# Trade-offs in Monitoring Social Interactions

*Aleksandar Matic, Venet Osmani, and Oscar Mayora, CREATE-NET*

## ABSTRACT

Social interaction is one of the basic components of human life that impacts thoughts, emotions, decisions, and the overall wellbeing of individuals. In this regard, monitoring social activity constitutes an important factor in a number of disciplines, particularly those related to social and health sciences. Sensor-based social interaction data collection has been seen as a groundbreaking tool, having the potential to overcome the drawbacks of traditional self-reporting methods and revolutionize social behavior analysis. However, monitoring of social interactions typically implies a trade-off between the quality of collected data and the levels of unobtrusiveness and privacy respect, aspects that can affect spontaneity in subjects' behavior. In this article we discuss the challenges of automatic monitoring of social interactions; then we provide an overview of the current automatic monitoring concepts and the associated trade-offs. We finally present our approach of using non-visual and non-auditory mobile sources that mitigate privacy concerns and do not interfere with individuals' daily routines, while providing a reliable platform for social interaction data collection.

## INTRODUCTION

*"Man is by nature a social animal; an individual who is unsocial naturally and not accidentally is either beneath our notice or more than human. Society is something that precedes the individual. Anyone who either cannot lead the common life or is so self-sufficient as not to need to, and therefore does not partake of society, is either a beast or a god."*

Aristotle (384 BC–322 BC)

Despite the fact that the interest of humanists in understanding social behavior dates back to the times of ancient civilizations, it is significant to note that first incidences of scientific data collection on human interactions took place in the beginning of the 20th century, relying on surveys or engaging a human observer who was taking notes about social interactions within monitored groups. Nowadays, a century later, the same methods for analyzing social behavior are still prevalent in social and health sciences, although they have a number of shortcomings. Periodical surveys, diaries, and similar self-reporting methods suffer from memory dependence, recall bias, and high end-user effort for continuous long-term monitoring [1]. Moreover, they correspond poorly to communication patterns as recorded by independent observers [2]. Albeit a more reliable method, relying on a human observer to record social interactions in groups is inefficient, particularly if the size of the group is large, the interactions occur in various physical locations, or the study requires longitudinal data collection [3].

The advent of sensor-based instruments for recording social activity of individuals is considered to be a critical point in the evolution of social behavior analysis, exhibiting the potential to overcome the limitations of self-reporting and observational methods [1]. Buchanan [4] envisioned that sensors will transform social sciences as much as microscopes transformed medicine in the 18th and the 19th century. Undoubtedly, pervasive computing paradigms have already enabled new findings on social interaction phenomena by providing automatic recognition of social encounters as well as insight into domains that are difficult or impossible to record by hand-annotating methods. However, despite the rapid development of technology, health and social scientists still do not rely on automatic tools to a great extent.

The drawbacks of the current sensor-based methods can shed light on why self-reports are still prevalent for collecting social interaction data. Existing solutions for recognizing social interactions mostly require expensive infrastructures, which spatially constrain applications, involve devices that are often not available off the shelf, provide limited accuracy in gathering real-time data with spatial and temporal granularities, or make use of microphones/cameras, the activation of which may raise privacy concerns in monitored subjects. Furthermore, acquiring high-quality social interaction data typically requires use of more invasive methods that tend to affect the natural behavior of subjects and consequently the reliability of measurements. When monitoring social interactions there is a trade-off between the quality of collected data and the level of attaining real-life conditions in experiments.

In this article, we analyze aspects of sensor-based approaches for monitoring social interactions with the focus on trade-offs of approximating natural experimental settings: the level of obtrusiveness, respecting the subject's


0163-6804/13/$25.00 © 2013 IEEE

privacy, and spatial restrictions. We consider social interactions that occur on a small spatio-temporal scale (i.e., collocated face-to-face conversations) to which we refer in the rest of this article. We provide an overview of the current approaches to collecting social interaction data and discuss the main trade-offs with respect to different monitoring solutions. Furthermore, we propose the concept of sensing social interactions by using non-visual and non-auditory mobile sources that do not capture privacy-sensitive data, do not spatially limit the applications, and minimize interference with typical daily activities.

## The Challenges of Automatic Monitoring of Social Interactions

*"To observe is to disturb."*
Werner Heisenberg (1901–1976)

Monitoring social interactions represents an important aspect of social behavior analysis, a domain that has a wide-reaching multidisciplinary impact. These disciplines range from medicine, where quantitative evaluation of social activity represents a tool in coaching and diagnosis, to economics where social relationships are used to model both micro- and macroeconomic phenomena, to anthropology, which analyzes differences in social behavior across different cultures, to epidemiology, which examines interpersonal contacts as the main cause behind the spread of an epidemic, to social psychology, which studies how individuals' thoughts, feelings, and behaviors are influenced by the presence of other people. It is of interest to all these disciplines to capture and analyze spontaneous social interactions that occur in natural conditions, which pertains to recording people as they freely go about their lives [5]. The ultimate goal is to develop an automatic method that provides the highest precision in collecting social interaction data that is fully privacy respecting and entirely unobtrusive for users. In practice there is typically a trade-off between these aspects — the more privacy respecting and unobtrusive the approach is, the more limited are the possibilities of acquiring social interaction data [6].

### Obtrusiveness

One of the main challenges in the research domain of automatic sensing social interactions and, in general, human behavior is performing data collection in a manner invisible from the subjects' perspective. Having visible sensors, moreover ones that may interfere with daily activities, reminds subjects that they are being monitored, which can influence their behavior, and consequently affect the reliability and objectiveness of measurements. Therefore, extracting the most information out of the least obtrusive sources is the objective when collecting social interaction data. However, this is a challenging problem since noninvasive methods typically result in output that is difficult to process effectively; vice versa, invasive methods provide more detailed information that is easier to process but tends to affect the behavior of monitored subjects [6].

### Privacy

In addition to physical obtrusiveness, monitoring of human behavior is often closely linked to disturbing one's privacy. Privacy issues relate to an array of ethical norms that need to be addressed. All subjects in the study should always know that they are being monitored; moreover, they must have the right to authorize the use and diffusion of the collected data [6]. If monitoring involves audio or video archives, they can be partially or totally deleted by subjects, while recording uninvolved parties without their consent is considered unethical and illegal [5]. However, despite addressing all the ethical norms, people are prone to change their behavior if they have concerns about the method of monitoring, which negatively affects the reliability of the collected data. In particular, the presence of audio/video data analysis becomes an issue to consider. Even though privacy sensitive recording techniques can be applied, the fact that a microphone or a camera is activated may still raise concerns. This often depends on the technical education and cultural background of monitored subjects, which can affect their perception of privacy [7, 8]. On the other hand, protecting privacy often implies discarding sociologically useful information [5], which is not always an acceptable compromise.

The common challenges of automatic monitoring of social interactions (i.e., obtrusiveness and privacy respect) illuminate a well-known trade-off between the spectrum/quality of collected data and enabling natural conditions, where the solution reflects the trade-off. In the following section, we discuss the most common sensor-based concepts for monitoring social interactions and the associated trade-offs.

*The ultimate goal is to develop an automatic method that provides the highest precision in collecting social interaction data, and is fully privacy respecting and entirely unobtrusive for users. In practice, there is typically a trade-off between these aspects.*

## Overview of Existing Sensor-Based Approaches for Collecting Social Interaction Data

A steady decrease in device form factor, coupled with an increase in computational capabilities, has enabled automatic monitoring of many aspects of social behavior, from quantifying dynamics of social activity to extracting various nonverbal behavior cues expressed during social interactions. The choice of sensors and their arrangement in experimental settings determines the level of privacy and obtrusiveness, and the spectrum of interaction data that can be extracted. The use of video/audio infrastructure, wearable dedicated hardware, or mobile phones provide different trade-offs between the quality of collected data and the constraints for experimental settings.

### Video/Audio Infrastructures

Video/audio infrastructure refers to the equipment installed in a specified area for a specific scenario (rather than for a longitudinal study), in order to track social interactions and extract behavioral cues for the analysis. Automatic video/audio analysis of face-to-face social interactions extracts an ample spectrum of information that can provide high scientific and technological

The challenge is how to address monitoring of specific activities relying on existing sensing technologies that are embedded in mobile phones, which is the issue not encountered when using purpose-manufactured devices that already have dedicated sensors incorporated.

value. Since subjects are not required to wear sensors, such systems allow monitoring in a physically non-intrusive manner (except for cases when microphones with headsets need to be attached to the subjects). However, the use of video/audio systems typically implies mobility restrictions to the monitored subjects since video analysis requires a direct line of sight between subjects and cameras, while the audio data is captured from microphones situated within the area of interest. In addition, video and/or audio data can contain privacy sensitive information, thus creating additional issues when monitoring social interactions.

### Wearable Devices

***Dedicated Hardware*** — As opposed to video/audio infrastructures, wearable solutions are mostly used for recording occurrences of social interactions and quantifying dynamics of social activity on a long-term scale. To achieve high accuracy in detecting the occurrence of face-to-face social interactions in a mobile way requires knowledge of both the proximity of subjects and their speech activity status. In order to infer speech activity status, dedicated wearable devices typically involve audio analysis, which can face ethical issues and privacy concerns. Besides, most dedicated devices for inferring face-to-face contacts require a direct line of sight between two units, which imposes a specific position on the body for their placement; therefore, such approaches are prone to affect the natural behavior of the subjects since they can interfere with daily activities.

One way to address the issue of stigmatizing subjects is to utilize the sensing capabilities available in one of the most familiar devices: the mobile phone.

***Mobile Phone Sensing*** — The rapid adoption of mobile phones brings the opportunity for unobtrusive and continuous monitoring of social interactions and, in general, individuals' behavior [1]. The challenge is how to address monitoring of specific activities relying on existing sensing technologies that are embedded in mobile phones, which is an issue not encountered when using specific-purpose-manufactured devices that already have dedicated sensors incorporated.

Current work on mobile phone sensing to detect social interactions has relied mostly on using Bluetooth to sense nearby mobile phones. Using Bluetooth as a proximity sensor to reconstruct social dynamics on a large scale has been extensively investigated under the umbrella of the reality mining initiative [9]. Since the Bluetooth communications range is on the order of 10 m, this approach provides only coarse spatial granularity in recognizing interpersonal distances; therefore, knowledge of the proximity of individuals is used to model the dynamics of social interactions on a large scale rather than to detect each single social encounter that takes place on a small spatio-temporal scale.

In order to address the limitation of Bluetooth scans to detect actual face-to-face proximity between subjects, the Virtual Compass project [10] estimates interpersonal distances using received signal strength indicator (RSSI) analysis of Bluetooth and Wi-Fi signals. By applying empirical propagation models, the approach achieves the median accuracy between 0.9 m and 1.9 m; however, the lack of subjects' orientation information and of speech activity might not be sufficient for a highly accurate detection of face-to-face social interactions. As an alternative approach, recent research work [5] is extracting audio data features using microphones from a pair of collocated mobile phones in order to detect who was speaking and when, thus detecting face-to-face interactions. The algorithms usually do not capture raw audio data but a set of features that does not contain verbal information. However, the limitations of this approach include:

- Sensitivity to false positives, since conversations occurring in close proximity of the monitored subjects in which they are not involved can be incorrectly classified.
- Activating a microphone can negatively affect the perception of privacy in monitored subjects while also requiring the consent of surrounding individuals uninvolved in the study.

## Our Approach: Collecting Social Interaction Data Using Non-Visual and Non-Auditory Sources

It is interesting to note that the current systems for automatic sensing of face-to-face social interactions have mostly relied on the same senses as human observers, the visual and auditory, thus capturing video and/or audio data. However, as previously discussed, the use of microphones and cameras can negatively affect the subjects' perception of privacy; moreover, video systems constrain the movements of monitored subjects into areas covered by machine vision systems. Therefore, the question is *whether face-to-face social interactions can be reliably detected without using visual and auditory sources.*

### The Main Concept

One can estimate whether two persons are having a face-to-face conversation by simply observing them from a relatively long distance in an unobtrusive manner and judging solely by the mutual position of their bodies. In order to ascertain if they are talking or just facing each other but not interacting, it is necessary to obtain evidence about speech activity. However, getting within earshot of monitored subjects may raise privacy concerns and consequently affect their natural behavior. Detecting speech activity while not affecting perception of privacy and observing the mutual position of subjects' bodies unobtrusively would lead toward capturing the natural behavior of subjects. This principle was followed to develop our approach, which is intended for continuous monitoring of face-to-face social interactions while not using visual or auditory sources. In the following, we describe our concept of exploiting advantages of sensors that, unlike human senses, are able to detect interpersonal spatial settings and speech activity using neither visual nor auditory information.

### Inferring Interpersonal Spatial Settings

Our first task is to infer spatial settings between subjects, described by parameters of interpersonal distance and relative body orientation. This is because setting appropriate spatial settings is a prerequisite for carrying out a face-to-face conversation. In particular, Groh *et al.* [11] demonstrated that these two parameters provide sufficient evidence to detect the occurrence of social interaction, but using a highly precise camera-beacon system with the accuracy of <1 mm and <1°. In our approach we have avoided the use of a camera, which may contain sensitive information. We demonstrated in [12] that spatial settings parameters can be extracted by using mobile phone sensing mechanisms with a sufficiently high precision to indicate social encounters. In the following, we provide the main concepts of our method for inferring interpersonal spatial settings.

***Distance Estimation*** — Existing solutions for distance estimation between two mobile phones exploit either acoustic components or mechanisms for transmitting/receiving radio signals. There have only been a few solutions based on the former approach, which used ultrasound [13] (which is not available in standard mobile phones) or acoustic signals emitted from the speaker [14] (which require devices to be within earshot/non-noisy environments and also can cause privacy concerns due to microphone activation). The current literature mostly reports the use of electromagnetic transmitting/receiving mechanisms to sense the presence of nearby mobile phones (e.g., Bluetooth scans [15, 16]) or to infer proximity based on collocation (e.g., NearMe [17]). However, both approaches have been shown to provide distance estimation accuracy on the order of 10 m, which does not suffice for detecting the occurrence of social interaction on a small spatio-temporal scale.

Our approach for estimating distance between two mobile phones is based on RSSI analysis, which has already been shown to be a promising solution for indoor positioning. The RSSI-based method is not limited to line of sight like infrared sensors, and is not privacy-sensitive in comparison to capturing audio data. In contrast to the approach of building a generic empirical model (regardless of the phone used, as implemented by Virtual Compass), we map RSSI values to distances relying on supervised learning, thus trading off between the accuracy in distance estimation and the user effort in signal fingerprint collection. The approach was tested using Wi-Fi signals (setting the transmitting power to the minimal value of 1 mW); however, other radio transmitting/receiving mechanisms with accessible RSSI (e.g., FM [18] or Bluetooth) available in mobile phones could be used for the same purpose or in combination with WiFi. We estimated the accuracy by applying a cross-validation method: an RSSI pattern captured in one out of six different environments (measuring Wi-Fi signal strength at different distances while using two mobile phones carried on a body, one in transmitting, the other in receiving mode) was used for building the model (i.e., a training set), while measurements from the five remaining environments were used for testing. In this manner, the procedure was repeated to cover all the combinations regarding distinct training and test sets across six environments, which demonstrated a median distance estimation accuracy of 0.5 m (Fig. 1).

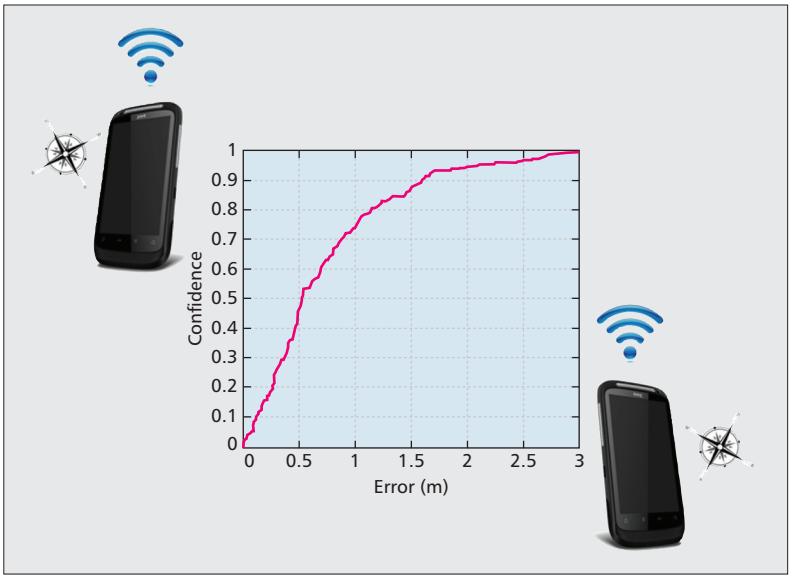


**Figure 1.** *Spatial settings detection.*

Considering the fact that RSSI patterns depend on a wide array of factors including (but not limited to) receiver's characteristics and the characteristics of the environment, repeating the training phase would be required often to prevent accuracy degradation. However, unlike time-consuming measurements typically required for fingerprinting methods, our approach decreases the user effort to only a couple of minutes for calibrating the phone signal at one distance (e.g., 1 m) while estimating the rest of the training set by applying the signal propagation model. This resulted in accuracy comparable to a full fingerprinting method (the median accuracy was again 0.5 m). Unexpectedly, calibrating the phone and testing in the same environment provided similar accuracy as in the case of performing calibration and testing in different environments (which was evidenced across all six environments). This may be indicative that the predominant factor that influences RSSI pattern lies in a receiver's characteristics, in our case captured through a fast calibration process. The less prevalent impact of environmental conditions may be explained by relatively short distances, and no obstacles between receiver and transmitter that could affect the signal propagation. In addition, due to relatively short distances, calibrating the phone signal only at one distance was proven to be sufficient to provide the above-reported accuracy (in both indoor and outdoor conditions); however, we would expect higher discrepancies at distances above 8 m (which are not relevant for detecting social interactions). The details of our approach and experiments can be found in [12].

***Relative Body Orientation Detection*** — Relative body orientation refers to the angle between the orientations of torsos considering two subjects who are facing each other. In order to estimate relative body orientation, we used the compass sensor embedded in modern mobile phones. Knowing the

*In order to prevent a negative impact on the perception of privacy in monitored individuals, our approach is based on identifying a manifestation of speech different than voice: the vibration of vocal chords.*

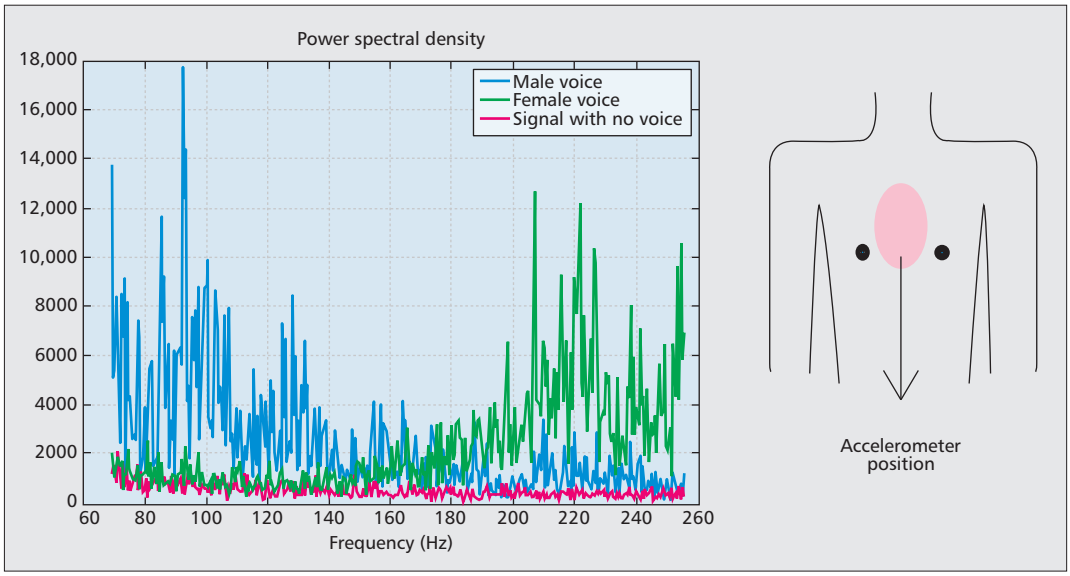


**Figure 2.** *Speech detection.*

relationship between the body's and the phone's orientation is the fundamental condition in order to recognize the individual's body orientation and the relative body orientation between subjects. Once the relationship between the body's and the phone's orientation is determined, calculating the relative body orientation requires a simple processing of azimuth, pitch, and roll values acquired from a pair of phones. The on-body position of the mobile phone can be either reported by subjects or automatically detected through existing algorithms such as [19]. Our study [12] also suggested the use of standard deviation of relative body orientation as a suitable feature for social interaction analysis, representing an index of stable relative position of participants in a social encounter. The experiments demonstrated that such an index contributes to recognizing not only whether a social interaction is taking place, but also the type of social interaction, distinguishing between formal and informal social contexts. Calculating the standard deviation of relative body orientation does not require users to carry the phone at a predefined position on the body or using complex algorithms to estimate the phone position.

However, spatial settings alone do not always provide enough evidence for inferring the occurrence of social interaction [5] (e.g., in the case of two subjects sitting across from each other in an office and not engaging in an interaction). Therefore, the second task is acquiring knowledge about the speech activity of collocated subjects.

## Speech Activity

Although people, consciously and unconsciously, communicate in a nonverbal way, speech is still considered to be the main modality of a conversation and its direct manifestation [20]. Looking from the perspective of a human observer whose task is to collect interaction data, annotating the occurrence of a conversation pertains to witnessing the speech activity, while most sensor-based systems for detecting social interactions rely on audio data analysis. In order to prevent a negative impact on the perception of privacy in monitored individuals, our approach is based on identifying a manifestation of speech different than voice: the vibration of vocal chords. In this regard, we use an external off-the-shelf accelerometer intended to infer speech activity by detecting vibrations at the chest level that are generated by vocal chords during phonation (details of our method are provided in [21]). Although a microphone embedded in the mobile phone could be used for speech detection (as in [5]), our system involves an additional sensor for several reasons. First, despite privacy sensitive techniques, activating a microphone may raise privacy concerns for subjects, thus affecting their natural behavior. Second, nearby conversations in which the monitored subjects do not participate can be unintentionally picked up by the microphone. Finally, in a number of situations (e.g., in public spaces or in the case of monitoring patients), audio data cannot be captured due to legal or ethical norms.

The concept of using an accelerometer for recognizing speech activity is based on detecting phonation-caused vibrations at the chest level, targeting a frequency range between approximately 100 and 200 Hz (which is the predicted fundamental frequency range of vocal chord vibrations for adults over the age of 20). Figure 2 shows distinct examples of frequency spectra for samples containing no voice, a male voice, and a female voice captured in our experiments. It can be noted that daily physical activities are not expected to overlap with vocal chord vibrations in the frequency domain since they typically occupy frequency ranges lower than 20 Hz [22].

On the other hand, when using an off-the-shelf accelerometer that is not manufactured specifically to detect chest vibrations, it is important to examine if there are potential sources in everyday life that produce components in the same range of frequencies that can be confused with speech activity. This concern refers mostly to low amplitudes of the chest wall vibrations that may be similar to noise level, while also the engines of vehicles such as cars, buses, trains, and airplanes can provide components in higher frequency

ranges that may result in false positives for speech detection. In our experiments, we achieved an accuracy in recognizing speech activity of 93 percent when using a Shimmer accelerometer. However, intense physical activities and traveling on a bus increased the rate of false positives up to 30 percent, which is an issue that can be addressed by using a different type of accelerometer.

### Value of the Extracted Information

The two modalities, spatial settings recognition and speech activity detection, provide complementary information for the analysis of social interactions. We have applied these two sensing modalities both separately and in fusion in a set of different experimental settings, which will be described in the following.

First, we investigated if solely sensed spatial settings provide sufficient information for detection of small-group face-to-face social interactions. Relying on mobile phone sensing we were capturing feature vectors composed of interpersonal distances ($d$), relative body orientation ($\alpha$), and standard deviation of relative body orientation ($\sigma$). When applying probabilistic classifiers on feature vectors ($\alpha$, $d$), ($\sigma$, $d$), and ($\sigma$, $\alpha$, $d$) every timeframe of 10 s, the goal was to distinguish social interactions from non-existing social situations. The experiments included 43 participants (not connected to this study) engaged in 42 social interactions (5.9 ± 4.0 min) monitored in both indoor and outdoor conditions, which resulted overall in 3500 collected timeframes of 10 s analyzed through the above-described feature vectors. In 18 social interactions, participants were asked to communicate, while the remaining 24 interactions were captured while occurring spontaneously and voluntarily (the details of this study can be found in [23]). In order to assess the potentials of using spatial parameters to distinguish existing and nonexisting social interactions, it was necessary to also create a solid corpus of data that does not correspond to social interactions. This included measurements from previously described experiments which included subjects that were in concurrent social interactions and in close proximity (within 5 × 5 m space). Table 1 shows the results for the three types of feature vectors; the highest accuracy was achieved using the 3-feature vector ($\sigma$, $\alpha$, $d$), which resulted in 89 percent successfully classified vectors corresponding to social interactions and 26 percent false positives. Using the model based on the 2-feature vector ($\sigma$, $d$) provided accuracy of 79 percent with a relatively high rate of false positives; however, it should be noted that this model does not require users to carry the phone at a predefined/known position on the body as in the case of ($\alpha$, $d$), which resulted in lower rates of both true and false positives.

As previously mentioned, in certain situations spatial settings do not provide enough evidence for inferring the occurrence of social interaction, thus also requiring knowledge of speech activity, as in the case of two subjects sitting across from each other in the office and not engaging in an interaction. In order to evaluate performance of our system in a continuous and challenging experimental scenario, we recruited four subjects who share the same office to carry the mobile phone and to wear the accelerometer (for speech activity detection) for seven days. Situations in which subjects hold the position that indicates a conversation, albeit not interacting, resulted in a higher rate of false positives (as expected), particularly in the case of using distance and the standard deviation of relative body orientation as a classification feature (($\sigma$, $d$), Fig. 3). The issue of false positives can be resolved by including the knowledge of speech activity status, which we used to confirm or reject the occurrence of a social interaction suggested by inferred spatial settings (Fig. 3 — ($\sigma$, $\alpha$, $d$) + speech). Therefore, the most accurate interaction detection (approximately 90 percent) was achieved by relying on the fusion of inferred speech activity status and spatial setting parameters.

In addition to recognizing the occurrence of face-to-face social interactions, the proposed sensing platform provides a set of nonverbal cues for the analysis of social interactions related to interpersonal spatial settings and speech activity. For instance, the amount of time that each participant spent talking during a social encounter, relative body orientations, an index of stable relative position of participants, and interpersonal distances. We showed in [23] that the extracted parameters provide meaningful information for interpreting the social context. In particular, we demonstrated the high predictive power of spatial settings parameters (up to 81 percent) in classifying the type of social interactions, perceived by subjects as formal or informal.

We envision that our approach to monitoring social interaction can provide a foundation for developing a mobile instrument for gathering rich and large-scale data, thus supporting research in social interaction analysis.

| | ($\alpha$, d) | ($\sigma$, d) | ($\sigma$, $\alpha$, d) |
|---|---|---|---|
| | SI/NonSI | SI/NonSI | SI/NonSI |
| SI | 74% 26% | 79% 21% | 89% 11% |
| NonSI | 24% 76% | 31% 69% | 26% 74% |

**Table 1.** *Classification results.*

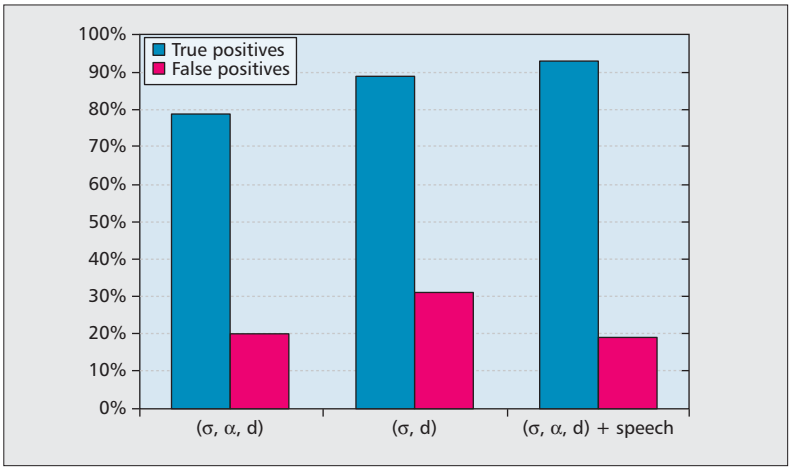


**Figure 3.** *Accuracy in detecting social interactions.*

*We envision that our approach for monitoring social interaction can provide a foundation for developing a mobile instrument for gathering rich and large-scale data thus supporting the research in social interaction analysis.*

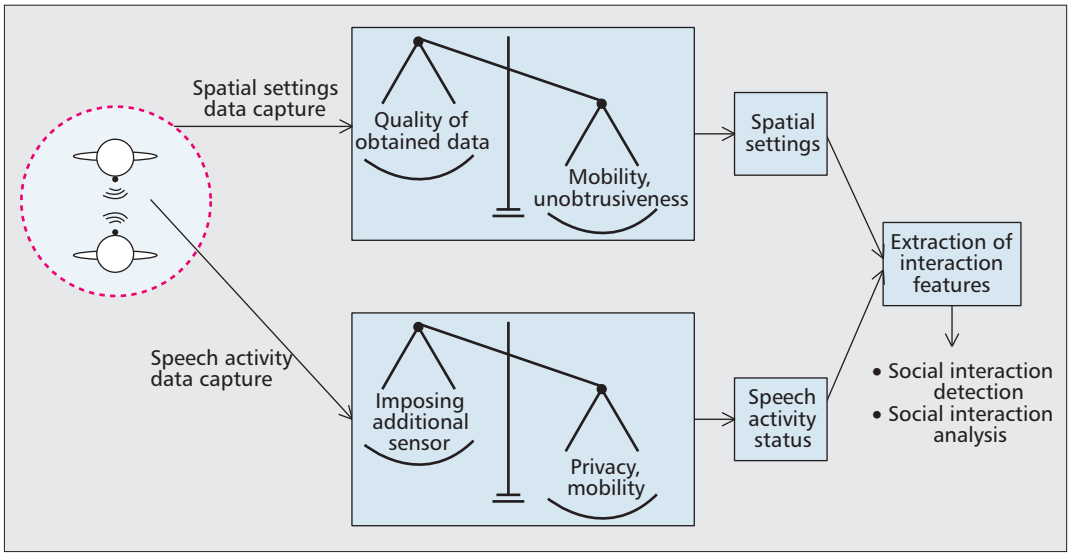


**Figure 4.** *Trade-offs of our approach.*

## Trade-off of Our Approach

As previously discussed, monitoring social interactions implies a trade-off between the quality of collected data and the levels of unobtrusiveness and privacy respect, aspects that can affect fidelity in subjects' behavior. It was demonstrated that our system provides reliable detection of social interactions as well as the possibility to extract a unique set of non-verbal cues in a mobile way. Relying on non-visual and non-auditory sources allows for a solution that is not expected to provoke privacy and ethical issues — while being based on wearable sensors, the proposed system does not spatially restrict its applications. On the other hand, the trade-offs are reflected in taking on the challenges of interpreting noisy data (to provide a mobile and unobtrusive solution for inferring spatial settings) and involving an accelerometer as an additional sensor (to provide a privacy respecting and mobile approach) (Fig. 4). These trade-offs are discussed below.

### Inferring Spatial Settings: Trade-offs

Our method of inferring spatial settings among subjects relies on sensing capabilities available in one of the most familiar and widely used wearable devices: the mobile phone. The fact that people habitually carry mobile phones makes this device a suitable source for unobtrusive and continuous monitoring of social interactions. However, being a device that is not dedicated to inferring face-to-face social interactions, the mobile phone does not provide interpersonal distances and body orientations natively, in contrast to a specifically designed camera system. The mobile phone requires a complex interpretation of noisy data obtained from available embedded sensors. Thus, such an approach trades off the quality of acquired information for allowing a mobile and minimally obtrusive solution (Fig. 4). However, we demonstrated [12, 23] that spatial settings parameters can be extracted using mobile phone sensing mechanisms with sufficiently high precision to indicate social encounters and provide meaningful information for further analysis of social interactions.

### Speech Detection: Trade-offs

The accelerometer-based approach does not require obtaining sensitive information; on the other hand, wearing a sensor at the chest level may be perceived as obtrusive, and consequently it may stigmatize monitored subjects. However, this issue occurs even in the case of using a microphone-based approach, since the microphone needs to be mounted close to the mouth to achieve higher accuracy in detecting speech. The obtrusiveness of the accelerometer, while currently a concern, is expected to be mitigated, as accelerometers are increasingly becoming widely adopted in both research and everyday life. The shape and size of already accepted commercial accelerometer-based solutions can also suit speech recognition purposes (e.g., Fitbit [24], an accelerometer device for tracking well-being aspects of individuals' behavior). Therefore, relying on an accelerometer as an alternative to the use of a microphone can be a compromise for preventing privacy concerns in subjects as well as ethical issues of monitoring in public while providing a mobile solution for continuous monitoring of speech activity (Table 2).

## Conclusions

The work presented in this article has analyzed aspects related to the trade-offs between the quality of collected data and enabling natural conditions when monitoring social interactions. Technology provides ample opportunities for acquisition and processing of a variety of information having the potential to overcome the drawbacks of survey-based methods; however, the challenge remains for the researchers as how to use these new instruments to conduct studies that approximate real-life situations. In this regard, we provided an overview of the current sensor-based methods for monitoring social interactions with a focus on the trade-offs between the quality of collected data on one hand and level of obtrusiveness, respecting subjects' privacy and spatial restrictions on the other hand — aspects that directly affect the natural behavior of subjects.

| | Accelerometer | Microphone |
|---|---|---|
| Privacy concerns | Not expected | Expected |
| Accuracy in detecting speech | Up to 93% (in our experiments) | Up to 95% [5] |
| False positives | Intense activity, some vehicles, coughing | Nearby conversations |
| Obtrusiveness | High (can be mitigated by using already accepted designs of accelerometers) | Strongly depends on the position (the higher accuracy required, the closer to the mouth an accelerometer should be mounted, the more obtrusive approach becomes) |
| Other advantages | Detection of physical activities | Omnipresent sensor |

**Table 2.** *Speech activity detection: accelerometer versus microphone.*

Furthermore, this article provided a concept for monitoring social activity by using non-visual and non-auditory mobile sources that preserve privacy, do not spatially limit applications, and minimize interference with typical daily activities.

The drawbacks of the current sensor-based methods may be the rationale behind self-reports still being prevalent for collecting social interaction data. Neither the current systems nor the approach presented in this article are a suitable replacement for the gold standard surveys for a number of studies. However, addressing shortcomings of the current sensor-based collecting methods for monitoring social interactions and decreasing negative effects of the observation will lead toward their wider acceptance.

## References


[1] N. N. Eagle, "Machine Perception and Learning of Complex Social Systems," MIT, 2005.

[2] H. R. Bernard *et al.*, "The Problem of Informant Accuracy: The Validity of Retrospective Data," *Annual Review of Anthropology*, vol. 13, no. 1, 1984, pp. 495–517.

[3] T. Choudhury, "Sensing and Modeling Human Networks," MIT, 2004.

[4] M. Buchanan, "The Science of Subtle Signals," *Strategy+Business*, 2007, pp. 48:68–77.

[5] D. Wyatt, T. Choudhury, and J. Bilmes, "Inferring Colocation and Conversation Networks from Privacy-Sensitive Audio with Implications for Computational Social Science," *ACM Trans. Intelligent Sys. and Technology*, vol. 2, no. 1, 2011.

[6] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social Signal Processing: Survey of an Emerging Domain," *Image and Vision Computing*, vol. 27, no. 12, Nov. 2009, pp. 1743–59.

[7] Y. Lee and O. Kwon, "Information Privacy Concern in Context-Aware Personalized Services: Results of Adelphi Study," *J. Info. Sys.*, vol. 20, no. 2, 2010.

[8] D. H. Nguyen, A. Kobsa, and G. R. Hayes, "An Empirical Investigation of Concerns of Everyday Tracking and Recording Technologies," *Proc. 10th Int'l. Conf. Ubiquitous Computing*, 2008, p. 182.

[9] N. Eagle and A. (Sandy) Pentland, "Reality Mining: Sensing Complex Social Systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, Nov. 2005, pp. 255–68.

[10] N. Banerjee *et al.*, "Virtual Compass: Relative Positioning to Sense Mobile Social Interactions," *Pervasive Computing*, 2010, pp. 1–21.

[11] G. Groh *et al.*, "Detecting Social Situations from Interaction Geometry," *IEEE Int'l. Conf. Social Computing/IEEE Int'l. Conf. Privacy, Security, Risk and Trust*, 2010.

[12] A. Matic *et al.*, "Multi-Modal Mobile Sensing of Social Interactions," *6th Int'l. Conf. Pervasive Computing Technologies for Healthcare*, San Diego, CA, May 21–24, 2012.

[13] M. Hazas *et al.*, "A Relative Positioning System for Co-Located Mobile Devices," *Proc. 3rd Int'l. Conf. Mobile Systems, Applications, and Services*, 2005, p. 177.

[14] C. Peng *et al.*, "Beepbeep: A High Accuracy Acoustic Ranging System Using COTs Mobile Devices," *Proc. SenSys '07 Proc. 5th Int'l. Conf. Embedded Networked Sensor Systems*, 2007.

[15] N. Eagle and A. (Sandy) Pentland, "Reality Mining: Sensing Complex Social Systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, Nov. 2005, pp. 255–68.

[16] N. Eagle, A. S. Pentland, and D. Lazer, "Inferring Friendship Network Structure by Using Mobile Phone Data," *Proc. National Academy of Sciences of the United States of America*, vol. 106, no. 36, Sep. 2009, pp. 15,274–278.

[17] J. Krumm and K. Hinckley, "The NearMe Wireless Proximity Server," *Proc. Int'l. Conf. Ubiquitous Computing*, 2004, pp. 283–300.

[18] A. Matic *et al.*, "FM Radio for Indoor Localisation with Spontaneous Recalibration," *J. Pervasive and Mobile Computing*, vol. 6, no. 6, 2010, pp. 642–56.

[19] Y. Shi, Y. Shi, and J. Liu, "A Rotation Based Method for Detecting On-Body Positions of Mobile Devices," *Proc. 13th Int'l. Conf. Ubiquitous Computing*, 2011.

[20] D. B. Jayagopi, "Computational Modeling of Face-to-Face Social Interaction Using Nonverbal Behavioral Cues," EPFL, Switzerland, 2011.

[21] A. Matic, V. Osmani, and O. Mayora, "Speech Activity Detection Using Accelerometer," *34th Annual Conf. IEEE Eng. in Medicine & Biology* Society, 2012.

[22] M. J. Mathie *et al.*, "Accelerometry: Providing an Integrated, Practical Method for Long-Term, Ambulatory Monitoring of Human Movement," *Physiological Measurement*, vol. 25, no. 2, Apr. 2004, pp. R1–R20.

[23] A. Matic, V. Osmani, and O. Mayora, "Analysis of Social Interactions through Mobile Phones," *Mobile Networks and Applications*, 2012.

[24] "Fitbit," http://www.fitbit.com/, accessed: 10-Mar-2012.


## Biographies


Aleksandar Matic (Aleksandar.Matic@create-net.org), Ph.D., has been a researcher at CREATE-NET since 2008 and he is a lecturer at the ICT Doctoral School of University of Trento. His research focuses on ubiquitous and mobile computing, pervasive healthcare, and social computing. He serves as a reviewer for a number of ubiquitous computing related conferences and journals. He established the MindCare Workshop on pervasive computing paradigms for maintaining and improving mental health, which is now being organized on a yearly basis.

Venet Osmani (Venet.Osmani@create-net.org), Ph.D., leads the research group on User Behaviour and Mobility at CREATE-NET and lectures at the University of Trento. His main research interests are in data mining techniques that can be used to analyze human behavior and how this information can be applied in healthcare applications, specifically patient monitoring and correlation with disease. He is a Steering Committee member of the Pervasive Health Conference and the current Program Chair.

Oscar Mayora (Oscar.Mayora@create-net.org), Ph.D., is head of the Ubiquitous Interaction Group in CREATE-NET. He is a Senior Member of the ACM and SIG-CHI, and is a former President of ACM SIG-CHI for Mexico. he is the founder and a permanent member of the Steering Committee of the Pervasive Health Conference. Currently he is the Scientific Project Coordinator of the MONARCA project on personal health systems for bipolar disorder treatment.