions between subjects using a mobile phone. In this regard, Microsoft's Virtual Compass [12] is the mobile-phone based system which, to the best of our knowledge, achieves the highest accuracy in sensing interpersonal distances by using transmitting/receiving mechanisms of radio waves. Through the analysis of both Wi-Fi and Bluetooth signal strengths, Virtual Compass estimates the position of subjects in 2D plane with the median accuracy of 0.9m for two nodes and of 1.9m for positioning of nine nodes. However, the lack of subjects' orientation information and the lack of the knowledge of speech activity might not be sufficient for modeling social interaction occurrence (for example, the case of two subjects

sitting across from each other in the office or in the bus, and not engaging in an interaction). On the other hand, the Electronically Activated Recorder (EAR) detects social situations through identifying talk, laughter and arguments among individuals in sampled ambient sounds. Regarding healthcare applications, EAR was used to observe depression symptoms in rheumatoid arthritis patients [13] and to detect depressive behavior in women coping with an illness [14]. Two recent approaches utilized a microphone embedded in mobile phones to detect the amount of social activity with the goal of modeling and promoting the wellbeing of individuals. Lane et al. proposed the BeWell platform [2], which associates the level of detected ambient noise with social interactions [2], which was used in the study as a wellbeing index in addition to physical activity and sleep quality. Also aiming to improve wellbeing, Rabbi et al. [5] proposed a more advanced method for identifying conversations among subjects which includes an analysis of voice segments acquired from a pair of mobile phone microphones. The limitations of these approaches include sensitivity to false positives since nearby conversations can be unintentionally picked up. Moreover, activating microphone typically raises privacy concerns, even with methods that do not require continuous voice recording thus affecting natural behavior of subjects. Furthermore, in a number of situations (for example, in the case of public spaces or often when monitoring patients) audio data cannot be obtained due to legal or ethical issues [15].

In this work, we aim to detect face-to-face social interactions on small spatio-temporal scale relying on off-the-shelf and widely available sensing technologies. The approach recognizes both speech activity and spatial settings among subjects, while it does not capture privacy-sensitive data. The main contributions of this paper are the following:

1. Detection of spatial setting, described through parameters of interpersonal distances and relative body orientations,, using mobile phones
2. Recognition of speech activity using off-the-shelf accelerometers;
3. Evaluation of the two modalities' contribution to inference of social interactions.

These two modalities provide complementary information about social interactions constituting a multi-modal mobile sensing system. Evaluation of each modality is described in Section II and III respectively, while the performance of the fusion of both modalities is presented in Section IV. Lastly, we discuss the applicability of the proposed approach in health-related domain and we draw the main conclusions.

II. MEASURING INTERPERSONAL SPATIAL PARAMETERS

A. Estimating Interpersonal distances

Similarly to indoor positioning systems that use signal fingerprinting, our method for distance estimation is based on analyzing RSSI values between a phone that acts as a Wi-Fi access point (a feature called “Portable Hot Spot” in Android phones and “Personal Hot Spot” in iPhone) and a client Wi-Fi phone. However, unlike time-consuming measurements typically required for fingerprinting methods, we decrease the user effort by fast-calibrating the model based on measurement

of a few RSSI values on a number of distances, coupled with a propagation model. Before delving into the fast calibration approach we analyze the potential of Wi-Fi RSSI for distance estimation between two smart phones.

As a first step in evaluating the feasibility of using mobile phone sensing to estimate distances between two units, we analyzed the RSSI dependence on distance for various different transmitting power levels, from the minimal of 0dBm (1mW) to 32dBm (1.6 W) which is the maximal power level. The measurements were carried out in the same environment using HTC Desire smart phone, while recording 300 samples with the sampling rate of 1Hz for each of the distances following a grid of 0.5m. The transmitting power of 0dBm provided the smoothest and the most monotone characteristics thus proving to be the best fit for short distance estimation. With the 0dBm power level we analyzed RSSI patterns in seven different environments (indoor areas ranging from 30m² to 90m²) and recorded 300 samples every 0.5m up to the distance of 8m. The cumulative RSSI patterns are presented in Figure 1. where small circles represent outliers, thick horizontal lines correspond to median values, bottom and top of each box corresponds to the first and the third quartiles of distribution and the whiskers extend up to 1.5 times the interquartile range (IQR). It can be seen that the RSSI shows relatively monotone characteristics across different environments, demonstrating the instability and fluctuations of the Wi-Fi signal, typically due to environmental factors [16] requiring the use of pre-processing and classification algorithms for distance estimation.

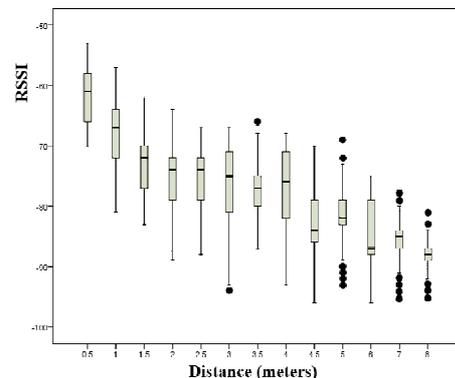


Figure 1. RSSI dependence on the distance

To acquire the training set we used two phones, one in transmitting mode (HotSpot) the other in receiving (client) mode, to carry out Wi-Fi signal measurements. We used different distances following the same grid of 0.5m thus collecting RSSI training values. RSSI values were segmented by grouping every 10 consecutive samples (corresponding to 10 seconds) and calculated signal characteristics for each segment. This mitigated the effects of Wi-Fi signal instability in a short time frame [17]. Since RSSI distribution varies according to its mean, we selected the mean value as a candidate parameter to represent the RSSI pattern. It turned out that, among other tested signal characteristics (such as standard deviation, minimum and median), the combination of the mean and maximal value was proven to provide the highest accuracy in distance estimation.

Overall, six smart phones (with Android operating system) were used including three different models, namely *HTC Desire*, *Samsung Nexus S* and *HTC Nexus One*¹ that used a modified firmware to allow adjustment of transmitting power. Different phone units were distinguished by their MAC addresses. Measurements were taken in three offices with dimensions of 12×8 m, 6×5 m and 6×3 m, a balcony of 12×2.5 m and a meeting room of 10×8 m. Figure 2. shows the system’s accuracy when using the same phone model acting as a receiver in both training and test phase. We tested Naïve Bayes classifier with Kernel Density Estimation (KDE) and Gaussian Process (GP) regression. The median estimation error (50th percentile) of approximately 0.5m was achieved using both machine learning techniques. Naive Bayes with KDE showed a slightly better overall performance, providing distance estimation with a 50% percentile error of 0.5m and 95th percentile error of around 2m.

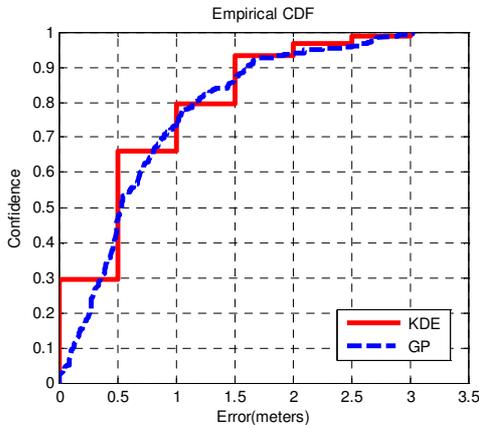


Figure 2. Distance estimation accuracy (same phone mode)

However, when *different* phone models were used for training and test phase, the system’s accuracy significantly degraded to 1.8m (50th percentile). This is due to the fact that RSSI patterns, among other factors, highly depend on the receiver characteristics [17] which are likely to be different across different phone models. In order to tackle this problem, while avoiding the repetition of RSSI measurements, we calibrated only one point by measuring RSSI for a couple of minutes on a fixed distance of 1 m. Once the RSSI is captured, the rest of the training set is estimated by applying the following propagation model [18]:

$$P(d)[dBm] = P(d_o)[dBm] - 10n \log\left(\frac{d}{d_o}\right)[dBm] - X[dBm] \quad (1)$$

where n is the path loss exponent, $P(d_o)$ is the signal power at the reference distance d_o from the transmitter phone (in our case 1m) and d is the distance in which RSSI is estimated by applying the model. X is a component that reflects the sum of losses induced by each wall between the transmitter and receiver. We have found that the best suited value for the

coefficient n is 1.5, while for X is zero (there are no walls or other obstacles between points).

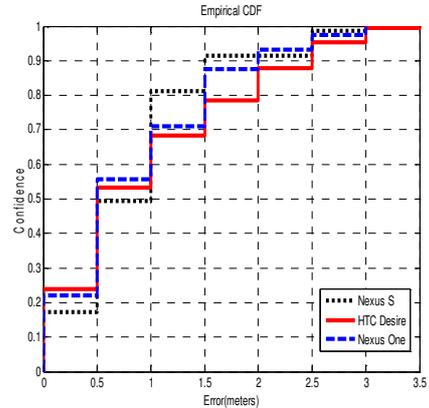


Figure 3. Distance estimation accuracy using training set generated by applying propagation model

Figure 3. presents a cumulative distribution function of distance estimation errors, showing again 50th percentile error of 0.5m and 95th percentile error between 2m and 2.5m (for each of the phone model that we tested applying cross-validation method across different environments). Gaussian processes regression achieved also median error of 0.5m while 95th percentile of 2.5m but for brevity reasons we presented only classification method (Bayesian with KDE). Therefore, by investing a minimal effort in performing the calibration for a couple of minutes at a single distance, we demonstrated that it is possible to achieve comparable performance to acquiring a full training set (Figure 2. and Figure 3.).

B. Relative Body Orientation

Relative body orientation refers to the angle between the orientations of torsos [19] considering two subjects that are facing each other. To recognize the relative body orientation of subjects carrying mobile phones, we use the embedded orientation sensor that provides the following values (expressed in degrees): Azimuth – the angle between the magnetic north direction and the y-axis, around the z-axis (0° to 359°); Pitch – the rotation around x-axis (-180° to 180°) with positive values when the z-axis moves towards the y-axis; and Roll – the rotation around y-axis (-90° to 90°) with positive values when the x-axis moves towards the z-axis. Knowing the relative position between the body and the phone orientation is a fundamental condition in order to recognize the individual’s body orientation and the relative body orientation between subjects. Once this relationship is determined, calculating the relative body orientation would require relative processing of *azimuth*, *pitch* and *roll* values. In our experiments, we were always aware of the exact position where participants carried the phone. However, in the recent study, Shi et al. [20] demonstrated that it is possible to automatically detect on-body position of the mobile phone by utilizing the fusion of accelerometer and gyroscope.

It will be shown that spatial setting detection can be used as an indicator of an ongoing social interaction. However, in order to ascertain whether or not two people are actually having a conversation it requires also knowledge about speech activity, which we describe in the next section.

¹ In the experiments that will be presented in Section V we tested also three additional models, Samsung Galaxy S, HTC Desire S and HTC ChaCha that provided similar accuracy.

III. ACCELEROMETER TO DETECT SPEECH ACTIVITY

Vocal chords (also known as vocal folds) are muscles within larynx that vibrate when air from lungs passes through thus producing voice [21]. The fundamental frequency of vocal chords vibrations depends on a variety of factors including age, gender and individual differences [22]. After the age of 20 the predicted fundamental frequency remains approximately 100Hz for male and 200Hz for female adults [22]. Our approach to recognizing verbal communications is based on detecting vibrations generated by phonation using off-the-shelf accelerometers. Since mounting sensors on the neck (close to the larynx area) may be obtrusive, we opted for the chest as a suitable body position, which is already being used for placing various sensors including cardio, respiratory and kinematic sensors. Sundberg [23] identified a number of factors that contribute to the chest vibrations during phonation and examined the distribution of displacement amplitude over the chest wall surface, demonstrating that the vibrations can be detected all over the chest with the highest displacement amplitude located in the central part of the sternum, which is the area that we chose to place the sensor on (Figure 4.). This position is also convenient for attaching a sensor with an elastic band (similarly to attaching respiratory or cardio sensors) minimizing the interference with typical daily routines.

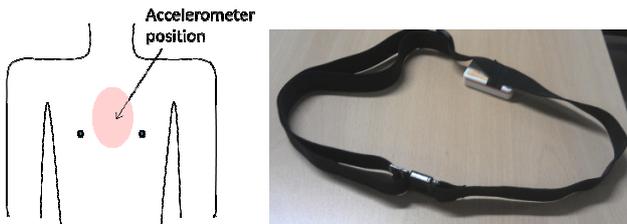


Figure 4. Area on the chest for placing the accelerometer (left); Shimmer accelerometer on the elastic band (right)

Specialized accelerometers for analyzing vibrations have found their applications in a number of fields such as sensing oscillations in bridges, unbalance in rotating machines, fluid flow or acoustic energy input. Using these kinds of sensitive accelerometers to detect the amplitudes of the chest vibrations would provide highly accurate detection of phonation. However, our goal was to detect chest wall vibrations (and hence speech activity) using only off-the-shelf accelerometers. In addition, accelerometers are used in a wide variety of applications (such as physical activity monitoring, for example), which could benefit from including speech recognition, thus providing new insights and more comprehensive description of daily activities without additional sensing equipment. Olguin and Pentland [24] demonstrated that it is possible to recognize some of the most common activities using a single accelerometer attached to the chest. Furthermore, the approach of using accelerometer for detecting speech does not include recording of sensitive data (such as voice) thus addressing the aspect of privacy concerns.

A. Our Approach

During bare foot walking 99% of the acceleration power is below 15 Hz, during running it is below 18 Hz while most of other daily movements occur between 0.3 Hz and 3.5 Hz [25]. Therefore, physical activities are not expected to overlap with vocal chords vibrations in the frequency domain. However, our

investigation is focused on the following two aspects: 1) whether the characteristics of off-the-shelf accelerometers (i.e. not specifically designed for detecting small vibrations) are sufficient for recognizing speech activity and discriminate it from other components in the frequency spectrum – this concern refers mostly to low amplitudes of the chest wall vibrations [23] that may be similar to noise level, imperfect contact between the sensor and the chest, and physiological and acoustic differences between genders [22] and also across all individuals; 2) whether other sources of vibrations encountered in everyday life including elevator, car, bus, train or airplane, whose engines provide components in higher frequency ranges that may result in false positives for speech detection.

To evaluate our approach of detecting speech activity based on analyzing frequency spectrum of data acquired from an off-the-shelf accelerometer attached to the chest (Figure 4.), we used the accelerometer produced by Shimmer, which is not specifically adapted to detect small vibrations (the sensor used in our experiments comes as a part of an ECG device [26]). The specifications are the following: the range of ± 1.5 and $\pm 6g$, sensitivity of 800mV/g at 1.5g and a maximal sampling rate of 512Hz. According to the Nyquist-Shannon sampling theorem, the ceiling boundary frequency component that can be detected using this accelerometer is 256 Hz, which fulfils the requirements for our application (since the fundamental frequencies of vocal chords are approximately 100Hz for males and 200Hz for females). To analyze the frequency domain of acceleration time series (square roots of the sum of the values of each axis x , y and z squared), we performed the Discrete Fourier Transform (DFT) defined² for a given sequence x_k , $k = 0, 1, \dots, N-1$ as the sequence X_r , $r = 0, 1, \dots, N-1$ [27]:

$$X_r = \sum_{k=0}^{N-1} x_k e^{-j2\pi rk/N}$$

Frequency spectrum was analyzed in Matlab applying the Fast Fourier Transform (FFT) to calculate the DTF and then we computed the power spectral density.

As we expected, low amplitudes of the chest wall vibration were similar to the noise level and we were unable to distinguish accelerometer readings that contained speech from those that contained noise, only by analyzing the frequency spectra. In order to tackle the problem of noise, we applied a simple noise cancelling strategy [28] of summing frequency spectra in time which is based on the assumption that the signal components are always focused in the same frequency range in contrast to noise that is, in this case, more random. Considering time frames for performing power spectral density analysis, we found that the best accuracy was achieved by analyzing a sum of power spectral densities computed separately for five consecutive 2-second long time series (corresponding to 1024 samples in our case). Hence, each 10-seconds frame was represented with the power spectral density that was a sum of spectral densities computed for each 2 seconds. Therefore, our goal was to recognize the presence of spectral components that correspond to speech with the resolution of 10 seconds. Processing data in 10-second time frames resulted in the highest accuracy regardless of the

² Some conventions place the factor $1/N$ in front of the transform [27]

duration of the speech i.e. whether there was only one word spoken or a continuous talk of 10 seconds. Decreasing the resolution corresponded to lower ratio between speech amplitudes and noise levels while processing data in longer time units was more likely to fail in detecting shorter durations of speech.

We investigated various classification algorithms (namely SVM, Naïve Bayes, Naïve Bayes with kernel density estimation and k-NN) and parameters for characterizing the spectral density (namely mean, maximal, minimal, and integral values regarding different frequency ranges). It turned out that Naïve Bayes with kernel density estimator applied on the two parameters – integral and mean values of the components between 80 Hz and 256 Hz, provided the highest classification accuracy. Here we do not further elaborate on the classification selection, a choice of signal parameters, frame size for calculating power spectral density and the resolution since they strongly depend on the accelerometer’s characteristics. However, considering the predicted fundamental frequencies of vocal chords, the requirement for the accelerometer is sampling rate of minimum 200 Hz for detecting male voices and 400 Hz for detecting female voices. In the following, we report the accuracy of our approach.

B. Experiments and Results

In total, 21 subjects participated in the speech activity detection experiment (11 males, 10 females; 31.8 ± 7.6 years old). Each subject was asked to read out loud the article from the latest newspapers for at least two minutes, while having the Shimmer accelerometer attached to the chest with an elastic band. We evaluated the performance of our approach separately for each subject through cross-validation of two sets, one including the frequency spectra of 10-second frames containing subject’s voice and the other including only spectra of accelerometer data samples recorded during mild physical activities without voice. 10 out of 11 male and 9 out of 10 female voices were successfully recognized, demonstrating that in large majority of the cases the accelerometer was sufficient to distinguish the frequency spectra of readings with and without voice despite the imperfect skin-sensor contact and individual subjects’ characteristics.

In addition, we created a set of accelerometer data that contained speech activity of 19 subjects excluding 2 subjects that were not previously detected (overall, 2 minutes each subject, that is 38 minutes, divided in 10-second time frames) and accelerometer readings that contained physical movements without voice (approx. 2 hours of accelerometer readings that included sitting, standing and normal speed walking in 10-second data resolution). This was done so that we can build a generic speech detection model. The voice recognition accuracy was estimated through leave-one-out method of sequentially selecting accelerometer readings that corresponded to one subject/one activity as a test unit while using the rest of the set for building the model (training set for Naïve Bayes with KDE classification). The voice was correctly recognized in 93% of cases while mild physical activities without voice induced false positives in 19% (TABLE I. a). The same model was used to test accelerometer readings acquired in more intensive activities such as fast walking or running which resulted in 29% rate of false positives (TABLE I. b). Furthermore, we investigated whether some sources that may

be encountered in everyday life including elevator (5min of data), car (30min of data), bus (30 min of data), train (20min of data) or airplane (1 hour of data) whose engines provide components in higher frequency ranges result in false positives in speech detection. It turned out that elevator, train and airplane do not present an additional issue for the speech recognition, causing the same rate of false positives as physical movements performed in normal conditions (TABLE I. a and TABLE I. c) while travelling in a car or a bus increases the occurrence of false positives to the rate of 32%. It should be mentioned that intense physical activities and the transportation vehicles did not affect the recognition of speech i.e. the rate of true positives and false negatives remained above 90%.

TABLE I. A) VOICE/MILD ACTIVITIES, B) VOICE/INTENSIVE ACTIVITIES, C) VOICE/SOURCES OF HIGHER FREQUENCIES

a)		b)	
	Voice Detected	No Voice Detected	Fast Walking or Running
Voice	93%	7%	71%
Mild Activities	19%	81%	29%

c)		Elevator	Bus/ Car	Train	Airplane
No voice detected (true negatives)		80%	68%	81%	79%
Voice detected (false positives)		20%	32%	19%	21%

In addition to phonation there are other causes of vocal chords vibrations, which can be incorrectly classified as speech activity such as coughing or mumbling; however, their occurrence is less frequent and typically negligible in comparison to speech.

Our approach demonstrates that the speech activity can be reliably detected in typical daily situations except in vehicles (such as car or bus) whose engine frequencies may result in a higher rate of false positives. However, this may be mitigated by using a different type of the accelerometer. In the following section we apply this approach in continuous experimental trial and assess the fusion of speech activity detection with spatial settings detection in order to recognize social interactions.

IV. DETECTING SOCIAL INTERACTIONS

Through several studies, Groh et al. developed probabilistic models for detecting social interactions based on interpersonal distance and relative body orientation pairs [19], speech activity [29], or the combination of various parameters [30]. The authors were mainly focused on developing probabilistic models while using existing solutions for detecting the underlying parameters: microphone from MP3-player worn around the neck to identify speech activity [29] and to recognize interpersonal distances and relative body orientation they relied on a commercial, multiple-camera infrared beacon-based tracking system, where the cameras were mounted on the floor and ceiling and beacons attached on the body [19]. Groh et al. demonstrated that using the abovementioned parameters was sufficient for providing high accuracy in detecting social interactions, but emphasized the importance and potential benefits of using mobile sensing paradigms instead of the camera/beacon-based system due to obvious infrastructure costs.

We have developed mobile modalities for sensing parameters relevant for social interactions and in this section we will assess whether the achieved precision of detected

parameters is sufficient to identify existing social interactions. For the rest of this section, we first analyze a relative predictive power of using detected spatial parameters to identify co-located, face-to-face social interactions. Afterwards, we investigate the potential of fusing speech activity and spatial settings detection.

A. Detecting social situations through spatial settings

The study of proxemics [31] investigated how people unconsciously organize space around them corresponding to different degrees of intimacy. A number of other factors also affect absolute measures of interpersonal distances (denoted as d) and relative body orientations (denoted as α) such as cultural background, gender, or types of personalities [32]. Detection of social interactions that will be presented in this section is based on analyzing spatial parameters between a pair of subjects that carry mobile phones. If more than two subjects are involved in the same conversation, our method recognizes other participants by examining information for each pair of individuals involved in the social interaction. On the other hand, as the number of participants increases, interpersonal distances expand and the angles become wider, thus putting constraints on developing a single model of social interactions, regardless of the number of participants. However, these effects (such as changes in angles) are typically neglected in the literature since practical experience suggests that when there are more than four or five individuals, they frequently split up into sub-situations [19] [33]. Therefore, the experiments that follow were conducted under the assumption that in real-life setting a number of individuals which actively participate social interaction is limited to four of five (usually referred to as a “small-group” interaction) [19] [33].

We chose the time frame of 10 seconds to process data as suggested by [30] in order to capture dynamic changes in social interactions while at the same time to discriminate between existing and non-existing social interactions. Therefore, interpersonal distances were estimated using a sequence of Wi-Fi RSSI values for every 10-second frame while body orientation is averaged for every 10 seconds (i.e. 10 samples). Relative body orientation of subjects was considered only if the standard deviation of the samples was less than or equal to 10 degrees for each subject (regarding the 10-second time frame), otherwise the current frame of samples was left out. This was done in order to analyze situations in which subjects held stable relative orientation, such that random body movements are removed as a source of orientation uncertainty. The threshold of 10 degrees was confirmed to be a trade-off between decreasing the standard deviation of the estimated relative body orientation (proportional to decreasing threshold) and decreasing the amount of discarded data (proportional to increasing threshold). Overall, approximately 20% - 25% of unstable orientation data was discarded. We installed the application in five phones, two *HTC Desire*, two *HTC Desire S* and one *Samsung Galaxy S* with synchronized clocks to ensure correct data aligning which was crucial, considering a short time frame of 10 seconds for data analysis. Focusing on small-group co-located face-to-face social interactions, we performed the experiments in two types of scenarios:

1) Scenario 1

Participants, that partially knew each other, were asked to communicate for an amount of time of their choosing, while

carrying the mobile at a known place. The first trial involved 6 participants (4 males, 2 females, age: 31 ± 4 years) that were talking to each other, maximum four at a time, at 14 randomly selected locations, including 12 indoor and 2 outdoor environments. The duration of these interactions was 5.6 ± 3.8 minutes. The second trial was conducted in a meeting room and it consisted of two 15-minute sessions in each involving 4 people (6 males, 2 females, age: 29 ± 4 years) who were let to communicate freely as they wanted. This experimental trial resulted in 1300 pairs of relative body orientation and interpersonal distance (α, d) with a time frame window of 10 seconds for processing and averaging data.

2) Scenario 2

The break room is the place where employees in our research center typically socialize. This created the opportunity to monitor social interaction in a natural setting. When people were coming to the break room, we asked them to place the phone in a case attached on the right hip and to continue their interaction. Overall, we recorded 15 interactions of duration 6.2 ± 3.5 minutes that included 24 different people. This experimental trial resulted in 1300 orientation, distance (α, d) pairs.

In order to assess the potential of using spatial (α, d) parameters to distinguish existing and non-existing social interactions, it was necessary to create also a solid corpus of the pairs that do not correspond to social interactions. Four subjects that attended a fair called “Researchers Night” were monitored while being asked to report any social encounter among them. Measurements from one-hour period in which they reported no social interactions was extracted as a suitable data set containing overall 1400 (α, d) pairs for creating non-existing social interaction corpus; being at the stand implied their constant proximity and random relative body orientations (while sitting/standing/moving). In addition, we added measurements from previously described Scenario 1 (2nd trial) that included subjects that were in concurrent social interactions and in a close proximity (all social encounters occurred within 5x5m space). Certainly, there are a number of scenarios which include subjects that are not interacting while being in close vicinity, thus making it challenging to classify between occurrence and non-occurrence of social interaction. The example of one such a case involves colleagues that share the office, sitting at their desks across each other at a short distance, but not having a conversation. This case will be evaluated in the continuous experimental trial, presented in the next section.

TABLE II. presents the results of distinguishing existing (denoted as SI) and non existing social interaction cases (denoted as NonSI) represented with a feature-vector (α, d) by applying Linear Classification and Naïve Bayes with KDE techniques. The classification performance was evaluated using 10-fold cross validation.

TABLE II. CLASSIFICATION RESULTS (ORIENTATION, DISTANCE)

	Naïve Bayes (KDE)		Linear Classifier	
	SI/ NonSI	SI/ NonSI	SI/ NonSI	SI/ NonSI
SI	79%	21%	73%	27%
NonSI	24%	76%	12%	88%

The results demonstrate the accuracy of 79% in detecting social interactions based on interpersonal distance and relative body orientation. Naïve Bayes with KDE performed slightly better in identifying social interaction pairs while Linear Classifier provided lower rate of false positives. A contributing factor to a relatively high accuracy is also a simple method of taking out of the consideration (α , d) pairs corresponding to the situations in which subjects did not hold a stable relative body orientation, thus eliminating the source of uncertainty created in most cases by random body movements. However, instead of using the standard deviation (SD) of relative body orientation for identifying “unstable” (α , d) pairs, we attempted to use it also as a classification feature that can be considered as an index of holding stable relative position of participants in a social interaction. SD of relative body orientation (denoted with σ) was also calculated for each 10-second frame (i.e. for 10 samples) and combined with distance d and averaged relative body orientation α , constituting 2-feature vector (σ , d) and 3-feature vector (σ , α , d). TABLE III. shows the results of 10-fold cross validation.

The combination of interpersonal distance and SD of relative body orientation provided higher accuracy in comparison to the previous case of using relative body orientation angle (TABLE II.). This may be due to the fact that feature-vector (σ , d) does not discriminate classes based on the absolute angle between body orientations in social interactions thus allowing for more situations to be included in the model in comparison to feature-vector (α , d). As expected, this resulted in a higher rate of false positives that occurred mostly when subjects were in a close proximity, having a stable body orientations but not interacting (for instance, sitting or being in concurrent social interactions). The highest accuracy was achieved using 3-feature vector (σ , α , d) that resulted in 93% of successfully classified vectors corresponding to social interactions and 26% of false positives.

TABLE III. CLASSIFICATION RESULTS (2-FEATURE AND 3 FEATURE VECTORS)

	(σ, d)				(σ, α, d)			
	Naïve Bayes (KDE)		Linear Classifier		Naïve Bayes (KDE)		Linear Classifier	
	SI/ NonSI	SI/ NonSI	SI/ NonSI	SI/ NonSI	SI/ NonSI	SI/ NonSI	SI/ NonSI	SI/ NonSI
SI	89%	11%	76%	24%	93%	7%	90%	10%
NonSI	31%	69%	29%	71%	26%	74%	19%	81%

The results demonstrate that the accuracy of estimating interpersonal distances and relative body orientations achieved with mobile phone sensing was sufficiently discriminative to identify social interactions on a small spatial scale. Note that the position of the phone does not affect SD of relative body orientation thus the model based on 2-feature vector (σ , d) does not require users to carry the phone on a pre-defined/known position on the body (or using algorithms for estimating the phone position [20]).

In the next section, we analyze the performance of detecting social interactions in more challenging conditions - continuous monitoring of co-located subjects.

B. Detecting social interactions using multi-modal mobile sensing

We recruited four subjects that share the same office (3 males and 1 female, 29.0 ± 1.4 years) for 7 working days. Each day, they were carrying the mobile phone at a known and fixed position on the body, typically between 11h and 17h. Accelerometer (embedded in Shimmer device [26]) was attached on the chest using an elastic band which was comfortable for all of the subjects except one that asked to put the sensor over a t-shirt. Overall, there were 40 hours of measurements, resulting in 452 hours of sensor data, 113.0 ± 20.4 hours per person. In order to avoid recording (audio/video) or inquiring participants to label social interactions, ground-truth was annotated by a human observer, a colleague that shares the office with the participants, hence minimizing intrusion in typical daily routines of monitored workers. The observer manually noted each participant’s speech activity and ongoing social interactions while marking all the periods in which the notes were not reliable (usually due to the lack of his presence). Labeling structure of ground-truth was divided in the two categories: 1) participation in social interactions and speech activity (present/not present) for each participant annotated every minute (52% of experimental data), 2) the existence of an ongoing social interaction, without a minute-by-minute description, including identification of participants and duration of interaction (27% of data). The rest of the data (21% of overall measurements) were lacking labels. The second category of labeling ground-truth corresponded mostly to longer discussion, conversations during lunch or coffee breaks and similar occasions during which it was cumbersome for the observer to take precise notes. The annotations were taken also for social interactions with non-monitored subjects.

The accelerometer data was processed providing a binary result (1/0) for every 10-second frame that indicated speech activity or not. TABLE IV. shows the results for each participant separately, analyzing only the portion of the data that was precisely annotated (minute-by-minute) – overall 30.4 ± 9.5 hours per person. According to the structure of annotations, true positives denote the percentage of speech-labeled minutes in which speech activity was detected in at least one 10-second frame within that minute while false positives represent the percentage of minutes in which the speech was detected but not annotated in the observer’s notes. It can be seen that the accuracy is slightly worse than in the case of shorter experimental trials (Section III). This was expected considering that due to daily activities the elastic band can move thus causing slipping or detaching sensor from the skin surface. However, our prototype was an improvised elastic band (not purpose-manufactured) and an accelerometer that was not designed to be stuck to the skin. The lowest accuracy corresponds to the participant that was wearing the band over a t-shirt (Subject 1) that may have mitigated the chest vibrations.

TABLE IV. SPEECH ACTIVITY DETECTION

	Subject 1	Subject 2	Subject 3	Subject 4
	SI/Non-SI	SI/Non-SI	SI/Non-SI	SI/Non-SI
SI	67%/33%	77%/23%	73%/27%	73%/27%
Non-SI	18%/82%	21%/79%	29%/71%	25%/75%

Unfortunately, the resolution of 10 seconds for speech activity status was not sufficient for detecting turn-taking patterns [30] in order to identify conversation between two individuals while decreasing the frame length for processing accelerometer data resulted in speech recognition accuracy degradation. Therefore, by relying only on speech detection it was not possible to differentiate if two or more subjects have a conversation or participate in concurrent social interactions. Rather, speech activity detection was used complementary with spatial settings recognition that will be described in the paragraphs that follow.

Regarding spatial settings, we have built a classification model for social interactions using the social interactions that occurred during the experiments. In the test phase we excluded the whole day to which the tested social encounter belonged, hence applying a 7-fold cross-validation according to 7 days of measurements. We have also built and tested the model for non-existing social interactions in the same way while excluding from the model the situations in which subjects were sitting at their desks. The reason is that due to the office layout (Figure 5.) where the experiment took place, the corresponding feature-vectors labeled as non-social interaction would have yielded a large number of false positives. However, since the situations in which subjects were sitting at their desks corresponded to the periods annotated as non-existing social interactions, they were included in the test phase with all other situations in which no interaction was reported.

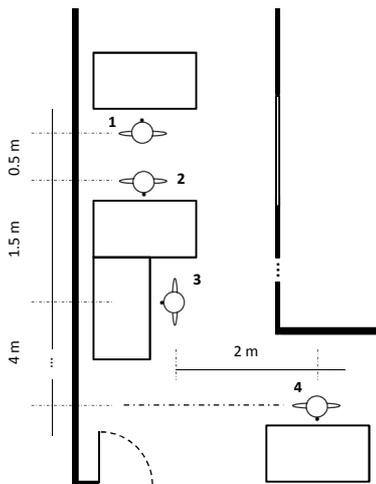


Figure 5. The office layout

Overall, there were 6 hours of social interactions that, according to the annotations, occurred at several locations including office (taking into account all social encounters in the office except previously mentioned “desk-to-desk” conversations), break room, meeting room, and corridors and involving two, three or four monitored workers at a time. On the other hand, 25 hours of non-existing social interactions data were annotated regardless of the location. Approximately 13% of the overall annotated data was discarded due to sporadically missing samples either from compass sensor or Wi-Fi RSSI. The experiments resulted in 1872 and 7420 feature vectors corresponding to existing and non-existing social interactions respectively.

The results are presented in TABLE V. when Naïve Bayes classification and (σ, α, d) model were applied, which previously provided the highest accuracy. 89% of true positives and 11% of false negatives in detecting existing social interactions presents the accuracy across all the pairs of subjects. Since no major differences were witnessed for different pairs of subjects regarding true positives, we reported the accumulated accuracy. Although at an initial look (TABLE V.) the fusion of the two modalities yielded no improvement in the rate of true positives, speech activity detection was used to confirm occurrence of social interactions through the presence of voice of the participants. Otherwise, if no speech was detected, even though spatial settings suggested an occurrence of social interaction, the event was categorized as a non-existing social interaction. This strategy particularly improved the identification of non-existing social interaction when relying solely on spatial settings, which resulted in a very high rate of false positives for subjects pair 1 and 2 and pair 2 and 3 (Figure 5.). A higher rate of false positives was mainly due to the small distance and fixed body orientation of subjects in the office. Particularly for the pair of subjects 1 and 2, the feature vector of the interpersonal distance, relative body orientation and its stability (σ, α, d) , was similar when sitting in the office to the feature vector describing face-to-face social interaction. This is due to the limitation of the proposed spatial settings detection system which has difficulty discriminating between a back-to-back and face-to face position of subjects. However, in our experimental scenario, the fusion of the two sensing modalities significantly improved the overall results. A portion of false positives was resolved by checking speech activity status whenever spatial settings analysis indicated an existing social interaction. In these cases, if the speech activity was not recognized for both subjects during an arbitrarily selected time frame of 5 minutes, the system indicated non-existing social interaction. For the pair of subjects 1 and 2, the results for false positive showed the drop from 76% to 34% and for the pair 2 and 3 from 39% to 29%. The fusion of the two sensing modalities also contributed considerably in resolving false positives for other pairs of subjects that occurred mostly due to their random daily movements. It should be mentioned that this method for resolving false positives did not negatively affect the recognition rate for existing social interactions.

TABLE V. RESULTS OF MULTI-MODAL SENSING OF SOCIAL INTERACTIONS

	Spatial		Spatial + Speech	
	SI/ NonSI	SI/ NonSI	SI/ NonSI	SI/ NonSI
SI	89%	11%	89%	11%
NonSI				
Sub 1&2	76%	24%	34%	66%
Sub 1&3	19%	81%	11%	89%
Sub 1&4	17%	83%	15%	85%
Sub 2&3	39%	61%	29%	71%
Sub 2&4	15%	85%	15%	85%
Sub 3&4	17%	83%	14%	86%

V. DISCUSSION

The level of social activity typically reflects the overall wellbeing status in individuals [5], while social isolation can lead to myriad of health implications even presenting a risk

factor for mortality [1]. Measuring daily amount of social interactions is linked to various disorders including clinical depression, mania and bipolar disorder for which the level of social activity represents an important evidence for diagnosing the current episode or the progress in disorders. Therefore, social interactions become an important aspect for monitoring and assessing, both for healthy subjects and patients. Encouraging healthier lifestyle through the monitoring mobile applications is based on increasing the individuals' awareness of their own daily routines and consequently their wellbeing, while for patients, evaluating the participation in social interactions represents a tool for coaching and diagnosis.

The ultimate goal is to develop a method with the highest precision in collecting interaction data which is fully privacy respecting and invisible from users' perspective, while not restricting the application to a limited number of scenarios. However, this problem relates to a well-known trade-off between the spectrum of collected data and enabling natural experimental conditions, for which there is never a straightforward solution. Furthermore, monitoring patients involves additional constraints with respect to the symptoms of a certain disorder and typically rigorous ethical norms. The proposed solution for gathering interaction data consists of two technologies, namely mobile phone and accelerometer, which creates a unique set of concerns for the adoption in the healthcare setting.

The fact that people habitually carry the mobile phone makes this device a suitable source for unobtrusive and continuous monitoring of their behavior. The applicability of a mobile phone has been demonstrated in monitoring both patients and healthy subjects. However, regarding our design, an accelerometer attached on the chest level may be perceived as obtrusive and consequently it may stigmatize monitored subjects. This issue, while currently a concern, is expected to be mitigated, since accelerometers are increasingly becoming widely adopted both in healthcare and everyday life. In our experiments, in order to test the concept of using an accelerometer for speech detection, we used prototype which was an improvised elastic band (not purpose-manufactured) and the accelerometer not designed to be stuck to the skin. However, the same concept can be applied in a less obtrusive manner, such as embedded in a t-shirt. Moreover, the shape and size of already accepted commercial accelerometer-based solutions can suit also the speech recognition purpose (such as Fitbit [34] – an accelerometer device for tracking wellbeing aspects of individuals' behavior). In the healthcare setting, the chest surface is widely used for attaching cardio and respiratory sensors, thus being an adopted position by patients not only suffering from heart/lung related diseases but also mental disorders such as psychological stress [35] and bipolar disorder [36], for which cardio/respiratory parameters and the social activity are important aspects to assess. However, regarding certain health populations, wearing an additional sensor on the body may be physically and/or psychologically burdensome with respect to the symptoms of a particular disorder. In addition, the accelerometer-based speech detection does not provide a reliable solution for users that prevalently suffer from symptoms which may cause non-phonation related chords vibrations such as mumbling or coughing. Otherwise, relying on accelerometer as an alternative to the use of microphone can be an acceptable compromise to prevent privacy concerns

while providing a mobile solution for continuous monitoring of speech activity.

VI. CONCLUSIONS

Classical methods of measuring social interaction include diaries, questionnaires and similar self-reports tools that are limited in detecting social interactions which take place on spatial scales of meters and time scales of minutes while also suffering from recall bias issue. The goal of this work is sensing small group social interactions that occur on small temporal and spatial scale relying on widely available technologies – accelerometer and smart phone.

We exploited compass embedded in smart phones to measure body orientations and Wi-Fi transmitting/receiving mechanisms for distance estimation. These parameters were further combined in feature vectors demonstrating a high predictive power of social interaction detection. The study also suggested the use of standard deviation of relative body orientation as a suitable classification feature representing an index of stable relative position of participants in a social encounter. Such measure does not require users to carry the phone on a pre-defined position on the body or using complex algorithms to estimate the phone position. Mobile phone is typically a well-adopted technology both in everyday and healthcare settings, allowing continuous data collection with a minimal effect on the users' behavior and consequently, their social interaction patterns.

By using an off-the-shelf accelerometer, we demonstrated that speech activity can be reliably recognized by detecting vibrations at the chest level that are generated by vocal chords during phonation. This approach is not expected to raise privacy concerns, in comparison to typical microphone-based methods while being easily applicable and cost effective solution for recognizing speech activity considering that accelerometers are widely used sensors (being already commonly used for detecting a number of activities). However, wearing an accelerometer at the chest level can be burdensome for certain health populations which should be considered when designing the system.

The two proposed sensing modalities provide complementary information about social interactions constituting a multi-modal mobile sensing system. It was demonstrated that the system provides the accuracy of 89% when tested in continuous, real-life experiments over 7 working days. Considering the fact that the level of participation in social interactions has been shown to have an impact on various aspects of health, the proposed system provides a tool for monitoring one component of wellbeing of healthy subjects. Although this work so far has focused on healthy subjects, we envision its applicability in various healthcare settings.

VII. ACKNOWLEDGEMENTS

We would like to acknowledge Iacopo Carreras and Piret Saar for their assistance and development of the mobile phone application.

This work was supported by EU FP7 MONARCA project – <http://monarca-project.eu>

REFERENCES

- [1] J. S. House, K. R. Landis, and D. Umberson, "Social relationships and health," *Science (New York, N.Y.)*, vol. 241, no. 4865, pp. 540-5, Jul. 1988.
- [2] N. D. Lane et al., "BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing," in *5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth2011)*, 2011.
- [3] V. Isaac, R. Stewart, S. Artero, M.-L. Ancelin, and K. Ritchie, "Social activity and improvement in depressive symptoms in older people: a prospective community cohort study," *The American journal of geriatric psychiatry official journal of the American Association for Geriatric Psychiatry*, vol. 17, no. 8, pp. 688-696, 2009.
- [4] H. B. Bosworth, J. C. Hays, L. K. George, and D. C. Steffens, "Psychosocial and clinical predictors of unipolar depression outcome in older adults," *International Journal of Geriatric Psychiatry*, vol. 17, no. 3, pp. 238-246, 2002.
- [5] M. Rabbi, T. Choudhury, S. Ali, and E. Berke, "Passive and In-situ Assessment of Mental and Physical Well-being using Mobile Sensors," in *13th International Conference on Ubiquitous Computing (UbiComp'11)*, 2011.
- [6] H. R. Bernard, P. Killworth, D. Kronenfeld, and L. Sailer, "The Problem of Informant Accuracy: The Validity of Retrospective Data," *Annual review of anthropology*, vol. 13, no. 1, pp. 495-517, 1984.
- [7] T. Choudhury and a. Pentland, "Sensing and modeling human networks using the sociometer," *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.*, no. 1997, pp. 216-222, 2004.
- [8] C. Cattuto, W. V. den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani, "Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks," *PLoS ONE* 5(7): e11596. doi:10.1371/journal.pone.0011596, 2010.
- [9] A. Barrat, C. Cattuto, V. Colizza, L. Isella, A. E. Tozzi, and W. V. D. Broeck, "Wearable sensor networks for measuring face-to-face contact patterns in healthcare settings," in *3rd International ICST Conference on Electronic Healthcare for the 21st century*, 2010, pp. 1-4.
- [10] N. Eagle and A. (Sandy) Pentland, "Reality mining: sensing complex social systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255-268, Nov. 2005.
- [11] T. M. T. Do and D. Gatica-Perez, "Contextual grouping: discovering real-life interaction types from longitudinal Bluetooth data," *infoscience.epfl.ch*.
- [12] N. Banerjee, S. Agarwal, P. Bahl, R. Chandra, A. Wolman, and M. Corner, "Virtual compass: relative positioning to sense mobile social interactions," *Pervasive Computing*, pp. 1-21, 2010.
- [13] M. L. Robbins, M. R. Mehl, S. E. Holleran, and S. Kastle, "Naturalistically observed sighing and depression in rheumatoid arthritis patients: a preliminary study," *Health psychology official journal of the Division of Health Psychology American Psychological Association*, vol. 30, no. 1, pp. 129-133, 2011.
- [14] M. L. Robbins, E. S. Focella, S. Kastle, A. M. López, K. L. Weihs, and M. R. Mehl, "Naturalistically observed swearing, emotional support, and depressive symptoms in women coping with illness," *Health psychology official journal of the Division of Health Psychology American Psychological Association*, vol. 30, no. 6, pp. 789-92, 2011.
- [15] M. Cristani, A. Pesarin, and A. Vinciarelli, "Look at who's talking: Voice activity detection by automated gesture analysis," in *Workshop on Interactive Human Behavior Analysis in Open or Public Spaces*, 2011.
- [16] A. Popleteev, "DISI - University of Trento Indoor positioning using FM radio signals," no. April, 2011.
- [17] P. Bhagwat, B. Raman, and D. Sanghi, "Turning 802.11 inside-out," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 1, p. 33, Jan. 2004.
- [18] A. Matic, A. Papiatseyeu, V. Osmani, and O. Mayora-Ibarra, "Tuning to your position: FM radio based indoor localization with spontaneous recalibration," *2010 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 153-161, Mar. 2010.
- [19] G. Groh, A. Lehmann, J. Reimers, M. R. Frieß, and L. Schwarz, "Detecting Social Situations from Interaction Geometry," in *IEEE International Conference on Social Computing/IEEE International Conference on Privacy, Security, Risk and Trust*, 2010, pp. 1-8.
- [20] Y. Shi, Y. Shi, and J. Liu, "A rotation based method for detecting on-body positions of mobile devices," in *Proceedings of the 13th International Conference on Ubiquitous Computing*, 2011.
- [21] "Medicine Net." [Online]. Available: <http://www.medterms.com/script/main/art.asp?articlekey=6224>. [Accessed: 15-Nov-2011].
- [22] I. Titze, "Physiologic and acoustic differences between male and female," *J. Acoust. Soc. Am.*, pp. 1699-1707, 1989.
- [23] J. Sundberg, "Chest wall vibrations in singers," *Journal Of Speech And Hearing Research*, vol. 26, no. 3, pp. 329-340, 1983.
- [24] D. Olgu and A. S. Pentland, "Human Activity Recognition : Accuracy across Common Locations for Wearable Sensors," *Networks*, pp. 5-7, 2006.
- [25] M. J. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler, "Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *Physiological Measurement*, vol. 25, no. 2, p. R1-R20, Apr. 2004.
- [26] "Shimmer - Wireless Sensor Platform for Wearable Applications." [Online]. Available: <http://www.shimmer-research.com/p/products/sensor-units-and-modules/wireless-ecg-sensor>. [Accessed: 15-Nov-2011].
- [27] "Linear Systems, S.M. Tan, The University of Auckland, Chapter 9 The Discrete Fourier transform," pp. 1-8.
- [28] B. Widrow, J. G. Jr, and J. McCool, "Adaptive noise cancelling: Principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 105-112, 1975.
- [29] G. Groh and C. Fuchs, "Combining Evidence for Social Situation Detection," *in.tum.de*.
- [30] G. Groh, A. Lehmann, and M. D. Souza, "Mobile Detection of Social Situations with Turn Taking Patterns," in *WAC2011*, 2011.
- [31] E. Hall, *The Hidden Dimmesion*. New York: Double Day Anchor Books, 1966.
- [32] G. Paggetti, A. Vinciarelli, I. Italiano, and G. It, "Towards Computational Proxemics: Inferring Social Relations from Interpersonal Distances."
- [33] D. Gatica-Perez, "Automatic nonverbal analysis of social interaction in small groups: A review," *Image and Vision Computing*, vol. 27, no. 12, pp. 1775-1787, Nov. 2009.
- [34] "Fitbit." [Online]. Available: <http://www.fitbit.com/>. [Accessed: 10-Mar-2012].
- [35] "Interstress project." [Online]. Available: <http://interstress.eu/>. [Accessed: 10-Mar-2012].
- [36] "Monitoring, treatment and prediction of bipolar disorder episodes (MONARCA)." [Online]. Available: <http://www.monarca-project.eu/>. [Accessed: 10-Mar-2012].