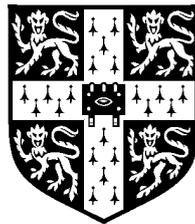


Devising and evaluating wearable technology for social dynamics monitoring

Alessandro Montanari



Girton College
University of Cambridge

May 2018

This dissertation is submitted for
the degree of Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

This dissertation does not exceed the regulation length of 60 000 words, including tables and footnotes.

Devising and evaluating wearable technology for social dynamics monitoring

Alessandro Montanari

Summary

The importance of studying social interactions has been proven useful in several fields. In the workplace, studies have found that allowing mixing among different groups could improve team coordination and productivity. Architectural studies have analysed how physical spaces can potentially increase unplanned interactions. Other areas such as epidemiology have also benefited from tracking face-to-face contacts to study the spread of disease. Although technology has progressed significantly, the automated and accurate measurement of human interactions with mobile devices is still lagging. The main shortcomings have to do with accuracy of the captured data and with the communication modalities considered. Additionally, non-verbal behaviours during social interactions (e.g. body posture, orientation and interaction distance) have been often neglected, with a few exceptions, even if traditional sociology has highlighted their importance. In this dissertation we address these challenges by developing two wearable research platforms to monitor different dimensions of social interactions.

First, we study the extent to which Bluetooth Low Energy could detect proximity in indoor environments. We analyse all the relevant protocol parameters and measure their impact on power consumption, on custom as well as on commercial devices. We assess its accuracy with a 4-week long deployment illustrating its sustainability for social dynamics studies. With the contacts and mobility data collected during the deployment we study the relationship between social contacts and space design, focusing on a modern architectural concept, Activity-Based Working (ABW). We uncover several patterns and we show how they could be the result of the correct adoption of ABW principles. However, we also discover that the employees might not have fully embraced the ABW concepts entirely, leading to mismatches between principles and actual space usage.

Given the importance of studying non-verbal behaviour during social contact we then devise a novel wearable device that, by exploiting near-infrared signals, is able to capture accurate information about distance and angle of interaction between people. We show how we design the device to be robust to ambient light changes and short occlusions by leveraging inertial measurement units. With extensive testing we evaluate its accuracy and robustness. We then explore its potential to study creative processes by deploying it to capture non-verbal cues during a creative task. We show how data about the relative orientation between people and their interpersonal distance could be used to predict the role they have during the interaction and the status of the task.

The platforms developed and the insights drawn in this dissertation provide evidence to support the use of wearable technologies to monitor social interactions at an unprecedented level.

Acknowledgments

I would like to express my heartfelt gratitude to my supervisor Cecilia Mascolo. She guided me during my PhD with invaluable feedback and lessons, giving me, at the same time, the freedom to pursue what I liked the most. Thanks to her I was given plenty of opportunities to learn and grow, personally and professionally. I will be always grateful.

I am grateful to all my talented collaborators who worked with me for making this dissertation possible. Thanks to Sarfraz for supporting me with his expertise during my first year. I am grateful to Kerstin, Rosie and people at Spacelab for helping me in planning and organising the deployment that became an important part of my research and for the brainstorming sessions about data analysis. A sincere thanks to Xia, Zhao, Elena and Ben for the great work we did together during the last year of my PhD. Thanks also to Robert Harle and Alan Blackwell for their relevant feedback and for helping me shaping this dissertation. A particular thanks to Brian Jones who was always present when I needed someone to brainstorm ideas and solutions about hardware and electronics. Thanks to Lise Gough, Gillian Gill, Nicholas Ward and Jennifer Roberts for the administrative support during my PhD and in particular for the trip to Africa. I am also grateful to Ettore Ferranti, Yvonne-Anne Pignolet and Afra Mashhadi for introducing me to research and for convincing me to pursue a PhD; it was worth it.

These three years flew by in the blink of an eye thanks to all the amazing people who made my journey unforgettable: Sandra, David, Petko, Desi, Tassos, Andrea, Xiao, Andreas, Dimitris, Krittika, Dionysis, Miles, Ellen, Alice, Fredrika, and others. Thanks to them my period in Cambridge has been full of adventures and funny moments. Thanks also to Hugo, Andrea, Matt, George, Duncan and Chris for making me enjoy my break from research while working at ARM.

I wish to express my greatest gratitude to my family, my parents, my brother and my grandparents for their great moral support throughout my studies. They never doubted that I would be successful in everything I did.

Finally, I will be forever grateful to my wife Lorena. She always supported me and believed in my strengths, especially during the darkest and hardest moments when I had a lot of doubts. I love you.

Contents

1	Introduction	7
1.1	Limitations of traditional methods for social interaction monitoring	8
1.2	Challenges in detecting social interactions with mobile devices	9
1.3	Thesis and substantiation	10
1.4	Contributions and chapter outline	10
1.5	List of publications	13
2	Related Work	15
2.1	Ethnographic methods for measuring social interactions	15
2.2	Technology for social interaction monitoring	18
2.3	Application areas	23
3	Exploring wearable sensing for office analytics	31
3.1	Introduction	31
3.2	Proximity sensing with BLE	33
3.3	Proximity sensing on commercial devices	41
3.4	Workplace deployment	46
3.5	Discussion	59
3.6	Conclusion	65
4	Detecting emerging Activity-Based Working traits through wearable technology	67
4.1	Introduction	67
4.2	Dataset description	69
4.3	Analysis of ABW principles impact	71
4.4	Discussion	82
4.5	Conclusion	85
5	Measuring interaction proxemics with wearable light tags	87
5.1	Introduction	87
5.2	A case for light-based tags	90
5.3	Protractor design	93
5.4	Protractor prototype	101

5.5	System evaluation	103
5.6	Discussion	111
5.7	Conclusion	113
6	Studying proxemics behaviour in small groups	115
6.1	Introduction	115
6.2	Study setup	117
6.3	Dataset structure and features	120
6.4	Classification of video coded labels	123
6.5	Results	124
6.6	Discussion	127
6.7	Conclusion	129
7	Conclusion and future work	131
7.1	Summary of contributions	131
7.2	Future directions	133
	Appendices	135
A	Decision Trees parameters	137
B	Proximity detection results with original dataset and with under-sampling of majority class	139
C	Wearable device user experience online survey	141
D	Protractor schematic	143
E	Random Forest parameters	145
	Bibliography	147

Chapter 1

Introduction

Research has shown that in-person social interactions play a significant role in different contexts. In the workplace, serendipitous interactions between members of different groups have been demonstrated to be a key factor for team coordination, cohesiveness and productivity [98, 36, 109]. Pentland et al. have explored this idea by looking at the level of *Engagement* and *Exploration* employees have in their interactions [161]. *Engagement* reflects the energy devoted by team members to their own team while *Exploration* represents how teams interact with one another. Productive and creative teams are the ones that are able to strike a balance between the two: they periodically seek new ideas and perspectives outside their team and then they bring them back. Architects have studied how to increase this sort of unplanned interaction by changing the layout and design of physical spaces. Informal and high-traffic spaces such as coffee areas and photocopiers have been proven to encourage inter-group serendipitous meetings and their location inside a building is crucial [33, 195, 65, 111, 209]. However, the design of efficient workplaces is not the only application of social interaction monitoring: other areas of study, such as epidemiology, have also benefited from tracking face to face interactions. Several studies and deployments have been conducted in order to develop an understanding of the spread of disease [219, 220].

Furthermore, social interactions could be studied at different levels. When people interact, the verbal part of the communication (speaking to each other), is not the only modality used. Other non-verbal channels, such as body language, facial expressions or characteristics of the speech, are combined to influence the conversation and its participants. For example, the outcome of a job interview could improve if both verbal and non-verbal skills are trained [137]. Patient satisfaction is affected by the physician's expressiveness which includes non-verbal behaviours such as more leaning forward, more nodding, more gestures and more gazing [134]. This means that increasing the awareness that a person has about her own non-verbal behaviour could be beneficial for her because she could improve her outcome in certain situations which involve interpersonal interactions (e.g., during job interviews and relations with colleagues and managers). There are also other disciplines that

demand accurate monitoring of non-verbal behaviours during social contacts. For example, the related fields of Affective Computing [166] and Social Signal Processing [204], which aim to make machines capable of recognising and managing human social signals (e.g. affect, empathy, turn taking, agreement, etc.), would greatly benefit from advancements in the tools used to detect and monitor non-verbal cues. In this context, capturing accurate and meaningful data about social interactions represents the first fundamental step in the detection and analysis of social signals and, as noted by Vinciarelli et al. [204], one of the main challenges regards *passiveness*. In other words, the ability to monitor individuals unobtrusively and without affecting their behaviour.

1.1 Limitations of traditional methods for social interaction monitoring

Despite the importance of detecting and studying social interactions, sociologists, architects and health researchers have suffered from a crucial limitation: the lack of reliable and scalable means of tracking social contacts. Scientists using ethnographic research techniques, such as participant observations and surveys, often encounter several limitations.

Usually during observations in indoor settings a researcher considers an area of the building for a certain period of time annotating all the interactions that take place during that period. In a similar approach, called shadowing, the annotations are taken while following a particular participant, who is the focus of the observation, for a certain period of time [51]. Typical guidelines require observing a certain area or person repeatedly and at different times to ensure data validity and accuracy [76]. Participant observations offer the advantage of being able to collect additional information, along with the occurrence of the social contacts. For example, it is possible to record whether the participants are sitting or standing, if the conversation is work related or not, the exact location of the contact, and other subtle behaviours. However, it also presents several drawbacks. The main issue lies in the fact that the observations are only snapshots and are temporally limited; this raises concerns regarding the data not being representative and therefore not generalisable [25]. A significant number of additional observations would be required to guarantee validity but given that they are extremely time consuming, repetitions are not always feasible. Moreover, observers intervene in the environment and participants might change their behaviour [211].

Surveys scale better, especially if administered online, and reduce interviewer bias and improve comparability by using standardised questions [25]. However, they are subject to other biases as participants might not remember their behaviour correctly or because they prefer to provide socially desirable responses [30, 177]. Participants could also interpret the questions differently, diminishing data validity. Further, it might be difficult to obtain a reasonably high number of responses [25].

In Section 2.1 we provide a more extensive review of ethnographic methods to study social interactions and offer a discussion about their benefits and drawbacks.

1.2 Challenges in detecting social interactions with mobile devices

Several mobile technologies have been proposed to overcome these issues but the automated and accurate measurement of human interactions is still lagging. The main shortcomings relate to the accuracy and resolution of the captured data. For example, many systems use Bluetooth transceivers included in modern smartphones to detect proximity: these are usually power hungry and do not offer fine spatial and temporal granularity as they sample every few minutes (typically 5 minutes or more) [4, 37, 129, 220, 40]. Similarly, from the spatial point of view, current systems offer a resolution in the order of a few meters, which although sufficient in urban environments, presents important limitations in indoor settings. Other systems can provide better accuracy, but rely on dedicated hardware and require the instrumentation of the building which hinders their widespread adoption [32, 47, 87]. In particular, non-verbal dimensions of social contacts have been studied in very specific contexts reproduced in contrived settings (e.g. job interviews or public speeches) and with the use of cameras which limits the flexibility of the system [23, 22, 52, 39, 7]. While other works have addressed the analysis of speech-related non-verbal signals in real environments [49, 154, 153], very little knowledge is available for other aspects like body language, distance and angle of interaction in realistic settings.

The scalability and reliability of these monitoring systems are also critical aspects. The ability to collect longitudinal fine grained data can provide important insights on how human relations evolve over time and on how organisational structures change. Effort needs to be focused on making the technology easy to deploy and accurate in order to guarantee a large adoption and collect rich data that could provide a better understanding of social interaction dynamics. Many of the current systems require instrumentation of the building which can raise logistic and privacy issues. Some of them also need time consuming calibration procedures in order to work accurately. The recent development of wearable devices [170], especially wrist worn fitness trackers and smart watches, offers interesting opportunities for sensing social interactions which could overcome some of the issues. These devices are becoming more and more ubiquitous, powerful and enriched with many sensors and have the advantage of always being co-located with the user. However, their scarce resources and need to serve the user for other tasks apart from monitoring her interactions make devising efficient solutions that capture fine grained data about social interactions a major challenge.

1.3 Thesis and substantiation

We have seen how different fields could benefit from the automatic collection of social interaction data and what challenges and limitations are encountered when traditional ethnographic methods or technology are employed for this task. Our thesis is as follows: *to support diverse applications relying on social interaction detection we need to consider and devise wearable technologies capable of accurately monitoring different dimensions of social contact and evaluate their benefits for the understanding of human behaviour.* We corroborate this statement by firstly evaluating the potential of existing wearables for behavioural sensing, both as data collection platforms and the utility of the data they gather. Subsequently, by designing a novel data collection platform we demonstrate the benefits of monitoring and analysing non-verbal cues of social interactions. In particular, this dissertation addresses the following three research questions:

- **Research Question 1.** How can we take advantage of radio communication interfaces embedded in many commercial wearable devices (i.e., Bluetooth Low Energy) for the efficient detection of social contacts in very dynamic environments?
- **Research Question 2.** How can we leverage data gathered automatically with wearable devices to analyse team dynamics and the strength of employees' interpersonal ties in relation to space usage and organisational hierarchy?
- **Research Question 3.** How can we devise a wearable sensing technology suitable for the fine granularity detection and analysis of non-verbal cues during social interactions?

To address these questions we developed two wearable research platforms to gather data and study social contacts. First, we analysed the potential of Bluetooth Low Energy (BLE) as a relatively simple and widespread technology to monitor people's proximity in indoor environments. Second, we deployed the BLE platform in a very dynamic office space adopting Activity-Based Working (ABW) principles to study how sensing techniques can be used to detect behavioural traits and relate them to space design principles and organisational hierarchy. Third, we devised a novel wearable device able to accurately and robustly measure some non-verbal aspects of social contacts (i.e., distance and relative orientation) and we analysed its potential to support the study of creative processes in small teams.

1.4 Contributions and chapter outline

This dissertation analyses how mobile sensing techniques can be applied to the study of in-person social interactions. We first studied the potential of emerging wearable devices with the objective of lowering the data collection burden on users. We discovered how this technology can shed light on the relationship between social contact and the design of

office spaces. We then moved beyond detection and analysis of occurrence of social contact by considering non-verbal dimensions. We proposed a novel research platform that can be reliably adopted to support organizational science studies by enabling the collection of non-verbal behaviour during intense problem-focused discussions. The state-of-the-art on techniques for social interaction monitoring is summarised in Chapter 2 while the rest of the dissertation answers the three main research questions outlined in the previous section and makes the following three major contributions:

Contribution 1: Exploring wearable sensing for office analytics

In Chapter 3 we report on the first study of Bluetooth Low Energy radios on a wearable platform for proximity monitoring and provide insights for social interaction sensing applications. This technology has been chosen because it is available in all current mobile devices and it has the potential to simplify large deployments. Moreover, as a short range low-power radio, it offers several transmit power levels which makes it ideal for fine-grained interaction sensing by periodically transmitting and listening for beacons without draining the battery too quickly. We analyse in detail the most common wearable platforms available: Android Wear and Tizen Wearable. With detailed experiments we study which BLE parameters can be controlled on these devices and their effect on power consumption. This study leads to the conclusion that despite the fact that the hardware used in modern wearables offers the key functionality for interaction detection (i.e. the ability to detect nearby devices and be detected by alternating between transmitting and scanning), existing firmware and software stacks allow only limited control over the BLE interface.

In order to experiment with Bluetooth Low Energy more freely we build a prototype wearable platform which can sense proximity between devices and detect the coarse location by using static beacons in the environment. By deploying this platform we are able to evaluate BLE in a working environment. We gather data about 25 employees of a very dynamic company for a period of four weeks. Through data post-processing we investigate the achievable performances if our system was to run on off-the-shelf wearable devices and understand their strengths and weaknesses. Even if currently commercial wearable devices do not allow the developer to control every detail of the BLE interface (e.g. it is not possible to set specific values for certain parameters) they can be employed to detect proximity with high accuracy (F1 score between 0.81 and 0.97) with a 10-second granularity. We conclude the chapter by offering guidance to Operating System (OS) developers and manufacturers on the impact of the limitations of their software stacks and informing application developers on the flexibility of off-the-shelf wearables.

Contribution 2: Detecting emerging Activity-Based Working traits through wearable technology

In Chapter 4 we focus on the analysis of the data collected with the BLE research platform described in the previous section. A recent trend in corporate real-estate is Activity-Based Working (ABW). The ABW concept removes designated desks but offers different work settings designed to support typical work activities¹. In this context there is still a need for objective data to understand the implications of these design decisions. We aim to contribute by using automated data collection to study how ABW's principles impact office usage and dynamics.

Toward this aim we analyse team dynamics in relation to space usage and organisational hierarchy using data collected by wearable devices in a company adopting ABW principles. In particular we focus on two core aspects of ABW: (1) absence of allocated desks which allows employees to flexibly use the office space and (2) freedom of interaction and collaboration across team boundaries by designing an office which stimulates serendipitous social contacts among people in different groups.

Our findings show that the office fosters interactions across team boundaries and among the lower levels of the hierarchy, suggesting strong lateral communication. Employees also tend to have low space exploration on a daily basis which is instead more prevalent during an average week and strong social clusters seem to be resisting the ABW principles of space dynamics which should instead motivate people to move inside the office to select the best workstation for the current task. With the availability of two additional data sets about social encounters in traditional offices we highlight traits emerging from the application of ABW's principles. In particular, we observe how the absence of designated desks might be responsible for more rapid dynamics inside the office.

Contribution 3: Automatic measurements of interaction proxemics

In Chapter 5 we introduce our efforts to devise an automatic system capable of recording non-verbal cues during social interactions. Proxemics of social interactions (e.g., body distance, relative orientation) influences many aspects of our everyday life: from patients' reactions during interaction with physicians to success in job interviews, to effective teamwork. Traditionally, interaction proxemics has been studied via questionnaires and participant observations, imposing a high burden on users, low scalability and precision, and potential biases. Technology employed for this task has mostly relied on cameras deployed in the environment which could raise potential privacy issues and complicates data collection due to the instrumentation effort required.

Chapter 5 presents a novel wearable technology for measuring interaction proxemics with fine granularity as part of non-verbal behaviour cues. Our approach employs near-infrared

¹Section 2.3.1 provides an overview of ABW's core principles.

light to monitor both the distance and relative body orientation of interacting users. We leverage the characteristics of near-infrared light (i.e., line of sight propagation) to accurately and reliably identify interactions; a pair of collocated photodiodes aid the inference of relative interaction angle and distance. We achieve robustness against temporary blockage of the light channel (e.g., by the user’s hand or clothes) by designing sensor fusion algorithms that exploit inertial sensors to obviate the absence of light tracking results. We fabricate wearable tags and conduct real-world experiments. Results show the accuracy of our system in tracking body distances and relative angles. The framework achieves less than 6° error 95% of the time for measuring relative body orientation and 2.3-cm – 4.9-cm mean error in estimating interaction distance.

In Chapter 6 we provide an initial exploration of the possibilities offered by our novel device in the understanding of complex and often abstract processes, comprising multiple, interrelated sets of human actions such as creativity in an organizational environment. To this aim we deployed our tags to track users’ non-verbal behaviours when conducting collaborative group tasks. In particular we explored the possibility of predicting, using only proxemics information (i.e., angle and distance between pairs of participants), two aspects of team dynamics: (1) task role, the verbal role assumed by each participant, (2) task timeline, the different procedural phases of the creative task. Results with 64 participants show that distance and angle data can help assess individual’s task role with 80% accuracy, and identify the task timeline with 92% accuracy.

The last chapter of this dissertation (Chapter 7) reflects on the results and insights provided and outlines potential developments for future research.

1.5 List of publications

Some of the work related to this dissertation has been published in peer-reviewed journals, conferences and workshops as listed below.

Works related to this dissertation

- [151] A. Montanari, Z. Tian, E. Francu, B. Lucas, B. Jones, X. Zhou, and C. Mascolo. Measuring interaction proxemics with wearable light tags. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, volume 2, pages 25:1–25:30, New York, NY, USA, Mar. 2018. ACM
Zhao Tian, PhD student at Dartmouth College, contributed equally to the research work reported in this paper.
- [146] A. Montanari, C. Mascolo, K. Sailer, and S. Nawaz. Detecting emerging activity-based working traits through wearable technology. In *Proceedings of the*

ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), volume 1, pages 86:1–86:24, New York, NY, USA, Sept. 2017. ACM

- [149] A. Montanari, S. Nawaz, C. Mascolo, and K. Sailer. A Study of Bluetooth Low Energy performance for human proximity detection in the workplace. In *Proceedings of the 15th IEEE International Conference on Pervasive Computing and Communications (Percom'17)*, pages 90–99. IEEE, 2017 [Best Paper Nominee]
- [148] A. Montanari, S. Nawaz, C. Mascolo, and K. Sailer. Experience in deploying wearable devices for office analytics. In *Proceedings of the 1st International Workshop on Quantified Workplace. Co-located with CSCW 2016.*, 2016
- [144] A. Montanari. Multimodal Indoor Social Interaction Sensing and Real-time Feedback for Behavioural Intervention. In *Proceedings of the 7th Workshop on Wireless of the Students, by the Students, & for the Students*, pages 7–9. ACM, 2015

Other works

During my studies I also published other works:

- [147] A. Montanari, F. Kringberg, A. Valentini, C. Mascolo, and A. Prorok. Surveying Areas in Developing Regions Through Context Aware Drone Mobility. In *Proceedings of the 4th ACM Workshop on Micro Aerial Vehicle Networks, Systems, and Applications (DroNet18). Colocated with ACM MobiSys18.* ACM, 2018
- [147] A. Montanari, A. Mashhadi, A. Mathur, and F. Kawsar. Understanding the privacy design space for personal connected objects. In *Proceedings of the 30th International BCS Human Computer Interaction Conference (British HCI '16)*, pages 18:1–18:13, Swindon, UK, 2016. BCS Learning & Development Ltd
- [150] A. Montanari, Y. A. Pignolet, and E. Ferranti. Rest assured, we manage your microgrid. In *Proceedings of the 5th IEEE International Conference on Smart Grid Communications (SmartGridComm '14)*, pages 284–289, Nov 2014

Chapter 2

Related Work

In the previous chapter we highlighted the importance of monitoring and studying social interactions in different fields. In this chapter we delve deeper into the techniques used for such task. We provide a review of ethnographic research approaches (Section 2.1) as well as of automatic techniques developed recently (Section 2.2). We distinguish the automatic approaches in the ones capable of detecting the occurrence of the contacts and their location indoors and the ones capable of recording more subtle non-verbal behaviours. We conclude the chapter (Section 2.3) with a discussion of two application areas that have been considered in this thesis: the study of how social contact relates to the design of office spaces and the analysis of non-verbal behaviour in small groups.

2.1 Ethnographic methods for measuring social interactions

In order to test hypothesis and refine theories, social scientists use a variety of methods to measure human behaviour and in particular social interactions. These methods typically include surveys, participant observations and audio/video recordings,. These approaches allow to record different modalities (visual, verbal, non-verbal, etc.) and have different requirements in terms of time and cost. Additionally they provide information at potentially different levels of granularity and quality. Once the data has been collected it needs to be annotated in order to transform the collected information into systematic data that can be used for the subsequent steps of analysis and interpretation. This is commonly a lengthy operation that could take several weeks or months and it is usually performed by multiple coders. We will now focus on the measurement techniques and briefly discuss each of them.

With *surveys* or *self reports* researchers gather information about social dynamics by asking

participants to complete a predefined set of questions which could be more or less structured and quantified [138]. Questions asked depend on the behaviour that the researchers are interested in studying and can include information about frequency of different kind of interactions (e.g., face-to-face planned or unplanned, work or not work related), strength of relationship with others or their influence on work activities [177], location where most of the interactions happen, relative orientation and preferred interpersonal distance [104], and more. This strategy is commonly retrospective in the sense that surveys are administered some time after the occurrence of social contacts. The period that is covered by the survey could vary across studies but usually it does not consist of a single contact but rather a considerable period in the past (e.g., one week or one month). This highlights one of the potential drawbacks of self reports: recall bias. Participants might recall better most recent events and therefore weigh these more compared to older ones, which might be forgotten completely [138]. This is particularly problematic in environments where contacts are short and frequent. People might be engaged in many interactions during the day and hence fail at correctly remembering every single event or important information about them (e.g., location inside the building, topic, etc.). In some situations, participants might provide answers that they perceive as socially desirable for example by emphasising the perceived favourable behaviour or by diminishing the unwanted behaviour, negatively affecting the analysis and interpretation of the results [30, 177]. Additionally, participants might not interpret the questions as the researchers expected, causing again potential problems in the analysis of the data [25].

Another strategy used in social sciences is *participant observations* which consists in having observers (typically more than one) watch participants involved in social interactions and record useful information to study their behaviour [138]. Similarly, *shadowing* focuses on a particular participant who is followed for a period of time and is the only subject of the observation [51]. Information recorded by this approach include the exact start/end times of social contacts, the people involved, the location inside the building, the nature of the contact (social or work related), the participant's individual contribution to the conversation and much more. The presence of observers might be known to the participants or they might be hidden from them. The recorded information could include many (or potentially all) of the participants' actions and behaviours or it might focus only on specific events which might never happen during the observation period. Through observations researchers are able to collect fine grained information that are difficult, if not impossible, to capture with retrospective surveys. For example, these include subtle behaviours like changes in the tone of voice or slight body movements. This however requires extensive training of the observers in order to equip them with the required knowledge to produce high quality coded information as the social contacts are happening. The requirement of having multiple observers is another measure to ensure valuable and correct information is recorded. Nevertheless, multiple observers could potentially have a negative impact on the participants which might lead to a change in their behaviour because they are aware of being observed [211]. Hidden observers on the other hand, might raise ethical issues. The rich information provided by observations is however temporally limited and restricted to

few sessions [25]. This is a severe drawback of this strategy because their high cost usually prevents long and repeated sessions.

Video and audio recording is a method used to study human behaviour that gained popularity with the lowering cost of technology, despite being used already in the 70s [138]. Video cameras allow the comprehensive recording of what happens in their field of view and the stored data could then be coded and analysed by researchers at a later stage. This allows flexibility in the kind of coding performed given that the same recording could potentially be used to extract different information (e.g., verbal and non-verbal behaviour). Additional benefits include the possibility of pause and rewind the recording enabling more complex coding schemes compared to participant observations where most of the coding needs to be done “online” by the observer. Moreover, it simplifies the coding from multiple people since they are not required to be all present at the time of recording. Issues faced when using this approach are similar to the ones encountered with participant observations. For example, unless multiple cameras are installed, the recorded data will contain only one point of view, similarly to when a single observer is present (although the observer can typically move freely in the environment). Participants might change their behaviour because they are aware of being video recorded and they know that the data could potentially be stored for a long time [211]. This also raises privacy concerns which should be taken into consideration when installing the equipment. Similarly to participant observations this method could be expensive and could also incur in equipment breakdowns with consequent data loss. However, it is typically more convenient to cover a large area (e.g. a large office space) with cameras rather than with observers for a long period of time.

For studies that target non-verbal behaviours, other strategies have been used in addition to the ones discussed above. Sommer [193] in reviewing works on personal space, lists several techniques used for the measurement of these behaviours. For example, in field studies, strangers approach participants in natural settings (unaware of being recorded) at different distances or in various situations and observers record the participants’ reactions. In simulations instead participants are aware of being tested and they are asked to place human figures (e.g., photographs, silhouettes, dolls or manikins) as to resemble social contacts, or they are asked to stop at a comfortable distance while approaching a third person or a person surrogate.

The reviewed techniques rely typically on intensive manual labour which represent an important challenge for social scientists [124]. This is particularly problematic when studying non-verbal behaviours due to the additional time required to annotate subtle behaviours. The limited time and funds available often results in limiting the range of behaviours studied [124]. Researchers have highlighted how social scientists would benefit from automatic ways of detecting and recording human behaviours, simplifying longer data collection sessions in natural settings [204, 9, 124]. In the next section we review technologies developed to support the study of social contacts by providing automatic systems to detect and monitor social interactions.

2.2 Technology for social interaction monitoring

In this section we focus on related work that employ technology to study social contacts. First, we analyse methods and technologies that detect and record the occurrence of contacts in Section 2.2.1. Second, in Section 2.2.2 we review research that developed techniques to automatically detect non-verbal behaviours during social interactions.

2.2.1 Automatic detection of social contacts

The detection of fairly long lived interaction (in the order of several tens of seconds) has been accomplished by technology reasonably successfully. Bluetooth Classic has often been at the basis of these platforms mainly due to its availability on consumer devices, which makes it extremely suitable to large deployments [4, 220, 88, 59]. However, several works [37, 129] have tried to improve the temporal and spatial granularity of traces collected with Bluetooth Classic. In fact, Bluetooth's main drawbacks reside in the high power consumption and low granularity of the traces. Usually, it is sampled every few minutes [4, 88, 59] to avoid draining the battery too quickly and the range of transmission is around 10m [220].

Other technologies have also been proposed, for instance, based on IEEE 802.15.4 low power radio standard [68], RFID [44], Zigbee radio [132], infrared sensors [47, 154] and hybrid approaches with radio and ultrasound sensors [158, 87]. These devices offer better performance (e.g. temporal granularity from 20s to 2s with still reasonable battery consumption) but are not suitable for wide scale and long term adoption because they rely on dedicated hardware which needs to be deployed just for the purpose of the study. In fact, problems have been reported with the usability of these devices [177].

Cabrera-Quiros et al. used a custom wearable device to detect social contacts in various mingling events [40, 38, 41]. The device is capable of recording triaxial acceleration at 20Hz and binary proximity to other devices using a radio interface. However, the authors do not provide any detail about the hardware implementation, the communication protocol used or the energy requirements, making a comparison with the data collection platforms developed in this dissertation not feasible.

Another source of data used to sense social interactions with consumer devices is the microphone. Lee et al. used the microphone in smartphones to monitor the conversation between several people by matching the volume signature captured by each phone with a topography database built during a learning phase [120]. This approach has the benefit of monitoring the actual interaction and conversation between people rather than only their co-location but it is limited to capturing only relatively long interactions because it requires an initial training phase which is proportional to the number of members and it needs to be re-trained if the position of the phones or the participants change. Zhang et al. exploited the Doppler effect to detect the trajectories of approaching people and

adopted voice profiling to confirm the occurrence of a conversation [223]. By monitoring the actual conversation, however, these approaches are sensitive to false positives if other nearby users are in a different conversation. Also, the use of microphones might raise ethical and privacy issues, preventing the wide adoption of the system. Tan et al. are able to detect co-location in a privacy friendly way by using audio silence patterns [196]. The power consumption, however, is an important problem which prevents the system to be run continuously. Additionally, co-location detection works only in the presence of sufficiently loud acoustic events which contributes to generating the pattern of silence, and this is not always the case if we consider particular contexts such as libraries or exam rooms.

Several works have combined different existing technologies to detect social interactions and collect multimodal datasets. In the SALSA dataset four cameras and Sociometric badges were used to record 18 participants for 60 minutes during a poster session and a cocktail party [7]. The authors conducted several experiments on the manually annotated dataset, such as people tracking and pose estimation from visual data, speaker recognition and f-formation detection. For the f-formation task the authors report a marginal improvement in F1 score when visual information is combined with proximity data from the Sociometric badges. Similarly cameras and Sociometric badges were used by Zhang et al. to monitor a team of six people for four months during a space exploration simulation and study social cohesion [224]. The MatchNMingle dataset is a larger multimodal dataset which includes data from 92 participants recorded in natural settings for about 2 hours [39]. The multimodal data includes acceleration and binary proximity from wearable devices used by the participants and videos recorded by several cameras deployed in the environment (including audio). The wearable devices used for data collection are based on the MyriaNed wireless sensor nodes [57]. They use a Nordic nRF23L01+ radio module operating in the 2.4GHz spectrum, hence having similarities, at the physical level, with Bluetooth. However, this module does not provide any received signal-strength indicator (RSSI) and therefore proximity is determined only by the reception of packets from nearby devices without the possibility to estimate the distance from them. The combination of mobile devices and cameras allows to collect rich information about social contacts, including also subtle non-verbal cues. However, the presence of cameras might raise potential privacy issues among participants and limit the data collection coverage given that data can be recorded only in previously instrumented areas.

Approaches based on Bluetooth Low Energy

Relatively fewer works have specifically used BLE to collect data about human behaviour. Townsend et al. tested 4 different smartphones to assess if they could detect each other using BLE [201]. Boonstra et al. deployed an Android and iOS app to 14 participants for a period of one working week to collect data about social contacts [28]. However, the authors offered a limited evaluation of their system by using only two meetings during the study period to validate their methodology and they did not collect participants' locations, which

is a valuable piece of information when studying social dynamics. Other works instead have used simple wearable BLE tags, capable of transmitting only, to study mobility patterns of large gatherings [93, 94]. Radhakrishnan et al. have analysed BLE characteristics on Android mobile phones for indoor localisation [169]. They implemented a BLE-like duty cycling on top of the Android BLE stack which already performs duty cycling in accordance to the BLE specification. In a very recent work Katevas et al. investigated a multi-modal approach to detect stationary social contacts employing data from the BLE radio, accelerometer and gyroscope embedded on smartphones [100]. The authors found that modern mobile operating systems (especially Apple devices¹) do not allow to transmit BLE advertisements when the screen of the mobile device is off and the app is in the background. To overcome these limitations the authors asked the participants to carry a small, battery-powered BLE beacon in addition to the smartphone. They found that the features computed from BLE data is the most important in discriminating between interacting and non-interacting participants, while the motion-related features contribute to a lesser extent. Lederman et al. introduced the Open Badges framework [117] later renamed Rhythm [118] consisting in wearable badges and online applications to measure interaction patterns in co-located and remote teams. The badges are built on top of the same chip we used in this dissertation to study the BLE parameters (i.e. Nordic NRF51822). The authors used the BLE interface to capture proximity between people (every 60 seconds) and added a microphone to recognise when a user is speaking. The proximity detection technique uses a fixed threshold on the RSSI values of packets received from nearby devices [118].

In the last years very small and ultra-wearable devices started becoming popular. In 2018, the most prominent manifestation of these devices is wireless earbuds with embedded sensing and communication capabilities (e.g. Apple AirPods or Samsung Gear IconX). Commercial devices include a dual-mode Bluetooth/BLE transceiver but usually do not allow application developers to use the BLE interface for proximity sensing by alternating transmission and scan or by controlling its parameters. In 2018 the only earbud research platform which could be used for proximity detection is eSense [103]. Despite not being capable of scanning for beacons due to the limited battery capacity, eSense allows to continuously transmit BLE beacons (even during a connection) and to control the beaconing rate [103, 141]. This allows other devices with larger energy budget (e.g. smartwatches and smartphones) to detect the presence of eSense devices in the environment.

In Chapter 3 we offer an analysis of the low level BLE parameters (as defined in the standard) to understand if BLE can be employed to collect fine grained and accurate encounter traces. We then analysed if the approach could be adopted on commercial wearable devices to free proximity-based systems from the need for custom devices, which is usually one of the limiting factors of long term studies. Additionally, the availability of our prototype allowed us to test the impact of these parameters on a large scale deployment.

¹Apple devices are typically not the first choice for data collection applications due to the limitations imposed by the operating system on services running in the background.

Our study of the BLE parameters provides useful insights not only for wrist-worn devices, as we focus in Chapter 3, but also for devices with other form factors which use BLE as their main communication medium.

While BLE specification v4.0 is currently ubiquitous in wearable and mobile devices, the next version of the specification, marketed as Bluetooth 5 [27], started being supported in commercial devices in 2017/18 (e.g., Samsung Galaxy S8, Apple iPhone X and HomePod). The new specification will allow to transmit eight times more data in advertising packets and achieve longer communication range. These are features that could be useful from a sensing perspective to enable low latency, connection-less communication among a large number of devices. Additionally, the Bluetooth SIG included in the specification the capability of measuring the angle of arrival [173] and angle of departure of packets using multiple antennas on the same device. This permits devices to detect the direction from which packets are received and potentially the orientation of the device or the user carrying it. This functionality could complement the infrared technology we devised in this dissertation (Chapter 5) and other technologies developed for the automatic detection of non verbal cues, as we will review in the next section.

2.2.2 Automatic detection of non-verbal behaviours

In Section 2.2.1 we reviewed the technologies used to capture meaningful data about social interactions, like physical proximity and conversation. In this section we focus on previous works that targeted the automatic detection of non-verbal behaviours during human contacts. We place particular emphasis on the technical solutions used for this task, while we will focus on the analysis of the collected data in the following sections.

The automatic recognition of these behaviours has been mainly addressed with cameras and microphones [204, 205]. Many works focused on speech related cues, like turn-taking and prosody related features [182, 181] or simple hand-picked features [20] (e.g. total amount of speech spoken, amount of speech overlapped with others, etc.). Similar audio cues have also been employed in other works that combined them with visual cues [96, 185, 95, 89, 221]. Researchers relied almost exclusively on visual cues extracted from tracking motion in videos. For example Jayagopi et al. used motion vectors and residual coding bitrate to estimate visual activity, which is a binary variable indicating if a person is visually active or not [96]. On top of this low level feature the authors then computed aggregated features like total visual activity length and total visual activity turns. Other works used the same techniques and similar aggregated features to estimate head and body activity individually [185, 95] or hand fidgeting (i.e. tapping on the table or playing with glasses) [221]. Alameda-Pineda et al. computed head and body orientation from images by calculating Histograms of Oriented Gradients (HoG) features for head and body bounding boxes and then learning a separate classifier for each pose [7, 8]. The two classifiers consider 8 classes which correspond to an angular resolution of 45° for head and body orientation.

Recent works relied on more sophisticated cameras, like Microsoft Kinect. Batrinca et al. focused on public speaking training and analysed the correlation between presenters' automatically extracted non-verbal behaviours (speech characteristics, gestures, and gaze) with experts' assessment of the presentation [22]. A Microsoft Kinect, two webcams and a microphone were used to capture the non-verbal behaviours of each participant while they were giving a presentation in front of a projected virtual audience. Baur et al. [23] and Damian et al. [52] also used a Microsoft Kinect to record and then analyse non-verbal behaviours. In the first work, non-verbal signals, which include posture of the upper trunk, legs and arms configurations, were captured and analysed by the proposed system and, after the interaction, statistics were presented to the user in an aggregated format allowing her to reflect on her behaviours. Damian et al. used similar techniques to analyse the person's non-verbal behaviours (speech rate, body energy and openness) during public speaking, but they also devised a mechanism to provide real-time feedback to the presenter through a Head Mounted Display [52]. In the last few years, deep learning models have started to automatically detect and track body joints from images without the need for specialised cameras [200, 42, 78].

Although camera-based approaches capture semantic rich data of social contacts, they face several limitations. First, they require considerable effort in instrumenting buildings [43, 221, 89] to enable data collection. An infrastructure-free solution is superior from this point of view since it allows to collect data even in areas that cannot be instrumented, like public spaces or during large events. Second, in general the analysis of non-verbal behaviours through video recording requires a considerable amount of storage space and processing power, limiting the usability of the system. Finally, cameras raise privacy concerns from the user and from people being recorded without their consent. This is particularly true for wearable cameras that are always with the user, even during private or intimate moments [86].

Relative device positioning

Another related line of work is on sensing the relative position and orientation between devices. Particularly for short-range positioning, existing work has explored the use of ultrasound (18 – 20 kHz, or 40 kHz) and infrared. Ultrasound methods measure time-of-flight of acoustic signals to position devices by multilateration [84, 167, 143, 128], estimate device orientation by measured phase offset [168] or positions of multiple devices [85]. These systems, however, require either additional RF radio [84, 167, 143], or the aid of multiple nodes (pre-deployed anchor nodes with known locations [128, 168] or multiple peer nodes [85]). Ultrasound has also been combined with RF signals to measure distances, using the time-difference-of-arrival technique. As examples, the iBadge [158] applies this principle to capture interactions between kids, teachers and objects in a kindergarten classroom; Opo [87] further boosts the ranging accuracy (5-cm accuracy) with a temporal fidelity of 2 seconds. However, despite offering high accuracy in a small device (especially

for Opo) these devices are capable of measure only distance between nearby devices and not their relative orientation.

Infrared-based systems have commonly been used in robotics, which measure the reflected infrared light to detect surrounding obstacles and distances [24], or use static stereo-cameras to track moving objects that carry active tags emitting infrared signals [15, 5]. Similarly to our work, Frantal et al. measured infrared incident angle using 12 photodiodes each facing a different direction [69]. Its resulting form factor however, makes it not suitable in our context where a smaller and more portable device is preferable.

In Chapter 5 we present a system that enables wearable tags to continuously and unobtrusively track each other without any infrastructure support and without relying on potentially privacy invasive cameras. We show how we measure both distance and angle of contact directly with data exchanged between wearable devices with higher accuracy compared to previous technologies. We demonstrate how we design our system to ensure reliability and energy efficiency.

2.3 Application areas

In this section we review previous work that examined the two application areas we consider in this dissertation: 1) analysis of social dynamics in workplaces using data gathered with automatic methods and 2) the automatic analysis of proxemics in small groups. Before focusing on the two application areas, in the following section we provide an overview of the core design principles of Activity-Based Working (ABW). This creates the foundation for the work we present in Chapter 4 where we analyse how the application of two core ABW's principles might be responsible for social and mobility dynamics captured with wearable devices.

2.3.1 Activity-Based Working overview

Understanding the communication and collaboration patterns of employees is critical for the efficient and effective operation of the organization and could lead to improvement in productivity and exchange of innovative ideas [11, 110, 161, 36, 195]. For this reason, in the field of architecture, increasing effort is invested into the design of spaces that could potentially promote more frequent and serendipitous face-to-face contacts [12].

In recent years, the increase of knowledge-intensive firms led to the emergence of Activity-Based Working (ABW) concepts to design offices that better support modern workforces. The concept of ABW is complex and each organization can adapt it to its specific needs and possibilities. However, three core principles are common to different implementations of the concept: (1) absence of allocated desks, (2) availability of diverse spaces and settings including those for concentration and collaboration and (3) allowing interaction and

collaboration to spread across team boundaries. The idea is that employees can choose the workstation that best matches the current task they have to complete and their personal preferences, possibly even switching between workstations during the day [13]. As a result, offices designed with ABW principles usually consist of a mix of different types of areas: isolated and quiet workstations for focused individual work, large and open settings where serendipitous interactions can flourish and meeting rooms for private discussions. In Chapter 4 we focus on these central principles and introduce a methodology based on technology and analysis techniques which is able to help in understanding the degree of effectiveness of these principles.

Activity-Based Working with unallocated desks is still an exception rather than the norm in corporate workplaces: a one-year study of working environments in 2016 showed that only 4% of the surveyed workplaces embraced ABW [121]. Given that this “agile” working style is on the rise, however, our work takes a further step towards understanding its impact on workplaces. Some previous work has analysed companies adopting ABW principles using traditional ethnographic methods of participant observations and surveys. In a recent study where more than 500 workplaces were surveyed (with ABW and without), Leesman found that ABW environments deliver performance improvements only when the employees correctly embrace the central principle of mobility. However, most of the employees who work in an ABW office still keep habits typical of traditional workplaces and present rather static work styles [121]. Appel-Meulenbroek et al. surveyed and observed four organizations with ABW and similarly found that most of the people use up to two different types of spaces, never switch work location during an average day and concluded that the offices are not always used as intended [13]. By contrast, Meijer et al. focused on workers’ health and productivity and found that ABW had some positive effects on general health in the long term [140]. With the work presented in Chapter 4 we aim to contribute to the study of this new kind of workplaces and work practices.

2.3.2 Analysis of social contacts in the workplace

Automatic systems have been used to study human behaviour in the workplace. We have reviewed the technologies and approaches employed in previous works in Section 2.2.1. In this section we cover works that focused on the analysis of data gathered with automatic methods.

Brown et al. deployed RFID-based systems to collect face-to-face contacts to study how different cultures interact with others in different job roles and the impact of physical space on social interactions [32, 33]. The work presented in [33] is the only one we are aware of that analysed, using wearable sensors, the effect of vertical structure on social contacts in a research facility without ABW. The authors reported no significant impact of the management structure on social connectivity, confirming the need to study these dynamics in various settings to better understand the generalizability of the findings.

The Sociometer, first introduced by Choudhury et al., was a very prolific wearable device used for many deployments [47]. Olguín et al. continued the development and presented the SocioMetric badges which rely on similar sensors and have been employed in several organizations to study interaction patterns and peoples' behaviour [154, 155]. Other deployments of the same technology investigated how social interactions can affect productivity [209] and how they relate to electronic communication [215]. Lepri et al. employed the SocioMetric badges to collect a multilayer dataset comprising different information sources (sensor data, surveys and experience sampling) about fifty-three employees of an Italian research centre [126]. Do et al. used the same dataset to develop a model to automatically discover and label social activities (e.g. coffee breaks and meetings) starting from social contacts and location information [56]. In a different context, Zhang et al. combined cameras and Sociometric badges to monitor a team of six people for four months during a space exploration simulation [224]. The authors studied the group task and social cohesion also using daily surveys. While the participants were working as a team, the isolated and confined environment they were living in constitutes a peculiar scenario which is hardly representative of classic office settings.

While all of the above mentioned studies investigate settings with fixed desk assignments (e.g., research laboratories, call centers and banks), in Chapter 4 we focus on a very dynamic office where employees have flexibility in where, how and when they work, resulting in an unusually dynamic workplace environment. To the best of our knowledge, only one work used technology to study this kind of workplaces. Ianeva et al. used RFID tags embedded in employees' badges to monitor occupancy of spaces [91]. The authors found that three kinds of areas (cafeteria, private booths and meeting rooms) were consistently under-occupied, revealing a mismatch between intended and actual use of these areas. With our work in Chapter 4 we go one step further in the analysis of ABW workplaces by including data about proximity contacts between employees. This allows us to study ABW principles concerning communication and collaboration and to link usage of space with interpersonal contacts. In Table 2.1 we offer a closer comparison with related works that not only used a similar data collection methodology but also studied environments similar to the one we considered (companies and research institutes) and conducted alike analysis (e.g. collaborations patterns and role of office space). We omit an in-depth comparison with other works that have studied completely different environments such as conferences, schools, museums and hospitals. The table highlights that we are studying a company that adopts ABW principles and has a dynamic working style, while the majority of previous work focused on traditional office spaces with allocated desks. Our aim with Chapter 4 is to provide an understanding of peoples' behaviour in this kind of company. By contrast, most previous work focused on the relationship between social contacts and productivity.

Other contexts have also been studied with automatic methods, for example learning environments [177, 59], high schools [191, 135], clinical wards [130], family communities [4], conferences [192, 133, 44, 88] and mingling scenarios [7, 39, 8]. These are very different settings from the office environments we consider in this dissertation: they present particular patterns of proximity and dynamics centred around the peculiar characteristics of the

environments (e.g. classrooms and class schedules or visits to patients).

2.3.3 Automatic analysis of proxemics behaviour in small groups

In this section we provide an overview of the vast literature in the area of automatic proxemics behaviour analysis, also referred to as Social Signal Processing (SSP) [160]. Several social aspects have been considered in previous work, for example, social emotions, role recognition and social attitudes [204]. We concentrate on the body of work that tackled the automatic recognition of roles during social interactions and meeting phases given the similarity with the objective we set in this thesis and the work we present in Chapter 6.

Role recognition

The analysis of speech and lexical choices are among the main approaches employed for role recognition [205]. Several works consider *formal* settings, like radio news and talk-shows, where roles are defined by the function of each individual, for example, anchorman, interview participant or guest. Salamin et al. were able to achieve a considerable accuracy (from 76% to 99%) in classifying these roles in about 50 hours of audio recording. The authors' approach relied on turn-taking behaviour (accounts for how people participate in conversations) and prosodic behaviour (i.e., the way people talk) [182]. Vinciarelli used two different approaches to automatically recognise the role of speakers during radio news bulletins [203]. The author identified six roles that were used to label all speakers. The first approach was based on Social Network Analysis while the second relied on the distribution of speakers' interventions. The best accuracy was achieved by combining the two approaches which resulted in 85% of the recording time correctly classified with speaker's role. Salamin et al. extended and improved the same two methods and analysed three datasets: two included a collection of radio news bulletins and radio talk-shows while the third consisted of simulated corporate meetings where the participants acted different roles other than the ones played in their real lives [181]. The authors were able to achieve an overall accuracy of 85% (percentage of recording time correctly labelled) for the first two datasets and 45% for the third one. The main differences between these works and our approach presented in Chapter 6 concern the collected data and the setting of focus. In our study we relied exclusively on physical non-verbal cues (i.e., relative orientation and distance between participants) and we focused on spontaneous interactions where people could assume any role and change it during the meeting. We attempted to predict the various roles a person might assume and not only a single label assigned to the participant. The setting we considered (a creative challenge) despite being constrained in terms of the task that has to be completed, represented a more *informal* meeting with fewer constraints on the behaviour of people as opposed to news broadcasts and talk-shows.

Considering works that try to predict roles that are dynamic during the meeting and not predefined for the entire duration, Banerjee et al. proposed a taxonomy of meeting states

and participant roles and used C4.5 Decision Trees to predict them during meetings between faculty, staff and students at an American university [20]. The objective of predicting meeting states was similar to our Task Timeline prediction, however with different target labels. The authors adopted empirical features, like the number of speaker changes, the number of speakers talking during a given time interval, the total time spoken, etc., and achieved an accuracy of up to 51% for the meeting state and up to 53% for the participant role prediction. Contrary to our study, the authors relied exclusively on speech-related features and had a lower number of target labels: 5 for the participants roles and 4 for the meeting states. In our case instead we predicted 5 and 9 labels for roles and timeline (similar to meeting state) respectively.

Few works have taken a multi-modal approach to role recognition. Zancanaro et al. used cameras and microphones to analyse the roles played by team members in relation to the tasks the group had to face (“Task Area”) and in relation to the functioning of the group (“Socio-Emotional Area”) [221]. Using the behavioural traits of speech activity (presence or absence) and fidgeting (“the amount of energy in a person’s body and hands”, e.g. tapping on table) the authors were able to predict the manually coded “Task Area” and “Socio-Emotional Area” roles of 10 participants with accuracy between 65% and 68% (F-score between 0.52 and 0.55) using Support Vector Machines. The work was later extended by Dong et al. to use influence models, obtaining an increase in accuracy up to 75% [58]. These works present similarities with ours. First, the task chosen by the authors (the Survival Task [82]), in which participants need to reach consensus on how to survive in a disaster scenario, resembles our creative challenge because the participants have a clear goal and need to co-operate to achieve it. Despite the fact that the dataset used by the authors included data about a person’s body and hands energy (automatically gathered with computer vision techniques) the participants did not need to move to complete the task therefore physical non-verbal behaviour was probably less prominent. Second, similar to our approach, the authors did not assume that the same participant would play the same role for the duration of the task and predicted the instant roles as they varied during the discussion. The best accuracy, of 75%, in the detection of roles was achieved in the second version of the work [58] using influence models. In the Task Role detection we achieved better results with an accuracy of 84.9% and F-score ranging from 0.42 to 0.88 (Table 6.5).

Other analysis

Other works relied on non-verbal behavioural cues to study other aspects of social interactions. Jayagopi et al. studied conversational group dynamics (e.g., conversational topics, leadership styles) [95], and group dominance [96], using non-verbal cues extracted from an existing dataset with 100-hr meeting recordings [43]. In [89], group cohesion was studied using hours of audio-visual group meeting data. Hung et al. estimated group formations in crowded environments using a graph clustering algorithm [90]. The analysis was based

on video footage of over 50 people presenting scientific work in a poster session. Sanchez-Cortes et al. inferred emergent leaders using non-verbal cues extracted from audio and video channels [185]. These prior studies commonly collected richer data (e.g., speaking turn and prosodic cues, head and body activity) in both visuals and audio, while our tags (Chapter 5) collect body distances and orientation.

Table 2.1: Comparison between our work and previous work that adopted sensors to study office spaces.

Paper	Context	Main Aspect Studied	Main Finding
Wu et al. [215]	Company with assigned desks	Effect of social structures on knowledge transfer and productivity	Network cohesion in face-to-face networks is positively correlated with higher productivity in particular when executing complex tasks.
Olgún et al. [154]	Company with assigned desks	Effect of physical activity and speech on job performance	Positive correlation between task completion time and physical activity and speech level
Olgún et al. [155]	Company with assigned desks	Employees' job Satisfaction	More communication leads to less satisfaction. More central positions in the organisation leads to less satisfaction.
Waber et al. [209]	Company with assigned desks	Effect of social groups on productivity	Productive employees belong to the strongest social groups. Achieve stronger social groups by changing breaks schedule.
Brown et al. [33]	Company with assigned desks	Effect of spatial layout on social interactions	Purposely designed offices can promote inter-team mixing. Central food areas host many inter-team contacts.
Brown et al. [32]	Research institute with assigned desks	Effect of individuals' cultural differences and spatial layout on serendipitous interactions	Cultural dimensions (Power Distance, Individualism and Masculinity) affect the likelihood of contacts between people with different roles. Printers and kitchens are likely places for serendipitous inter-group contacts.
Génois et al. [71]	Company with assigned desks	Analyse social contacts for epidemiology studies	The spread of infectious diseases is hindered by the sparsity of the contacts. Vaccinations of linkers between communities can prevent epidemic outbreaks.
Ianeva et al. [91]	Company adopting ABW principles	Monitor and understand the building occupancy rate and use of shared areas	Several office areas are under-occupied highlighting a miss match between intended and actual use of the office
Our work	Company adopting ABW principles	Analyse core aspects of ABW and comparison with traditional offices	High degree of inter-group contacts, especially in the lower levels of the hierarchy. The principle of not having allocated desks seems not completely adopted. Different network characteristics between traditional offices and offices adopting ABW principles.

Chapter 3

Exploring wearable sensing for office analytics

3.1 Introduction

In Chapter 1 we discussed the importance of studying social interactions and highlighted its challenges. In this chapter we focus on interactions in the workplace, which play an important role in team performance and productivity [109, 98]. For example, informal inter-team interactions have been shown to be an important trait of successful teams [161]. Work interactions can also influence the design of physical spaces [33] or help in developing an understanding of the spread of diseases [219, 220]. Researchers have relied on surveys and observations for years to gather data about these phenomena. However, the cost of observations is high as they usually involve long hours of monitoring. They might also not scale to high numbers of participants and can therefore only be applied for limited time periods. Surveys instead scale better but offer a much coarser grained view as people might forget to report [30, 25, 177].

Usually two contrasting needs have to be satisfied when trying to capture such interactions automatically: (1) the need to collect accurate and reliable data and (2) the need to have large deployments to get a clearer picture of human behaviour. Existing solutions usually tend to tackle one problem or the other. Bluetooth-based systems for example can rely on widespread adoption but are usually power hungry and do not offer fine grained data [4, 37, 129, 220]. On the other hand, systems based on custom built devices can provide fine granularity but require dedicated hardware which hinder adoption [32, 47, 87]. The recent interest in wearable devices [170] has brought us to question if those devices are able to fulfill both needs. In particular we directed our attention towards Bluetooth Low Energy (BLE) which is embedded in all current wearables. We envision an interaction sensing system that can be easily installed on a wearable device like a smart watch thus extending its functionality to interaction sensing and offering widespread adoption. The

system will be able to both gather data about interactions for offline analysis but also data which can be analysed in real-time and fed as recommendation or temporal statistics to workers. However, before this can become a reality there are fundamental questions which need to be answered. Namely: How accurate could BLE proximity detection be? What could be the expected lifetime of this system on an off-the-shelf device? How can it be employed for social interaction sensing and space occupancy monitoring?

In this chapter, we analyse the potential of BLE to monitor people proximity as a first step towards a social interaction sensing system. The objective is to assess its capabilities, first by analysing its parameters and their impact on both accuracy and power consumption, and then, from a practical perspective, with a large user study in a real workplace. The specific scenario we consider is the one that takes into consideration office based social interactions. In such a setting, serendipitous interactions, where, say, a user glances from an office doorway, might be meaningful and indicative of productivity [161, 36]. This is, of course, in addition to prolonged and repeated interactions.

Current hardware, available in modern wearables and smart watches, offers the key functionality for proximity detection: the ability to detect nearby devices and be detected by them by alternating between transmitting and scanning. While the manufacturers have recently updated device firmware and software stacks to support this kind of behaviour there are still several limitations that prevent an accurate study of all the key factors involved in proximity monitoring. In particular, mobile operating systems do not allow the application developer to freely control all the BLE parameters. Thus, we build and use a custom made wearable prototype in which we are in control of all parameters. This allows us to study in detail the interplay of all the BLE parameters and their impact on power consumption. Using our prototype we collect proximity traces in a commercial organisation with 25 participants. We are then able, through data post-processing, to investigate the achievable performance if our system were to run on off-the-shelf wearable devices and understand their strengths and weaknesses. This leads us to the important conclusion that large scale proximity studies are viable, even at the accuracy level required by domain scientists, with off-the-shelf devices.

To the best of our knowledge, this is the first study of BLE radios on a wearable platform for proximity monitoring which provides useful insights for social interaction sensing applications. This chapter also offers guidance to operating systems (OS) developers and manufacturers on the impact of the limitations of their application program interfaces (APIs) and informs application developers on the flexibility of off-the-shelf wearables. Our contributions are:

- a detailed analysis of BLE parameters that play a central role in proximity detection;
- the first analysis of BLE capabilities and limitations on commercial wearable devices (Android Wear and Tizen);
- an extensive experimental validation with lab experiments and a longitudinal user study with 25 participants in an office environment. We confirm the BLE suitability

for accurate proximity monitoring with detection F1 score between 0.81 and 0.97 with a 10-second granularity. Ground truth observation for around 19 hours was performed to support our evaluation.

- a discussion on the restrictions imposed by OS developers on the use of BLE for proximity detection.

Chapter Outline. Section 3.2 introduces the wearable platform we developed in order to experiment with BLE and describes how the different BLE parameters affect proximity detection and their impact on power consumption. In Section 3.3 we analyse to what extent BLE parameters can be controlled on two popular wearable operating systems and the resulting power consumption. Section 3.4 describes the deployment of our wearable platform to evaluate the capabilities of custom and commercial devices for proximity detection. In Section 3.5 we discuss the results achieved during the deployment and provide guidelines for devices' manufacturers. Section 3.6 concludes the chapter summarising our contributions.

3.2 Proximity sensing with BLE

In this section we first provide a brief introduction about the BLE modes of operation and introduce our wearable platform. We then discuss the different BLE parameters and we present a detailed analysis of their impact on proximity detection accuracy and energy consumption using our platform.

3.2.1 BLE modes of operation

BLE provides two modes of communication: *connection based* and *broadcast based* [26]. The first requires two devices to establish a connection before exchanging data. This is not suitable for proximity sensing as it can introduce delays and is also restricted to a limited number of devices.

By contrast, the broadcast mode allows a *Broadcaster* to send data to several *Observers* simultaneously without establishing a connection. The Broadcaster periodically sends data on three predefined BLE advertisement channels (37, 38 and 39) as shown in Figure 3.1. The standard advertisement packet contains a 31-byte payload (maximum size) which describes the Broadcaster. Thus, the time required to transmit it on the three channels is on the order of few milliseconds. To adjust the transmission frequency, the BLE specification defines a parameter called Advertising Interval. It specifies the time between the start of two consecutive advertisements. The Advertising Interval can vary from 20 ms to 10.24 seconds (advertisements are perturbed in time using a pseudo-random value between 0 and 10 ms). On the other hand, the Observer listens on a advertising channel for the duration of the Scan Window at every Scan Interval. At each Scan Window, the Observer listens

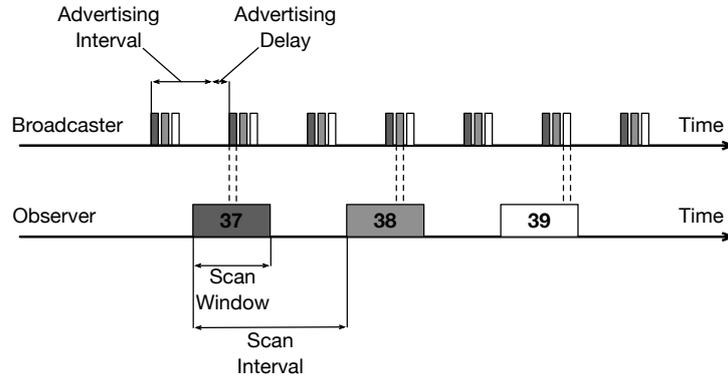


Figure 3.1: Operation in Broadcaster and Observer roles according to BLE Specification v4.0 [26]. The Broadcaster sends the advertising data on the three advertising channels (37, 38 and 39). The Observer detects the advertised data when the scanning channel is aligned with the advertising channel.

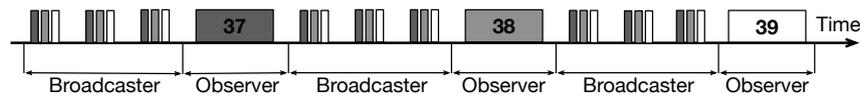


Figure 3.2: Alternate between Broadcaster and Observer roles. The device alternates between advertising data and listening for incoming data.

on a different advertisement channel, until all three are used and then repeats. When the current scan channel is aligned with the current advertising channel of another device, the Observer receives the advertisement packet from the Broadcaster and thus detects its presence.

The key for proximity detection resides in the fact that each device should be able to alternate between the Broadcaster and Observer roles periodically. For example, Figure 3.2 depicts the desired behaviour where the two roles are interleaved periodically: when a device is in Broadcaster role, it transmits an advertisement that can be detected by other devices and when it is in Observer role, it can detect other devices by listening for advertisements.

Although BLE is supported by most wearables, including smartwatches, OSs running on these devices prevent complete access to all BLE parameters. In order to analyse the effect of BLE parameters on sensing accuracy, we developed a prototype that allows us to freely control every parameter.

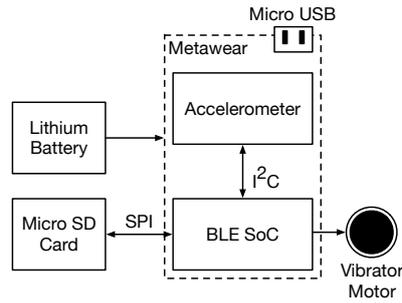


Figure 3.3: Wearable platform components block diagram. Only the components actually used for its development have been reported and not all available on the Metawear boards.



Figure 3.4: Electronics (left) enclosed in 3D printed box (3x4x1.5cm) and complete prototype attached to velcro wristband (right).

3.2.2 Wearable platform prototype

Our prototype is based on the Nordic nRF51822 BLE SoC that includes a 32bit ARM-M0 CPU and a 2.4GHz radio transceiver. We use a developer board from Mbitlab Inc. that contains the main SoC along with the associated circuitry, a Freescale MMA8452Q 3-Axis Accelerometer, an RGB LED, a push-button switch and a vibrator motor. Figure 3.3 shows a block diagram of our prototype. The entire prototype is powered by a 100mAh 3.7V lithium battery that can be recharged through a micro USB interface. We attach an SD card to log the list of nearby BLE devices. Figure 3.4 shows the current prototype. We designed a 3D printed box (3x4x1.5cm) to contain the device and we used velcro straps to wear the device on the wrist as to emulate a commercial smartwatch.

We use the S110 SoftDevice BLE stack by Nordic [152] for the Broadcaster role and to run the Observer role concurrently we use an open source library [189]. This library uses the Concurrent Multi-protocol Timeslot API to give access to the radio resource concurrently with the SoftDevice.

Data collection

Each device collects several pieces of information about other nearby BLE devices and about the participant wearing it. For each device in the vicinity, it logs the MAC address, the Received Signal Strength (RSS) and the channel on which a packet from the other device has been received (37, 38 or 39). The information is timestamped. The location information is provided by additional BLE devices (*static beacons* from now on) deployed in the building which are static and are associated with a certain area, usually a room or a desk. These devices continuously transmit a unique identifier of the area they are associated with, which is then used by the wearable devices to infer the current location. The recent increase in availability of static beacons in cities and retail spaces makes them perfect for localisation without the need to install additional infrastructure.

The embedded accelerometer is used to detect steps taken by the user. The 3-axis raw data is processed on the device with a step detection algorithm [225] to detect whether the user wearing it is stationary or walking. The SD card stores the number of steps taken from the moment the user starts walking to the moment she stops as well as the timestamps of beginning and end of walks.

3.2.3 BLE parameters

The parameters that characterise a BLE-based proximity sensing system are:

- *Advertising Time*: the time to send an advertisement on three channels.
- *Advertising Interval*: the time between each advertisement.
- *Scan Interval*: the time between scans.
- *Scan Window*: the duration of each scan.
- *Transmission Power*: the transmit power for each advertisement.

Advertising Time, Advertising Interval, Scan Interval and Scan Window affect how quickly a specific device can detect other devices in the vicinity and is detected by them. Intuitively, it is necessary to advertise more frequently than scanning in order to ensure that at least one advertisement will be captured during a scan and the Scan Window should be long enough to capture at least one advertisement on one channel. These parameters are interdependent and dictate the actual packet reception rate achieved by a device. It is not possible to achieve a higher rate and thus higher temporal granularity by simply advertising more frequently because it is also necessary to scan frequently and for longer periods. The Transmission Power is the only parameter available to control the maximum distance at which a contact can be detected: it allows changes in the range at which other devices can still correctly receive a packet. We now systematically inspect each parameter individually.

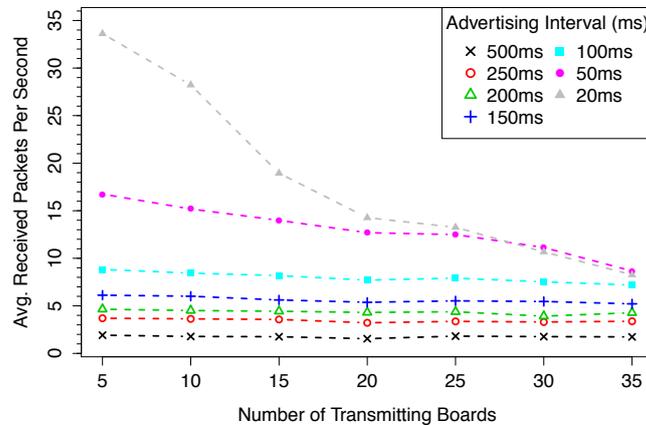


Figure 3.5: Average number of received packets changing the Advertising Interval and the number of transmitting devices.

Advertising Time: Advertising Time is the time required to send an advertisement packet on the three channels (Figure 3.1). It cannot be controlled directly as it depends on the packet’s payload. To keep this time to minimum, it is necessary to advertise a very limited amount of data. For example, our prototype platform achieves an Advertising Time of around 3ms for all three channels with 14 bytes of payload that includes Bluetooth flags¹, Transmission Power and custom information for identification, diagnostics and time keeping. However, Advertising Time does not have a significant impact on the packet reception rate at the receiver side or energy consumption at the transmitter side.

Advertising Interval: this controls how frequently advertisements are transmitted and thus it affects how quickly a device can be detected by other nearby devices. Assuming that an Observer device is scanning continuously, the time between the reception of two advertisements should, on average, be equal to the Advertising Interval under ideal conditions. However, packet loss due to collisions and environmental factors can affect how frequently advertisement packets are received. We, therefore, devised an experiment to understand the effect of Advertising Interval and the number of transmitting devices on the number of received packets. We configured one device to scan continuously and every 5 minutes we added 5 devices transmitting with a fixed interval, up to a total of 35 devices. The experiment was repeated for 7 different intervals.

Figure 3.5 shows that the number of transmitting devices affects the number of received packets, especially at high rates (advertising interval of 20ms and 50ms). At these rates, every time another set of devices was added, the average number of received packets dropped resulting in around 10 packets per second when 35 devices were transmitting at the same time. On the other hand, at lower rates (10Hz and less) the number of received pack-

¹<https://www.bluetooth.org/en-us/specification/assigned-numbers/generic-access-profile>

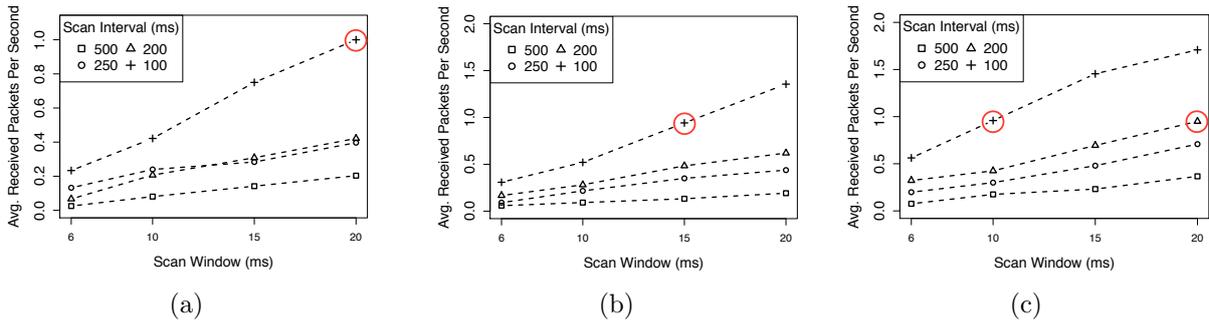


Figure 3.6: Average number of received packets changing the Advertising Interval, the Scan Interval and the Scan Window. (a) Adv. Interval 200ms. (b) Adv. Interval 150ms. (c) Adv. Interval 100ms.

ets remains constant even when the number of devices increases. We, therefore, chose 100ms as the lower bound advertising interval. This experiment shows that small advertising intervals do not necessarily lead to high reception rates (considering that the scan rate and window are constant) especially with high density of devices. Moreover, it can be detrimental for the battery lifetime as packets lost due to collisions represent wasted energy.

Scan Interval and Window: in the previous paragraph, the Observer was scanning continuously, but as we will show in Section 3.2.4, continuous scanning has a significant impact on battery life. Therefore it is necessary to duty cycle the scan operation using the Scan Interval and Window parameters. To study the effect of these parameters on the receive rate we ran several experiments where one device transmits at one of the Advertising Intervals tested in the previous paragraph and a second device performs scans with a particular Scan Interval and Window. For each Advertising Interval, we used the values in these sets, {100ms, 200ms, 250ms and 500ms}, {6ms, 10ms, 15ms, and 20ms}, respectively for Scan Interval and Scan Window, combining them in each possible way.

Figure 3.6 shows the results of these experiments. It demonstrates: (1) the interplay between Broadcaster and Observer parameters, and (2) how Scan Interval and Window can be combined to obtain specific receive rates. These results show that it is not possible to consider the Broadcaster and Observer roles in isolation when designing a proximity-based system. For example, an average receive rate of 1 packet per second can be achieved with four different combinations of the three parameters (red circles in Figure 3.6). We will show how this can be used to optimise the power consumption of the system in Section 3.2.4. Due to space constraints we do not report results for the other Advertising Intervals analysed in Figure 3.5, however, they follow the same trend.

Transmission Power: BLE transceivers usually offer the functionality to control the transmission power. This makes it possible to adjust the range and thus the spatial granularity of the system. In order to study the effect of transmission power on the communi-

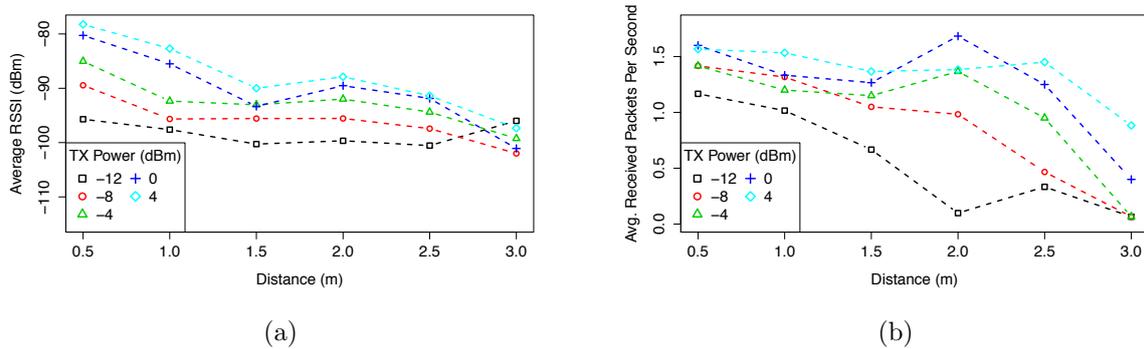


Figure 3.7: Received Signal Strength (a) and receive rate (b) at different distances and with different TX Power Levels. The power level values tested are some of the ones available on the Nordic nRF51822 chip and they do not correspond to the actual power emitted by the antenna.

ation range, we performed an experiment where two users wear our prototype devices on their wrists. We then varied the distance between them from 0.5 up to 3 meters. The two devices were configured to receive an average of 1.5 packets per second in close proximity with a transmit power of -4dBm. Figure 3.7 shows the average RSSI and the average number of received packets per second for this experiment. It shows that the maximum transmission range of a typical transceiver is of the order of a few meters. This makes them ideal for proximity sensing.

The graphs show how reducing the transmission power affects the BLE communication range: the signal strength and the number of received packets decrease with distance and with reduced power. These are expected results and similar to the ones observed in other studies regarding Bluetooth Classic [129] and BLE [64]. Even if the RSSI and packets per seconds trends are expected to be similar to the ones presented here, in other environments different absolute values could be observed because other factors could affect the radio communication range, such as the device’s antenna and its orientation. Therefore the system designer should test the target platform for the actual achievable range before deploying it.

3.2.4 Parameters’ impact on power consumption

We have seen that a specific receive rate can be achieved with different combinations of the parameters. Thus, it is important to consider the effect of each parameter on power consumption to select the combination that provides the desired receive rate and the least power consumption.

Advertising Time: We discussed how the advertising time depends on the size of the

Table 3.1: Impact of payload size on power consumption.

Payload Size (bytes)	Average Power (mW)
4	1.41
14	1.48
31	1.59

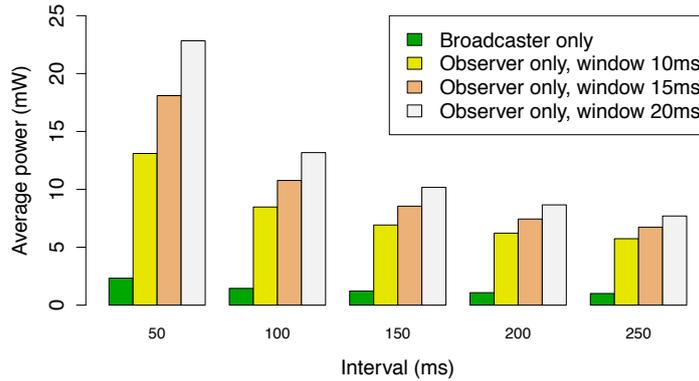


Figure 3.8: Average power consumption in isolation by Broadcaster and Observer roles for different combinations of Advertising and Scan Interval (x-axis) and Scan Window (last three colours).

transmitted packet. We, therefore, measured the power consumption of our prototype using the Monsoon Power Monitor² as it transmits advertising packets of different sizes. We tested packets with 3, 14 and 31 bytes (the maximum payload allowed in BLE advertisements). The device was configured in Broadcaster mode only with an Advertising Interval of 100ms and a transmit power of -8dBm. Table 3.1 shows the results for this experiment. The power consumption of the device increases slightly to 1.59mW from 1.41mW as the packet payload varies from 3 bytes to 31 bytes. This difference could reduce the battery life by more than 25 hours on a 100mAh battery.

Advertising Interval, Scan Interval and Window: These parameters provide greater control on the temporal granularity of a proximity system. However, the Scan Interval and Window are the parameters that affect the power consumption the most. To study the impact on power consumption we configured our device in Broadcaster only mode first and then in Observer only mode, with different combinations of the three parameters and we measured the average power.

Figure 3.8 shows the results of these experiments. Even at the same rates, Observer role has a greater impact on power consumption as compared to Broadcaster role. Therefore,

²<http://www.monsoon.com/LabEquipment/PowerMonitor>

to achieve a certain desired receive rate, it is better to scan with a low frequency and for short periods and transmit more frequently in order to have a lower impact on the power consumption. Advertising with an high frequency has significantly less impact on the power consumption than scanning with high rates and duty cycles. However, as explained in the previous section, high transmission rates can lead to collisions if the density of devices is high: this must be kept in consideration when designing a proximity-based system for crowded environments.

Transmission Power: To study how this parameter affects the power consumption, six different power levels (from -20dBm to 4dBm with steps of 4dBm), which can be selected in software on the Nordic nRF51822 chip were tested. We configured our prototype in Broadcaster only mode with an Advertising Interval of 100ms and packet's payload of 14 bytes. The average power difference between the highest (4dBm) and the lowest (-20dBm) power level is around 0.37mW. This variation translates to an estimated difference in battery life of 58 hours for a 100mAh battery. This shows that even the selected transmit power could affect the battery life of the system but, this being the only parameter available to control the transmission range, it might not be possible to optimise it for power consumption.

All these measurements will be used later in the chapter to estimate the battery lifetime of the device when both the Broadcaster and Observer role are enabled at the same time.

3.3 Proximity sensing on commercial devices

After the study of BLE parameters using our prototype, we now analyse to what extent the same parameters can be exploited on commercial wearables. These devices are equipped with a BLE chip used for communication with the user's phone and being always co-located with (worn by) the user they offer a great advantage for proximity sensing. The platforms we used for our analysis are Android Wear 5.0 [75] and Tizen Wearable 2.3.1 [199]. While it was already possible to implement the Observer role, the Broadcaster role has been enabled in recent releases (e.g., September 2015 for Tizen).

The actual devices we used for our experiments are a Samsung Gear S2 [184] for Tizen and a Samsung Gear Live [183] for Android Wear. We developed an application for each device that allows us to change the parameters and start/stop the advertising and scan operations. Both devices are able to transmit and scan in the background and when they are connected to a phone, so they can still receive notifications from the paired phone as during normal operation. In all the following experiments, both watches are connected to an Android phone. Similarly to what we did with our custom device in Section 3.2.3 we now systematically inspect the parameters individually to understand capabilities and limitations of BLE on commercial devices.

Advertising Time: The only way to control this parameter is to vary the number of bytes

included in the advertising packets. Both operating systems expose APIs to configure the content of the BLE packet (i.e. device name, service and manufacturer data, etc.). The only difference is that Tizen offers the possibility to set the appearance of the device and solicitation UUIDs while Android Wear does not. With these APIs it is possible to reduce the Advertising Time, consequently reducing the power consumption.

Advertising Interval: This parameter can be adjusted on both platforms in similar ways. Three values are allowed: (1) *Low Latency*, (2) *Balanced* and (3) *Low Power* (called *Low Energy* in Tizen). However, the resulting behaviour is different for the two operating systems. In Tizen the three values correspond to an Advertising Interval respectively of 150ms, 500ms and 1s. By contrast, the Android Wear watch, regardless of the value set, starts advertising with a 30ms interval for about 150/180 seconds and then switches to an interval of 1280ms. This shows that for the Android Wear platform the Advertising Interval in practice cannot be controlled and the only usable value is 1280ms.

Scan Interval and Window: For what concerns the Observer role, the Tizen OS does not allow to set any parameter, it only allows the developer to start and stop the scan operation. Android Wear, on the other hand, does not permit the configuration of the Scan Interval and Window individually, but it allows to choose among three global values for the scan operation: (1) *Low Latency*, (2) *Balanced* and (3) *Low Power*. To test the actual achievable receive rate with these three values, we configured one of our custom devices in Broadcaster only mode with an Advertising Interval of 100ms and we let the Android Wear watch scan. If the scan is configured in *Low Latency* mode the watch scans continuously, achieving in this case an average receive rate of around 9Hz. In *Balanced* mode, instead, the average receive rate is halved (around 4Hz) and in *Low Power* is one-tenth (around 1Hz).

Figure 3.9 shows the packets received over a 2 minute scan with each setting. Looking at the grey lines overlaid on the graph, which represent the instants when a packet has been received, it is possible to observe the duty cycle applied on the scan operation. While in Low Latency the packets are uniformly distributed across the scan period, in Balanced and Low Power modes the watch performs a scan around 12 times in a minute. Therefore we assume the Scan Interval is roughly 5 seconds. Moreover, in Balanced mode the duration of the scans (Scan Window) is larger than in Low Power mode. Looking closely at the received data we notice that the Scan Window is around 2s in Balanced mode and 500ms in Low Power mode.

As mentioned earlier, in Tizen it is not possible to select any setting for the scan operation. Performing the same experiment with the Android Watch (with a device transmitting at 10Hz), we discovered that for the Tizen watch the average receive rate is around 1Hz and the Scan Interval and Window are equal to the ones adopted by Android Wear in Low Power mode (scan for 500ms every 5 seconds).

Transmission Power: Tizen does not provide any API to control the transmit power. Therefore it is not possible to control the transmission range. The transmit power level

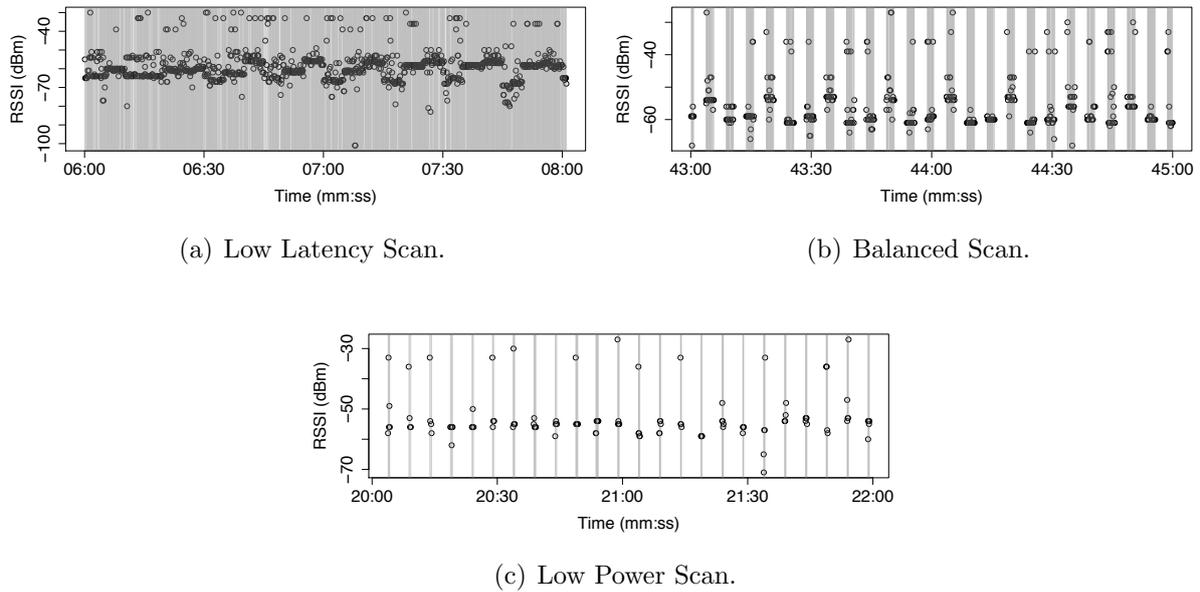


Figure 3.9: Packets received by an Android Wear watch with the three possible scan settings. The transmitting device was configured with an Advertising Interval of 100ms. Each gray line overlaid on the graph represent the instants when a packet has been received by the watch.

Table 3.2: Summary of control possibilities on Android Wear and Tizen. The asterisk character (“*”) indicates that the APIs offer the possibility to set different values but they have no effect on the watch we tested.

Parameter	Gear Live (Android Wear)	Gear S2 (Tizen)
Advertising Time	Yes	Yes
Advertising Interval	No*	Yes
Scan Interval and Window	Yes	No
Transmission Power	No*	No

included in the advertisement packets is 12dBm and the average RSS at 1 meter is around -78dBm.

By contrast, Android Wear offers an API that permits choosing between four different values: *High*, *Medium*, *Low* and *Ultra Low*. However, regardless of the value set, the watch we tested uses the same power level and includes the value -21dBm in the advertisement packets. This is also confirmed by the fact that even by setting a different value, there is no substantial difference in the RSS we measured at 1 meter and its average is always around -66dBm.

Table 3.3: Impact of payload size on the power consumption of the Gear Live and Gear S2.

Payload Size (bytes)	Average Power Android Wear (mW)	Average Power Tizen (mW)
6	10.58	7.79
14	10.84	8.21
31	10.92	9.65

3.3.1 Power consumption

In this section we analyse the impact of the different adjustable parameters on the watches' power consumption. All the measurements have been taken with the watch connected to an Android phone and with the screen off.

Android Wear

The Gear Live has a 3.7V, 300mAh battery and the power consumption when idle with the screen off is 10.29 mW. The only parameters that can be controlled are the Advertising Time and the combination of Scan Interval and Window.

As expected, the Advertising Time has a limited impact on power consumption. For example, the power difference when transmitting 6 or 31 bytes is around 0.19mW which gives a difference in lifetime of only 1.6 hours for a 300mAh battery. These measurements have been performed after the initial period (150/180 seconds) in which the watch transmits at a high rate, because that period does not represent the normal transmission rate (see Section 3.3 for more details). Table 3.3 reports the power measurements data in comparison with the Gear S2 described in the next section.

Regarding the Scan Interval and Window, we tested the three possible global values, *Low Latency*, *Balanced* and *Low Power*. We observe a more substantial effect on power consumption. Table 3.4 shows that when the *Low Latency* mode is selected, which corresponds to the watch scanning continuously, the power consumption is very high and in this case the expected battery life would be only around 5 hours. This would make it impossible to deploy a proximity-based application in a workplace environment because it could not cover the standard 8 hours of work. Similarly, the *Balanced* mode would result in an expected battery lifetime of slightly more than 8 hours. The only mode that would enable this kind of deployment is the *Low Power*. In this mode, the watch has an estimated battery life of more than 13 hours but during a typical working day the proximity detection system would remain active for 8 hours only. This means that in the remaining part of the day the power consumption will be lower because the watch would not scan and advertise periodically. This should guarantee enough energy for normal usage.

Table 3.4: Power consumption of Scan and Advertising modes for the two smartwatches. The packet used for the Tizen experiments is 14 bytes long.

Mode	Average Power Android Wear (mW)	Average Power Tizen (mW)
Scan Low Latency	227.25	-
Scan Balanced	124.32	-
Scan Low Power	82.18	-
Advertising Low Latency	-	8.21
Advertising Balanced	-	6.61
Advertising Low Power	-	6.19

Tizen Wearable

The Samsung Gear S2 (3.8V, 250mAh battery) allows the developer to control Advertising Time and Interval but not the scan parameters and the Transmit Power. This watch consumes 5.81mW when idle and with the screen turned off.

The first measurements we analyse are in relation to the Advertising Time. Even if the Tizen APIs offer a function to specify that the transmit power level should not be included in the advertisement packets, this API is not working and the power level is always included. Therefore the smallest packet that we were able to test is 6 bytes long (Bluetooth flags and power level). Table 3.3 reports the details of the power measurements while the watch advertises in Low Latency mode (every 150ms). In this case we observe a greater impact on power consumption. Indeed we estimate that when advertising 31 bytes the battery would last 23 hours less compared to transmitting just 6 bytes.

As opposed to Android Wear, Tizen OS permits the developer to select one of three different Advertising Intervals. In this case, as it is possible to see in Table 3.4, the average power consumed, even at a relatively high transmission rate (i.e. *Low Latency*), is limited and it is considerably lower than during the scan operation, which for this watch is 55.56mW.

To summarize, we have observed that both systems do not give complete freedom on the setting of the parameters, rather they allow the developer to choose between predefined values. Android Wear offers APIs to control all the BLE parameters but only two of them work on the watch we tested (Advertising Time and Scan related parameters). On the other hand, Tizen offers APIs only to modify Advertising Time and Advertising Interval. Table 3.2 summarises the parameters that can be controlled on the two platforms. We have also found the Tizen Wearable watch we examined to be more energy efficient than the Android Wear device. The Tizen watch in fact consumes less energy when idle and when transmitting or scanning for beacons.

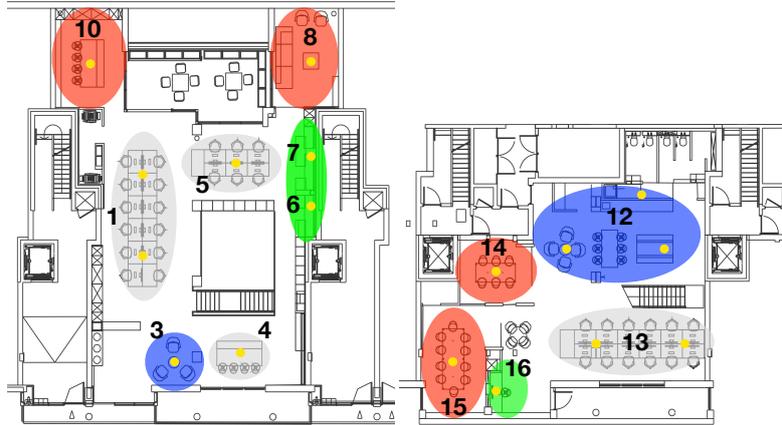


Figure 3.10: Floor plans of the first floor (left) and lower ground floor (right) of the office space studied in this work. The yellow circles represent the location of the static BLE beacons. The coloured shadows represent the category of the locations.

3.4 Workplace deployment

We now evaluate the overall performance for proximity sensing with a deployment in a workplace environment. We begin by explaining our experimental method and the study environment. We then describe how we extracted the proximity information and the location traces from the raw data. Finally, we present the results for our experimental platform and for the two analysed smartwatches.

3.4.1 Experimental method and testbed

A proximity sensing system is characterised by many parameters that affect performance. As shown in the previous sections, these parameters are often inter-dependent. Therefore, to evaluate different parameter combinations in a real environment multiple deployments would be necessary.

Our approach instead was to deploy our prototype, which offers greater flexibility, and then test different combinations of the parameters by post-processing the collected data. In particular, we are interested in knowing how a proximity detection system would work on wearable off-the-shelf devices.

Our testbed consists of an architecture company (Spacelab Ltd.) which employs more than 35 people. The company occupies a building which consists of two floors with a staircase opening in the middle (Figure 3.10). The two large open spaces host different workstations where several employees share the same large table. There are meeting rooms on both floors, while the kitchen and break out area are on the lower ground floor. The

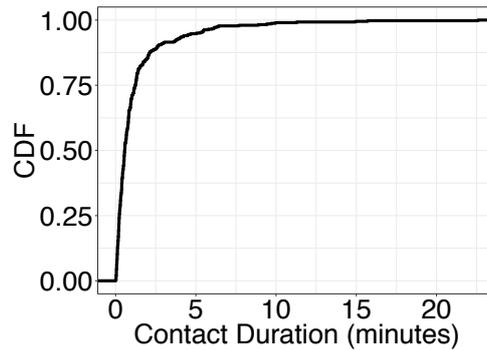


Figure 3.11: Cumulative Distribution Function of the contact durations recorded by the observer who collected the ground truth during the study.

company has a very dynamic and flexible working style. Employees do not have assigned desks, the work tasks are fluid and people have considerable interaction. Before beginning with the deployment our work has been approved by the University of Cambridge ethics committee³. All the participants consented to take part in the study after being informed of the purposes of the study. All collected data are anonymous and make no reference to the individual participants.

3.4.2 Participants

We recruited 25 participants (15 females) aged 21-44 ($\mu = 31$) for a period of four weeks between September and October 2015. The company is structured into five teams: Architecture (4 participants), Interior Design (10 participants), Workplace Consultancy (6 participants), Project Management (1 participant) and Administration (4 participants).

The vertical structure comprises 7 levels, from the top level (1), to the bottom (7). At the top of the hierarchy is one of the two Partners who works mainly with the Architecture team. The second highest level are Directors and the Project Manager (4 participants). The third and fourth levels consist respectively of Associates (2 participants) and Senior architects, designers and analysts (7 participants). At the fifth and sixth levels there are architects, designers and analysts (4 participants) and Assistants (3 participants). The Administration team was counted as a seventh level for consistency, although it would be fair to consider it as external to the hierarchical levels.

3.4.3 Ground truth

In order to collect ground truth data, one researcher performed observations for three days during the study. During each observation, the researcher followed a person and annotated

³Our agreement with Spacelab does not include the publication of the collected dataset.

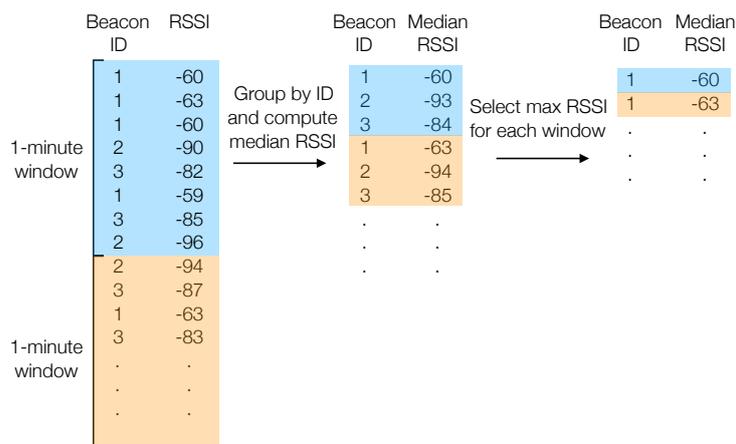


Figure 3.12: Visual representation of the processing applied to the data received from the static beacons to compute the user’s approximate location. The raw data from the static beacons (left) is first segmented into 1-minute windows. Within each window, all the received packets are grouped by beacon ID and for each ID the median RSSI value is computed. At this point the ID with the highest RSSI is chosen as the current location of the user for each window. The colours represent the aggregation windows used to segment the data.

all the social interactions the person had. Since we are interested in detecting fine grained proximity between people, the researcher recorded only those interactions that happened in close proximity, i.e. up to a distance of 3 meters between people. For each interaction event the researcher recorded the start time, the end time, the location inside the office and the unique ID of the people involved. In total we observed 18 different participants who have been chosen in order to represent the teams in the company. This resulted in 19 hours of observations during which we captured 401 interactions. On average an interaction is 1 minute and 13 seconds long and 70% of the interactions are shorter than 1 minute while only 5% are longer than 5 minutes. The largest interaction captured involved a group of 5 people. The distribution of the contact durations is reported in Figure 3.11.

3.4.4 Static beacons and location traces

As shown in Figure 3.10 seventeen BLE static beacons were deployed in the building with the purpose of giving coarse grained (at the desk level) location information about the participants. One beacon was placed on each desk or, if the desk was too big, two beacons were used. The beacons were configured with a beacon rate of 5Hz with a range of about 4 meters. We highlight that the static beacons have been used in this work for evaluation purposes, however our wearable prototype could be used for proximity detection even when those beacons were not available (e.g. outdoor).

Beacon ID	Walking		Beacon ID
1	0		1
1	1		1
2	1		2
2	0		2
2	0		2
3	0	Correction →	2
1	0		2
2	0		2
2	0		2
2	0		2
2	1		2
3	1		3
4	1		4
5	1		5
⋮	⋮		⋮
⋮	⋮		⋮
⋮	⋮		⋮

Figure 3.13: Visual representation of the correction applied to the approximate location computed from the static beacons. After we compute the initial approximate location we align the sequence of locations with the walking activity of the user (represented by 1 if the user is walking or 0 otherwise). Changes in locations that correspond to walking events are preserved while variations in location without walking activity are discarded. For example, the picture shows that there is an apparent movement from location 2 to 3 and then 1, however since the user was not walking at that time the change is discarded and the last valid location (2 in this case) is used instead. The other location changes (marked in green) are preserved because the user was walking at the same time.

To associate the current approximate location to the participants at each point in time, the data received from the static beacons (containing their ID and the packets RSSI) was grouped into non-overlapping windows of 1 minute. We then computed the median value of the received signal strengths (RSS) from the different beacons. This process removes high frequency variations in the data which might ruin the location inference. We chose the location for each time period by selecting the beacon with the strongest median RSS, which represents the closest one to the user. Figure 3.12 shows a visual representation of this process. To improve the location estimation we used the accelerometer data. With a step detection algorithm [225]⁴ we detected when the participants were walking and given that a person changes location only when she walks, we could remove spurious changes in location (which might be due to reflections in the radio signals) if the user was not walking at that time. Figure 3.13 illustrates how this correction works with a visual example. Using the location traces we inferred the desk each participant used each day by selecting the one where the person spent most of the time.

The different locations were grouped in 6 semantic categories represented by the colours in Figure 3.10:

⁴We empirically evaluated the accuracy of the step detection algorithm and we found it to be within 5-10 steps.

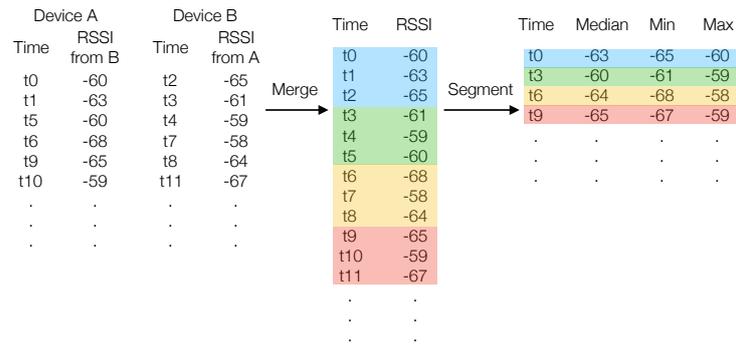


Figure 3.14: Visual representation of how data from two devices is merged together and then segmented into non-overlapping windows. RSSI-based features are then computed over the windows (median, min and max). The colours represent the aggregation windows used to segment the data.

Open space workstations: shared workstations that can accommodate several people and represent the main areas of work in the company (colour grey).

Meeting rooms: four meeting rooms are present in the building, two on the first floor and two in the lower ground floor (colour red) and all of them have a table in the middle.

Private workstations: small areas for individual work or maximum for two people (colour green).

Breakout areas: a relatively large open kitchen is present in the lower ground floor and a small table with magazines on the first floor (colour blue).

Circulation spaces: this is not an exact location because we did not deploy beacons in the space around desks. However, this label is used to tag contacts between people that are not close to the same static beacon. When this happens one of the participants is co-located with one beacon and the other participant with a different nearby beacon, in this case the contact between the two is tagged as happening in Circulation.

Outside office: the absence of location information while the device was in use is interpreted as if the user was outside.

We will use this semantic organisation of the static beacons in Chapter 4 where we study the social dynamics inside the company.

3.4.5 Wearable devices and proximity traces

Each participant was asked to wear, on the wrist and only when inside the office, our wearable prototype. We provided a charging station where all devices were recharged during the night and where some spare devices were stored as replacement in case of

failures. Every night an Android Phone collected all the data from the wearables and uploaded it to our servers.

The devices were programmed with an Advertising Interval and Scan Interval of 100ms and a Scan Window of 20ms. The transmit power was set at -8dBm. This configuration allowed us to achieve an average receive rate of 2.15Hz and a range of around 4–5 meters and was selected because it represented the best compromise between battery lifetime (around 20 hours to cover a working day) and granularity of the collected data. We specifically chose a higher receive rate than the ones achievable by the watches because this would enable us to post-process the data and match the smartwatches' rates.

The raw data collected by the Bluetooth devices had to be processed in order to classify the contact events as proximity or not. We adopted a supervised machine learning approach where we trained a binary classifier with a set of examples labelled as “proximity” or “non-proximity”. We were not interested in measuring the actual distance between the participants but only if they were close to each other as during a conversation.

Using only the data collected during the three days of participant observations we built a data set where the positive examples (“proximity” label) were labelled with the observed communication events. We recall that the researcher was instructed to record only interactions that involved close proximity between the participants, assuming that social interaction events are examples of close proximity. Instead, the negative examples (“non-proximity” label) have been labelled using the static beacons. From the logged data we computed the beacon with the strongest signal strength (i.e. the closest one) at each point in time and we co-located the participant with that beacon (as described in the previous section). For each pair of participants it will happen that for some time periods they would be co-located with the same beacon (e.g. when they were sitting at the same desk), and for other periods they would be co-located with different beacons (e.g. when they were at different locations in the building). We selected those periods where the two participants were at different locations and we used them as “non-proximity” examples.

For each pair of individuals, A and B, we extracted from their devices the stream of raw data relative to the other device. This consists of the received timestamped packets with MAC address and RSSI value. In other words, from A's device we extracted all packets received from B's device and vice versa. Given that the logging of packets is symmetrical (each device transmits and receives packets) we merged the two streams into one in order to have more data points between the two participants. This stream of data is then split into non-overlapping windows of different sizes (from 1 to 60 seconds). For each window we computed the following features: median RSSI, min RSSI and max RSSI. Figure 3.14 shows this approach visually. The main purpose of those features is to mitigate multipath interference which produces high frequency variations in the RSSI measurements. Median and max features have also been identified by Faragher et al. to provide the best multipath mitigation effect in the context of indoor localisation with BLE [64, 63]. Empirically we determined that adding also the min RSSI feature improves the detection accuracy while adding the number of received packets and the standard deviation of RSSI values produces

worse results.

When two people were very far from each other (e.g. in different floors of the building) the two devices would not receive any packet and this would result in missing values in the data set. In our context, those missing values indicate that the two devices were not in proximity and could be used by a machine learning algorithm to correctly classify them. For this reason we replaced the missing values with the value -110 which represents a very low RSS which is below the minimum detectable power by our device (-105dBm). At this stage of the processing we knew the missing values are only due to the fact that the devices were not in range because we had already filtered the devices for which data was not recorded (e.g. malfunction or forgotten at home). Once we segmented the data streams for each pair of participants in windows and computed the aforementioned features we aggregated all windows (which represented examples for the supervised machine learning algorithm) into a single data set. At this point we overlapped the participant observations and we labelled each window with “proximity” or “non-proximity”.

3.4.6 Training and Evaluation

For the classification we adopted Decision Trees (C4.5) and we trained and evaluated them with two strategies: 1) stratified 10-fold cross-validation and 2) splitting the dataset into training and test set ensuring dyad independence between the two sets. In the second strategy we reserved roughly 80% of the data for training and 20% for testing ensuring that data from the same dyad would not be included in both sets. 10-fold cross validation has been chosen because, despite it could produce over-optimistic models, it is a widely used technique and provides an understanding of the base performance of the model, especially when dealing with small datasets. The second strategy instead, where dyad independence is maintained between train and test set, could provide a better understanding of the generalisation power of the model.

In both cases the resulting dataset presented class imbalance because for each pair of people we labelled the positive examples from the interaction events that have been observed, which represented a limited period of the day, but we derived the negative examples from the times when they were at different locations in the building and these could cover longer periods of the same day. Thus we over-sampled the minority class generating synthetic examples using the SMOTE technique [45] in order to balance the two classes. For the second evaluation strategy (i.e., with dyad independence) the minority class has been oversampled only in the training set.

We favoured the selection of the Decision Trees classifier because we considered the possibility of implementing the proximity detection classifier on our wearable prototype (Section 3.2.2) to enable online proximity detection. However, the limited amount of computational power and memory available on the Cortex-M0 (i.e., 16MHz clock speed and 16KB of RAM) severely limits the choice of algorithms. Decision Trees are simple to im-

plement, fast and efficient at inference time [17, 6] making them a reasonable candidate for applications on constrained devices. Also, the work from Liu et al. [129] which used multiple, manually calibrated thresholds on RSSI values to estimate face-to-face proximity motivated us further on the use of Decision Trees to automatically learn the thresholds from data. Additionally, Decision Trees have been successfully used on a variety of human behaviour classification tasks (e.g., activity recognition [21, 190], social relationship classification [142] and transportation mode detection [210]). More complex algorithms could be selected if hardware with more capabilities is available.

The algorithms have been taken from Weka version 3.7.13 [131]. The list of parameters used for the Decision Trees is reported in Appendix A. For the performance metrics, we used the average F1-score [46] and the average area under the Receiver Operating Characteristic (ROC) curve [83].

3.4.7 Results

Before analysing the classification results we will report the metadata about the study. During the 4-week long study we lost 10.8% of the total amount of data that we were expecting to collect due to failures. These failures were due to different causes: device malfunctions, devices out of battery, devices forgotten at home or lost (two participants reported that they lost their devices due to problems with the plastic box). Malfunctions were identified by the fact that the data we were expecting to collect from the devices (one file per day) was completely missing in some cases. In these situations the devices had to be replaced (3 devices out of 25 have been replaced during the study). Devices out of battery and forgotten at home actually fall in the same category which represent a situation where the participants forgot to re-charge the device at the end of the working day. By logging the battery voltage every 30 minutes, both during charging and battery-powered operation, we could identify when a device was let deplete all its power and then charged again after some time (usually the next working day). 30% of the data did not contain any contacts because, although the devices worked properly, they were not in use but they were charging at the charging station⁵. This could be due to the fact that the working style is very dynamic and people are often outside to visit construction sites and forget to wear the device.

We now present the classification results for the raw data collected with our prototype and we show the results that would have been achieved by the two off-the-shelf wearable platforms. We do this through down-sampling of the raw data to match the wearable devices' rates. We consider two down-sampling strategies: 1) where we uniformly remove data points to match the data rates achievable by the commercial watches and 2) where we emulate the watches' scan behaviour by keeping only the first 500ms of data every 5 seconds and then applying uniform down-sample, if necessary, to ensure that the resulting

⁵This situation was also detected thanks to the battery voltage logging every 30 minutes.

Table 3.5: Parameter configurations and expected battery life for the two considered wearable platforms and for our wearable prototype when the Broadcaster and Observer role are enabled at the same time.

Configuration Name	Advertising Interval (ms)	Scan Interval (ms)	Scan Window (ms)	Average Receive Rate (Hz)	Expected Battery Life (Hours)
Our Wearable Prototype	100	100	20	2.15	19.33
Android Wear Low Power	1280	5000	500	0.08	13.74
Tizen Low Latency	150	5000	500	0.62	14.95
Tizen Balanced	500	5000	500	0.19	16.36
Tizen Low Power	1000	5000	500	0.1	17.02

average receive rate matches the one of the different watches’ configurations.

In the following we present results achieved with 10-fold cross validation and dyad independent training and testing when we over-sample the minority class in our dataset as described in Section 3.4.6. For completeness, in Appendix B we provide also the results when we train a model with 10-fold cross validation without any oversampling or with down-sampling of the majority class.

Table 3.5 summarises the different configurations identified for the two platforms and the configuration we used on our device. It also reports the expected battery life achievable by each device when the Broadcaster and Observer role are enabled simultaneously. We decided not to include the configurations *Android Wear Low Latency* and *Android Wear Balanced* because, as observed in Section 3.3.1, they present an excessive power consumption for our target environment.

Our Wearable Prototype

We begin by looking at the results achieved by our wearable prototype in Table 3.6 for 10-fold cross validation and Table 3.7 for dyad-independent training and test sets. Firstly, we notice that, in both cases, when increasing the window size the F1 measure and the area under the ROC increase. This is because the RSS data has high-frequency noise which is increasingly attenuated by computing the features over a larger number of data points. However, this impacts the granularity of the detected proximity events. For example, when using a 10-second window it is impossible to say if the proximity event was 5 or 8 seconds long. From Table 3.6 we notice that already with a 10-second window our prototype achieves an F1 score of 0.97 with very little (or no) improvement for larger windows. Instead the model evaluated on the dyad-independent test set (Table 3.7) shows lower F1 and ROC scores especially for larger windows. This is explained by the fact that

Table 3.6: Average F1 Measure and average area under ROC curve (AUC) for different windows when the raw data is down-sampled uniformly (without emulation of commercial device behaviour) applying 10-fold cross validation on the entire dataset. The configuration names refer to Table 3.5. The data from our wearable prototype has not been post-processed.

Window Size (s)	Our Wearable Prototype		Tizen Low Latency		Tizen Balanced		Tizen Low Power		Android Wear Low Power	
	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC
1	0.79	0.85	0.67	0.71	0.54	0.60	0.46	0.55	0.44	0.54
5	0.94	0.97	0.90	0.94	0.81	0.85	0.72	0.77	0.70	0.73
10	0.97	0.98	0.96	0.98	0.91	0.95	0.84	0.90	0.81	0.86
20	0.97	0.98	0.98	0.98	0.97	0.98	0.95	0.97	0.93	0.96
30	0.97	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.97	0.98
40	0.97	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.98
50	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
60	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98

Table 3.7: Average F1 Measure and average area under ROC curve (AUC) for different windows evaluated on the dyad-independent test set. The raw data is down-sampled uniformly (without emulation of commercial device behaviour). The configuration names refer to Table 3.5. The data from our wearable prototype has not been post-processed.

Window Size (s)	Our Wearable Prototype		Tizen Low Latency		Tizen Balanced		Tizen Low Power		Android Wear Low Power	
	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC
1	0.79	0.68	0.79	0.63	0.78	0.56	0.76	0.54	0.76	0.52
5	0.85	0.84	0.85	0.77	0.83	0.74	0.82	0.68	0.82	0.66
10	0.86	0.87	0.85	0.81	0.86	0.81	0.84	0.75	0.84	0.75
20	0.87	0.85	0.88	0.84	0.86	0.84	0.87	0.83	0.85	0.81
30	0.86	0.88	0.86	0.87	0.88	0.85	0.86	0.85	0.86	0.82
40	0.86	0.85	0.86	0.83	0.88	0.86	0.86	0.81	0.88	0.87
50	0.90	0.91	0.88	0.87	0.88	0.91	0.84	0.84	0.84	0.87
60	0.92	0.91	0.90	0.91	0.85	0.78	0.86	0.87	0.86	0.87

this model is evaluated on data not seen at train time and therefore it is more difficult to correctly classify. While in the 10-fold cross validation scheme data from the same dyad might be present in both training and test set when the folds are created. Considering only results from our wearable prototype, which include all data collected from the deployment, we notice that the performance difference between the model evaluated with 10-fold cross validation and the one with the dyad independent test set is fairly limited and it ranges between 0.06 and 0.11 for the F1 score and between 0.07 and 0.17 for the ROC AUC (comparing Table 3.6 and 3.7).

To confirm that our devices can capture social dynamics accurately, we compared the devices' data with other datasets collected with a similar radio technology (i.e., RFID).

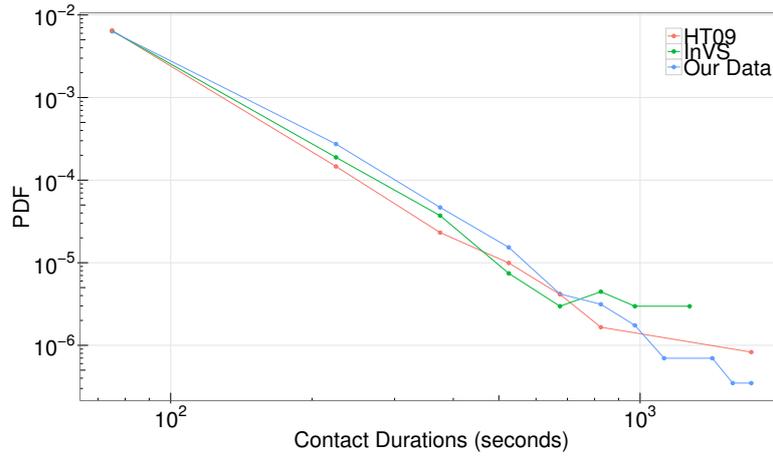


Figure 3.15: Probability Distribution Function of contact durations from our study and from two other studies that employed RFID tags: a conference (HT09) [92] and a workplace (InVS) [71]. Our data has been aggregated with a 20-second window given that the same resolution is used in the other two datasets.

We used open datasets available online from previous works and we plot (Figure 3.15) the Probability Distribution Function of contact durations for two datasets (HT09 and InVS) in relation to our dataset (Our Data). The two datasets have been collected with the Sociopatterns tags [44] and they are discussed by Isella et al. (HT09, conference) [92] and by Génois et al. (InVS, workspace) [71]. The technology used in these tags (i.e., RFID) is comparable to the BLE we used in our platform: this makes the comparison with our data relevant. As it is possible to see from Figure 3.15, the data collected during our study has a very similar contact duration distribution to the data collected in other settings. While this confirms the goodness of our data, we highlight that our work on agile workplace presented in Chapter 4 offers a variety of novel and additional findings around how groups use the space over time and how the groups continue to interact despite hotdesking. This goes beyond the aggregate contact duration distribution shown in Figure 3.15.

Commercial Devices

To assess the proximity detection capability of the commercial devices we considered, we post-processed the data collected with our wearable prototype by removing data points uniformly to match the watches' data rates. Tables 3.6 and 3.7 report their classification results. Firstly, we notice that similarly to what we observed with our custom device, when increasing the window size the metric scores increase due to the attenuation of high-frequency noise in the RSS values. However, given that the commercial devices have a lower average receive rate compared to our prototype (Table 3.5), they require a longer window to reach the highest F1 and ROC scores. For example, for the configuration Tizen Low Latency, characterised by an average receive rate of 0.62Hz, the best scores

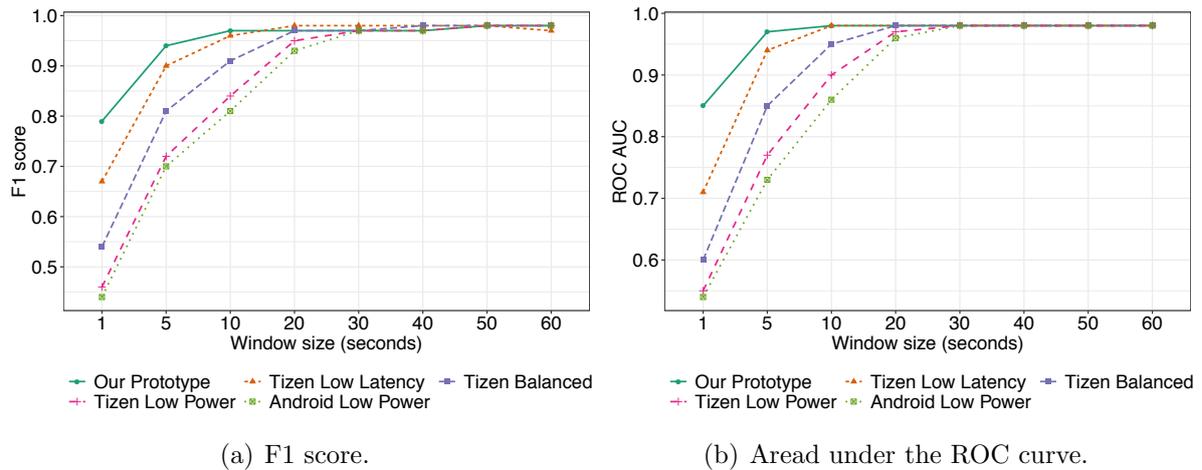


Figure 3.16: Performance metrics comparison for different window sizes when the raw data is down-sampled uniformly (without emulation of commercial device behaviour) applying 10-fold cross validation on the entire dataset (data from Table 3.6). The data from our wearable prototype has not been post-processed.

are first achieved with a 20-second window while for the configuration Android Wear Low Power (0.08Hz average receive rate) a 50-second window is required. This is because even if Android Wear offers the same Tizen’s duty cycle for the scan operation, its large Advertising Interval does not allow us to obtain a receive rate that is high enough. This results in the need to use larger window sizes to improve the accuracy. Figure 3.16 shows how the performance metrics increase when larger windows are used.

Now we consider a different way to post-process the data which is more similar to how the watches perform the scan operation. Watches in fact scan every 5 seconds and just for 500ms, therefore in this case we emulate this behaviour by keeping the first 500ms of data every 5 seconds and then making sure that the average receive rate matches the one achievable with each configuration applying uniform down-sampling if necessary. In this case we could not consider the configuration *Tizen Low Latency*. When choosing the deployment parameters for our device we had to find the best compromise between device lifetime and data collection rate in order to ensure a realistic scenario were people would wear the device for at least 8 hours a day. This resulted in a rate (2.15Hz) that prevented us from post-processing the data to match the *Tizen Low Latency* configuration. Tizen version 2.3.1, which was the first version to support BLE advertisements, was released after our deployment, therefore we could not account for its receive and transmit rates when configuring our device for the deployment.

From Table 3.8, which reports the results for the 10-fold cross validation evaluation, we notice that as the window size increases the metric scores increase and approximates the one of our custom device. This is because with larger windows the fact that the watch scans with a low duty cycle is mitigated. Indeed, with a larger window, data from subsequent

Table 3.8: Average F1 Measure and average area under ROC curve (AUC) for different windows when the raw data is post-processed to emulate the watches’ scan behaviour (scan for 500ms every 5 seconds) applying 10-fold cross validation on the entire dataset. The configuration names refer to Table 3.5. The data from our wearable prototype has not been post-processed.

Window Size (s)	Our Wearable Prototype		Tizen Balanced		Tizen Low Power		Android Wear Low Power	
	F1	AUC	F1	AUC	F1	AUC	F1	AUC
1	0.79	0.85	0.43	0.57	0.41	0.53	0.41	0.53
5	0.94	0.97	0.70	0.83	0.65	0.68	0.65	0.68
10	0.97	0.98	0.82	0.93	0.76	0.81	0.77	0.81
20	0.97	0.98	0.94	0.97	0.89	0.94	0.89	0.94
30	0.97	0.99	0.97	0.98	0.96	0.97	0.96	0.98
40	0.97	0.99	0.97	0.98	0.97	0.98	0.97	0.98
50	0.98	0.99	0.97	0.99	0.97	0.98	0.97	0.98
60	0.98	0.99	0.97	0.99	0.97	0.99	0.98	0.99

Table 3.9: Average F1 Measure and average area under ROC curve (AUC) for different windows evaluated on the dyad-independent test set. The raw data is post-processed to emulate the watches’ scan behaviour (scan for 500ms every 5 seconds). The configuration names refer to Table 3.5. The data from our wearable prototype has not been post-processed.

Window Size (s)	Our Wearable Prototype		Tizen Balanced		Tizen Low Power		Android Wear Low Power	
	F1	AUC	F1	AUC	F1	AUC	F1	AUC
1	0.79	0.68	0.76	0.52	0.76	0.52	0.75	0.51
5	0.85	0.84	0.82	0.65	0.82	0.63	0.81	0.62
10	0.86	0.87	0.83	0.76	0.82	0.70	0.83	0.68
20	0.87	0.85	0.83	0.82	0.85	0.78	0.84	0.80
30	0.86	0.88	0.86	0.85	0.85	0.82	0.85	0.80
40	0.86	0.85	0.87	0.85	0.86	0.84	0.85	0.84
50	0.90	0.91	0.85	0.86	0.83	0.84	0.85	0.83
60	0.92	0.91	0.85	0.86	0.86	0.87	0.88	0.88

scans is considered in the same window and this increases the accuracy at the expense of granularity. Instead, with shorter windows the accuracy declines drastically because there might be more windows with no data at all (between two consecutive scans for example) which might be misclassified.

When we compare results obtained with dyad-independent training and test sets for both uniform down-sampling (Table 3.7) and with the emulation of watches’ scan behaviour (Table 3.9), we again observe an overall decrease in performance, with limited difference in metric scores, as noticed for our wearable prototype in the previous section.

3.5 Discussion

The study has confirmed that BLE is an appropriate technology for the automatic detection of individuals' proximity in the workplace. Our device is able to reach a considerable F1 score (0.97) with a relatively small time window of 10 seconds when evaluated with 10-fold cross validation and 0.86 when evaluated on dyad-independent data with the same aggregation window. The lower performance of the model evaluated on dyad-independent data is expected because in this case the model is tested on data not seen at train time and therefore it is more difficult to classify. This model potentially generalises better to unseen data compared to the model trained with 10-fold cross validation where dyad-independence between train and test set is not guaranteed. However, this might not necessarily make the model robust to new environments given that systems which rely on radio signals typically need to be re-trained when the surrounding environment changes significantly [64, 62]. This is because radio signals are heavily affected by the environment conditions. For example, signals in one context (e.g., large open space) might have very different characteristics compared to another setting (e.g., narrow space with cement or metal walls).

While technologies like the Sociometric badges [47] or Opo [87], provide more directional information about face-to-face contacts compared to pure RF systems, they might be more sensitive to false negatives in cases where people are side-by-side or in large groups with more distance between the participants. Accurate social interaction monitoring is still a hard problem especially because there are different aspects that can be captured (e.g., proximity, distance, angle of contact, communication and content).

In the following sections we discuss the implications of our work for manufacturers of commercial wearables and Operating System designers, we discuss the challenges we faced during our deployment and we report the feedback we received from our participants regarding the device we deployed and the way the study was conducted.

3.5.1 Implications for commercial device manufacturers

Our work explores how current commercial wrist-worn devices would perform in such deployments. We highlight that the concurrent use of Broadcaster and Observer roles on these devices does not affect their usability, except for an increased battery consumption. Even when advertising and scanning periodically both watches allow the same functionalities they are meant for: receiving notifications from the phone, use vocal commands, etc.. Advertising and scanning are operations rarely used on a watch: the first one is used when it needs to be connected to a phone while the second one when the user wants to connect an accessory to the watch (e.g. headphones). In normal operation the watch is connected to the phone and does not perform any of the two. This is probably one of the reasons why the OSs give only limited options to control advertising and scan. In general, the OS vendors tend to be conservative in terms of energy consumption in order to provide a

satisfiable experience to the end-users. For this reason they limit the configurable options to the ones that least impact the battery consumption. This is the case for Tizen OS which permits the developer to set different Advertising Intervals but offers no options for the scan operation. Android Wear instead allows us to scan quite aggressively in *Low Latency* and *Balanced* modes. However, we showed in Section 3.3.1 that these settings result in excessive power consumption which would make an office deployment infeasible.

The Samsung Gear S2 could well support proximity-based applications because it allows proximity detection with an F1 score of around 0.90 with a 5-second window, when the data is down-sampled uniformly (Table 3.6). By contrast, Android Wear would require an increased transmit rate or a bigger scan window. In fact, the 1280ms Advertising Interval is too large to achieve a useful receive rate. Our results show that with an Advertising Interval of around 100/200ms an Android Wear watch would also be able to capture short-lived proximity events as Tizen does. However, the Android Wear watch we tested suffers from high power consumption which should also be addressed in order to make longitudinal studies with this platform feasible.

By comparing Tables 3.6 and 3.8 one notes that by having more uniform data it slightly improves the granularity and accuracy, even when using the relatively low receive rates achievable by the watches. In fact, when the raw data has been post-processed with uniform down-sampling (Table 3.6), the accuracy is higher even for smaller windows compared to when the data is post-processed to emulate the current operation of the watches (Table 3.8). This suggests that OS vendors could improve the proximity detection on wearables (although only to a limited extent) by allowing their devices to scan on a more regular basis. At the moment in fact, the few seconds of gap with no data, due to the scan being performed every 5 seconds, is detrimental to detection accuracy. The OS should allow more frequent scanning but for shorter time in order to obtain more uniform data while keeping the current rate and similar power consumption. We also note that giving more control to the developer on the Advertising Interval setting would not be counter-productive in terms of end-user experience: this parameter is the one that least affects the power consumption of the wearable device. Another factor to consider is the context where this technology has to be deployed. The office we used for our evaluation represents a very dynamic environment where people have many short interactions throughout the day. This means the system would require small processing windows and higher receive rates to capture all the proximity events. On the other hand, in an environment with a slower pace, like for example a research lab, the interaction events are likely to be longer and less frequent, therefore a higher receive rate would not give significant returns.

The Transmission Power instead is a parameter that is non-adjustable on both tested watches (at least Android Wear offers the API so we can speculate that it may be enabled in future releases). This could be a limiting factor for proximity-based applications. Too high power would require filtering on the RSS values to remove data corresponding to devices that are too far away or could create collision problems in crowded environments, while too low power could result in missing contacts. Again, since this parameter does not

affect the battery life dramatically, allowing the developer to adjust it would not impact the user experience excessively.

3.5.2 Deployment Challenges

In this section, we briefly discuss the three main challenges we tried to address while designing the wearable device and we provide considerations about large deployments. The challenges are related to the form factor of the device and the measures we took to guarantee its correct operation.

Wearability and comfort

One of the main challenges faced by this kind of deployments is participation. After an initial period of excitement, participants tend to stop wearing the device, especially if it is obtrusive. To address this issue we tried to make our device as comfortable as possible. Our objective was to keep it small and discreet in order to maximise participation. However, this limited the size of the battery we could use and therefore the maximum achievable battery life. We tuned the BLE parameters to obtain an expected battery life of around 20 hours. Although this decision required to recharge the device every day, it allowed us to gather fine grained proximity and location data, which was useful to study the performance of BLE for proximity detection and to analyse social dynamics in the office (Chapter 4). Additionally, to remind people to wear the device while in the office we also sent emails every two days to all participants.

Timekeeping

In order to timestamp all the data logged on the SD card, the device maintains an internal clock. The devices were programmed to maintain the current real time with a resolution of 250ms. The correct time was provided by two Android phones twice a day in order to compensate for drifts in the devices' internal clock. Additionally, every time a device resets, it advertises that it does not have the correct time (see next section for more details). On receiving these advertisements, the phones send a synchronization beacon to allow the device to timestamp the data correctly. The two Android phones were plugged into wall outlets and therefore could operate continuously. When post-processing the data after the deployment, the synchronization beacons received from the phones were used to compensate for the inevitable drifts and re-align the timestamps to the correct real times.

Device diagnostic

To make sure all the devices worked properly, we implemented two simple diagnostic features in our devices. In the BLE advertisement packets transmitted by the wearable devices, we included one bit to signal problems with the time keeping and one bit for problems regarding the SD card. These bits were checked by two Android phones (the same ones responsible to send time synchronization beacons) that scanned continuously to detect these anomalies.

The first bit is set to 1 when the device does not have the correct real time. This happens every time a device resets, in case for example of an internal error or when the battery is completely drained and then re-charged. In this case, the Android phones re-transmit a time synchronization beacon.

The second bit is used to inform that the device is not able to write data to the SD card. This can happen due to errors in the code or because the SD card is faulty or it has been pulled out of its socket. We discovered that for simple errors a reset of the device would solve the problem. Therefore, we implemented a way to remotely reset each device. The Android phone that detects the problem connects to the wearable device and resets it by writing a value into a Bluetooth GATT characteristic. If the same device reports a problem with the SD card more than 10 times, it is an indication of a major problem with the SD card or with the wearable device itself. When this happens, the Android phone reports it to the researchers by sending an e-mail in order for them to replace the device.

This diagnostic mechanism has been proven useful during the deployment where 3 devices have been reported to have major problems and have been replaced with minimal loss of data.

Large deployments considerations

We have shown how sensing could be used to gather spatio-temporal information in the workplace. This information can easily create the foundation for ubiquitous applications. However, the applications of this technology are definitely not limited only to workplaces, other settings, such as large events, could also benefit from it.

Large deployments pose other challenges and issues. First of all, to cover a large office space, possibly split over multiple buildings, more localisation static beacons have to be deployed and maintained. One possibility in this situation would be to use the existing WiFi and BLE infrastructure to locate participants and deploy dedicated beacons only in specific areas where more resolution is necessary or where WiFi/BLE coverage is not optimal. In terms of mobile devices, one potential issue regards the possibility of radio signals collisions that might prevent correct detection when many devices are in range. However, this is likely to be a real problem only during large events where many employees would attend and can be avoided by correctly tuning the transmission frequency to reduce

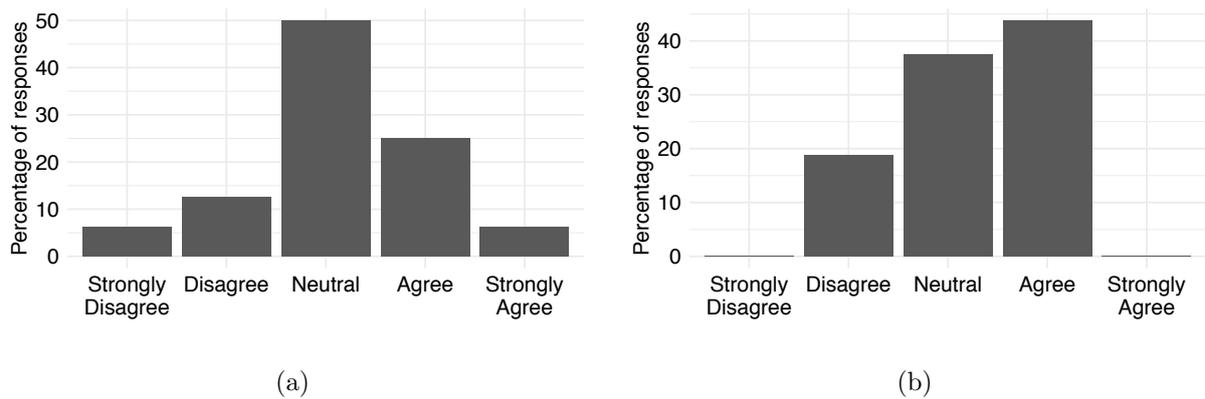


Figure 3.17: Agreement with the statements “The device was comfortable to wear all day” (a) and “The device was easy to use (recharge, put it on, ...)” (b).

collision probability. More realistic is the challenge of adoption. In fact, the burden of carrying an additional device might be too big for employees and this could result in limited adoption inside the company. In case a dedicated device has to be used, conceivably because special sensors have to be employed, it is important to pay particular attention to its comfort and ease of use, especially if the study is planned to run for a long period. In particular, the battery life is one of the most important aspects. The device should be built and tuned to allow the data collection at a reasonable rate without disturbing the user with the need for frequent charges. In alternative, it would be possible to use devices that people already carry with them (e.g. smartphones and smartwatches) even if this might result in limited data resolution and accuracy. Additionally, if smartphones are used another aspect to consider is the fact that people do not always carry their phone when indoor, therefore alternative strategies have to be devised (e.g. use the smartwatch when the phone is not with the person). Moreover, if long deployments are planned it is crucial to engage the participants with the study and data collection by, for example, providing statistics about the study or make the devices useful beyond data collection (e.g. the device could double as access card). This should motivate people to use the device resulting in higher quality data.

3.5.3 Participant feedback

To examine how our devices were perceived by the participants and how their suggestions might be used to improve the current version we asked our participants to complete an online survey (available in Appendix C) with closed and open-ended questions. We received 16 responses. We also interviewed seven participants who were asked to comment freely on their experience with the device and the deployment. In total we received feedback from 20 different participants.

Duration of the Deployment. The majority of the responses (68.8%) indicated that the

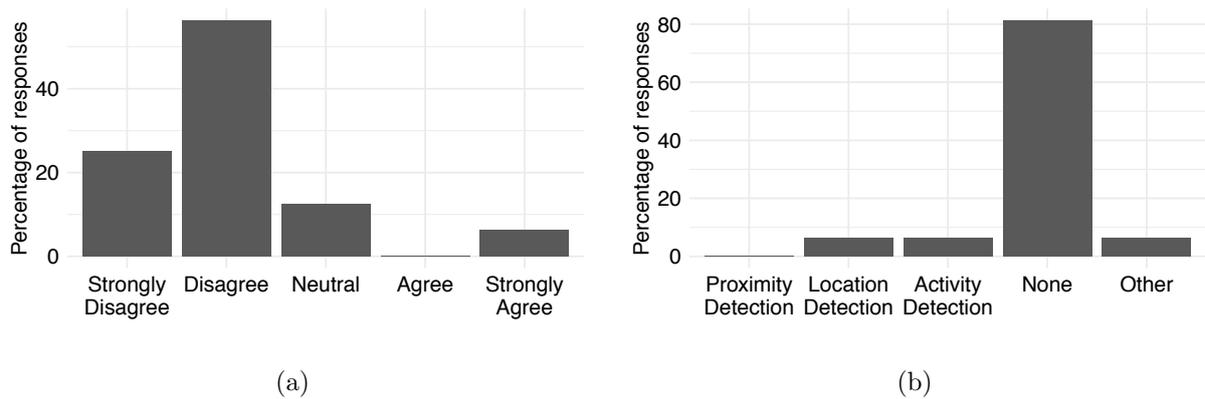


Figure 3.18: Agreement with the statement “I am concerned that the device can threaten my privacy” (a) and responses to the question “Which aspect of the device makes you more concerned?” (b).

device was in use for half of the intended period (four weeks) or more. However, in 62.5% of the responses and during most of the interviews it was declared that the deployment duration was too long. Most of the participants felt that a period of one or two weeks would be more suitable and some of them asked for some sort of incentive to remember to wear the device (e.g. gamification).

Wearable Device Comfort. The wearable device was not perceived as completely comfortable. In fact, when we asked the participants to rate their agreement with the statement “The device was comfortable to wear all day” from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) the average of reported responses was 3.13 ($\sigma = 0.96$). Figure 3.17(a) reports detailed data for each level. In terms of ease of use the participants showed a little higher scores ($\mu = 3.25$ and $\sigma = 0.77$) when asked to rate their agreement with the statement “The device was easy to use (recharge, put it on, ...)” (Figure 3.17(b)). The most common complaint regarded the plastic box that contained the electronic components. In fact, it was detaching from the velcro band quite often and this caused discomfort for the participants. Two devices were lost due to this issue. Additionally the devices were not equipped with a status LED so the participants were not sure if the device was working or charging correctly. Some of the participants ($n = 6$) reported that the velcro band was not comfortable and thought it should be softer (e.g. rubber band).

In general, the participants were not bothered by the fact that the device needed to be recharged every day but some of them ($n = 3$) asked for the possibility of using the device without the need to re-charge it for one or two weeks.

Privacy Concerns. From a privacy point of view, our participants did not appear to be concerned with the data collected by our device. We asked them to rate their level of agreement with the statement “I am concerned that the device can threaten my privacy” in a scale from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*) and the average of the reported

responses was 2.06 ($\sigma = 1$). Detailed data is reported in Figure 3.18(a). When we asked which one of the three kinds of data collected makes them more concerned (i.e. proximity, location or activity), only two responses reported concern, one with the activity detection and the other one with the location detection. The other responses reported no concern (Figure 3.18(b)).

Although this study did not raise any particular concern in the participants, it is known from past research that privacy concerns in the workplace are also related to the working environment [18]. This suggests the integration in our device of privacy protection techniques, such as the possibility for the participants to stop the data collection at any time.

3.5.4 Limitations

Limited Sample of Commercial Devices. In our study of BLE capabilities on commercial devices we examined only two operating systems and two particular devices. However, we do not expect substantial differences across devices when the same operating system is used because it is likely that the operating system would uniform hardware variations and expose the same functionalities to the developer. Smartphones instead might allow for higher rates of the transmission and scan of BLE packets given their larger energy budget due to bigger batteries. The two operating systems we evaluated, Android Wear and Tizen Wearable, represent the two largest systems, by market share, after watchOS from Apple [14]. The evaluation of Apple hardware and software would complement our analysis. However we are aware of severe limitations in the handling of beacon transmission and scan while the application is in background which would likely make this kind of study impossible. Other versions of the operating systems we examined might offer different functionalities in the future following advancements in hardware and software.

Technological Accuracy. We acknowledge that the particular technology we used for this work (BLE wearable devices) is not capable of providing an exact detection of social contacts. Our devices include a larger range of contacts, compared for example to participant observations, because they record every time people are in close physical proximity even if they are not engaged in a conversation. Our findings (including the ones in the next chapter) suggest that although it does not capture the exact type of interaction (i.e., it is not possible to know if two people were actually talking or not), it provides usable information to aid the study of behavioural patterns and dynamics in the workplace.

3.6 Conclusion

This chapter highlighted the feasibility of workplace interaction studies using commercial BLE wrist-worn devices. It offered a detailed analysis of BLE parameters that play a

central role in proximity detection through a prototype platform on which BLE could be used without constraints. This allowed us to conduct the first analysis of BLE capabilities and limitations on commercial wearable devices (Android Wear and Tizen). We showed how the parameters are interdependent, how their combination could be used to collect data at the desired rate and their corresponding power utilisation.

We conducted an extensive experimental validation with lab experiments and a longitudinal user study with 25 participants in an office environment. We confirmed the BLE suitability for accurate proximity monitoring with 0.97 F1 score and 0.98 ROC AUC at 10 seconds granularity. Ground truth observation for around 19 hours was performed to support our evaluation. We hope this study can offer guidance to developers and hardware producers regarding their APIs and specifications.

While here we focused on data gathering options and their issues, we have yet to analyse what useful knowledge could be extracted from the data and how this could support modern companies. This is the topic of the next chapter.

Chapter 4

Detecting emerging Activity-Based Working traits through wearable technology

4.1 Introduction

In the previous chapter we have analysed the potential of commercial wearables to collect data about social dynamics in an office environment. We focused on the data gathering without considering which insights could be generated from its analysis. In this chapter we explore how we can relate social contact with the design of office spaces through the data analysis of our deployment. The common facet of studies looking at how space layout affects human interaction in the workplace is that they considered traditional offices where employees have assigned desks and static routines. However, recently, several design principles have been emerging with the objective of realizing dynamic and agile working environments that can better support knowledge workers.

The Activity-Based Working (ABW) concept is one of these principles. It aims at designing the office architecture based on the activities the employees have to perform daily [13]. At the foundation of ABW there is the freedom for employees to choose where and when they work. This translates into absence of allocated desks with the assumption that employees will move within the office by choosing the best functional work setting for the tasks to complete and that best matches their preferences, thus improving productivity. As a side effect, ABW generally reduces costs due to a lower requirement on total floor space [53]. It is also likely to increase communication between groups and foster knowledge sharing and collaboration given the mobility that derives from having unallocated desks. Even if ABW is not a new concept, its adoption has recently been increasing. However its benefits are not yet well understood [13, 213]. Some works have analysed ABW offices using traditional ethnographic methods of participant observations and surveys to study

productivity, health and satisfaction [140, 13, 121] or have provided a theoretical model of the benefits and risks of ABW [213]. We provide an overview of Activity-Based working in Section 2.3.1.

Observing agile working through traditional methods (such as surveys, or participant observations) requires considerable effort, given the high mobility and dynamics of the setting. This is evidenced by the limited number of existing research studies on ABW offices. A typical approach is, for instance, to investigate patterns of occupancy and space usage [179], which are then averaged across teams. With individuals enjoying free choice of where to work for any given point in time, occupancy patterns of teams would be almost impossible to track. Another relevant research insight - the distance dependency of frequent interaction [12, 175, 176] would be impossible to repeat in an ABW environment with traditional research methods, since fixed desk locations in traditional offices are typically used to calculate distances between co-workers. In contrast, non-assigned desks lead to constantly changing patterns of proximity and co-presence, since members of staff sit next to different people all the time.

In this chapter we focus on and analyse two core aspects of ABW, flexible use of office space and collaboration opportunities, relying on data automatically collected through wearable devices as described in Chapter 3. The work, which exploits the advantages of technology in automating the collection of fine grained temporal data of a number of individuals, shows how our methodology is able to detect behavioural traits and relate them to ABW core principles. We exploited the solution based on Bluetooth Low Energy (BLE) and 3-axis accelerometer described in Chapter 3, which was deployed in a company office. We captured a data set of close proximity contacts, location traces and physical activity of 25 employees for a period of 4 weeks. The office had been intentionally designed with ABW principles in mind and offers the employees several opportunities for adopting flexible working practices. This study allowed us to investigate social ties in relationship to the hierarchy and roles of employees and their use of office space in a specific kind of work environment that has not been thoroughly studied before.

The specific contributions that this chapter offers are as follows:

- We show how, in this company, interactions easily cross team boundaries, in line with ABW principles. We also find that a good amount of these inter-team contacts happen in the kitchen and in circulation areas suggesting a more serendipitous nature than the ones happening at the workstations. However, the mix among different layers of the organizational hierarchy is not as strong, with more contacts among people in the lower levels. It could be hypothesised that ABW concepts might facilitate lateral communication (compared with reports on non ABW studies).
- We show how the absence of assigned desks, at the core of ABW and flexible office spaces, relates to a good usage of different spaces in the building. This however emerges when we consider a larger temporal scale of an average working week while it is less prevalent within a day.

- We further find relevant spatial and locational effects in the data. Spatial preferences arise from strong contact ties: pairs of individuals connected with a strong tie (defined as a higher than average number of contacts taking place in locations away from desks) are more likely to choose to sit at the same workbench. This points towards a possible mismatch between ABW principles and the actual use of space where people seem to choose working spaces based on the presence of other colleagues at the same location rather than exclusively based on the task they have to complete. However, this might also be related to the nature of work done as team affiliation played a strong role in this.
- We discover differences in the temporal contact patterns between the company we studied and other more traditional offices by comparing our data set with two other ones. We discover how our participants have on average shorter contacts and how the network has the potential to better enable quicker communication of information (both in times and hops) compared with traditional offices. A potential effect of the ABW principles adopted.

In more general terms, we show that our methodology allows to study the effects of ABW principles applied in offices. This work goes one step further than previous works on ABW [91] by analysing not only occupancy data but also contacts between participants and by showing the potential of technology in gathering contact data in a challenging environment where a traditional observation modality would struggle. The comparative analysis with more traditional offices also contributes to the understanding of the behavioural characteristics emerging from the application of ABW principles.

Chapter Outline. Section 4.2 describes the dataset used in our analysis and participants' self assessment of their mobility inside the office. In Section 4.3 we detail our analysis of the ABW principles and Section 4.4 discusses our findings and their implications and Section 4.5 concludes the chapter.

4.2 Dataset description

The dataset we analyse in this chapter is the same one described in Chapter 3, Section 3.4. However, for this analysis we decided to aggregate the data with a relatively long time window (1 minute) with the aim of focusing on potentially more meaningful contacts without losing the quick dynamics that might characterise the company. We applied the same techniques described in Section 3.4.4 and 3.4.5 to extract location traces and proximity contacts with a 1-minute aggregation window. For the following analysis we used the model evaluated with 10-fold cross validation which achieved an F1 score of 0.98 as reported in Chapter 3, Table 3.6.

In line with what was reported by the participants, who felt the study was too long (Section 3.5), we noticed that the devices have been mostly used during the first two weeks of

Table 4.1: Mobility profiles we asked our participants to identify with. Source: Leesman’s study 2016 [121].

Profile Name	Mobility Profile
Profile 1	I perform most/all of my activities at a single work setting and rarely use other locations within the office.
Profile 2	I perform the majority of my activities at a single work setting but also use other locations within the office.
Profile 3	I perform some of my activities at a single work setting but often use other locations within the office.
Profile 4	I use multiple work settings and rarely base myself at a single location within the office.

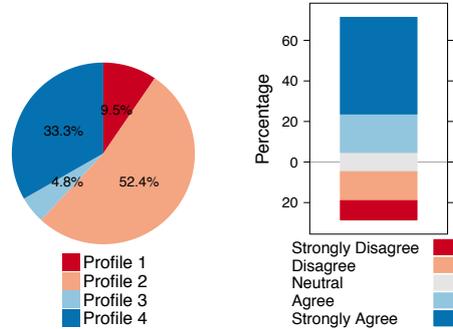


Figure 4.1: Percentage of participants identified with each mobility profile (left) and agreement with the statement “I believe I move more at Spacelab than in other companies I worked” (right).

the study. Therefore we decided to consider only those two weeks for our analysis. From the collected raw data for the first two weeks we extracted 2190 proximity contacts with a temporal resolution of one minute.

As mentioned earlier, the company adopts ABW principles and has a very dynamic and flexible working style. To understand how the participants perceived the mobility level of the workplace we administered a survey to the company employees where we asked them to identify their mobility profile with one of the four used by the Leesman’s study [121] of workplaces that adopt Activity-Based Working (reported in Table 4.1). We received 21 responses and we found that 38% of the participants identify themselves with the two profiles that describe more mobility (Profiles 3 and 4) and only 10% identified with Profile 1. Instead, in Leesman’s study only 27% of the employees identified themselves with Profile 3 and 4 and 32% in Profile 1. This indicates that the working style in the company might be more dynamic compared to other more traditional offices. We also asked our participants to rate their agreement with the statement “I believe I move more at Spacelab than in other companies I’ve worked at” on a likert scale from 1 to 5, and more than 60% of

the participants responded with “Agree” or “Strongly Agree”, showing that this company might present different dynamics than other offices. Figure 4.1 reports the data collected in both parts of the survey.

4.3 Analysis of ABW principles impact

4.3.1 Organizational structure and interactions

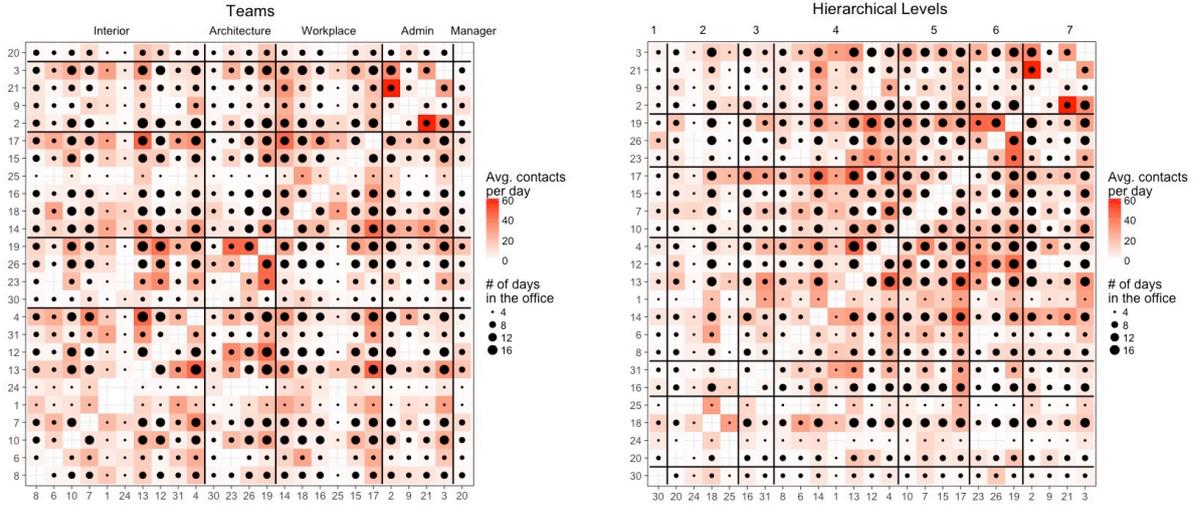
In this section we study how the organizational structure of the company relates to the contact patterns of the employees to identify the ABW impact. In particular, we examine two aspects of the organization: the horizontal structure where people are arranged into teams and the hierarchical vertical structure of who reports to whom. Previous work has highlighted the importance of interactions between members of different teams as a source of new ideas and a way to increase productivity [11, 110, 36, 161] and ABW principles are certainly based around these aims.

Beginning with the analysis of the horizontal structure, Figure 4.2(a) shows a netgraph reporting the normalized number of contacts for each pair of participants aggregated over the entire duration of the study. Contacts have been normalized by the number of days both participants were in the office at the same time (overlaid circles) in order to account for the fact that some people were in the office more than others. The ordering of the participants on the axis is such that adjacent participants belong to the same team. With this ordering, the contacts along the diagonal from bottom-left to top-right represent intra-team interactions. From Figure 4.2(a) we observe that different teams have a good amount of contacts with one another even if some of the strongest contacts are between members of the same team.

Netgraphs are a visualisation technique introduced by Varghese and Allen to represent communication networks [72]. They are similar to adjacency matrices and are typically used as an analysis tool in organisational and architectural studies [175, 178, 12]. Brown et al. also used them to analyse proximity contact patterns measured with wearable tags [33]. The power and versatility of netgraphs consists in rearranging the individuals along the two axis based on interesting variables (e.g., role in the organization, project affiliation or demographic information). This helps in identifying structures and patterns in the contact network through an easily interpretable visualisation¹.

To better quantify the relation between inter and intra-team contacts we computed the number of these contacts for each team normalizing the results by the sizes of the groups involved to account for the number of possible pairs. Figure 4.3 shows the results. The intra-team contacts are obviously higher, as expected when people work together, but we

¹Thanks to their versatility we use again netgraphs in Section 4.3.3 to study how contact patters relates to the work location chosen by the participants and their team affiliation.



(a) Participants sorted by the team they belong to. (b) Participants sorted by their level in the organizational structure.

Figure 4.2: Normalized number of contacts for each pair of participants for the entire study. Each cell represents the average number of contacts per day. The horizontal and vertical lines separate the participants in the different groups. The size of the circles overlaid represents the number of days that both participants in each pair were in the office at the same time.

observe that the number of inter-team contacts is similar for all groups. We verified the similarity between the inter-team contacts for the four groups with the TOST Equivalence procedure using the Wilcoxon Rank-Sum Test [188, 212]. We found that the similarity is significant (p -value < 0.05) within $[-18, 18]$ equivalence bounds for all pairs of groups except for Architecture and Workplace which is significant within $[-25, 25]$ and Architecture and Admin within $[-22, 22]$ equivalence bounds. This workplace shows a large number of opportunities for interaction across teams (proximity contacts) which could be a result of the implementation of ABW principles. Figure 4.2(a) appears almost randomly distributed, if compared to examples of traditional open-plan workspaces [175], where team clustering is much more prevalent.

To gain an insight about the nature of the inter-team contacts we looked at the total number of contacts happening at each location. We discovered that while most of the inter-team contacts happen at Workstation #1 (38% of the contacts), the second and fourth locations for number of inter-team contacts are Circulation and Kitchen, respectively, with 28% and 8% of the contacts. The open space workstations are the main locations where work is done in the company hence it is expected to see a high number of inter-team contacts. However, the contacts in Circulation and Kitchen might represent more spontaneous ones. Both are in fact highly integrated into the spatial system of the office and research shows that integrated spaces attract more activities [159]. The kitchen also acts as an attractor

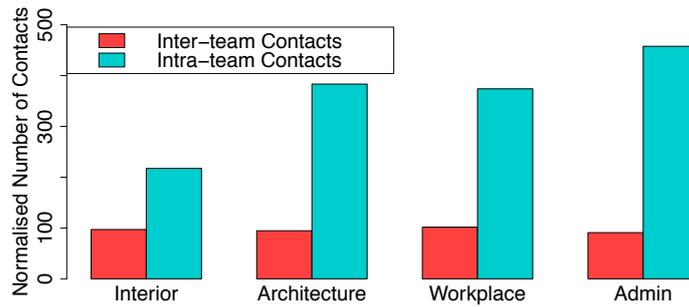


Figure 4.3: Normalized number of inter and intra-team contacts for each of the larger teams.

in line with previous findings, highlighting the role of social spaces in fostering inter-team contacts [33].

The second aspect of the organizational structure that we analysed is the vertical division into hierarchical levels. Figure 4.2(b) shows the aggregated number of normalized contacts with the participants ordered by their level in the hierarchy. The main pattern that emerges is that there are fewer contacts among the upper levels (i.e. 1, 2 and 3) than among the lower levels. In fact, the plot shows darker and denser regions from level 4 to 7 going towards the upper right corner. We also noticed that there are several pairs formed by a person from an high rank role (levels 1, 2 or 3) and a person from a low rank role where, despite both being in the office for several days, the average number of contacts is low. In contrast, there are pairs with both people from low rank roles that were together in the office for few days but had more contacts, on average. This shows that even when high rank people are in the office, they have less contacts with others and this might be related to the kind of work they have to do. To test the significance of the patterns we observed, we first looked at the pairwise intra-level contacts within low (4, 5, 6 and 7) and high levels (1, 2 and 3) and found a significant difference in the two distributions (Kruskal-Wallis rank sum test, p -value < 0.05). We also found a significant difference (p -value < 0.01) in the distributions of the pairwise inter-level contacts within low (4, 5, 6 and 7) and high levels (1, 2 and 3). This supports the conclusion that participants in the lower levels have more contacts than the ones in the upper levels.

These results point to a strongly networked type of organization, where the way work gets done does not resemble the formal organizational hierarchy [109]. Instead, strong lateral links emerge among the lower ranks of the hierarchy across reporting lines and team affiliation. Burns and Stalker [35] have argued that this type of organization provides a suitable structure in dynamic organizational environments. It could be hypothesized that the ABW principles applied here do not hinder lateral communication and they might even facilitate it instead (given comparisons with reports of communication in non ABW studies [175]).

4.3.2 Demise of allocated desks

One of the central principles in ABW is the absence of allocated workstations. The assumption is that employees will change from one work location to another in order to best match the needs of the current task and personal preferences. In this section we attempt to understand to what extent this principle is implemented in the company we studied and if our participants adapted to this working style.

Jayarajah et al. [97] used Cumulative Distribution Functions (CDFs) to compare behavioural traits of individuals vs. groups using location traces from phones. The authors considered groups of different sizes (small between 2 and 3 members, medium between 3 and 7 and large more than 7) and analysed mobility patterns, responsiveness to calls/SMSs and application usage looking at the CDFs distributions and using the Kolmogorov-Smirnov test to determine the significance of differences observed between individual and group behaviour. Inspired by this work we decided to analyse the mobility patterns of our participants using the same techniques. The analysis of CDFs represents a proper tool to study different data and visually compare their distributions with one another and then verify their difference with a significance test.

Figure 4.4 shows the Empirical Cumulative Distribution Functions of the number of distinct locations visited by each participant averaged per day, per week and the total for the entire duration of the study. We selected three thresholds on the dwell times to understand if there is a difference in the number of locations visited based on time spent at each location. This allowed us also to filter out very short dwell times that are due to people walking inside the office. Two-sided Kolmogorov-Smirnov tests ($\alpha = 0.05$) performed on each pair of distributions, for each aggregation period (day, week and study), show that the differences among the distributions are significant (test statistic D ranging from 0.4 to 1 and p -value ranging from $2.778e-11$ to 0.03663)².

Looking at the distributions for dwell times larger than or equal to ten minutes we see that participants visit almost half of the monitored locations (13 locations excluding “Outside office”) during an average day and almost all locations if we consider the entire study, suggesting a great level of mobility in the office. However, when we consider longer dwell times, participants explore significantly fewer locations. In particular, Figure 4.4(a) shows that on an average day the employees visit slightly more than one location for one hour or more, meaning that people rarely use more than one location per day for long tasks. Looking closer at the data for dwell times equal to or longer than 1 hour, we observe that the maximum number of work locations used in any day of the study is 2 and 52% of the participants worked at least once in 2 different locations in any day of the study. However, when we aggregate the data per week, these figures rise to a maximum of 4 distinct locations in any week of the study and 80% of the participants worked at least once in 2 or more distinct locations in any week. We point out that only two employees,

²We performed the Kolmogorov-Smirnov test because it does not assume a specific underlying distribution.

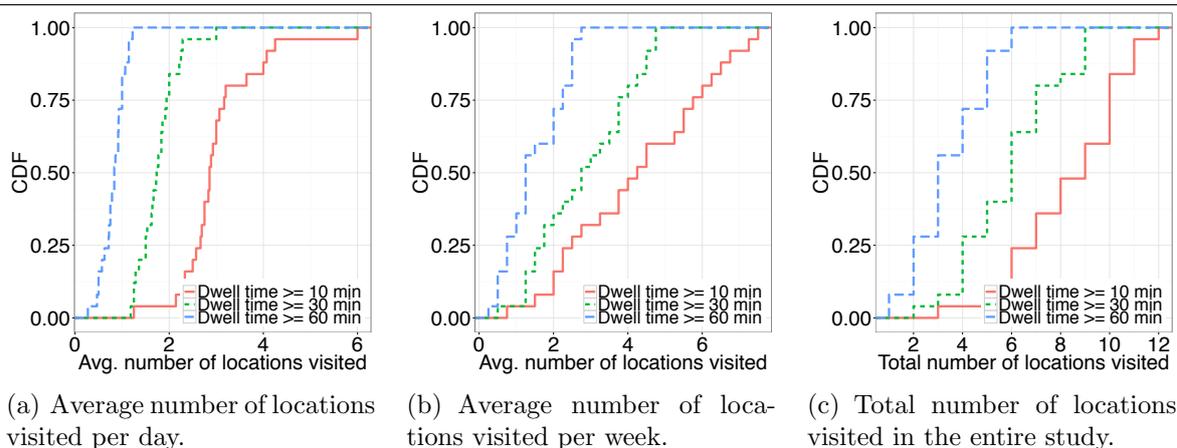


Figure 4.4: Empirical Cumulative Distribution Functions of the number of locations visited by each participant averaged per day, averaged per week and the total number for the entire study for different dwell time thresholds. The “Outside office” location is excluded because not every employee is required to carry out work tasks outside.

belonging to the Administration team, have assigned desks.

To understand which are the locations used to carry out longer tasks we computed the average dwell time per location (Figure 4.5(a)). We observed that the locations with the longest average dwell times are also the biggest workstations on the two floors (workstations 1 and 13) and two of three non-private workstations that have computers for the employees. We further observed that private workstations 6 and 16 also present long dwell times probably for individual and focused work. By contrast, meeting room number 10 seem less used than the others probably because it does not have a door to close the room and therefore it would be difficult to have a meeting in isolation from the rest of the office. The kitchen also seemed to be used primarily for short periods of time. Figure 4.5(b) shows the distributions of certain locations that are representative of the room types we considered. It is possible to observe that open space and private workstations have longer tails and are used for longer periods while the kitchen and the meeting room #10 have different distributions and host people for shorter periods of time.

These results suggest that the employees might not have completely adopted the ABW principles. In fact, switching settings within a day for work related tasks (≥ 60 minutes) is not very prevalent. Similar results have been reported by Appel-Meulenbroek et al. where 68% of the employees surveyed never switched during an average day and only 14% switched once [13]. However, if we consider a weekly time scale we can see that choosing different settings is more likely. This behaviour is also highlighted by the employees’ mobility self-assessment (see Section 4.2) where they reported greater levels of mobility. So, while at a very fine-grained temporal scale ABW traits are not observed, they are indeed observed at a coarser temporal granularity, retaining the advantages related to serendipitous encounters and potential idea exchanges which are usually associated with

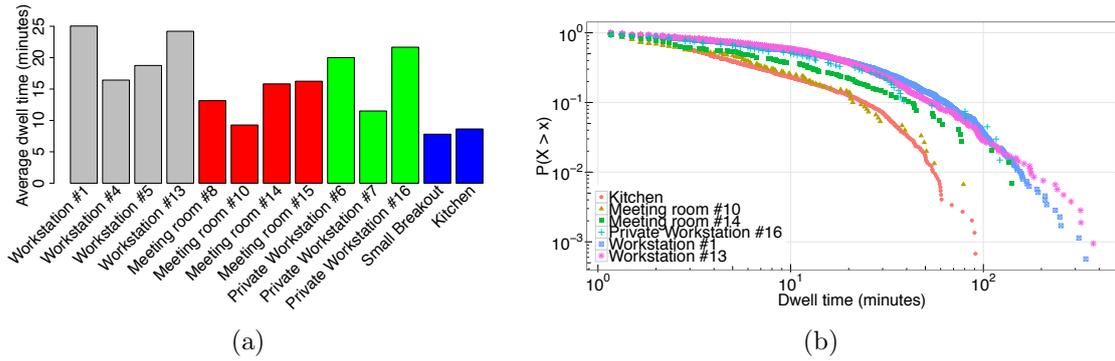


Figure 4.5: Analysing dwell times for different locations inside the office. (a) shows the average dwell times for each location. The “Outside office” location is excluded because not every employee is required to carry out work tasks outside. (b) shows the Complementary Cumulative Distribution Function (CCDF) of dwell times of certain locations representative of all the room categories considered (log-log plot).

this [161, 36].

4.3.3 Social ties and agile working

Given the previous results where we showed that people tend to use on average one work location for long tasks we try now to understand if this could be due to the fact that people work in teams. Towards this objective we first looked at the contacts for each day of the study. Figure 4.6 reports the netgraphs for two representative days as an example. The patterns for the other days are highly correlated with the ones we show here. Each square represents the number of contacts between a pair of people and a darker colour means more contacts. On the left the participants on the two axes are sorted by the locations they chose for the day while on the right by the team they belong to. Given that our devices capture only proximity, it is obvious that working at the same location increases contact intensity, therefore we intentionally removed all the contacts that happened at the main desks (1, 4, 5 and 13) and considered only contacts detected somewhere else in the office.

The plots show that there are more contacts among people sitting at the same workstation rather than people in the same team. In fact, the diagonal (from bottom-left to top-right), which represents pairs sitting at the same location or in the same team, has denser and darker colours in the plots on the left. By contrast, the plots on the right show a sparser and less defined pattern indicating that there are several contacts across different teams. A Kruskal-Wallis rank sum test [112] of the number of intra-team contacts and intra-location contacts (between people who chose the same desk for the day) for each day of the study shows that the differences among the distributions are significant (p-value < 2.2e-16)³. As

³We performed a Kruskal-Wallis test because the data is not normally distributed, therefore we had to

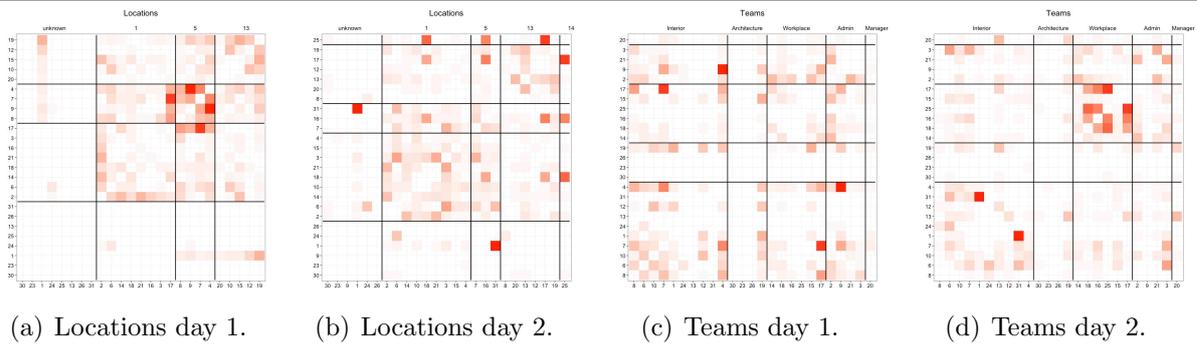


Figure 4.6: Number of contacts for each pair of participants for two days of the study. The contacts that happened at the main workstations have been omitted. In (a) and (b) the participants are sorted by the location they chose for the day while in (c) and (d) they are sorted by the team they belong to. The darker the colour the more contacts between the people. The solid horizontal and vertical lines separate the participants in the different groups. The location “unknown” represents participants that were not in the office that day or for whom we do not have location information (device malfunction).

reported in Figure 4.6 the contacts detected at the main workstations (1, 4, 5 and 13) have been removed.

To further study this difference we divided all the possible pairs of participants in two groups:

- **Strong Ties:** pairs that have a number of contacts greater than or equal to the average of the contacts happened away from the main workstations (1, 4, 5 and 13);
- **Weak Ties:** pairs that have a number of contacts less than the average of the contacts happened away from the main workstations (1, 4, 5 and 13).

For all pairs of participants we also determined the total number of days in which they chose the same desk. We found that the pairs that have strong ties, on average, choose to stay at the same desk more than the pairs that have weak ties. Strong ties, in fact, stay on average 2.682 days (median = 2) at the same desk while the weak ties only 0.819 days (median = 0). Figure 4.7 shows the Empirical Cumulative Distribution Function of the number of days the two groups of pairs spent at the same desks. Clearly, the pairs in the strong ties group prefer to stay at the same desk more than the people in the weak ties group. A Kruskal-Wallis rank sum test [112] shows that we can reject the null hypothesis that these two samples are drawn from populations with the same median values (p-value $< 2.2e-16$)⁴. An analysis of the teams of the participants revealed that 33.77% of the pairs in the strong ties group are made of people from the same team while for the weak ties only 14.09% of pairs. This suggests that some of the strong ties might be due to the fact

adopt a non-parametric test.

⁴We performed a Kruskal-Wallis test because the data is not normally distributed, therefore we had to adopt a non-parametric test.

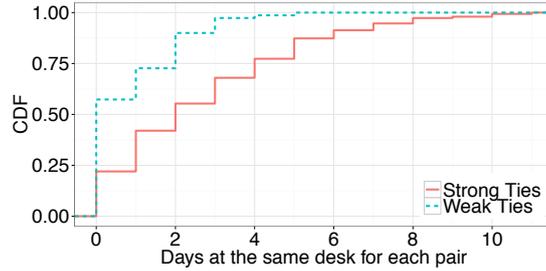


Figure 4.7: Empirical Cumulative Distribution Function of the number of days pairs of participants choose to stay at the same desk for the Strong and Weak Ties groups.

Table 4.2: Characteristics of the data sets.

Dataset	Participants	Study Period	Contacts
Spacelab	25	10 days	2190
Brown Old	39	10 days	683
Brown New	48	10 days	1065

that people work together in the same team and this might be one of the explanations why they choose the same desk more often. However, we cannot speak of causality. This result also highlights how intensely entangled spatial, behavioural and organizational phenomena are.

In summary, our results here suggest a mismatch between ABW principles and the actual use of space. In particular, the principle of allowing people free movement and choice of work location has not come to full fruition in the context studied. Pairs with high co-presence in locations away from desks (strong ties) also stick together at the same workstations much more than pairs with weaker ties. Team affiliation played a strong role in this. In short, people stick together with those they like or work with and choose their location not simply based on their task or the appropriate spatial setting for the day but with a social focus in mind. Relationships seemed to matter. Whether this process worked based on attraction (seeking like-minded people for one’s close proximity) or repulsion (avoiding those one does not like) would require further research. Either way, social clusters were formed based on preferences and thus resisted the randomizing effect ABW wishes to have in an ideal world.

4.3.4 Comparison with traditional offices

In this section we aim to better understand the impact of ABW principles on social dynamics inside the company by comparing our data set with two other data sets collected with the SocioPatterns badges [44] in two different offices [33]. SocioPatterns badges, like our BLE devices, use radio beacons to detect close encounters.

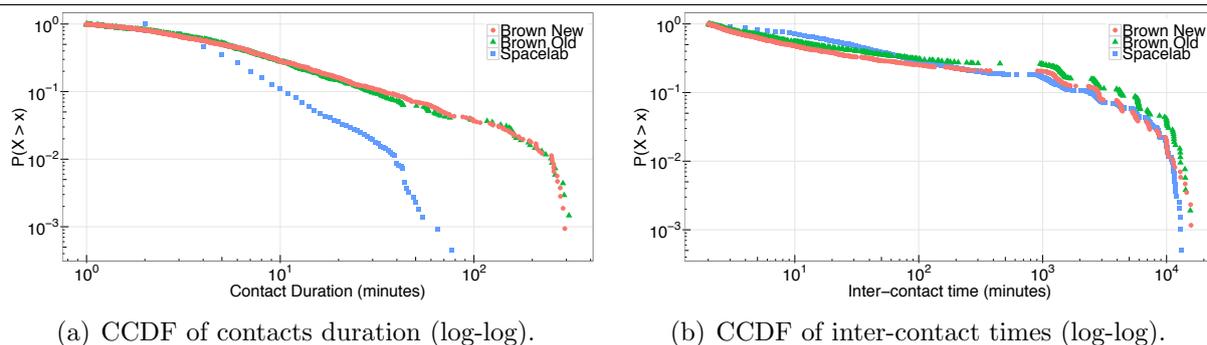


Figure 4.8: Complementary Cumulative Distribution Functions of contact durations and inter-contact times.

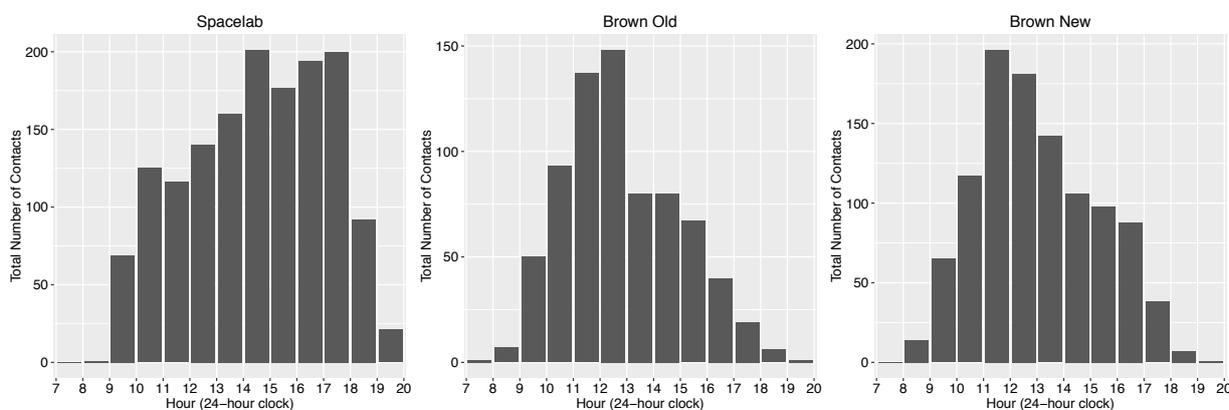


Figure 4.9: Total number of contacts aggregated in 60-minute windows by hour of the day.

The two data sets have been collected by Brown et al. in 2012 and 2013 in a research institution in UK that moved from one building to another during the study period (Table 4.2). The new building was designed for that specific research institution with the main aims of increasing the chances of serendipitous encounters among the employees and motivate an increased use of shared spaces. This was achieved by placing a central cafeteria on the ground floor and including larger lab spaces and more open areas. The old building represents an example of a traditional office with individual offices and few open or shared areas. The new building instead more closely reflects the ABW principles where diverse settings are offered to people for different kind of activities. However, in both buildings the employees had assigned desks and this might significantly impact the contact patterns we observe compared to our deployment at Spacelab.

We first analyse the contact durations and inter-contact times distributions shown in Figure 4.8. We observe that for the two buildings studied by Brown et al. the distributions have longer tails and longer contacts have been recorded compared to our deployment at Spacelab. This might suggest a successful implementation of ABW principles at Spacelab where employees tend to have more frequent but shorter interactions. However, this could

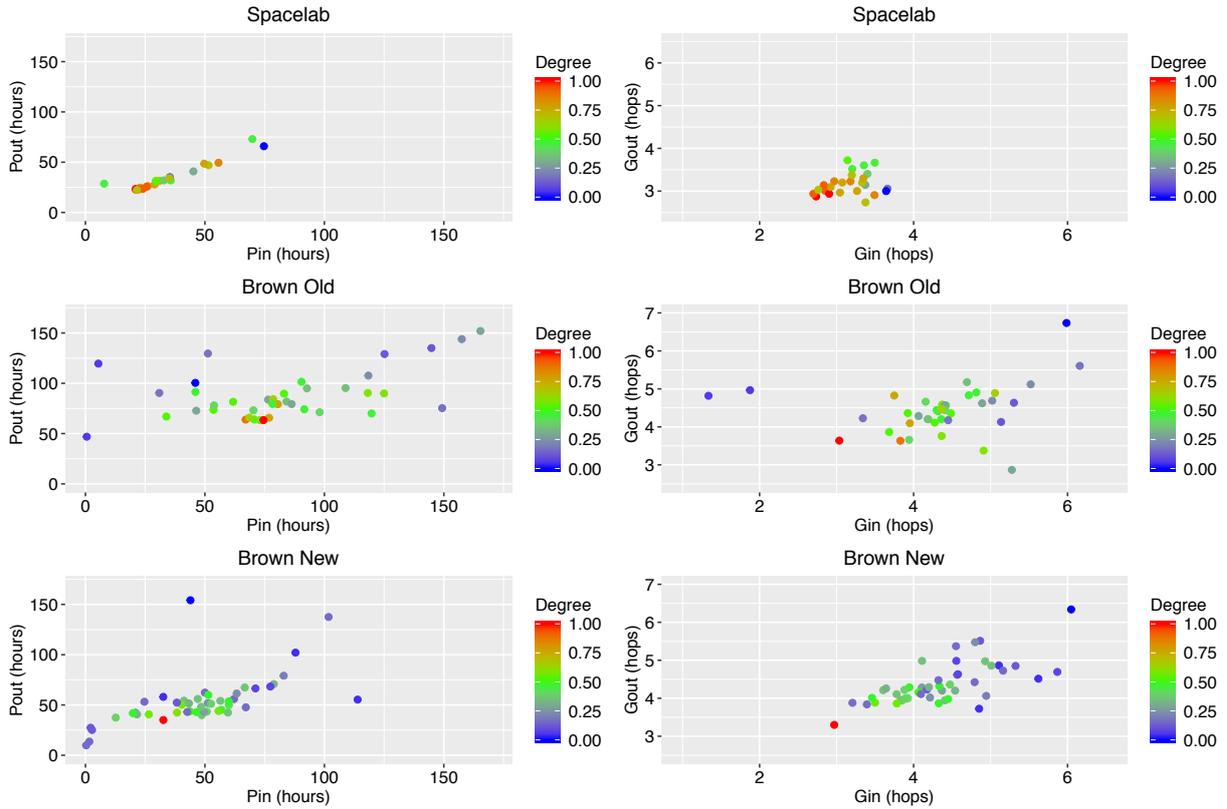


Figure 4.10: Correlation between P_{in} - P_{out} and G_{in} - G_{out} for the three data sets. Each point represents a participant and is coloured according to the vertex degree calculated on the aggregated networks. The vertex degree has been normalised in the range $[0, 1]$.

also be a reflection of the nature of work done in the different organizations. On the one hand, Spacelab is a young company with a very dynamic and flexible working style while, on the other hand, the research institution might have a more traditional working style. For what concerns the inter-contact times instead we observe similar distributions for the three data sets.

Taking advantage of the fact that the three data sets have temporal information, in Figure 4.9 we plotted the total number of contacts during different hours of the day for the entire duration of the study. While in the two data sets collected by Brown et al. there is a visible pattern where the number of contacts increases around midday, in the Spacelab data set we see less regularity in the number of contacts over time and more interactive afternoons than mornings. We also observed how there are, on average, more contacts at Spacelab than in the other two offices. We hypothesize that the differences we observe are due to the application of ABW principles at Spacelab and this results in a more even distribution of contacts across the day, as opposed to a more structured pattern in the other company where people primarily stay in their office, having most of the contacts during lunch time. However, there are many other factors that could contribute to this behaviour

and therefore we cannot speak of causality. Nevertheless, it is interesting to note that the office change observed in Brown et al. resulted in an increased number of contacts but it did not change the overall shape of the distribution of contacts during the day. We can only speculate that the additional contacts are an effect of the purposely built office and that a full adoption of ABW principles could possibly lead to a more substantial change in the daily patterns.

To further study the temporal characteristics of the data sets we borrowed the metrics *Average temporal proximity* ($P(X, Y)$) and *Average geodesic proximity* ($G(X, Y)$) from Kostakos [106]. The metrics measure the average time needed to go from vertex X to vertex Y and the average number of hops between X and Y respectively. The temporal proximity considers edge availability over time and takes into account possible wait times at one vertex before moving to the next one. The geodesic proximity instead counts only the number of hops from one vertex to another (without considering the time needed for the hop) but it is still subject to the temporal restrictions in the network. Kostakos also defines $P_{\text{in}}(X)$ and $P_{\text{out}}(X)$ as measures of “how quickly, on average, X is reached by the rest of the network” and “how quickly, on average, X reaches the rest of the network”. Similarly, $G_{\text{in}}(X)$ and $G_{\text{out}}(X)$ are defined as “the average number of hops needed to reach X from the rest of the network” and “the average number of hops needed to reach the rest of the network from X ”. Kostakos compared two datasets (corporate email communication and co-location in a public space) demonstrating that when temporal aspects are not considered the network structures described in the datasets present significant similarities. Instead, with the temporal metrics described above the author uncovered differences in the datasets which are peculiar of the two environments where the data has been recorded. This inspired us to adopt the same metrics to verify if similar dynamics could be found in our data and how they would compare across offices that have been designed following different design principles.

In Figure 4.10 we plotted the correlation between the *in* and *out* components of temporal proximity and geodesic proximity, colour coding each participant according to the number of its connections in the aggregated network. Above all we observe that people at Spacelab reach the rest of the network and are reached in less time compared to the other two offices, both in terms of time and number of hops. This could be attributed to the absence of assigned desks which brings employees into contact with a larger and more diverse set of people. Similarly, the presence of a central cafeteria in the new building studied by Brown et al. [33] might be responsible for the reduced time needed to reach the network and be reached that we observe when comparing the old and new building.

It is also interesting that the relation between P_{in} and P_{out} is more structured for Spacelab than for the other offices where there is more variability among the participants. The old and new office both present locally low-connected people that are quick at reaching the network or being reached by the network (blue points towards the bottom left corner) while for Spacelab the variation between people is much smaller.

These results give an indication, from a temporal perspective, of the possible effects

Activity-Based Working has in the workplace. Obviously, we cannot speak of causality since there are several other factors that play an important role, such as: the type and structure of the organization and the kind of work done, the culture inside the office and the personality of individuals.

4.4 Discussion

We have analysed the fundamental ABW principles that are common to implementations of the concept (flexible use of office space and collaboration opportunities). We have shown how technology can be used to collect data about human behaviour in a dynamic workplace and to allow reflection on how much ABW principles have been absorbed by the office settings. Gathering suitable data is usually very difficult in environments where behaviour tends to vary, change and evolve more than in traditional settings. This makes our technology-based solution even more key to these sort of validation studies. In the following we discuss our results with respect to implications for designers of ABW office spaces.

4.4.1 Theoretical and practical implications

From the collected data, we found that the different teams in the company present a considerable level of inter-group contacts which might be indicative of high collaboration. A different pattern instead can be observed for the vertical structure. In fact, higher levels of co-presence are visible among people at the lower levels of the hierarchy. These results show that ABW principles were realized to some degree: the aim of allowing interaction and collaboration to spread across team boundaries seems well achieved. The high level of inter-group contacts speaks of an equal, almost random spreading of contacts across the organization as a whole. Together with an open plan layout that connects both floors visually, the spatial layout in conjunction with agile working provides ample opportunities for co-presence. This allows communication to flow vertically along teams and reporting lines, and on the hierarchically lower levels (among people with more time available) provides social glue and creates a “networked” organization. Similar results have also been observed in other ABW offices where employees reported greater satisfaction for informal un-planned meetings, informal social interaction and collaboration on creative and focused work [121].

For what concerns mobility inside the office, we discovered that the ABW principle of not having allocated desks might not be well received by the employees, at least when a fine-grained temporal scale is considered. Our results show that desk selection seems to be constrained by strong team-related social clusters. On an average day, in fact, employees explore various locations for short periods but rarely change settings when longer dwell times are considered. However, more mobility is observed when we consider an average

week. Additionally, we discover that pairs of people with more contacts (strong ties) tend to choose nearby desks more often than people with less contacts (weak ties). This suggests that employees are driven by the presence of specific colleagues when choosing the work setting. Similar results have already been observed in other ABW environments but have never been related to proximity contacts as we do in this work. Leesman reported that, even within ABW offices, large numbers of employees fail to adopt Activity-Based behaviours and have rather limited mobility dynamics [121]. Also Appel-Meulenbroek et al. uncovered misuses concerning office areas, and found that most of the employees use up to two different types of space and never switch during an average day [13]. This represents a challenge for architects and office managers that have to deal with what seems like opposition to change by employees and habitually driven working styles. A possible solution could lie in the involvement of users in the design process with the objective to adapt ABW principles and implementations to users's needs and preferences. Additionally, training sessions might be useful to clarify the benefits of more mobile behaviours. This also raises interesting questions for future research. Efforts could be directed towards the understanding whether this behaviour is driven by attraction (seeking like-minded people for one's close proximity) or by repulsion (avoiding those one does not like) and if the choice of setting, and so who to sit next to, adds to satisfaction and increases happiness at work or not.

Our work also has practical implications for the future design of activity-based workplaces. The results presented here are indeed insightful, as they show the powerful effect of allowing employees completely free choice of where to sit. The company we studied had a good level of inter-team contact and a significantly higher number of co-presence events between participants in the lower ranks of the hierarchy. Previous research has shown that temporary co-location can increase the possibility of collaboration between scientists [29]. This could be used as a guideline in evidence-based design to encourage organizations who wish to become more collaborative and leverage knowledge-sharing across teams by introducing proximity to a wider range of people. Specifically, against the background that often organizations moving towards Activity-Based Working do not allow completely free choice of desks, but assign certain areas to certain teams, which could potentially limit the benefits of widespread contact patterns. However, our results have also shown that switching between work locations and desk selection seem to be driven by team-related and social preferences, so in effect there might be less randomization of contact than first appears. Clearly more research of different settings would be required to establish a clearer relationship between the amount of choice in agile working and the degree of dynamic contact and co-presence. Additionally, it would be interesting to explore the possibility of voluntarily disrupting peoples' habits (i.e., sitting always at the same location) to understand if it would be beneficial and in which way.

4.4.2 Drawbacks of ABW concepts

We have seen that the participants did not completely adapt to the ABW principles, especially when looking at the mobility and seating arrangement patterns. Our main objective was to uncover emerging office dynamics from a quantitative point of view, hence the study of why participants preferred certain behaviours is beyond the scope of our study and certainly relevant for future work. Here we discuss what other works have found in this area and relate them back to our results. De Been et al. studied employees satisfaction in 20 ABW environments with questionnaires and group interviews [55]. The open layout of the work environment has been acknowledged to stimulate more communication between different departments and to increase knowledge sharing. This is what we found at Spacelab where inter-group contacts seem to be quite substantial and where dynamics are different than traditional offices. Nevertheless, De Been et al. reported that the office layout also had negative effects, finding that people experienced lack of possibilities to concentrate, lack of privacy and unavailability of desired work points (i.e. waste of time finding a desk, attractive ones already occupied), with the latter also mentioned by Appel-Meulenbroek et al. [13]. However, we can argue that the lack of places for concentration and privacy could also apply to open plan offices with fixed seating allocation, so they do not necessarily relate to ABW exclusively.

In another study, De Been et al. compared different office types including flexible settings (ABW) [54]. They concluded that ABW environments do not support productivity, privacy or concentration as well as enclosed or open offices (with fixed desks). The authors provide one possible explanation for the difference between assigned and unassigned desks in terms of psychological identity. In ABW offices people are not assigned to specific workstations and they cannot personalise the spaces they use. This might result in people not feeling attached to their workplace anymore and therefore having a lower satisfaction. The study also mentions the difficulty of finding people as a potential drawback, which could explain why ABW offices are rated lower with regards to satisfaction with communication. In the company we studied, this is less of a problem since the office is small enough. However, usually, in larger settings ABW is often arranged with team zones and there is a dedicated area to go to, if one is looking for somebody, meaning that people can be found, if needed.

4.4.3 Limitations

Generalisation. Different companies, with distinct organizational structures and cultures might present contrasting interaction patterns which might lead to different conclusions to the ones presented here. The company we considered is an architecture firm and working style might be different in organizations operating in other sectors (e.g., commercial or scientific) or on different continents and with different cultures. More studies are needed to capture and understand dynamics that could be generalized more widely. The analysis we performed however is applicable to other organizations, facilitating the validation of

ABW principles despite the spatio-temporal challenges of the office dynamics.

Scalability. Bigger companies with more employees might show different interaction patterns: there might be inertia in interacting outside teams, unlike in our limited setting. Moreover, dynamics of team formation when people join or leave should be studied at a longer time scale. Effects of space and social ties on productivity also needs longitudinal efforts.

Technological Accuracy. Our devices are able to record proximity contacts and not actual communication. Therefore it is possible that our data overestimates the actual contacts because our measurements represent potential communication opportunities rather than real communication events. Similarly, for the location traces, some dwell times might be underestimated as it is possible that, even if a person is always sitting at the same location, the device could detect radio signals from other beacons and temporarily associate the person to that beacon. Nevertheless, for social psychology theories, physical proximity increases chances of interaction among people [74].

Comparison Validity. When comparing data collected in different buildings, organisation-specific variables such as structure of the organisation, its culture and peoples' personalities, might affect the validity of comparisons. However, by showing a comparative analysis with traditional offices we are able to gain insights into how ABW principles might affect social dynamics even if additional research is needed to better generalise the results. Likewise, the technology used to collect the data was slightly different; however this should not compromise the results as we have compared the data from the two technologies and found that they have overall similar properties [149].

4.5 Conclusion

We have shown how data from wearables can be used to offer insights about the adoption of ABW principles in companies trying to adopt them. We showed how some of the ABW principles, like the facilitation of inter-team contacts, seemed well received by the company we studied, while we uncovered a possible mismatch between the principles and actual use of space in other circumstances. Comparing our dataset with similar ones from traditional offices we highlighted how ABW principles could potentially be applied to enable quicker communication inside the office. In conclusion, our work offers a mechanism for space designers to reflect on the application of ABW principles and study its impact longitudinally.

In the two previous chapters we collected and analysed data on the occurrence of social contacts and their location inside a building. While this proved to be useful for the analysis of space utilisation, there is still a lot to learn from the analysis of the non-verbal channels involved in social interactions. We begin this study in the following chapter where we design and implement a novel wearable device capable of detecting non-verbal cues about

body orientation and interpersonal distance without the need for expensive infrastructure and calibration.

Chapter 5

Measuring interaction proxemics with wearable light tags

5.1 Introduction

In the previous chapter we have seen how the analysis of social contact occurrences could shed light on space usage and over a multitude of other processes such as team coordination and productivity [98, 36, 109].

Traditional sociology has placed high importance on observing the non-verbal aspects of social interactions such as interaction proxemics (e.g., interaction distance and relative body orientation). Non-verbal behaviour is the combination of speech-unrelated behaviour, such as facial expressions, hand and arm gestures, postures and body movements, and speech-related behaviour like speech rate, speaking time and tone of voice [104]. Non-verbal cues on interaction proxemics reveal user attitudes and emotions [16, 174] and are also crucial to understand epidemic infection rates [171, 214]. Observing these cues can facilitate many important applications. We list four specific examples:

- *team collaboration*: interaction details such as body distance and relative angles are important cues to study team collaboration (e.g., task timeline, individual roles) on creative tasks and assess a team's potential creativity [164];
- *job interviews*: non-verbal skills such as eye contact, energy level, and affect (expressed via hand gestures and body movements) can be the subject of training to improve the interview outcome [137, 50];
- *doctor-patient interactions*: patient satisfaction is affected by the physician's expressiveness that includes non-verbal behaviours like more forward leaning, nodding, gestures and gazing [134, 202];
- *marketing and sales*: the customer's engagement with the sales representative de-



Figure 5.1: Example of using Protractor to track team interaction when conducting a creative task (the Marshmallow challenge).

depends on his engagement abilities, which are therefore also important in sales training [125].

For all these examples, an accurate monitoring of body distance and relative orientation is crucial. The interaction distance between people has been estimated to be in approximately 7-cm intervals with a temporal granularity of 7 seconds in social interactions [172]. Angles of interactions are significant to study communicator’s attitude towards his interlocutor and should be estimated to the nearest 10° based on prior study [139].

To monitor interaction proxemics continuously, conventional approaches in behavioural sciences have relied on questionnaires, participant observations, or the use of non-human objects (e.g., life-sized photographs, miniature dolls or silhouettes) [104, 193]. Based on self-reporting, these approaches not only impose high burden on users and imply various biases, but also fail to provide behavioural information during a contact. Technology has progressed substantially in capturing fine-grained face-to-face interactions [48, 47, 87], however existing work still falls short: some either infer *only* user proximity [48, 47] or body distances [87, 223], or analyse speech-related non-verbal signals with no information on interaction distance and relative orientation [154, 181, 20]. Others focus on very specific contexts reproduced in the lab (e.g. job interviews, public speeches) and require cameras that bring privacy concerns and entail heavy environmental instrumentation, limiting the flexibility of the system [23, 22, 52].

The goal of this work, thereby, is to seek a more scalable and accurate approach to continuously measuring interaction proxemics as part of non-verbal behaviours during social interactions. To eliminate the need for infrastructure support, we consider a lightweight wearable tag resembling an access badge worn with a lanyard or clip (Figure 5.1). We leverage such tags to track *both the actual interaction distance and relative body orientation of users involved in a social interaction*. Specifically, each tag emits wireless beacons encoded with its tag ID and listens to beacons from other nearby tags. Based on the received beacons, the tag then identifies other tags/users within its sensing range, and estimates the relative angle and distance to each of these tags/users. These angle and distance numbers

are used to identify participants and recorded as their interaction proxemics during an interaction. At first sight, the problem appears to be a standard problem of relative device positioning. However, the context of tracking interaction proxemics presents three new challenges. We next review each challenge and our solution.

First, accurately identifying the participants in an interaction is challenging. Two users in close distance may not be in an interaction, as they may stand with other people in between them or are not facing each other (see examples in Figure 5.2). Thus, it is key to recognize both the line-of-sight proximity and the user's *relative* body orientation. To this end, as we have discussed in previous chapters, methods relying on radio frequency (RF) signals (e.g., Bluetooth, Wi-Fi) [33, 119] or microphones [120, 196, 223] are all prone to false positives, since RF signals and sound penetrate human bodies. Also, relative body orientation cannot be simply obtained by compass sensors, which measure only the absolute orientation of the user/tag itself, rather than how it relates to other tags, as shown in Section 5.3.1.

To reduce such false positives and enable accurate tracking without the need for expensive and cumbersome infrastructure, we choose near-infrared (NIR) light as the wireless medium for tags to transmit beacons. With wavelengths in nanometers, NIR light is imperceptible, directional, and cannot penetrate opaque macroscopic objects (e.g., human body). Thus, it is the ideal medium for measuring line-of-sight proximity in our context. Furthermore, to infer relative angles and distances to other tags, we leverage two collocated infrared photodiodes each pre-configured with a different orientation (Section 5.3.1). By analysing the difference of light intensity sensed by the photodiodes, we can compute the incident angle and distance to each sensed tag.

The second challenge lies in enabling reliable tracking using infrared light beacons. Light beacons can be accidentally blocked by user's hands, clothes, another user passing by, or other objects (e.g., book, paper) introduced during the interaction; the motion of user body can cause tags suddenly moving beyond each other's sensing range. In all these cases, the tracking results using NIR light can either become unavailable or have low fidelity. To deal with these artifacts and realize reliable tracking, we augment light-based tracking with inertial sensors (i.e., accelerometer, gyroscope). Although inertial sensors measure only the motion status (e.g., velocity, orientation) of the tag itself, we design a data fusion algorithm (Section 5.3.5) that leverages inertial sensor data to extrapolate missing relative angles and distances upon losses of light beacons.

The third challenge is to ensure that tags operate with low power to avoid frequent charging and to ease tag distribution for various studies. Certain components (e.g., NIR LED) consume relatively higher power than others, and directly detecting short (e.g., 1.8 μ s) NIR light pulses imposes an energy burden of high analog-to-digital (ADC) sampling (e.g., 500 KHz). To improve system energy efficiency, we design strategies (Section 5.3.6) for a tag to adapt its operation mode to the current context (e.g., presence of nearby tags, user's motion status). It selectively switches off more energy-demanding modules to save energy without much sacrificing sensing temporal granularity. We also judiciously design the NIR sensing circuit to eliminate the need of high ADC sampling (Section 5.4).

We have implemented our designs and fabricated wearable tags, which we name Protractor, using off-the-shelf, low-cost hardware. Each tag is measured $74 \times 54 \times 15$ mm in size and 40 g in weight. We have evaluated the efficacy of our tags in ranging and angle detection using both controlled experiments and on-body experiments in real-life interaction scenarios. Our main findings are as follows:

- Protractor achieves 2.2° mean angular error and 6° 95th percentile in estimating interaction angles and 2.3-cm – 4.9-cm mean error in ranging;
- Protractor is robust in diverse settings (e.g., tag height offsets, indoor lighting variations, reflections from nearby objects) and effectively mitigates occasional missing or unreliable NIR tracking results with data fusion;
- Protractor is capable of running continuously for 5 days with a single charge by switching into low power modes based on contextual information;

We see the potential of Protractor not only in the support of social research but also for practical applications (e.g., providing real-time behavioural feedback during interactions, novel human-computer interaction interfaces). In comparison to approaches using cameras, Protractor serves as a more lightweight and scalable alternative. Its unobtrusive nature and the wearable form factor could ease privacy concerns and potentially reduce biases for accurate behavioural monitoring. To examine its practical implications, we have further deployed our tags to track users’ interaction proxemics when collaborating on “The Marshmallow Challenge” [216] as a creative task. We will examine the results of this deployment in Chapter 6 while in this chapter we focus on the technical contributions.

Chapter Outline. In Section 5.2 we analyse different candidate mediums for the detection of distance and angle of interaction, discussing the strengths and weaknesses of each one. In Section 5.3 we describe the design of our system to measure distance and relative angle between interacting people. We detail how we extract these datapoints from a pair of collocated photodiodes and how we make our system robust by using IMU sensors to interpolate missing data. Section 5.4 describes Protractor’s implementation while Section 5.5 evaluates its performances in detection accuracy, robustness, scalability and energy consumption. Section 5.6 discusses the results we are able to achieve with Protractor and its potential applications and in Section 5.7 we conclude the chapter.

5.2 A case for light-based tags

Our design of the wearable tag starts with seeking a suitable wireless medium to transmit beacons, which are exploited to infer incident angles and distances to other tags/users in an interaction. The ideal medium should best facilitate the measure of line-of-sight distance and incidence angle, so that we can correctly identify participants in a contact. We now discuss three candidates: RF signals (e.g., Wi-Fi, Bluetooth), ultrasound, and light.

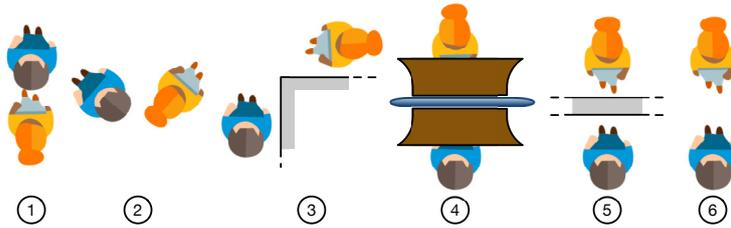


Figure 5.2: The first five are example scenarios with users in proximity but not in an interaction, where two users are 1) in line with one’s back facing the other; 2) in opposite directions; 3) on either side of a corner, without line of sight; 4) occupying opposite desks with a partition in the middle; 5) on either side of a wall when working in neighboring offices. Scenario 6 is an actual interaction with users facing each other at a distance of 1 m. We use the last scenario as a reference for comparison.

5.2.1 Radio frequency

Prior studies have utilized RF signals on wearable devices or smartphones to monitor social interactions [44, 4]. These systems examine the received signal strength (RSS) to infer if users carrying or wearing these devices are engaged in an interaction. However, RF signals are omni-directional, penetrate human bodies and objects, and are susceptible to multi-path effects. All these characteristics can make the identification of close encounters and relative orientation difficult and prone to false positives. We have discussed some of these issues in previous chapters where we used Bluetooth Low Energy for sensing (Section 3.5 and Section 4.4).

To verify this problem, we take Bluetooth Low Energy (BLE) as an example, and devise simple experiments that recreate realistic scenarios that involve two people in an indoor environment (Figure 5.2). These scenarios represent different combinations of people and objects between the transmitting devices. In each scenario, users wear a BLE device (Nordic nrf51822 SoC) on the chest, transmitting advertisement beacons at 10 Hz rate with -20 dBm TX power and scanning for beacons every 100 ms with each scan lasting 20 ms. We collect RSS traces from each BLE device for 60 seconds in each scenario. The experiments are conducted outside office hours to avoid the presence of moving people in the vicinity. However, the environment presents various surfaces that could reflect radio signals (e.g., walls, the floor, the ceiling) and there were also 5 Wi-Fi access points active (2.4 GHz).

Figure 5.3(a) shows box plots of RSS values in dBm in all scenarios, where a higher value indicates a higher received signal strength. Figure 5.3(b) shows the percentage of received BLE beacons. We make two main observations. First, as expected, in all scenarios (1 – 5) where users are not in a social contact, BLE packets can still be received even when two devices are not in line of sight. The reception ratio of BLE packets is below 30% because the device does not scan continuously but performs a 20ms scan every 100ms and thus

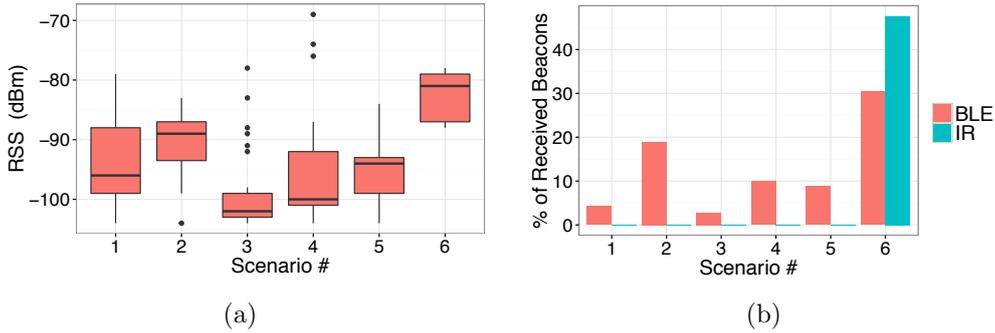


Figure 5.3: (a): Received signal strength of BLE packets for the six scenarios described in Figure 5.2. (b): Percentage of received beacons by BLE and Infrared for the six scenarios described in Figure 5.2.

misses advertisement beacons. Second, although users are stationary, RSS values vary significantly in a single setting and across different settings. This is because BLE uses three channels (separated by 2 MHz) to transmit advertisement beacons, resulting in fades at different spatial positions for different channels, even when transmitter and receiver are static [63]. Wi-Fi signals present similar characteristics. We conclude that while RF is suitable for omni-directional proximity detection, it is not the proper choice for accurate line of sight measurements.

5.2.2 Ultrasound

Next, we examine ultrasound for transmitting beacons. With wavelengths in millimeters, ultrasound has been shown to have line of sight propagation and be unable to penetrate objects. This has been exploited by earlier studies to sense interaction distances [87] or to position devices [167, 168, 143]. In our experiment, we modified the HC-SR04 [1] ultrasonic transducer (4.5-cm in diameter) with 40 kHz center frequency, commonly used by prior studies [87, 167, 168, 143]. It sends carrier bursts for 8 cycles periodically (1 transmission every 2.5 seconds in our experiment). These bursts are treated as pure pulses at the receiving end without any decoding, and we use an oscilloscope to inspect the signal and its amplitude. We repeated the experiment in the same scenarios in Figure 5.2. Our results confirmed that ultrasound cannot penetrate objects in scenarios 1, 4, or 5, whereas in scenarios 2 and 3, we occasionally observe weak pulses, possibly due to reflection and multi-path effect. Such pulses can trigger incorrect detection of social contacts if the appearances of pulses are used for ranging [87].

5.2.3 Light

We now move on to examining light as the final candidate. Specifically we considered NIR light rather than visible light¹, because NIR is imperceptible to the human eye and keeps the wearable tag sensing unobtrusive. We repeated the same scenarios in Figure 5.2, where users wear an NIR transceiver on their chest transmitting one NIR beacon per second. For each scenario we logged the received and decoded beacons for 60 seconds and computed the percentage of received beacons (Figure 5.3(b)). We observed that the NIR transceiver does not receive any beacons in scenarios (1 – 5) where the devices were not in line of sight. The beacon losses in scenario 6 were due to errors during the decoding at the receiver end, which prevented the identification of the correct beacon². NIR light propagates as a directional beam in a cone shape, thus it serves as a good medium to detect and monitor relative angle and distance of interacting people. Additionally, typical NIR emitters and receivers have a very small form factor (e.g., 5×5×7-mm), which is desirable for building a wearable device to be worn all day.

Based on all the above experiments, we decided to choose NIR light as the wireless medium for sensing non-verbal cues in social contacts.

5.3 Protractor design

The core of Protractor is to measure relative angles and distances of interacting users in an accurate and a reliable manner. Protractor achieves accuracy by exploiting the propagation characteristics of NIR light for precise angle detection and ranging. It ensures tracking reliability by fusing inertial sensors and NIR sensors to compensate for the occasional loss (e.g., light being blocked) of light tracking results. Above all, as a wearable tag, Protractor is designed to operate with low power. Next, we elaborate on each design component.

5.3.1 Angle detection and ranging

A face-to-face interaction can occur in various forms. Two important non-verbal interaction cues are the distance between any two involved users and their relative body orientation [80]. We define the latter as the *interaction angle*, which is the angle between the body normal and the line connecting the two users (Figure 5.4(a)).

¹A recent study [197] uses ultra-short visible light pulses to enable imperceptible communication. It can also be a candidate.

²Note that as a simple proof of concept, this experiment is comparing different media, rather than extensively analysing general success rates in decoding NIR beacons. Our tags achieve much higher success rates in decoding by regulating beacon transmissions and adding random delays. We will discuss our tag design in Section 5.4 and detailed experiments on its decoding robustness with multiple tags in Section 5.5.3.

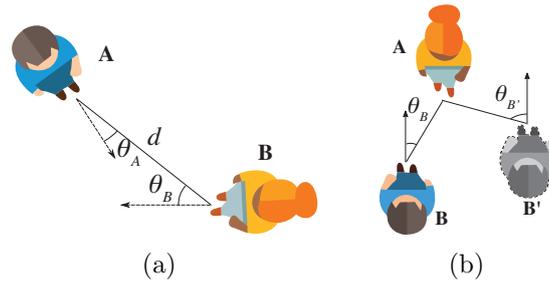


Figure 5.4: (a): Interaction distance d , and relative body orientation, i.e., interaction angles (θ_A, θ_B) , in a social contact. (b): The same absolute orientation of B can lead to different orientation relative to A .

At first glance, interaction angles can seemingly be obtained using the magnetometer/compass sensor, which measures the user’s absolute orientation. Then by exchanging the information with nearby users, one can estimate relative angles to others. However, knowing absolute orientation alone is inadequate to infer interaction angles. Figure 5.4(b) shows a simple example, where even if both users A and B ’s absolute orientations are known, their interaction angle still cannot be determined. Because B can be at location B' with the same orientation, which yet results in a different interaction angle $\theta_{B'}$. Adding the knowledge of A and B ’s distance does not help either (B' and B are at an equal distance to A). Such angle ambiguity can be resolved with A and B ’s absolute locations, obtained by existing user-centric indoor localization methods [127, 115, 222]. But still, the user’s 2D location coordinates indicate little about the actual occurrence of face-to-face contacts. As shown in earlier examples (Figure 5.2), nearby users can be separated by other indoor objects (e.g., a wall, desk partition) and thus not in a social contact.

Protractor overcomes the above problem by directly measuring the line-of-sight channel between two chest-worn Protractor tags using NIR light. Its key design elements are the NIR light beacons emitted by each tag, the detection of incident angle, and the estimation of line-of-sight distance.

5.3.2 NIR light beacons

A Protractor tag periodically emits NIR light beacons (1 beacon every 5 s in our implementation), each of which encodes the user ID. We choose the NIR wavelength of 940 – 950 nm for the beacon transmission. It is commonly used in consumer wireless infrared communication such as TV remote control. To encode data, an NIR emitter (i.e., LED) flashes at a carrier frequency (38 kHz) in bursts. Among various IR modulation/coding schemes, Sony IR coding [206] was chosen for its popularity. As illustrated in Figure 5.5, bit 1 is encoded as 1200 μs carrier frequency burst followed by an off duration (600 μs), while bit 0 is 600 μs carrier frequency burst followed by an off duration (600 μs). To reduce the power consumption, we decreased the LED’s duty cycle of the carrier to 7%.

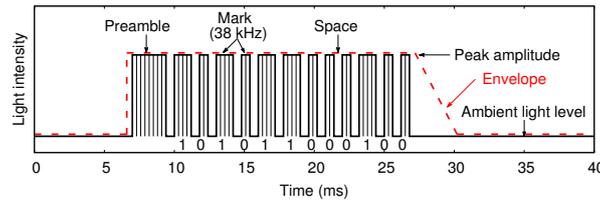


Figure 5.5: Time series of NIR light pulses of an example beacon.

To decode light beacons, we used an infrared receiver module [207], which outputs logic LOW continuously for carrier frequency (mark) and logic HIGH for off duration (space).

The micro-controller polls the receiver’s output every $50 \mu\text{s}$ to detect the duration of each mark and decode bits.

In addition to conveying the user/tag ID, the received signal strength (RSS) of a light beacon is utilized later for deriving interaction angles and distances. Here a light beacon’s RSS equals the peak amplitude of the light pulse minus the ambient light baseline (Figure 5.5). Measuring the RSS is challenging on a low-power wearable device because the common IR carrier frequency is 38 KHz, meaning that the light pulse can be as short as $1.8 \mu\text{s}$ (7% duty cycle). Detecting such short light pulses requires a sampling rate higher than 500 KHz, imposing a high energy overhead to the tag. To address this problem, Protractor leverages an envelope detector (Figure 5.5 and 5.10(b)) that holds the signal at its peak until the end of a beacon. It allows the micro-controller to sample the peak amplitude with much lower rates (1 kHz in our implementation).

5.3.3 Deriving interaction angle

Protractor reuses light beacons to derive the interaction angle from the user/tag that each received beacon corresponds to. In the RF literature, estimating the signal’s angle of arrival commonly relies on multiple antennas placed within known intervals to measure phase offset [187, 218, 73, 107] or mechanically rotating antennas [114, 113]. These methods are not applicable in our context, because of the tag’s small form factor. Also, since LED is an incoherent light source, there is no phase information as in RF technologies.

Instead, Protractor leverages the fact that an NIR photodiode responds to incoming light with different sensitivity depending on the light’s incidence angle, which is referred to as the photodiode’s *angular response*. Thus, if two collocated NIR photodiodes face different directions, incoming light from a given incident angle can result in different signal strength perceived by each photodiode. If we can obtain the one-to-one mapping between the light incident angle and the resulting signal strength pattern at photodiodes, we can then derive the incoming light’s incident angle based on measured RSS values at photodiodes.

Before diving into the detail of the above method, we first describe the optical channel

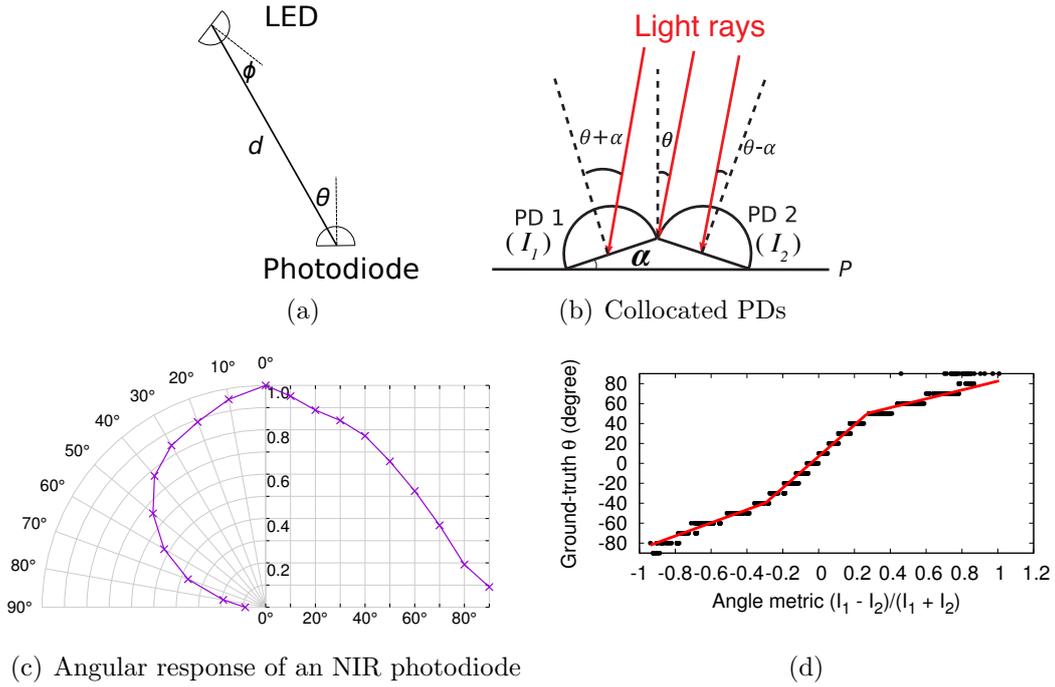


Figure 5.6: Estimating the interaction/incident angle θ using two collocated photodiodes (PD). (a) shows the optical channel between an LED and a photodiode, with irradiance angle ϕ at the LED and incident angle θ at the photodiode. (b) shows two collocated photodiodes facing different directions. Because of the photodiode's angular response (c), two PDs perceive different signal strength I_1, I_2 . The incident angle θ and the angle metric in Eq. (5.2) have a piecewise linear relationship, which can be used to estimate θ at runtime (d).

model characterizing the propagation of NIR light. For a LED and photodiode pair with distance d , assume that the LED's light ray with irradiance angle ϕ hits the photodiode with incident angle θ (Figure 5.6(a)), and I denotes the RSS at the photodiode. I can then be calculated as [105, 127]:

$$I = A \frac{F(\phi)G(\theta)}{d^2}, \quad (5.1)$$

where A is a constant determined by the transmit power and receiver's gain, $F(\phi)$ is the LED's irradiation pattern at irradiance angle ϕ , and $G(\theta)$ is the photodiode's angular response at incident angle θ .

Now consider two collocated photodiodes that are rotated clockwise and counter-clockwise respectively, by a pre-defined angle α with respect to the reference plane P (Figure 5.6(b)). Suppose θ is the interaction angle, i.e., the angle between the incoming light and the normal of P . Then for the first and second photodiodes, the light's incident angle is $\theta + \alpha$, and $\theta - \alpha$ respectively, resulting in a different RSS at each photodiode. Using the optical channel

model (Eq. (5.1)), we can compute the RSS at each photodiode as:

$$I_1 = I_0 G(\theta + \alpha), \quad I_2 = I_0 G(\theta - \alpha)$$

where $I_0 = AF(\phi)/d^2$. We considered the same I_0 for both photodiodes because ϕ and d are the same for both photodiodes, given that the distance from the LED to the photodiodes (e.g., 30 cm to 2 m for normal social contacts) is much larger than the photodiode size (5 mm in diameter). α is a known parameter, so the question now is how to derive θ after measuring I_1 and I_2 . A straightforward method is to exhaustively measure the photodiode's angular response at different incident angles and to seek the best-fit $G(\cdot)$ function. Then θ can be computed by solving the equation $I_1/I_2 = G(\theta + \alpha)/G(\theta - \alpha)$. This method, however, is ineffective. Because $G(\cdot)$ can be complicated (e.g., $\cos^m(\theta)$) or even without analytical form, there is no closed-form solution. Numerical methods, such as Newton's method, are too computationally intensive.

To circumvent the need to solve the complicated equation, we sought a metric that was computed based on I_1, I_2 and had a simple 1-1 mapping with θ . To this end, we defined an *angle metric* i as

$$i \equiv \frac{I_1 - I_2}{I_1 + I_2} = \frac{G(\theta + \alpha) - G(\theta - \alpha)}{G(\theta + \alpha) + G(\theta - \alpha)}. \quad (5.2)$$

Since the angular response of NIR photodiodes are typically symmetric (i.e., $G(\cdot)$ is an even function), the relationship between i and θ has the following properties: first, i is zero when $\theta = 0$, as $G(\alpha) = G(-\alpha)$; second, the relationship between θ and i is approximately linear, even when $G(\cdot)$ is non-linear, such as $\cos^m(\theta)$, based on our simulation, indicating that we can always apply linear regression to seek the relationship between θ and i .

To verify the relationship between i and θ , we conducted a benchmark experiment using two NIR photodiodes (OSRAM SFH 205 F [157]) with the measured angular response in Figure 5.6(c). We arranged the two photodiodes with $\alpha = 22.5^\circ$ (Figure 5.6(b)) on a table and moved the IR transmitter to emulate different interaction angles (-90° to 90°) and different distances (50 cm to 200 cm) (Figure 5.11(a)). At each location, the transmitter sent beacons for 30 seconds. We measured I_1 and I_2 at two photodiodes and computed the metric i (Eq. (5.2)). We then plotted all i values along with θ in Figure 5.6(d). We observed that θ is piecewise linearly³ related to i . With the linear relationship obtained offline through sample measurements, we could derive θ on the fly after computing i based on measured I_1 and I_2 .

5.3.4 Estimating interaction distance

Protractor estimates the interaction distance by leveraging the optical channel model

³We ran a linear regression at different intervals ($[-90^\circ, -30^\circ]$, $[-30^\circ, 30^\circ]$, and $[30^\circ, 90^\circ]$) to obtain the linear relationship. For photodiodes with single-slope linear angular response, the relation would also be single-slope linear.

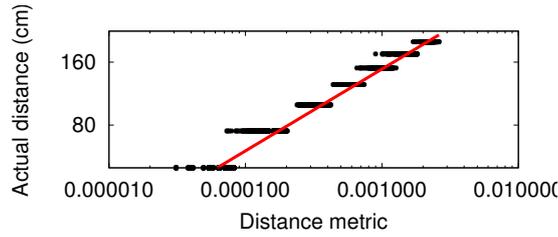


Figure 5.7: Distance measurement model. Both axes are in log scale.

(Eq. (5.1)) and derived interaction angles. Specifically, for a pair of tags m and n , each tag first detects its interaction angle with the other tag, i.e., θ_m, θ_n . Since the interaction/incident angle of a tag is also the irradiance angle of the other tag, we can compute the distance d_{mn} between m and n as $d_{mn} = \sqrt{A \frac{F(\theta_n)G(\theta_m)}{I_m}}$, where I_m is the RSS of light beacons from tag n measured at tag m .

Directly computing the above formula requires knowing the value of A . Instead, we define a *distance metric* l as $l = F(\theta_n)G(\theta_m)/I$ and rewrite d_{mn} as

$$\ln(d_{mn}) = a \ln(l) + b. \quad (5.3)$$

We computed the logarithm in the above equation because the exponent of the distance d is not exactly 2, as shown in our measurements. We calibrated parameter a and b using benchmark experiments, where we collected l values⁴ along with the ground-truth distance d_{mn} , and ground-truth interaction angles θ_m, θ_n . We then performed a linear regression to determine a and b . Figure 5.7 shows our benchmark experiment results and the linear model. With the trained linear model (Eq. (5.3)), we then computed interaction distances based on the derived interaction angles.

5.3.5 Sensor data fusion

While providing precision, NIR light tracking alone is not reliable for a number of reasons: light can be easily blocked by other objects (e.g., a waving hand, a book, a piece of paper) introduced in an interaction; or the chest-worn tags can occasionally move beyond each other’s sensing range, due to user’s body movement during a contact. To enhance the tracking reliability, Protractor leverages inertial sensors (i.e., accelerometer, gyroscope) to compensate for the low fidelity of light tracking results in those occasions. We chose inertial sensors because they are small in size (2.5×3 mm) and consume low power (e.g., 2.8 mW). They can be easily fit in the wearable tag and operate continuously in the background with negligible energy overhead.

⁴We estimated $F(\cdot)$ and $G(\cdot)$ based on sampled measurements.

The challenges of using inertial sensors lie in sensory measurement noise. Such noise is particularly troublesome when measuring small displacement (e.g., centimeter-level distance change)⁵. In a social interaction, users tend to remain static at their 2D locations while changing body orientation by a greater extent. Thus, we consider fusing only the gyroscope data and the estimated interaction angles, while using accelerometer to sense large location displacement for determining the start/end of a new sensor fusion process.

To fuse the NIR angle detection results and gyroscope readings, we adopted the Kalman filter algorithm [99, 61, 77] for its simplicity and efficiency. Specifically, we modelled the interaction angle as a discrete-time hidden Markov model (HMM):

$$\begin{aligned}\theta_t &= \theta_{t-1} + \Delta\theta_t + w_t, & w_t &\sim \mathcal{N}(0, \sigma_{w,t}^2) \\ \tilde{\theta}_t &= \theta_t + v_t, & v_t &\sim \mathcal{N}(0, \sigma_{v,t}^2)\end{aligned}$$

where θ_t is the hidden state (i.e., the actual interaction angle) at time t , $\tilde{\theta}_t$ is the observation (i.e., the estimated interaction angle using NIR measurements), $\Delta\theta_t$ is the orientation change measured by the gyroscope sensor, v_t denotes the Gaussian observation noise (i.e., the angle detection errors using NIR light), and w_t is the Gaussian noise of gyroscope readings. Given that it is a linear Gaussian Bayesian model, Kalman filters have been proven to seek the optimal solution recursively [99].

Our data fusion based on the Kalman filter recursively conducts two steps: prediction and updating. The prediction step produces the estimated mean and variance of the interaction angle at t , before the arrival of new NIR measurements at t . It predicts the interaction angle by:

$$\begin{aligned}\hat{\theta}_{t|t-1} &= \hat{\theta}_{t-1|t-1} + \Delta\theta_t \\ \sigma_{\theta,t|t-1}^2 &= \sigma_{\theta,t-1|t-1}^2 + \sigma_{w,t}^2.\end{aligned}$$

Upon the arrival of new NIR measurements and thus newly derived interaction angle $\tilde{\theta}_t$, the updating step then incorporates the new observation into the prior estimate and obtains improved posteriori estimates. It updates estimates as follows:

$$\begin{aligned}\hat{\theta}_{t|t} &= \hat{\theta}_{t|t-1} + k_t(\tilde{\theta}_t - \hat{\theta}_{t|t-1}) \\ \sigma_{\theta,t|t}^2 &= \sigma_{\theta,t|t-1}^2 - k_t\sigma_{\theta,t|t-1}^2\end{aligned}$$

where $k_t = \sigma_{\theta,t|t-1}^2 / (\sigma_{\theta,t|t-1}^2 + \sigma_{v,t}^2)$.

The update step can mitigate large accidental errors in NIR measurements, such as incorrect pulse amplitude detection due to ADC malfunction. The data fusion addresses the problem of occasional losses of NIR measurements, as its prediction step produces an

⁵Our experiments with inertial measurement unit Bosh BMI160 show non-zero sensor readings (e.g., 0.03 m/s² at x-axis) in the stationary mode even after removing the constant offset. It translates into 1.5-m location drift after only 10 seconds.

estimated interaction angle without new NIR measurements.

We start the fusion with an NIR measurement: $\hat{\theta}_{0|0} = \tilde{\theta}_0$, $\sigma_{\theta,0|0}^2 = \sigma_{v,0}^2$. We modelled the variance of noise v_t and w_t based on our experimental observations. Specifically, our experiments show that NIR angle detection errors tend to have a small variance when both photodiodes have large pulse amplitude readings. Thus, we modelled the variance $\sigma_{v,t}^2$ of the observation noise v_t as $\sigma_{v,t}^2 \propto 1/(I_{t,1} + I_{t,2})$. We modelled the gyroscope noise variance as $\sigma_{w,t}^2 \propto \Delta t$, because of the drifting problem of gyroscope sensor. The orientation change is an integration of the gyroscope readings and thus its error accumulates over time. We terminate the data fusion process when large location displacement is discovered from accelerometer readings [116], e.g., users walk away from their previous locations.

5.3.6 Adaptive sampling

Given our goal of continuously tracking social contacts, Protractor’s battery life is a critical aspect of our design. To ensure efficient use of available power while keeping the tag operational, we applied context-aware duty cycling. Succinctly, when no interactions are detected for a period or, when the tag is not being used, the more energy-demanding modules are switched off or reduced in their capability to save energy. The more energy-demanding modules are the angle detection module (mainly the transimpedance amplifier, Section 5.5.4) and the NIR LED, while the inertial measurement unit (IMU) and NIR receiver consume low energy. We thus use the IMU and NIR receiver to infer the current context (i.e., presence of other devices nearby and user’s motion status) and adapt Protractor’s operation accordingly.

We define three states that a Protractor tag could be in at any given time: 1) *High Power*: all modules are powered on, NIR beacons are transmitted every 5 s and the two photodiodes in the angle detection module are sampled at 1 kHz; 2) *Low Power*: the angle detection module is powered off and NIR beacons are transmitted every 20 s; 3) *System Off*: the angle detection module and the NIR transceiver are powered off.

In High Power and Low Power states, the IMU and NIR receiver are powered on and functional because they are used to trigger the state change. In System Off state only the IMU is powered on. The NIR transmission rate (one beacon every 5 s) is selected to reduce the probability of collisions in the presence of multiple tags. By contrast, we adopt a transmission period of 20 s in the Low Power state to save energy but be reactive in case of interaction and do not severely sacrifice the temporal granularity of collected data. Additionally, even though NIR transmission has a relatively high power consumption, its duration is short (i.e., few tens of milliseconds) and thus its impact on the overall energy consumption is limited (see power profiles of individual components in Section 5.5.4).

We define two rules for the state transition. *Rule 1*: no interaction has been detected in the last 20 minutes; *Rule 2*: no movement has been detected in the last 20 minutes. Rule 1 is to detect scenarios where people are not in interactions for long (e.g., when completing

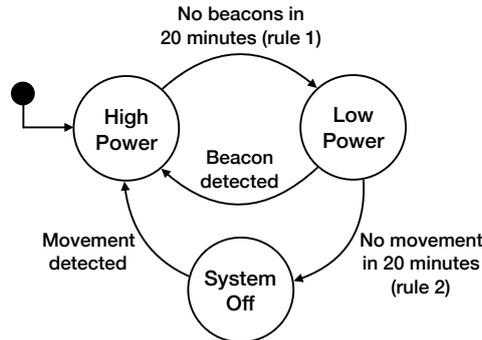


Figure 5.8: State machine diagram for the adaptive sampling technique implemented on Protractor.

individual work). Thus there is no need to monitor angle/distance with high granularity and the tag switches to the Low Power state. Rule 2 is to infer when the tag is not in use and triggers the transition to System Off state. Once any above condition is not met, the tag reverts to High Power state. The 20-min window from the last interaction is chosen to avoid missing short contacts with short intervals. Similarly, the 20-min window for body movements prevents the transition to System Off state when the user is stationary for a while with the tag still in use. Figure 5.8 shows the state machine diagram with the transitions between the states. In Section 5.5.4, we will examine the tag’s energy consumption and the benefits of adaptive sampling.

5.4 Protractor prototype

We fabricated 6 Protractor tags using off-the-shelf hardware contained in a 3D-printed case. The final assembled tag (Figure 5.9(a)) resembles an access badge that can be worn using or a clip. It measures $74 \times 54 \times 15$ mm in size and 40 g in weight (with a 560 mAh battery). Figure 5.10(a) shows its main internal components, including the NIR sensing module, the battery, the IMU, and the micro-controller. We will next describe three key components (NIR sensing module, IMU, and micro-controller) in detail.

NIR sensing

The NIR sensing components are the NIR transceiver and angle detection module, which are hosted by a customized printed circuit board (PCB) we design and fabricate (Figure 5.9(c) and 5.9(d)). For the NIR beacon transmitter, we choose the OSRAM SFH 4240 [156] as the NIR LED, because it provides $\pm 60^\circ$ 3dB beam angle that enables a wide sensing range. Its wavelength peaks at 950 nm. We use an NPN transistor to driver the LED. We choose the Vishay TSOP38238 as the NIR receiver, which includes both the

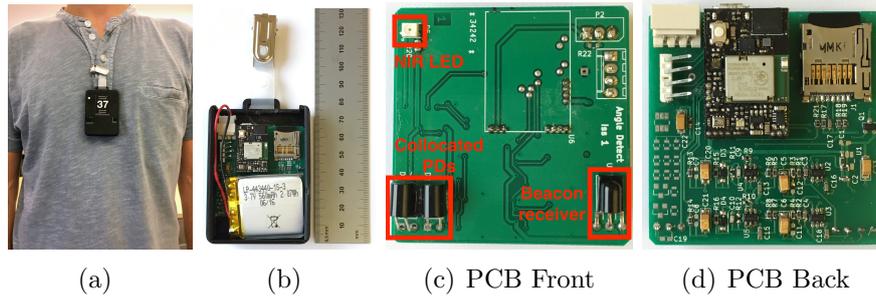


Figure 5.9: Protractor prototype. (a) and (b) show the assembled tag and its internal components. (c) and (d) show the two sides of the PCB we designed and fabricated to host the NIR transceiver and angle detection module.

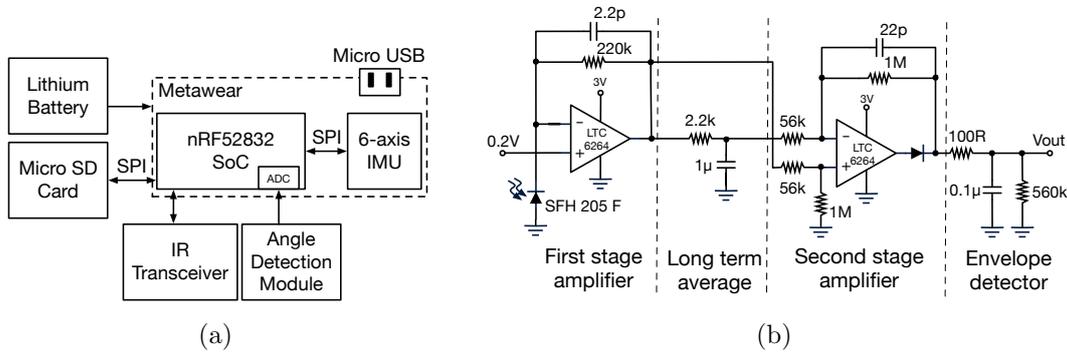


Figure 5.10: Protractor design. (a) is the block diagram of the components. (b) is the circuit design of the angle detection module, including a two-stage amplifier, a long-term average to remove the influence of ambient light, and an envelope detector.

photodetector and pre-amplifier. The receiver outputs low when it senses the carrier frequency 38 kHz. Its output signal is connected to the micro-controller for decoding. We use the Sony Serial Infra-Red Control (SIRC) protocol (12-bit) to transmit the tag ID every 5 seconds. We select 5 s as transmission period to balance power consumption and resolution of the collected data. To prevent collisions in case two or more devices have their transmissions synchronized, we perturb each transmission by adding a random delay (4 – 1020 ms). A collision of multiple NIR beacons makes beacon decoding impossible. In this case the beacon is discarded and it is not used to infer angle and distance, hence not affecting the inference accuracy.

The angle detection module has two NIR photodiodes (OSRAM SFH 205 F [157]) with spectral range of sensitivity from 800 nm to 1100 nm. They are arranged on a 3D-printed base and their orientations form a 45° angle.

Figure 5.10(b) shows this module’s circuit design including a two-stage amplifier and an envelope detector. We adopted a two-stage amplifier in order to detect the light beacons even in environments with high light levels. The first stage is a transimpedance amplifier with a relatively low gain to avoid saturation in bright conditions. The second stage is a differential amplifier which measures the difference between the average light level (RC

network between the two stages) and the instant light level and amplifies the signal further with a gain of 17.8. This configuration allows us to remove the ambient light level which is added to beacon signal and might cause the amplifier to saturate, preventing a correct measure of the amplitude of the signal.

Inertial measurement unit

We use the Bosh BMI160 6-axis IMU that embeds an accelerometer and gyroscope in the same package. The IMU operates with low power (around $950\mu\text{A}$ with accelerometer and gyroscope in full operation mode) and contains an on-board FIFO buffer where sensor readings can be accumulated without CPU intervention. This allows the micro-controller to sleep for longer periods, leading to a longer battery life. The accelerometer and gyroscope are sampled at 25 Hz.

Micro-Controller

All components are controlled by a Nordic's nRF52832 SoC that includes a 32-bit ARM-M4F CPU and a 2.4 GHz radio transceiver. We use a nRF52832 developer board from Mbienlab Inc. that contains the main SoC, the Bosh IMU and associated circuitry. We attach a micro SD card socket to the SoC using the Serial Peripheral Interface (SPI). The entire device is powered by a 560 mAh 3.7V lithium battery that can be recharged via a micro-USB interface.

The micro-controller samples the output of the two photodiodes (after the amplifier and envelope detection) every 1ms (1 kHz) using the on-board 14-bit ADC and logs the data on the SD card. The sampling is stopped during the transmission of NIR beacons to avoid the detection of false pulses from the same device. The entire schematic of Protractor's electronics is reported in Appendix D.

5.5 System evaluation

We evaluated the systems performance of Protractor prototypes, aiming to examine Protractor's accuracy in determining interaction angle and distance, the impact of practical factors (e.g., differences in user body heights, reflections, ambient light), its scalability with multiple tags, and its energy consumption. We will also examine the efficacy of data fusion in enhancing the tracking reliability.

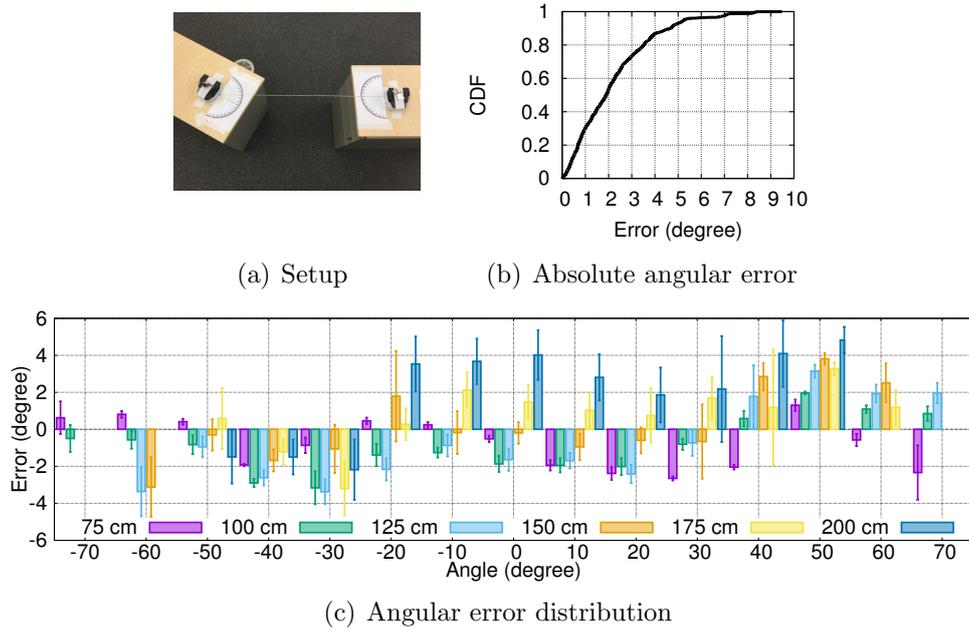


Figure 5.11: Protractor’s accuracy in estimating interaction angles.

5.5.1 Accuracy

Experimental setup

We conducted controlled experiments with two static tags to examine Protractor’s tracking accuracy using only NIR light. In particular, we placed each tag on a different table and supported each tag via a piece of foam to emulate the actual usage scenario where tags face each other (Figure 5.11(a)). The two tags were at the same height and we varied their distance and relative orientation. To obtain the ground truth on the distance d , we connected the tags with a string and measured the string’s length. To obtain the ground truth on interaction angles θ , we placed a printed angle meter under each tag to measure their relative orientation. To estimate angle $\tilde{\theta}$ and distance \tilde{d} , we computed the angular and distance error as $(\tilde{\theta} - \theta)$ and $(\tilde{d} - d)$, respectively. All experiments were indoor with normal lighting (300–400 lux, fluorescent lights).

Angle

We started with examining Protractor’s accuracy in angle detection. We rotated the table of a tag (tag 1) and kept the other table/tag (tag 2) fixed and facing tag 1. As a result, the interaction angle of tag 1 varied while the interaction angle of tag 2 remained 0° . We varied the interaction angle of tag 1 from -90° to 90° with 10° interval and the distance from 75 cm to 2 m with a 25-cm step. For each distance/angle combination, we let the

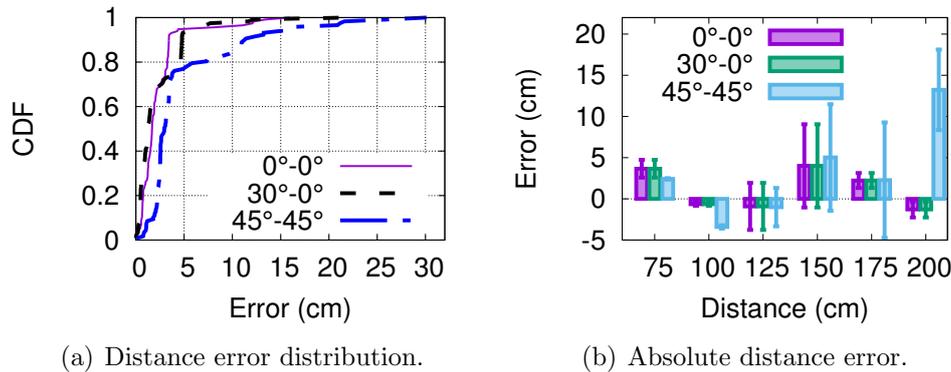


Figure 5.12: Accuracy of distance measurement.

tags transmit light beacons for one minute. We then computed the interaction angle of tag 1 using the method in Section 5.3.1.

We plotted the absolute angular errors (Figure 5.11(b)) and the error distribution under different angle/distance combinations, where error bars show the standard deviation (Figure 5.11(c)). Because we rotated each photodiode by 22.5° , one tag will not detect the other’s light beacons once the interaction angle exceeded 67.5° . Therefore, the tag’s angular sensing range spans approximately from -70° to 70° . We observed that within the sensing range, the mean error is 2.2° and the 95th percentile is 5.2° , expected to be sufficient for detecting interpersonal contacts. We observed that large errors occur at long distances and large angles (e.g., 2 m and 60°) with weak signal strengths. Since the ADC’s resolution is fixed, the ADC error ratio (error/pulse amplitude) is larger under weaker signals, leading to less precise RSS and larger angular errors.

Distance

We next examined Protractor’s accuracy in ranging. Instead of exhaustively testing all possible combinations of distance and relative angles ($\approx 1\text{K}$ test cases), we selected three representative interaction scenarios with different configurations on the two tags’ interaction angles: (1) face-to-face interaction ($0^\circ - 0^\circ$), (2) one person talking to many others ($30^\circ - 0^\circ$), and (3) two users discussing in front of a white-board ($45^\circ - 45^\circ$). In each scenario, we varied the tag distance from 75 cm to 2 m with 25-cm interval. We then measured the interaction angle at each tag and derived the interaction distance. We plotted the CDF of absolute distance errors in Figure 5.12(a). We observed that the three scenarios have similar mean errors (2.3 cm, 2.4 cm, and 4.9 cm respectively), while scenario (3) has a longer tail, with 11.4 cm as the 90th percentile compared to 3.4 cm and 4.7 cm in the other two scenarios. As we further examined the error distribution across distances for each scenario (Figure 5.12(b)), we found that the longer tail in scenario (3) is due to the error jump (10 cm) under 2-m distance. The error jumps in this case because the distance

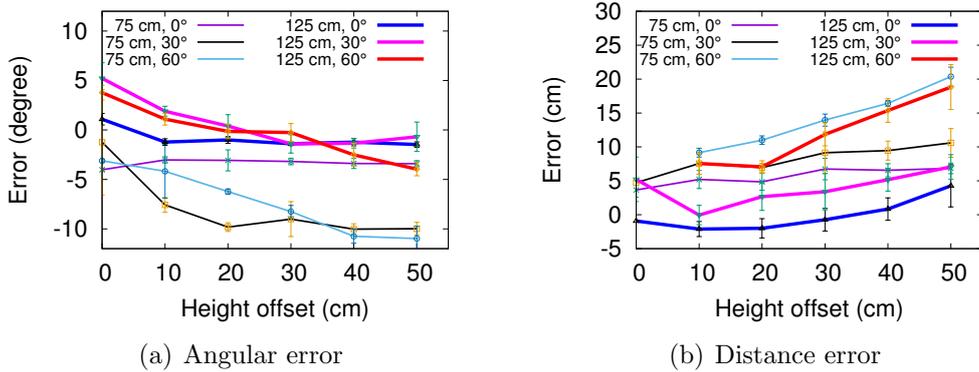


Figure 5.13: Influence of height offset.

(2 m) approached the sensing limit, and the interaction angle (45°) at each tag approached the half (3dB) viewing angle (60°) of our photodiode or LED. It resulted in weak RSS, increasing ADC error ratios and ranging errors.

5.5.2 Robustness

As a chest-worn tag, Protractor can be affected by various practical factors, such as height differences among tags, reflection of NIR light caused by nearby objects (e.g., walls), and ambient light. We now examine the impact of these factors on Protractor’s accuracy, using controlled experiments with the same setup as Figure 5.11(a).

Height offset

We first examined Protractor’s robustness when tags were at different heights. Such height offset can be caused by user’s body height difference, or the way users are wearing tags or interacting with each other (e.g., a sitting user talking to a standing user). For this purpose, we tested three settings of tags’ interaction angles ($0^\circ - 0^\circ$, $30^\circ - 0^\circ$, and $60^\circ - 0^\circ$) and two distances (75 cm and 125 cm). For each combination, we increased a tag’s height by raising its supporter and varied the height offset from 0 cm to 50 cm, which is approximately the height difference between a sitting user and a standing one. Figure 5.13(a) shows the angular errors in different combinations of interaction distance and angle. Our main observation was that angular errors did not exceed 10° even under 50-cm height offset, which demonstrates that Protractor’s angle detection is robust against tag height offset. The reason is that without any pitch rotation of the body, the vertical incident angle is the same for both photodiodes and thus has been cancelled out (similarly to the I_0 term) in our angular metric (Eq. (5.2)). The height offset, however, does affect ranging. As shown in Figure 5.13(b), Protractor increasingly overestimates the distance as the height offset increases. This is because we currently detect only horizontal interaction angles.

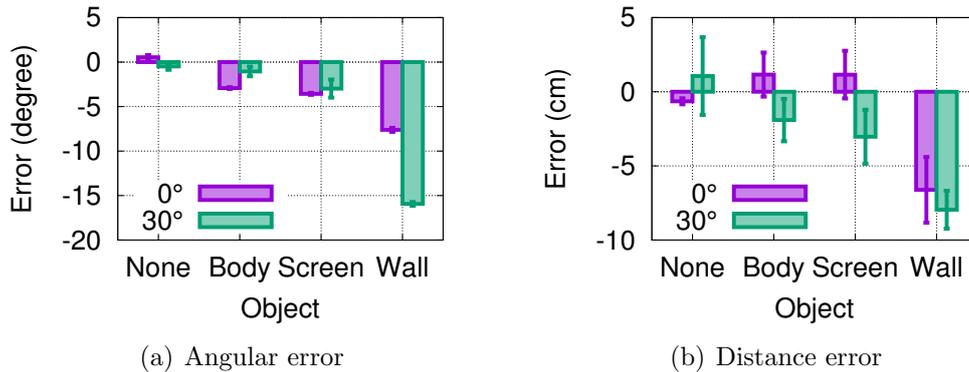


Figure 5.14: Influence of reflection.

Thus, height offset leads to a larger vertical angle and higher signal attenuation. Without knowing vertical angles, our method attributes the increase in attenuation to a longer distance. Overall, the maximum distance error caused by height offset is 20 cm. To diminish this error, we could add a pair of photodiodes to detect vertical angles, with the cost of a slightly bigger form factor and higher energy consumption. In this case, even if ambient light (e.g., office lighting) might affect the upper and lower photodiodes unevenly, it will not affect the angle detection because we subtract the background ambient light when extracting the beacon amplitude. We will leave this extension to future work.

Reflection

Next, we evaluated how Protractor’s performance was affected by NIR light reflection from nearby objects. In this experiment, we set two tags 1-m away. We then arranged another object in parallel to the line connecting the two tags at a 50-cm perpendicular distance. We tested two interaction angles (0° and 30°) for tag 1⁶ while keeping tag 2’s interaction angle at 0° . We tested three types of reflection objects: human bodies, screens, and walls. We conducted the experiment in a large office for the former two and in a corridor (1.8-m width) for walls.

In Figure 5.14, we plotted angular and distance errors for tag 1 under different types of reflection objects, where error bars show the standard deviation. As a reference, we also included the result when no reflecting objects are nearby. We made three main observations. *First*, reflection consistently caused underestimates of tag 1’s interaction angles. This is because reflection strengthens the RSS perceived by the photodiode closer to the reflection object, which biases the incident light towards the reflection objects. *Second*, among different reflection objects, walls better reflected NIR light and thus caused larger angular/distance errors, while reflections by human bodies and screens caused absolute

⁶For the 30° angle, we rotated tag 1 towards the reflection object.

errors of no more than 4° and 5 cm. *Third*, as for distance errors, wall reflection consistently caused underestimates, because it strengthens RSS and triggers our method to infer shorter distances. On the other hand, reflections by human bodies and screens were weaker and did not necessarily strengthen RSS, leading to possible overestimates. In summary, we observed that only strong reflections by nearby walls presented a challenge for Protractor, while smaller objects such as human bodies and screens introduce marginal effects.

Ambient light

We also examined the impact of ambient light on Protractor. From our experiments under different levels of indoor lighting, we observed that changes in indoor lighting did not affect Protractor’s accuracy in angle detection and ranging. The reason is twofold. First, indoor artificial lights (e.g., fluorescent lights) emit mainly visible light, whereas our NIR sensor [157] is not sensitive to visible light (390 – 700 nm), as its spectral sensitivity range is 800 – 1100 nm. Second, the measured RSS at each photodiode is the amplitude after subtracting the sensed ambient light (Section 5.3.1). Thus ambient light changes did not affect estimated angles/distances, as long as the photodiodes are not saturated.

Table 5.1: Closest working distance in various ambient lighting.

Ambient light (lux)	250	550	1220
Closest working distance (cm)	10	13	20

However, the saturation problem can occur under high ambient NIR light (e.g., bright sunlight through the window), which affects the closest working distance of our tags. Table 5.1 lists the closest working distance under different ambient lighting. The result demonstrates that our system works for common social interaction distance (longer than 20 cm) even in bright indoor environment (higher than 1000 lux). We also observed that tags cannot detect light beacons any more when its perceived illuminance exceeds 2500 lux. This level is well above the typical indoor illuminance that ranges between 300 and 500 lux [2, 3]. For comparison, in full daylight (not directed towards the sun) there is an illuminance between 10k and 25k lux [186]. We are able to achieve this robustness against variation in ambient light levels thanks to our two-stage amplifier which removes most of the ambient light from the beacon signal.

Occasional low fidelity in NIR tracking

We examined the efficacy of data fusion (Section 5.3.5) in compensating for occasional low-fidelity NIR tracking results. Using the setup in Figure 5.11(a), we set two tags 1.25-m away facing each other with 0° relative angle. We emulated two cases: occasional blockage of the light channel, and tags temporarily moving outside each other’s sensing range.

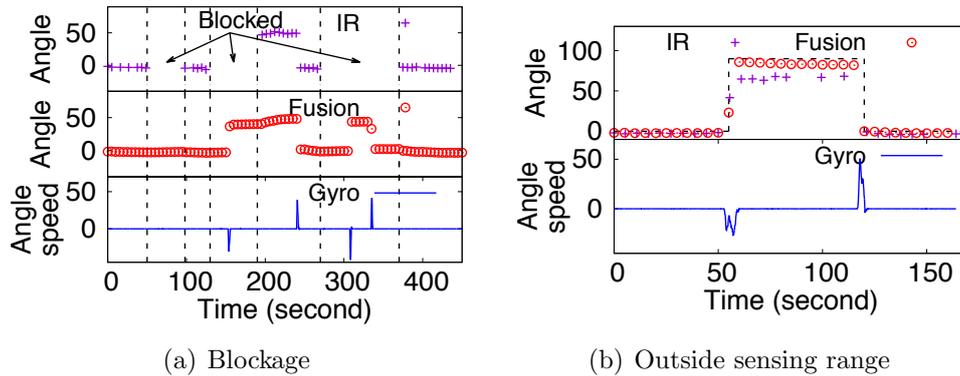


Figure 5.15: Data fusion of NIR and inertial sensors.

We tested three blockage scenarios by considering whether any tag changes its orientation during the blockage. Figure 5.15(a) shows three blockage periods: (1) 50”–90” : we placed a piece of cardboard between tags and then removed it; (2) 130”–190” : we placed a piece of cardboard, rotated a tag by 40°, and then removed the cardboard; (3) 270”–370” : we placed a piece of cardboard, rotated a tag by 40°, rotated it back, and then removed the cardboard. The tag orientation remained the same in period (1), while it changed once and twice in periods (2) and (3), respectively. We observed that although NIR angular results were absent during the blockage periods, our data fusion could immediately and accurately extrapolate missing angles using the prediction step⁷. Scenarios (2) and (3) also demonstrated the necessity of data fusion, which is capable of tracking the orientation change during the blockage. In comparison, methods such as using the most recent NIR angular result would completely miss the orientation change, which could be important non-verbal cues in a social contact.

We next tested the scenario when tags moved outside the sensing range. Our prior experiment (Section 5.5.1) showed that the maximum half sensing angle is 70° for our current prototype. Thus, we started with two tags directly facing each other, rotated a tag by 90°, and later rotated it back. Figure 5.15(b) plots the estimated angle with and without data fusion. We saw that estimated angles using NIR sensors alone were around 67°, translating into -23° error. With data fusion, the estimated angle was 85° with only a -5° error. Overall, our results validate data fusion’s efficacy in augmenting NIR tracking when NIR tracking is not available or reliable.

5.5.3 Scalability

After extensive experiments with two static tags, we then analysed the scalability of our design with more than two tags. The presence of more tags could increase the likelihood of

⁷We smoothed gyroscope data using a sliding window of length 25 samples (1 second).

NIR beacon collisions, during which signals from multiple NIR beacons add up, potentially causing errors in signal measurements and the decoding of NIR beacons. However, since our system discards collided beacons, beacon collisions do not affect the accuracy of ranging and angle estimation (our prior accuracy results with two tags hold); rather, they affect only the temporal granularity of the data.

To examine the efficacy of our system design – low transmission rate of beacons and random transmission delay (Section 5.4) – in reducing beacon collisions, we performed a test with six tags. We set up the tags on a table in two rows, where the front row is 80 cm away from the second row. Tags sent and received beacons with the configured transmission rate (0.2 Hz) for 21 hours. For each pair of devices (30 pairs in total), we computed the percentage of received beacons that are successfully decoded. Overall, we observed that the average success rate is 79.5% with 78.3% as the minimum and 80.8% as the maximum. The average duration between received beacons was 6.3 seconds. We conducted similar experiments with four tags and the average success rate in beacon decoding was 84.7%. In both experiments the distance and angle measurement errors are within the limits reported in Section 5.5.1. These results show that our system gracefully scales to larger number of tags by recording sufficient number of beacons and thus providing satisfactory temporal granularity. More sophisticated beacon designs (which we leave for future work) could be adopted in situations with a denser deployment of tags to limit collisions even further.

5.5.4 Energy consumption

Finally, we will report on the energy consumption of our prototype. We first analysed the power profile of each component using a Monsoon power monitor. Figure 5.16(a) shows the power trace of NIR beacon transmissions. For each transmission we repeated the same code 4 times to increase the chances of a successful decoding and to have enough data to infer distance and angle. A longer burst (i.e., > 4 beacons) would have provided more data for the angle and distance estimation but also increased the power consumption substantially.

Figures 5.16(b) and 5.16(c) show the power profiles of the ADC and IMU data logged on the micro SD card. The power consumed by the ADC during a conversion is low ($\approx 700\mu\text{W}$). The constant high power in Figure 5.16(b) is due to the transimpedance amplifier used to amplify photodiode signals in the angle detection module. This is the most power-demanding component in our prototype. To save energy, we buffer ADC and IMU readings (512-byte and 1024-byte respectively) and then log on the SD card only when the buffers are full. The power consumed by the NIR receiver is negligible in comparison, as it only entails the digital reading of a GPIO pin every $50\mu\text{s}$.

We also measured the average power consumed in each of the three power states (§ 5.3.6) and we obtain: 51.75 mW for High Power, 9.42 mW in Low Power and 7.96mW for System Off. The tag is powered by a 560 mAh (2.07 Wh) battery, however, the battery life of the tag depends on its usage pattern. To estimate the battery life, we computed the average

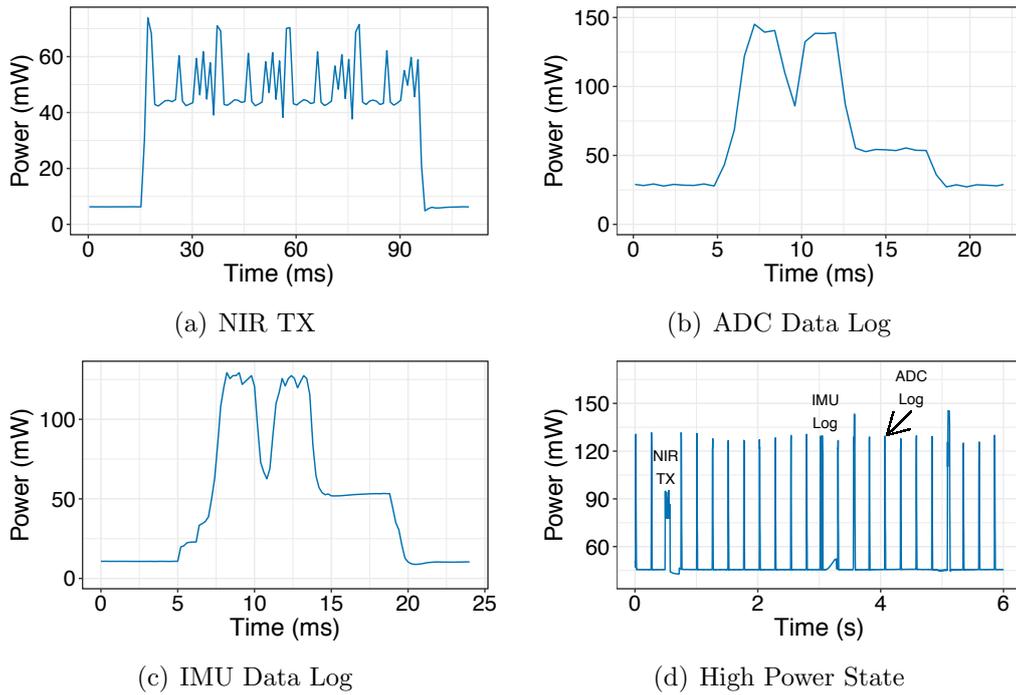


Figure 5.16: Power profiles of Protractor’s main operations and activities in the High Power state.

energy consumed per hour as:

$$P_{hour} = \frac{P_{high}t_{high} + P_{low}t_{low} + P_{off}t_{off}}{24}, \quad (5.4)$$

where t_{high} , t_{low} and t_{off} are the number of hours spent respectively in High Power, Low Power and System Off state while P_{high} , P_{low} and P_{off} are the respective power levels in each state. Assuming that on a normal working day a user spends 5 hrs interacting with people⁸ (i.e., tag in High Power state), 4 hours on individual work (Low Power state), and does not interact for the rest of the day (System Off state), we can compute the battery life by dividing the battery capacity (560 mAh) by $\frac{P_{hour}}{3.7V}$ and obtain an estimated lifetime of about 120 hrs (i.e., 5 days with a single charge). If the device was configured to stay in High Power state (9 hrs per day), without adaptive sampling, the battery would last 85 hrs.

5.6 Discussion

Protractor represents a step forward in the data collection to support social interactions

⁸Previous work found that university students spend on average 4.5 hrs per day in face-to-face conversation[120].

studies from two point of views: accuracy and scalability. The notation system designed by Hall to annotate proxemic behavior defines only 9 possible orientations and 8 distance ranges because the notations had to be simple given that the annotations were done manually [80]. Protractor instead enables data collection at fine granularity and with high accuracy, as required in previous works [139, 172]. Additionally, during manual annotations the observer needs to discriminate between the configurations available in the notation system with the possibility of introducing biases which are instead reduced with technology. The second strength of Protractor lies in the fact that it allows to automatically collect data and enables larger studies in the wild. In fact, it is not practical for observing a large number of people.

We see the potential of such a system not only to support organizational science research but also for other practical applications. Protractor tags could be used during job interviews or sales training sessions to collect data that can be later analyzed by the trainee in order to assess her behavior and to improve it over time. We also envision the possibility of using this system to provide behavioural real-time feedback during social interactions, similarly to what has been done in the past with cameras in controlled environments [52]. We believe the unobtrusive wearable form factor could ease privacy concerns and potentially reduce biases. The Protractor prototypes we built do not perform distance and angle measurements in real-time, on-device, however the algorithms used to estimate distance and orientation are sufficiently simple and lightweight to be easily implemented on the micro-controller we selected without exceeding its processing capabilities. As described in Section 7.2 we believe that the real-time availability of information regarding social contacts and non-verbal behaviour is crucial to take full advantage of wearable systems. In fact this would allow to deliver discreet feedback to the users when it is most required, for example during public speaking or job interviews. Therefore further research is needed in this direction.

Another area where Protractor could show its strengths is the design of novel human-computer interfaces based on people orientation and movements, also called *Proxemic Interactions*. In this field, expensive motion tracking systems (e.g., Vicon⁹) are used to create prototypes, but obviously they are not deployable in real applications at scale [208, 19, 34]. Protractor instead represents a viable, less expensive option to gather continuous user's orientation and motion without relying on invasive cameras. Protractor devices can be integrated into tangible objects and in the environment in order to gather accurate information about peoples' orientations in space and in relation to objects. The data generated by Protractor could then be used to offer innovative interaction paradigms for smart and connected objects. Our current prototype can also rely on the availability of Bluetooth Low Energy which could be used to locate users in the environment.

In our evaluation we have shown the maximum accuracy that we were able to achieve with our prototype. Given that different applications entail different requirements on accuracy and power consumption, our approach can be adapted to various requirements.

⁹<http://www.vicon.com/>

As an example, by tuning the beacon transmission frequency and ADC sampling rate it is possible to trade temporal granularity and accuracy for battery life. This is beneficial for applications that do not need continuous accurate angle and distance measurements but would prefer a longer operation period with a single charge. In this situation the beacons can be transmitted less frequently and the transimpedance amplifier can be switched off for longer periods of time.

5.6.1 Limitations

Technological Accuracy. Protractor focuses on the accurate measure of distance and angle of interaction, however it is not capable of detecting if people are actually interacting or not. Even if it is reasonable to assume that short interpersonal distance and small relative angles represent good proxies for social interactions, there are examples where this might not be the case. For example, in a library when two readers use the same reading table sitting one in front of each other or on a crowded bus. Additionally, Protractor does not track users' head movement, which can be useful information for understanding non-verbal behaviours.

Scalability. We showed how, in the presence of beacon collisions, only the temporal granularity of our system is affected but not its accuracy. The test was performed with 6 devices given the limited availability of Protractor prototypes. A larger number of devices in range might result in more packet loss due to collisions, up to a point where potentially no packets are received. Protractor should be evaluated in denser deployments to analyse this effect comprehensively.

Sensing Modalities. In the next chapter we will examine how interpersonal distance and angle of interaction could be used to study small groups dynamics, however the monitoring of other non-verbal cues could be useful for behavioural studies. For example, our device is not capable of detecting speech related cues, hands and head movements or gaze. We opted for a solution that would potentially result in fewer privacy concerns by the users and a more comfortable device to wear.

5.7 Conclusion

We have introduced Protractor, a system to accurately detect non-verbal cues in human interactions. The novelty of our approach lies in its ability to detect relative body orientation and distance via smart use of near-infrared light and sensor fusion algorithm exploiting inertial sensors. We showed how with a two stage amplifier we are able to drastically reduce the influence of ambient light and increase the robustness of our device. Our prototype experiments demonstrated Protractor's efficacy and its ability to reconstruct real-life interaction scenarios.

This technology significantly simplifies the study of non-verbal behaviours removing the need for expensive and invasive infrastructure, opening up the possibility to observe behaviours in real and informal settings. We explore this possibility in the next chapter where we deploy Protractor to study non-verbal cues of 64 participants while performing a creative task.

Chapter 6

Studying proxemics behaviour in small groups

6.1 Introduction

In the previous chapter we introduced a novel wearable device, called Protractor, to unobtrusively measure interpersonal distance and angle of interaction by relying on near-infrared beacons. In this chapter we explore the potential of Protractor in supporting studies in the context of organizational and social science.

Although human networks and social structures have been featured prominently in the fields of organizational behavior and human resources [136], recent research also highlights the importance of analysing actions and tasks to understand people working within organizations [79, 163]. This research however has tended to focus on higher-level perspectives such as organizational routines [66, 162], and not enough on leveraging the capacity of sensor technologies to examine micro-space and proxemic behavior as a basis for studying actions [80].

When studying social interactions, the context in which the contact is taking place affects the way people interact. For example, consider the differences between a corporate team discussing a new product or an informal coffee break. However, all social interactions seem to have one aspect in common, as described by Tischler: *“people do not interact with one another as anonymous beings. They come together in the context of specific environments and with specific purposes. Their interactions involve behaviors associated with defined statuses and particular roles. These statuses and roles help to pattern our social interactions and provide predictability”* [198]. This suggests that the roles people assumes while interacting could be essential for the understanding of social interactions. For example, in dysfunctional teams these roles are analysed by meeting facilitators who help the group to stay focused and mediate the discussion by providing feedback on individual

or group behaviour [165]. Other researchers have also found that changes in the distance and body orientation between interacting people might indicate different phases of the contact, for example beginnings, endings and changes in topic [60]. Previous works have used microphones and cameras to detect people's roles and meeting phases from non-verbal behaviours (e.g., speech characteristics or body and head movements) with the objective of realising a virtual facilitator that could provide feedback about the meeting automatically [165, 221, 20, 58].

With this chapter we propose to employ the measure of interpersonal distance and angle of interaction gathered by Protractor to recognise roles people take while working together in small groups and the different phases of the meeting. We conducted experiments in a controlled setting with 16 groups of 4 users each. We assigned participants a creative problem-solving task widely used for assessing teams' creative potential [217]. The intention was to simulate a team working together within an organizational environment (e.g. in new product development). The goal of this deployment was to provide an initial exploration of the possibilities offered by Protractor in the understanding of complex, and often abstract processes, comprising multiple, interrelated sets of human actions in an organizational environment. In particular we explored the possibility of predicting, using only proxemics information (i.e., angle and distance between pairs of participants), two aspects of team dynamics: (1) **task role**: the verbal role assumed by each participant, and (2) **task timeline**: the different phases of the creative task. We showed how Protractor is capable of supporting organizational science studies by providing objective data that could be used to predict the role a person assumes during a creative task as defined by her verbal communication with 84% accuracy, and the procedural phases of the task with 93% accuracy.

In this chapter we show how the spatial arrangements in small groups, similarly to other non-verbal cues that have been analysed in the past, reflect the role and attitude of participants (as defined by their verbal communication) and the various sections that characterise the encounter.

Chapter Outline. Section 6.2 describes the deployment setup reporting the collaboration task we used for the study, the demographics of the participants and the data sources used for the classification. The description of how we structure the dataset for the classification tasks is in Section 6.3. In Section 6.4 we detail the approach we adopted for training and evaluation of the classification methods we used and in Section 6.5 we present the results. We then discuss the implications and limitations of our results and approach in Section 6.6 and conclude the chapter in Section 6.7.

6.2 Study setup

In order to study team dynamics while collaborating toward a common goal, we employed an existing creativity task, “The Marshmallow Challenge” [216], which was designed to help teams experience fundamental collaboration dynamics in creative problem solving. The challenge consists in building the tallest free-standing structure from 20 sticks of spaghetti, one yard of tape, one yard of string, and one marshmallow which, most importantly, had to be supported by the free-standing spaghetti construction. Each group of 4 participants had 18 minutes to complete the structure. At the end of the allocated time the height of the structure was measured.

6.2.1 Participants

We recruited participants from the Computer Laboratory at the University of Cambridge (U.K.), and the Department of Computer Science at Dartmouth College (U.S.). We formed 16 teams of four participants ($n = 64$)¹. 90% of the participants were aged 18 to 29 years old and 79% of our participants were men. The participants were compensated by entering a raffle for an Amazon voucher (6 vouchers available valued £50 or \$50 each).

The teams were welcomed in the experiment room and then given the instructions and rules for the building process. All participants wore the Protractor and were video recorded throughout the entire building process.

6.2.2 Data sources

Three main sources of data were collected for this study: (1) angle and distance measurements recorded by Protractor for every pair of participants at approximately five-second resolution; (2) team members’ verbal interactions (i.e., their individual task role described by their verbal exchange); (3) the timeline of the teams’ building process, which we coded from the video recordings.

Task roles and task timeline were manually coded by one subject matter expert coder. The decision to employ a single trusted subject matter expert was deemed appropriate and safe given the novelty and preliminary nature of the study. Following, we describe the collected data with more details.

Angle and Distance. We gathered the raw angle values as detected by Protractor ranging from -90° to 90° for every A-B dyad approximately at five-second intervals (where 0° represents participants A and B facing each other, the negative interval indicates B to the left of A, and the positive interval indicates B to the right of A). We rescaled the raw

¹Ethical approvals have been obtained from both local institutions before the study.

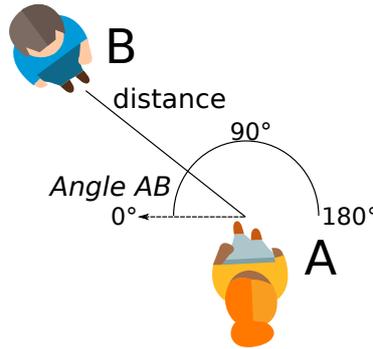


Figure 6.1: Illustration of angle and distance used in the classification.

data to the 0° to 180° interval to make the results more interpretable whilst preserving the left-right dichotomy. Thus, in the rescaled dataset, 90° represents participants A and B facing each other, 0° to 90° interval indicates B to the left of A, and 90° to 180° interval indicates B to the right of A.

Distance is measured in centimetres and captures the distance between dyads of participants and, just as the angle data, the resolution is approximately at five-second recorded for the duration of the experiment (Fig 6.1).

Task Roles. We coded the team members' verbal behaviours during the building process by using the Advanced Interaction Analysis (*act4teams*) video coding scheme from Lehmann-Willenbrock et al. [122]. This coding scheme has been employed to label verbal behaviours in video recordings of team interactions. The scheme covers four main categories² of statements, namely:

- Problem-focused (labelled *ProblFcs*): identifies communication directly related to the topics of the meeting. Problem-focused communication includes discussions about the problems, formulation of ideas and solutions and their analysis. This category includes statements of the following kind: identifying a problem, connections with problems, defining the objective, identifying a solution, describing a solution, problems with a solution, arguing for a solution, organizational knowledge, etc..
- Procedural: this kind of communication describes statements related to the structure and organisation of the discussion. Positive and Negative statements exists. Positive statements are the ones that are beneficial for the organisation of the discussion while the Negatives have a negative influence and lead to a loss of structure and loss of thought. In our dataset we labelled the two as *ProcedPos* and *ProcedNeg* respectively.
- Socio-emotional: captures the social relationships inside teams. Also in this case there are Positive (labelled *SocEmPos*) and Negative behaviours (labelled *SocEmNeg*). The Positive category includes statements used to show solidarity and support, release

²For a more detailed description of the four categories refer to the works from Lehmann-Willenbrock et al. [122] and Kauffeld et al. [101].

tension, or show agreement. On the other hand, Negative behaviours comprise self-promotion, criticizing, offending or interrupting others and having side conversations which demonstrate disengagement.

- **Action-oriented:** describes statements aimed at improving the team's work by showing willingness to take action. Positive statements (labelled *ActOrtPos*) show proactive behaviour, willingness to take responsibility or planning of concrete actions. By contrast, negative statements (labelled *ActOrtNeg*) manifest no interest in change, complaining, lack of initiative, seeking someone to blame or denying responsibility.

These categories describe solution-oriented behavior [102] and have been shown to help teams become aware of their dynamics in meetings and affect team and organizational success [10, 101, 123]. We used these labels (7 in total) to code our participants' individual verbal statements in the building process at five-second increments. We nominated starting points for each verbal code and assigned these codes to all subsequent time increments; as a new verbal behavior occurred, the new code replaced the previous code in subsequent time increments and so on.

Task Timeline. In the original design of the challenge [216], the building phases described were orient, plan, build, and ta-da or oh-no. Empirically, we adapted the phases to collect a more fine-grained taxonomy of the teams' building processes and we labelled our data with:

1. *Intro* for the introduction time before the actual discussion;
2. *Materials and logistics* for the discussions about the tools at hand, planning the building, and starting to put together pieces of structure or checking their strength and stability;
3. *Building levels one, two, three, and four* for assembling the materials and stacking them into the final structure;
4. *Consolidating level one* for reinforcing the base of the structure to ensure the structure is freestanding;
5. *Marshmallow on top* for the attempts to place the marshmallow on top of the structure to test the strength of the construction or to finalize it;
6. *Outro* for the time they finished building to the end of the allocated eighteen minutes.

Once we identified the start and end points of each phase we annotated the angle and distance data within a phase interval with the relative label.

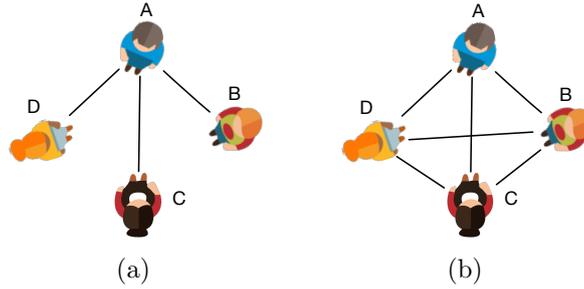


Figure 6.2: Features selection. (a): for predicting the role of each participant we are interested in looking at the relative orientation and distance between that participant and all other participants. (b): for the prediction of Task Timeline we wanted to capture the spatial arrangement of all members of the group at once, thus we considered the angles and distances between each possible pair.

6.3 Dataset structure and features

Our objective was to predict the task roles and task timeline labels, coded in the video, using the dyadic angle and distance measurements as input. We treated the input data as not formally sequence or temporally dependent, but rather by using all five second increments across all groups as separate instances for classification. We did this to examine the informational value of the angle and distance data collected by Protractor at the most basic level.

From Protractor we extracted the pairwise distance and angle measures between each pair of participants across all 16 groups in our study. The dynamics we aim to classify automatically (i.e. task role and task timeline) have different characteristics which have to be considered when selecting the features for their detection. In previous work on social contact monitoring, non-verbal features have been classified in: (1) *individual* which are derived only from the behaviour of each participant, without considering other people; and (2) *interpersonal* which instead are computed from the behaviour of the participants with respect to each other [180]. In this work we rely exclusively on interpersonal features because Protractor provides only relative distance and orientation between pairs of tags rather than absolute position and orientation in the environment. Protractor could be used to extract individual features by deploying static anchors in the environment allowing the detection of position and orientation of people with respect to fixed points rather than relative to other participants. However, this would undermine the benefit of having a wearable device which does not require any instrumentation of the environment. Data generated from the inertial measurement unit (accelerometer and gyroscope) could be adopted to derive individual features related to the participant movement. We leave the exploration of this possibility for future work. In the following, we analyse the specific characteristics of the two tasks we want to classify and describe how we chose the features for each of them.

Table 6.1: Class distribution of all (11454) instances for the classification of individual’s instant task role.

Label	Count
ProcedPos	5351
ProblFcs	3144
ActOrtPos	1591
SocEmPos	1138
SocEmNeg	110
ProcedNeg	75
ActOrtNeg	44

6.3.1 Task role

The Task Role classes represent the nature of the verbal communication that was taking place among participants during the construction of the structure. Task roles are defined at the individual level: each person assumes a role based on her verbal behaviour. With our deployment we are interested in exploring the potential link between spatial arrangements (i.e. variation in interpersonal distance and angle of interaction) and roles assumed by each individual. The intuition is that changes in roles could be reflected in variations in relative positions between people. For example, hostile or negative behaviours (like negative socio-emotional) could bring people to prefer longer distances and avoid direct confrontation, while collaborative actions (such as positive action-oriented) could pull people together. Based on this intuition, in order to recognise a participant’s role using spatial arrangements we needed to capture the participant’s position relative to all the others. To do that we use as features her angle towards each other group member (3 features), the angle of each other member toward her (3 features), and the distance between her and each other member (3 features). This gives a total of 9 features used to predict the instant role of a person.

In more detail, referring to Figure 6.2(a), for each participant A (with other participants in the group being B , C and D) each instance in the dataset has the following fields:

angle.AB, *angle.AC*, *angle.AD*, *angle.BA*, *angle.CA*, *angle.DA*, *distance.AB*, *distance.AC*, *distance.AD*, *role*

where, *angle.XY* identifies the angle between participant X and Y as captured by the device worn by X , *distance.XY* is the symmetric distance between X and Y and *role* is one of the Task Roles coded from the verbal behaviour as described in Section 6.2.2. We aggregate all instances for all participants across all groups leading to a total of 11454 instances³. The class distribution of these instances are listed in Table 6.1.

³The total number of instances theoretically is # of people per group \times # of groups \times # of 5-second intervals in 18 minutes = $4 \times 16 \times 216 = 13824$. We obtain a lower number of instances because some groups finished the challenges before the 18-minute mark and participants stopped working on the structure and interacting.

Table 6.2: Class distribution of all (3231) instances for the classification of task timeline phases.

Label	Count
Building level one	908
Consolidating level one	701
Materials and logistics	700
Building level two	539
Marshmallow on top	183
Building level three	106
Outro	53
Intro	28
Building level four	13

6.3.2 Task timeline

The Task Timeline classes represent stages in the building process followed by the participants. Contrary to Task Roles, the building phases during the challenge (i.e. Task Timeline) are not tied to each person individually, instead they are a characteristic of the group as a whole.

To predict the stage in the building process, we examined the configurations (i.e., relative orientations and distances) of all participants in the group, based on the rationale that these configurations vary across different stages of the building process. As examples, in the *intro* phase, participants might have longer distances from each other since they are not yet actively working; in the *materials and logistics* phase, they might come closer to one another and form sub-groups (pairs of people with short distance and angle close to 90°) while they get familiar with the materials or prototype a basic structure. To capture these configurations we concatenate angles and distances between all possible pairs of participants as the feature vector: this gives the current configuration of the entire group. Based on this consideration, identifying the members of a group with A, B, C and D, (Figure 6.2(b)) each instance in the dataset contains the fields:

angle.AB, angle.AC, angle.AD, angle.BA, angle.BC, angle.BD, angle.CA, angle.CB, angle.CD, angle.DA, angle.DB, angle.DC, distance.AB, distance.AC, distance.AD, distance.BC, distance.BD, distance.CD, task.timeline

where *angle.XY* and *distance.XY* are defined as in the previous section and *task.timeline* is one of the building phases introduced in Section 6.2.2. The resulting dataset contains 3231 instances, given the five-second resolution of Protractor data⁴. Table 6.2 lists the class distribution of these instances.

⁴The total number of instances theoretically is # of groups \times # of 5-second intervals in 18 minutes = $16 \times 216 = 3456$. We obtain a lower number of instances because some groups finished the challenges before the 18-minute mark and participants stopped working on the structure and interacting.

6.4 Classification of video coded labels

For the classification of Task Roles and Task Timelines we selected classifiers from the WEKA machine learning library (version 3.7.13) [131] and applied them as multi-class classifiers using a one-versus-all approach. Previous works in the area of automatic detection of non-verbal cues used several classifiers on features extracted mainly from audio and video data (e.g., Conditional Random Fields [182], C4.5 Decision Trees [20], SVM [221]). In this work instead we worked with a different kind of data therefore we opted for the Random Forest classifier because it has been found to perform well on many different datasets [67] and also for its simplicity and ability to minimize overfitting [31]. We run the classifier with 50 and 100 trees, referred to as RF50 and RF100 respectively henceforth⁵. For both classification tasks we followed these steps:

1. The input features were the angles of each dyad A-B (Angle AB and Angle BA) as well as the distance between A and B as described in Section 6.3. The data were then normalized to [0, 1] interval.
2. Next, we partitioned the data into 70/30 stratified splits for classifier training and testing. The split has been performed in a dyad-independent way, ensuring that data from the same dyad would not be included in the training and testing set. Model performance was assessed with reference to (a) sound precision, recall and F-measure scores across classification targets, (b) reasonable balance between these scores across targets, and (c) good overall model accuracy.
 - (a) For the classification of Task Roles we applied the SMOTE oversampling procedure [45] to create synthetic examples for the minority classes (*ActOrgNeg*, *ProcedNeg* and *SocEmNeg*) with the objective of balancing precision and recall across classes.
3. We further assessed model performance using 10-fold cross-validation with stratified sampling (without any oversampling).

To gain more insights into the role of the features, we also ran a simple forward feature selection loop using the same multi-class Random Forest Classifier with 50 trees (RF50 henceforth). Table 6.3 lists the results. For the Task Role classes, “distance.AB” contributes the most to the overall model accuracy (39.45%), followed by varied contributions from the angle features (the highest being “angle.AD” with 10.62%, and “angle.AC” with 11.32%), before peaking at eight of nine features (79.46%).

For the Task Timeline classes, “angle.BD” contributes 28.04% to the overall accuracy of the result, followed by gains of 15.88% (“angle.CD”), 16.60% (“angle.AD”), and 11.44% (“distance.BC”). Thereafter, gains are comparatively modest, peaking at fifteen of eighteen features (91.44%). The intention here is not to show a generalizable pattern of feature

⁵The classifier parameters used in this work are reported in Appendix E.

Table 6.3: Forward features selection results run with multi-class Random Forest classifier with 50 trees (RF50). The procedure involves adding features iteratively to the model with the goal of maximizing overall accuracy. Overall accuracy at any given feature shows the total model accuracy achieved up until that feature. Change shows the increase / decrease in model accuracy from one feature to the next.

Target	Order	Feature	Accuracy	Change	Target	Order	Feature	Accuracy	Change
Role	9	distance.a.d	79.40%	-0.06%	Timeline	18	distance.a.c	88.97%	-1.44%
	8	distance.a.c	79.46%	1.80%		17	distance.c.d	90.41%	-0.41%
	7	angle.d.a	77.65%	2.71%		16	distance.b.d	90.82%	-0.62%
	6	angle.a.b	74.95%	3.72%		15	angle.d.a	91.44%	1.86%
	5	angle.c.a	71.22%	5.62%		14	distance.a.d	89.59%	0.00%
	4	angle.a.c	65.61%	11.32%		13	angle.d.c	89.59%	-0.82%
	3	angle.a.d	54.29%	10.62%		12	angle.b.c	90.41%	1.24%
	2	angle.b.a	43.67%	4.22%		11	angle.a.c	89.18%	0.62%
	1	distance.a.b	39.45%	39.45%		10	angle.c.a	88.56%	1.13%
				9	angle.d.b	87.42%	2.89%		
				8	angle.b.a	84.54%	1.24%		
				7	distance.a.b	83.30%	2.27%		
				6	angle.c.b	81.03%	5.15%		
				5	angle.a.b	75.88%	3.92%		
				4	distance.b.c	71.96%	11.44%		
				3	angle.a.d	60.52%	16.60%		
				2	angle.c.d	43.92%	15.88%		
				1	angle.b.d	28.04%	28.04%		

contributions, but rather to shed light on the role of the distance and angle features together as markers of focal task related behaviours and interactions.

We now move to the analysis of the results of classifying the Task Roles and the Task Timeline labels using data collected with Protractor.

6.5 Results

We summarize the overall model accuracy results for RF50 and RF100 in Table 6.4. Overall accuracy results for cross-fold validation with RF100 are summarized in Table 6.5. Classes are ordered by F-measure. Next, we discuss the results for each classification task.

6.5.1 Task role

As shown in Table 6.4, the overall accuracy of classifying Task Role was 79.3% (RF50) and 80.7% (RF100) respectively. Cross-fold validation with RF100 achieved an overall accuracy of 84.9% (Table 6.5). Recall and precision are strong among all classes with ProcePos scoring the highest. SocEmNeg, ActOrtNeg, and ProceNeg score lower because they are minority classes with fewer instances, where SocEmNeg has 110 instances, ActOrtNeg has 44, and ProceNeg has 75 instances (Table 6.1). However, our use of SMOTE [45] to oversample minority classes has helped to improve the recall of minority classes, in

Table 6.4: Overall accuracy of predicting participants’ instant task role and groups’ task timeline. We show precision, recall, and F-measure scores, as well as the overall accuracy when using Random Forest with 50 trees (RF50) and 100 trees (RF100) respectively.

RF50 Multi-class Classifier					RF100 Multi-class Classifier				
Task Role	Recall	Precision	F1	Accuracy	Task Role	Recall	Precision	F1	Accuracy
ProcedPos	0.84	0.84	0.84	79.3%	ProcedPos	0.85	0.84	0.85	80.7%
SocEmPos	0.72	0.87	0.79		SocEmPos	0.74	0.87	0.80	
ProblFcs	0.81	0.72	0.76		ProblFcs	0.82	0.75	0.78	
ActOrtPos	0.69	0.78	0.73		ActOrtPos	0.70	0.81	0.75	
ProcedNeg	0.65	0.65	0.65		ActOrtNeg	0.64	0.75	0.69	
SocEmNeg	0.61	0.67	0.64		ProcedNeg	0.61	0.67	0.64	
ActOrtNeg	0.43	0.67	0.52	SocEmNeg	0.58	0.68	0.62		
Task Timeline	Recall	Precision	F1	Accuracy	Task Timeline	Recall	Precision	F1	Accuracy
Materialsandlogistics	0.96	0.94	0.95	91.1%	Buildinglevelfour	1.00	1.00	1.00	91.9%
Buildinglevelthree	0.91	1.00	0.95		Materialsandlogistics	0.96	0.94	0.95	
Buildinglevelone	0.93	0.90	0.92		Buildinglevelthree	0.88	1.00	0.93	
Buildingleveltwo	0.88	0.93	0.91		Buildinglevelone	0.94	0.91	0.92	
Consolidatinglevelone	0.90	0.88	0.89		Consolidatinglevelone	0.93	0.90	0.92	
Intro	0.78	1.00	0.88		Buildingleveltwo	0.88	0.93	0.91	
Buildinglevelfour	0.75	1.00	0.86	Intro	0.78	1.00	0.88		
Marshontop	0.84	0.87	0.85		Marshontop	0.84	0.89	0.86	
Outro	0.75	0.86	0.80		Outro	0.63	0.83	0.71	

Table 6.5: Model accuracy (recall, precision, and F-measure) of predicting participants’ instant task role and groups’ timeline using Random Forest with 100 trees (RF100) and 10-fold cross validation.

Task Role	Recall	Precision	F1	Accuracy	Task Timeline	Recall	Precision	F1	Accuracy
ProcedPos	0.96	0.82	0.88	84.9%	Buildinglevelfour	0.92	1.00	0.96	93.2%
ProblFcs	0.80	0.87	0.83		Materialsandlogistics	0.96	0.94	0.95	
SocEmPos	0.73	0.93	0.82		Buildinglevelone	0.96	0.93	0.95	
ActOrtPos	0.72	0.89	0.79		Consolidatinglevelone	0.96	0.92	0.94	
ProcedNeg	0.70	0.91	0.79		Buildingleveltwo	0.92	0.94	0.93	
SocEmNeg	0.38	0.98	0.55		Buildinglevelthree	0.88	0.97	0.92	
ActOrtNeg	0.27	0.92	0.42	Outro	0.87	0.96	0.91		
					Intro	0.75	0.91	0.82	
					Marshontop	0.72	0.89	0.80	

comparison to the result without any class balancing. This follows our objective of reaching a balance between the scores (precision, recall and F-measure) across classes.

As we further analyse the confusion matrix in Table 6.6 using RF100, we can see that despite our oversampling of the minority classes, the classifier is still slightly biased towards the majority classes (*ProblFcs* and the *positive* ones), resulting in more predictions of these classes. However, we did not want to oversample the minority classes further, due to the limited number of instances available. At this stage we considered satisfiable a precision of around 70% (or more) for all classes with a loss in recall for the less frequent *negative* classes (Table 6.4). Overall, our results show that task-role classes are quite distinguishable, meaning that the angle and distance data parallel verbal behavior.

Table 6.6: Confusion matrix of classifying the instant task role of each participant using RF100 while training on 70% of the data (model results in Table 6.4). We concatenate a participant’s instant angle and distance data to all other group members as the feature vector and predict her instant role. The rows represent the ground truth labels while the columns represent the labels predicted by the classifier. The matrix contains 3440 testing instances (30% of the total data).

	ProblFcs	ProcedPos	SocEmPos	ActOrtPos	ActOrtNeg	SocEmNeg	ProcedNeg
ProblFcs	776	123	8	31	1	2	3
ProcedPos	167	1371	23	36	1	7	1
SocEmPos	40	36	252	12	0	0	2
ActOrtPos	45	90	7	334	1	0	1
ActOrtNeg	1	3	0	1	9	0	0
SocEmNeg	2	11	0	1	0	19	0
ProcedNeg	7	2	0	0	0	0	14

Table 6.7: Confusion matrix of classifying timeline states, using RF100 while training on 70% of the data (model results in Table 6.4). We aggregate the dyadic angles and distances of all members in a group as group-level features to predict the current stage of the building process. The rows represent the ground truth labels while the columns represent the labels predicted by the classifier. The matrix contains 973 testing instances (30% of the total data).

	Intro	Materials and logistics	Building level one	Building level two	Marsh on top	Consolidating level one	Outro	Building level three	Building level four
Intro	7	2	0	0	0	0	0	0	0
Materials and logistics	0	203	6	2	0	0	0	0	0
Building level one	0	9	257	3	1	3	0	0	0
Building level two	0	0	7	143	0	12	0	0	0
Marsh on top	0	1	3	3	46	2	0	0	0
Consolidating level one	0	0	6	2	5	196	2	0	0
Outro	0	0	1	1	0	4	10	0	0
Building level three	0	0	4	0	0	0	0	28	0
Building level four	0	0	0	0	0	0	0	0	4

6.5.2 Task timeline

The overall accuracy of classifying Task Timeline is summarised in Table 6.4. We observed that the majority of tasks in the timeline can be distinguished well, achieving 91.1% (RF50) and 91.9% (RF100) overall (Table 6.4). Cross-fold validation with RF100 achieved 93.2% accuracy (Table 6.5). Recall and precision were strong for most classes, with *Intro* and *Marshontop* having a slightly lower recall (Table 6.5).

From the confusion matrix in Table 6.7, we observed that the classifier exhibit inter-class misclassification mainly between *building level one*, *building level two*, *materials and logistics*, *marshmallow on top* and *consolidating level one* (central section of the matrix). One possible reason for this is that the labels we selected might be too fine-grained and represent the same underlining action (e.g. *working on the structure*). A different coding scheme might account for these similarities and aggregate some of the labels we employed

for this work.

6.6 Discussion

6.6.1 Discussion of classification results

Overall, the results show that we can adequately distinguish between Task Roles and Task Timeline phases. This has been achieved with just angular and distance data showing promising preliminary results for Protractor.

Focusing on the results for Random Forest with 100 trees, which produced the best results (Table 6.4), we conclude that the Task Role class labels (created by coding the participants' utterances) are separable with an overall accuracy of 81%. This means that there is a link between the angle and distance of the participants and their verbal behavior. The classification of Task Timeline labels presents even higher overall accuracy (92%) with clearly distinguishable labels. In this case the result is more intuitive given the natural tendency of having different spatial arrangements given the current task (e.g., getting closer to one another when working on the structure). We observed such behaviour also during the coding of the video recordings of the challenges.

Delving deeper in the results, firstly, we observe that for the classification of Task Roles we were able to improve the balancing between precision and recall across the target classes by oversampling the minority classes *ActOrgNeg*, *ProcedNeg*, *SocEmNeg*, with SMOTE [45]. Without the additional synthetic samples, as shown in Table 6.5 for 10-fold cross validation, the recall of the classes *SocEmNeg* and *ActOrtNeg* is very low, meaning that the classifier is biased towards the classes with a larger number of samples. With the oversampling we lose slightly in terms of overall accuracy (81% against 85% without oversampling) but we obtain a less biased classifier. Clearly, the choice of optimising precision or recall depends on the particular context considered for the deployment of the classifiers. This consideration is beyond the scope of our work.

By contrast, for the classification of the Task Timeline phases we did not apply any oversampling technique and the classifier was still able to correctly distinguish the classes, even the ones with a limited number of samples. However, we observe that most of the misclassification is among classes that represent the same fundamental action, like for example *building level one*, *building level two* or *consolidating level one*. This is an indication that, despite the classes are well distinguishable, our coding scheme might be too granular and specific.

6.6.2 Implications for organizational science

The main methodological contribution of our study to the field of organizational science is in the combination of objective sensor-data and subjective assessments. We thus address recent calls for research in micro-meso level behavioural processes that have overwhelmingly been researched using retrospective self-reports [108]; known for being bias-prone and inaccurate, social scientists call for the supplementation of these tools with unobtrusive, data dense, and continuous measurement systems. By employing Protractor we were able to observe nuances in behavioural changes never captured before with a lightweight wearable system which does not require any building instrumentation (e.g., motion capture systems). We made and validated the conceptual link between fluid spatial arrangements (described by the variation in angle and distance between team members) and the communication's content.

Finally, Protractor would allow us to study the impact of culture on proxemics behaviours. Hofstede et al. have shown that cultural backgrounds can impact the way people think, feel, and act while working with others [70]. Cultural differences and personal preferences could alter the way people approach others in terms of interpersonal distance and relative orientation. Some cultures for example tend to have closer distances when interacting with strangers than other cultures [81, 194]. These differences could have a significant effect in today's highly international workplaces and would need to be factored in when studying non-verbal cues. We leave these aspects for future work.

6.6.3 Limitations

Generalisation. This study considered a relatively small number of participants ($n = 64$) resulting in a limited dataset. Additionally, our participants have similar backgrounds and occupations (e.g., students, PostDoc and researchers). A larger number of participants with more variability in their demographic could help refine and generalise our results more widely and create a more extensive dataset.

Experimental Setting. The analysis of human behaviours in controlled settings, although commonly adopted in previous work, might lead to different conclusions compared to considering real contexts, such as actual corporate meetings. Even if we considered a more natural setting compared to previous work where the roles were pre-assigned (e.g., radio talk shows), our study could benefit from the inclusion of data recorded in real-world scenarios. Additionally, our study required the participants to stand and move around a table, possibly making the individual non-verbal behaviours more visible. A different setting where people can not move so freely (e.g. sitting at a desk) might make the prediction of roles and meeting states more difficult.

Comparison Validity. The validity of comparisons with previous work might be affected by factors including the demographics and background of participants and the kind of

task they had to perform. Similarly, we adopted a coding scheme which was slightly different than previous work, choosing more granular labels to better capture different dynamics within the teams. This is another factor which might limit the validity of the comparison between our results and the ones reported by other researchers.

Video Coding Approach. Non-verbal behaviours are subtle, difficult to interpret and depend on many factors. This makes the coding of recorded videos a delicate and difficult task. We acknowledge the limitation of having a single coder to label the videos of the tasks. However, given the novelty and preliminary nature of the study we deemed it appropriate and safe to employ only one expert coder.

6.7 Conclusion

We have demonstrated how data about relative orientation and interpersonal distance could be employed for the analysis of behaviours within small teams and how non-verbal cues relate to the meeting's verbal content. The data was used to automatically predict the roles a person assumes during a collaborative task and the various phases of the task with 84% accuracy in the first case and 93% accuracy in the second one.

This contributes to the body of research that explores the automatic recognition of roles and meeting states by analysing, for the first time, people's spatial arrangements with objective and accurate measurements. The availability of rich information about meeting dynamics could facilitate practitioners, such as trainers or team facilitators to better understand teams' dynamics and intervene to support dysfunctional teams. Additionally, our device could be employed as a building block of a larger automatic system to provide such feedback.

Chapter 7

Conclusion and future work

At the beginning of this dissertation we highlighted the importance of studying social interactions and how the availability of accurate data about their dynamics could potentially affect many contexts. We discussed the limitations of traditional methods and current technology and we directed the focus of our work towards the use of wearable systems as platforms to ensure accuracy and ease of deployment. Consequently, we pursued the following thesis: *to support diverse applications relying on social interaction detection we need to consider and devise wearable technologies capable of accurately monitor different dimensions of social contacts and evaluate their benefits for the understanding of human behaviour.*

The previous chapters described and discussed the results of the research conducted to substantiate this thesis. In this chapter we will summarise the contributions of this dissertation and consider future directions.

7.1 Summary of contributions

In this section we will reflect on the research questions introduced in Chapter 1 and summarise the major contributions that support the thesis of this dissertation.

[Research Question 1] How can we take advantage of radio communication interfaces embedded in many commercial wearable devices (i.e., Bluetooth Low Energy) for the efficient detection of social contacts in very dynamic environments?

[Contribution 1] In Chapter 3 we analysed the potential of Bluetooth Low Energy (BLE) for proximity detection. We provided a comprehensive analysis of all the protocol parameters, considering their impact on detection accuracy and power consumption. We focused on two common wearable platforms (Android Wear and Tizen Wearable) showing their strengths and limitations. We presented a prototype platform which helped us in experi-

menting more freely with the BLE protocol. Additionally, using the platform we validated the use of BLE to sense people's proximity in a working environment, by deploying it to 25 employees of a very dynamic company. We found that despite the limitations of commercial devices they can be used to detect proximity with high accuracy. As a result of our analysis and deployment we provided guidance for manufacturers and applications developers on the flexibility and limitations of commercial platforms.

[Research Question 2] How can we leverage data gathered automatically with wearable devices to analyse team dynamics and the strength of employees' interpersonal ties in relation to space usage and organisational hierarchy?

[Contribution 2] In Chapter 4 we explored how Activity-Based Working principles affect office usage and people dynamics using the data collected during the deployment of the BLE platform. We studied how the two core ABW principles (absence of allocated desks and availability of diverse spaces) might be responsible for promoting interactions across teams and among lower levels of the organisational hierarchy. However, we also found that in terms of mobility inside the office, the ABW principles might not have come to full fruition. It seems that employees explore various locations for short periods but use only few settings for work-related tasks. Additionally, we discovered that social- and team-related clusters might play a significant role in desk selection. Through the comparison with other data collected in offices that do not apply ABW principles, we observed how the design of the office in accordance with those principles might be responsible for more rapid dynamics inside the office.

[Research Question 3] How can we devise a wearable sensing technology suitable for the fine granularity detection and analysis of non-verbal cues during social interactions?

[Contribution 3] In Chapter 5 we introduced a novel wearable device, Protractor, capable of accurately detecting angle and distance of interaction. We showed how we employed near-infrared light to infer relative interaction angle with 2.2° mean error and 2.3cm – 4.9cm mean error for the distance measurement. We reported how we coped with temporary blockage of the light channel and ensured a robust operation of the device in various conditions by exploiting inertial sensors with the use of sensor fusion techniques. We validated the device with extensive real-world experiments and in Chapter 6 we provided insights on how it could be used to study complex processes in an organisational setting. We explored the possibility of using angle and distance data between individuals engaged in a creative task to predict the role they assume towards others and the instant task phase they are currently in.

In summary, this dissertation provides evidence that wearable technologies represent an efficient and convenient platform to collect data and study social interactions. We have demonstrated how off-the-shelf wearables could be adopted to accurately study behaviour in very dynamic environments and generate insights on how space design and working styles, for example, could affect social dynamics. Moreover, we showed how technologies not typically designed for sensing purposes (e.g. infrared light transceivers) could be effi-

ciently employed in small, wearable devices to detect behavioural cues which have so far been studied only with invasive and costly methods. We expect the contributions of this dissertation to not be limited only to the results we presented. We believe, in fact, that our data collection platforms could be used in different contexts, possibly to study other phenomena, opening the door to new research on human behaviour sensing and analysis.

7.2 Future directions

The research presented in this dissertation could be expanded in several ways. In this section we overview some possible future directions.

System enhancements and optimisations. One important interaction modality which has not been considered in this dissertation is speech. While the two platforms we developed are able to accurately detect and monitor fine grained proximity contacts and their non-verbal cues, they are not capable of detecting if there was a meaningful exchange between the participants. Different sensing technologies could be researched to unobtrusively monitor conversations and understand what kind of novel insights and findings could be generated by the availability of this new source of data.

Several performance enhancements could be developed for Protractor, introduced in Chapter 5. These include lower-range solutions to improve accuracy in very bright settings and to better handle reflections, alternative NIR beacon designs to allow beacon decoding upon collisions, and adaptive duty cycling based on sliding-window average. It could also be interesting to explore the feasibility of generalizing our approach to other wireless media such as ultrasound.

Real-time feedback for social intervention. One aspect which was not the focus of this dissertation and which has not been investigated extensively in literature is the possibility of providing prompt feedback to the user while a social interaction is taking place. The offline analysis of large amounts of data is used to inform people about their habits, allowing them to change behaviour accordingly. This has been a common practice in social science where the results of a study eventually change people's behaviour a long time after the study. With technology, however, the intervention could be shifted to a previous moment, when the user needs it the most, during the data collection phase. In particular, for social interactions the need to change the behaviour quickly is important because, as highlighted in Chapters 5 and 6, our non-verbal behaviour could impact the outcome of the interaction. Therefore, the possibility to be discreetly advised in order to assume the ideal behaviour has the potential to improve social interactions. For this reason methods to provide real-time feedback to the user about her non-verbal communication behaviours should be developed. Important aspects such as the best form of feedback and the best moment in time to deliver it are crucial in this area.

Generalisation of findings. In Chapter 4 we have successfully used data automatically

collected to study social dynamics in an office. However, the actual reasons behind the patterns we observed remain unclear. Further investigation and additional deployments would be required to closely analyse the relationship between social contact and working styles, considering also diverse organisational structures with the objective of generalising the findings more widely. Similarly, it would be interesting to study how dynamics like team formation and productivity evolve over time with longitudinal deployments. Likewise, the results presented in Chapter 6 would benefit from longer-term user studies in real contexts rather than in controlled settings. In this context of social behaviour sensing and analysis an aspect that should not be overlooked is the privacy concerns of individuals. More research is needed in order to understand what could be the privacy concerns raised by the monitoring technology and develop ways to mitigate them.

Appendices

Appendix A

Decision Trees parameters

Table A.1 reports the parameters of the Decision Tree (C4.5) models used in Chapter 3, Section 3.4.7. The implementation of the algorithm has been taken from Weka version 3.7.13 [131]¹.

Table A.1: Decision Trees (C4.5) parameters used for the proximity detection models.

Parameter	Value
Batch Size	100
Binary Splits	False
Collapse Tree	True
Confidence Factor	0.25
Debug	False
Do not check capabilities	False
Do not make split point actual value	False
Min num obj	2
Num decimal places	2
Num folds	3
Reduced error pruning	False
Save instance data	False
Seed	1
Subtree raising	True
Unpruned	False
Use Laplace	False
Use MDL correction	True

¹The documentation regarding this classifier can be found at this link:
<http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/J48.html>

Appendix B

Proximity detection results with original dataset and with under-sampling of majority class

In this appendix we present the proximity detection results we obtain with the dataset collected during the deployment of our wearable prototype platform (Chapter 3) when we do not use SMOTE to over-sample the minority class. For these analysis we show only results achieved with the data collected with our wearable prototype, without emulating data collected by commercial devices, for ease of exposition. The original data collected by our devices is the best data we have from the deployment and emulating commercial devices would only show lower performance in general but the same trends.

We first trained our Decision Tree classifier using the dataset without any over-sampling or down-sampling using a 10-fold cross validation scheme. In this case we observe (Table B.1) that the classifier learns to predict only the majority class (*Non-proximity* label) therefore scoring a nearly perfect precision and recall for that class but a zero precision and recall for the minority class (*Proximity* label). The dataset in fact contains much more examples of the *Non-proximity* class and therefore it learns to always predict that class.

As second experiment we down-sampled the majority class randomly until the number of instances for the two classes was balanced. Table B.2 shows the results. We notice that in this case the performance of the model is much better, with an overall good precision and recall for both classes. However, this model did not reach the same level of performance of the one where we over-sampled the minority class presented in Section 3.4.7. For convenience, in Table B.3 we compare the F1 Measure of the two classes for these two models (with down-sample of the majority class and with over-sample of the minority class) where the difference is clearly visible.

Table B.1: Results obtained when no over-sampling of the minority class has been applied to the data collected with our wearable platform for different aggregation windows. Precision, Recall and F1 measure are reported for the two classes *Non-proximity* and *Proximity*.

Window Size (s)	Precision		Recall		F1 Measure	
	Non Proximity	Proximity	Non Proximity	Proximity	Non Proximity	Proximity
1	0.99	0	1	0	0.99	0
5	0.98	0	1	0	0.99	0
10	0.98	0	1	0	0.99	0
20	0.98	0	1	0	0.99	0
30	0.98	0	1	0	0.99	0
40	0.99	0.3	1	0.062	0.99	0.103
50	0.98	0.2	1	0.007	0.99	0.014
60	0.98	0.3	1	0.008	0.99	0.016

Table B.2: Results obtained when the majority class has been down-sampled to balance the dataset for different aggregation windows. Precision, Recall and F1 measure are reported for the two classes *Non-proximity* and *Proximity*.

Window Size (s)	Precision		Recall		F1 Measure	
	Non Proximity	Proximity	Non Proximity	Proximity	Non Proximity	Proximity
1	0.69	0.75	0.78	0.66	0.73	0.70
5	0.83	0.83	0.83	0.83	0.83	0.83
10	0.84	0.86	0.86	0.84	0.85	0.85
20	0.85	0.85	0.85	0.85	0.85	0.85
30	0.90	0.88	0.87	0.90	0.88	0.88
40	0.87	0.85	0.85	0.88	0.86	0.87
50	0.90	0.85	0.84	0.91	0.87	0.88
60	0.86	0.85	0.85	0.86	0.87	0.86

Table B.3: Comparison of F1 scores for the dataset when the majority class has been down-sampled and when the minority class has been over-sampled using SMOTE. Scores for both classes are presented.

Window Size (s)	Down-sample majority class		Over-sample minority class	
	Non Proximity	Proximity	Non Proximity	Proximity
1	0.73	0.70	0.82	0.77
5	0.83	0.83	0.94	0.94
10	0.85	0.85	0.97	0.97
20	0.85	0.85	0.97	0.97
30	0.88	0.88	0.97	0.97
40	0.86	0.87	0.97	0.97
50	0.87	0.88	0.98	0.98
60	0.87	0.86	0.98	0.98

Appendix C

Wearable device user experience online survey

In this appendix we provide the questions (Table C.1) we asked anonymously to the participants of the deployment described in Section 3.4. The analysis of the survey responses is provided in Section 3.5.3.

Table C.1: Online survey administered to the participants of the deployment described in Section 3.4.

Question	Allowed Answer
During the four weeks study, how long did you wear the device?	Percentage of time from 0% to 100%
Do you think the study was too long?	Yes No
If you said yes to the previous question, what do you think is the best duration and why?	Open-ended
The device was comfortable to wear all day. Specify to what extent you agree or disagree with this statement.	Likert scale: 1 (Strongly disagree) to 5 (Strongly agree)
Do you think other shapes would be more comfortable? (e.g. necklace, loop for the belt, ...)	Open-ended
Describe what shape you think is best and why.	
The device was easy to use (recharge, put it on, ...).	Likert scale: 1 (Strongly disagree) to 5 (Strongly agree)
Specify to what extent you agree or disagree with this statement.	
I am concerned that the device can threaten my privacy. Specify to what extent you agree or disagree with this statement.	Likert scale: 1 (Strongly disagree) to 5 (Strongly agree)
Which aspect of the device makes you more concerned? Select all that apply or none.	Proximity Detection (close to other people) Location Detection (your location inside the building) Activity Detection (stationary or walking) None Other (insert text) Yes
Would you prefer to have the same functionalities integrated in popular electronic devices (smartphones or smartwatches) as an app?	No I don't know Other (insert text)
Briefly motivate your previous answer.	Open-ended
When you are in the office, how often do you carry your phone with you?	Likert scale: 1 (Never) to 5 (Always) Yes
Do you plan to buy a smartwatch in the near future (Apple Watch, Samsung Gear, Motorola Moto 360, ...)?	No I don't know Other (insert text)
Positive Comments	Open-ended
Express here all your positive comments regarding the device and the study.	
Negative Comments	Open-ended
Express here all your negative comments regarding the device and the study.	
Improvements	Open-ended
Express here your ideas to improve the device.	

Appendix D

Protractor schematic

Figure D.1 shows the entire schematic of Protractor's design with connections between the Metamotion platform and the angle detection module, the IR LED used to transmit beacons, the beacon receiver and the micro-SD card slot.

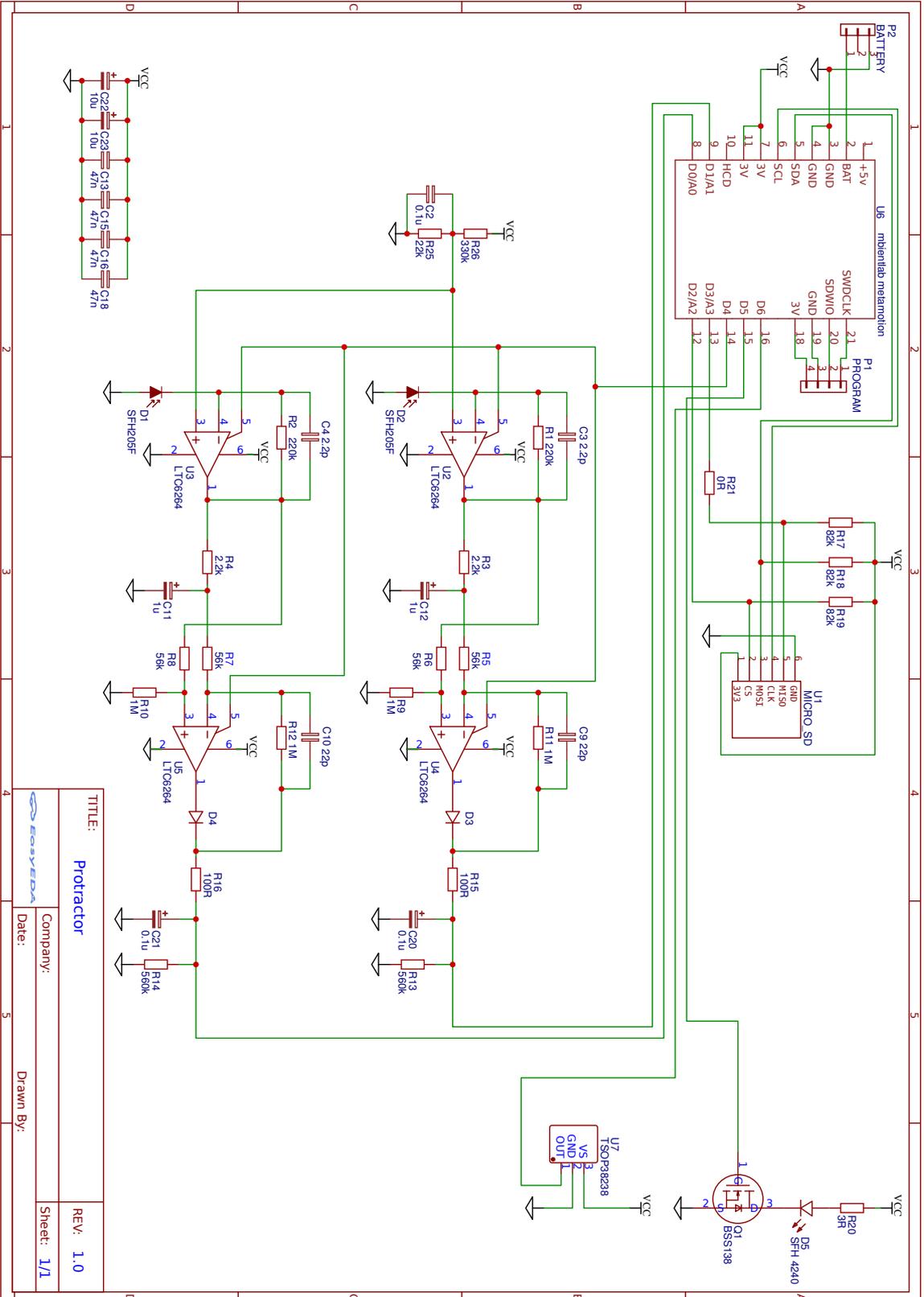


Figure D.1: Protractor schematic.

Appendix E

Random Forest parameters

Table E.1 reports the parameters of the Random Forest models used in Chapter 6, Section 6.5 for the classification of Task Role and Task Timeline. The implementation of the algorithm has been taken from Weka version 3.7.13 [131]¹.

Table E.1: Random Forest parameters used for the classification of Task Role and Task Timeline.

Parameter	Value
Batch size	100
Break ties randomly	False
Debug	False
Do not check capabilities	False
Do not calculate out of bag error	False
Max depth	0
Num decimal places	2
Num execution slots	1
Num features	0
Num trees	{50, 100}
Print trees	False
Seed	1

¹The documentation regarding this classifier can be found at this link: <http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/RandomForest.html>

Bibliography

- [1] Making a better HC-SR04 echo locator. http://uglyduck.ath.cx/ep/archive/2014/01/Making_a_better_HC_SR04_Echo_Locator.html. Last Accessed 014/09/2017.
- [2] Code of Federal Regulations 29 CFR standard 1926.56, 2010.
- [3] Standard UNI EN 12464-1, Lighting of workplaces, 2011.
- [4] N. Aharony, W. Pan, C. Ip, I. Khayal, and A. Pentland. Social fmri: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, 7(6):643–659, 2011.
- [5] E. Aitenbichler and M. Mühlhäuser. An IR Local Positioning System for Smart Items and Devices. In *Proc. of the 23rd International Conference on Distributed Computing Systems*, 2003.
- [6] F. Alam, R. Mehmood, I. Katib, and A. Albeshri. Analysis of eight data mining algorithms for smarter internet of things (iot). *Procedia Computer Science*, 98:437–442, 2016.
- [7] X. Alameda-Pineda, J. Staiano, R. Subramanian, L. Batrinca, E. Ricci, B. Lepri, O. Lanz, and N. Sebe. Salsa: A novel dataset for multimodal group behavior analysis. *IEEE transactions on pattern analysis and machine intelligence*, 38(8):1707–1720, 2016.
- [8] X. Alameda-Pineda, Y. Yan, E. Ricci, O. Lanz, and N. Sebe. Analyzing free-standing conversational groups: A multimodal approach. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 5–14. ACM, 2015.
- [9] J. A. Allen, C. Fisher, M. Chetouani, M. M. Chiu, H. Gunes, M. Mehu, and H. Hung. Comparing social science and computer science workflow processes for studying group interactions. *Small Group Research*, 48(5):568–590, 2017.
- [10] J. A. Allen, N. Lehmann-Willenbrock, and S. G. Rogelberg. *The Cambridge handbook of meeting science*. Cambridge University Press, 2015.
- [11] T. J. Allen. Managing the flow of technology: Technology transfer and the dissem-

- ination of technological information within the r&d organization, volume 1 of mit press books, 1984.
- [12] T. J. Allen and G. Henn. *The Organization and Architecture of Innovation: Managing the Flow of Technology*. Routledge, 2006.
- [13] R. Appel-Meulenbroek, R. Appel-Meulenbroek, P. Groenen, and I. Janssen. An end-user's perspective on activity-based office concepts. *Journal of Corporate Real Estate*, 13(2):122–135, 2011.
- [14] Apple. Watch OS. Last Accessed 17/02/2018.
- [15] Y. Arai and M. Sekiai. Absolute position measurement system for mobile robot based on incident angle detection of infrared light. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 1, pages 986–991, Oct 2003.
- [16] M. Argyle, V. Salter, H. Nicholson, M. Williams, and P. Burgess. The communication of inferior and superior attitudes by verbal and non-verbal signals. *British journal of social and clinical psychology*, 9(3):222–231, 1970.
- [17] A. Ashari, I. Paryudi, and A. M. Tjoa. Performance comparison between naïve bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 4(11), 2013.
- [18] K. Ball, E. M. Daniel, and C. Stride. Dimensions of employee privacy: an empirical study. *Information Technology & People*, 25(4):376–394, 2012.
- [19] T. Ballendat, N. Marquardt, and S. Greenberg. Proxemic interaction: designing for a proximity and orientation-aware environment. In *ACM International Conference on Interactive Tabletops and Surfaces*. ACM, 2010.
- [20] S. Banerjee and A. Rudnicky. Using simple speech based features to detect the state of a meeting and the roles of the meeting participants. In *Proceedings of International Conference on Spoken Language Processing*, pages 221–231, 2004.
- [21] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *International Conference on Pervasive Computing*, pages 1–17. Springer, 2004.
- [22] L. Batrinca, G. Stratou, A. Shapiro, L.-P. Morency, and S. Scherer. Cicero-towards a multimodal virtual audience platform for public speaking training. In *Intelligent Virtual Agents*, pages 116–128. Springer, 2013.
- [23] T. Baur, I. Damian, F. Lingensfelder, J. Wagner, and E. André. Nova: Automated analysis of nonverbal signals in social interactions. In *Human Behavior Understanding*, pages 160–171. Springer, 2013.
- [24] G. Benet, F. Blanes, J. Simó, and P. Pérez. Using infrared sensors for distance

- measurement in mobile robots. *Robotics and Autonomous Systems*, 40(4):255 – 266, 2002.
- [25] H. R. Bernard. *Social research methods: Qualitative and quantitative approaches*. Sage, 2012.
- [26] S. Bluetooth. Bluetooth core specification version 4.0. *Specification of the Bluetooth System*, 2010.
- [27] Bluetooth Special Interest Group (SIG). Bluetooth Core Specification v5.0. <https://www.bluetooth.com/specifications/bluetooth-core-specification>, 2016. Last Accessed 07/04/2018.
- [28] T. W. Boonstra, M. E. Larsen, and H. Christensen. Mapping dynamic social networks in real life using participants’ own smartphones. *Heliyon*, 1(3):e00037, 2015.
- [29] K. Boudreau, I. Ganguli Prokopovych, P. Gaule, E. C. Guinan, and K. R. Lakhani. Colocation and scientific collaboration: evidence from a field experiment. *Harvard Business School Working Paper*, 2012.
- [30] N. M. Bradburn, S. Sudman, E. Blair, and C. Stocking. Question threat and response bias. *Public Opinion Quarterly*, 42(2):221–234, 1978.
- [31] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct. 2001.
- [32] C. Brown, C. Efstratiou, I. Leontiadis, D. Quercia, and C. Mascolo. Tracking serendipitous interactions: How individual cultures shape the office. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW ’14*, pages 1072–1081, New York, NY, USA, 2014. ACM.
- [33] C. Brown, C. Efstratiou, I. Leontiadis, D. Quercia, C. Mascolo, J. Scott, and P. Key. The architecture of innovation: Tracking face-to-face interactions with ubicomp technologies. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp ’14*, pages 811–822, New York, NY, USA, 2014. ACM.
- [34] F. Brudy, D. Ledo, S. Greenberg, and A. Butz. Is anyone looking? mitigating shoulder surfing on public displays through awareness and protection. In *Proceedings of The International Symposium on Pervasive Displays*, page 1. ACM, 2014.
- [35] T. E. Burns and G. M. Stalker. The management of innovation. *University of Illinois at Urbana-Champaign’s Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*, 1961.
- [36] R. S. Burt. Structural holes and good ideas. *American journal of sociology*, 110(2):349–399, 2004.
- [37] J. M. Cabero, V. Molina, I. Urteaga, F. Liberal, and J. L. Martín. Acquisition

- of human traces with Bluetooth technology: Challenges and proposals. *Ad Hoc Networks*, 12, 2014.
- [38] L. Cabrera-Quiros. Towards multimodal analysis of human behavior in crowded mingling scenarios using movement cues from wearable sensors and cameras. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 399–404. ACM, 2016.
- [39] L. Cabrera-Quiros, A. Demetriou, E. Gedik, L. van der Meij, and H. Hung. The matchmingle dataset: a novel multi-sensor resource for the analysis of social interactions and group dynamics in-the-wild during free-standing conversations and speed dates. *IEEE Transactions on Affective Computing*, 2018.
- [40] L. Cabrera-Quiros, E. Gedik, and H. Hung. Estimating self-assessed personality from body movements and proximity in crowded mingling scenarios. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 238–242. ACM, 2016.
- [41] L. Cabrera-Quiros and H. Hung. Who is where?: Matching people in video to wearable acceleration during crowded mingling events. In *Proceedings of the 2016 ACM International Conference on Multimedia*, pages 267–271. ACM, 2016.
- [42] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1611.08050*, 2016.
- [43] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, I. McCowan, W. Post, D. Reidsma, and P. Wellner. The ami meeting corpus: A pre-announcement. In *Proc. of the Second International Conference on Machine Learning for Multimodal Interaction*, 2005.
- [44] C. Cattuto, W. Van 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, 07 2010.
- [45] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, pages 321–357, 2002.
- [46] N. Chinchor. MUC-4 evaluation metrics. In *Proceedings of the 4th conference on Message understanding*, pages 22–29. Association for Computational Linguistics, 1992.
- [47] T. Choudhury and A. Pentland. The sociometer: A wearable device for understanding human networks. In *Proceedings of the CSCW'02 Workshop: Ad hoc Communications and Collaboration in Ubiquitous Computing Environments*, 2002.
- [48] T. Choudhury and A. Pentland. Sensing and modeling human networks using the sociometer. In *Proc. of ISWC*, 2003.

- [49] T. Choudhury and A. Pentland. Characterizing social interactions using the sociometer. In *Proceedings of NAACOS*. Citeseer, 2004.
- [50] A. J. Cuddy, C. A. Wilmuth, A. J. Yap, and D. R. Carney. Preparatory power posing affects nonverbal presence and job interview performance. *Journal of Applied Psychology*, 100(4):1286, 2015.
- [51] B. Czarniawska-Joerges. *Shadowing: and other techniques for doing fieldwork in modern societies*. Copenhagen Business School Press DK, 2007.
- [52] I. Damian, C. S. S. Tan, T. Baur, J. Schöning, K. Luyten, and E. André. Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, pages 565–574, New York, NY, USA, 2015. ACM.
- [53] C. B. Danielsson and L. Bodin. Office type in relation to health, well-being, and job satisfaction among employees. *Environment and Behavior*, 40(5):636–668, 2008.
- [54] I. De Been and M. Beijer. The influence of office type on satisfaction and perceived productivity support. *Journal of Facilities Management*, 12(2), 2014.
- [55] I. De Been, M. Beijer, and D. Den Hollander. How to cope with dilemmas in activity based work environments-results from user-centred research. In *Conference paper 14th EuroFM Research Symposium. EuroFM research papers*, 2015.
- [56] T. M. T. Do, K. Kalimeri, B. Lepri, F. Pianesi, and D. Gatica-Perez. Inferring social activities with mobile sensor networks. In *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, 2013.
- [57] M. C. Dobson. *Low-power epidemic communication in wireless ad hoc networks*. PhD thesis, Amsterdam: Vrije Universiteit, 2013.
- [58] W. Dong, B. Lepri, A. Cappelletti, A. S. Pentland, F. Pianesi, and M. Zancanaro. Using the influence model to recognize functional roles in meetings. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 271–278. ACM, 2007.
- [59] N. Eagle and A. S. Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4), 2006.
- [60] F. Erickson. One function of proxemic shifts in face-to-face interaction. *Organization of behavior in face-to-face interaction*, pages 175–187, 1975.
- [61] R. Faragher et al. Understanding the basis of the kalman filter via a simple and intuitive derivation. *IEEE Signal Processing Magazine*, 29(5):128–132, 2012.
- [62] R. Faragher and R. Harle. Smartslam-an efficient smartphone indoor positioning system exploiting machine learning and opportunistic sensing. In *ION GNSS*, volume 13, pages 1–14, 2013.

- [63] R. Faragher and R. Harle. An analysis of the accuracy of bluetooth low energy for indoor positioning applications. In *Proceedings of the 27th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2014)*, Tampa, FL, USA, volume 812, 2014.
- [64] R. Faragher and R. Harle. Location fingerprinting with bluetooth low energy beacons. *Journal on Selected Areas in Communications*, 2015.
- [65] A.-L. Fayard and J. Weeks. Photocopiers and water-coolers: The affordances of informal interaction. *Organization studies*, 28(5):605–634, 2007.
- [66] M. S. Feldman, B. T. Pentland, L. D’Adderio, and N. Lazaric. Beyond routines as things: Introduction to the special issue on routine dynamics. *Organization Science*, 2016.
- [67] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim. Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research*, 15(1):3133–3181, 2014.
- [68] A. Forsys, K. Min, T. Schmid, W. Pettey, D. Toth, and M. Leecaster. Wrenmining: Large-scale data collection for human contact network research. In *Proceedings of First International Workshop on Sensing and Big Data Mining*, pages 1–6. ACM, 2013.
- [69] D. Frantal and D. Arbula. Mobile node localization using infrared angle of arrival sensor. In *International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015.
- [70] H. Geert, H. G. Jan, and M. Minkov. *Cultures And Organizations: Software of The Mind*. McGrawHill USA, 1991.
- [71] M. Génois, C. L. Vestergaard, J. Fournet, A. Panisson, I. Bonmarin, and A. Barrat. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. *Network Science*, 3(03):326–347, 2015.
- [72] V. P. George and T. J. Allen. Relational data in organizational settings: an introductory note for using agni and netgraphs to analyze nodes, relationships, partitions and boundaries. 1993.
- [73] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson. Phaser: Enabling Phased Array Signal Processing on Commodity WiFi Access Points. In *Proc. of MobiCom*, 2014.
- [74] E. Goffman. *Encounters: Two studies in the sociology of interaction*. Indianapolis, Bobbs-Marrill, 1961.
- [75] Google. Android v5.0 BLE. <https://developer.android.com/about/versions/android-5.0.html>. Last Accessed 28/11/2015.

- [76] T. Grajewski and L. Vaughan. Space syntax observation manual. *UCL Bartlett and Space Syntax Limited, London*, 2001.
- [77] P. D. Groves. *Principles of GNSS, inertial, and multisensor integrated navigation systems*. Artech House, 2013.
- [78] R. A. Güler, N. Neverova, and I. Kokkinos. Densepose: Dense human pose estimation in the wild. *arXiv preprint arXiv:1802.00434*, 2018.
- [79] T. Hærem, B. T. Pentland, and K. D. Miller. Task complexity: Extending a core concept. *Academy of Management Review*, 40(3):446–460, 2015.
- [80] E. T. Hall. A system for the notation of proxemic behavior. *American Anthropologist*, 65(5):1003–1026, 1963.
- [81] E. T. Hall. *The hidden dimension*. Doubleday & Co, 1966.
- [82] J. Hall and W. H. Watson. The effects of a normative intervention on group decision-making performance. *Human relations*, 23(4):299–317, 1970.
- [83] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36, 1982.
- [84] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster. The anatomy of a context-aware application. In *Proc. of MobiCom*, 1999.
- [85] M. Hazas, C. Kray, H. Gellersen, H. Agbota, G. Kortuem, and A. Krohn. A relative positioning system for co-located mobile devices. In *Proc. of MobiSys*, 2005.
- [86] R. Hoyle, R. Templeman, S. Armes, D. Anthony, D. Crandall, and A. Kapadia. Privacy behaviors of lifeloggers using wearable cameras. In *Proc. of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014.
- [87] W. Huang, Y.-S. Kuo, P. Pannuto, and P. Dutta. Opo: a wearable sensor for capturing high-fidelity face-to-face interactions. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*. ACM, 2014.
- [88] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket switched networks and human mobility in conference environments. In *Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking, WDTN '05*, pages 244–251, New York, NY, USA, 2005. ACM.
- [89] H. Hung and D. Gatica-Perez. Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Transactions on Multimedia*, 12(6):563–575, Oct 2010.
- [90] H. Hung and B. Kröse. Detecting f-formations as dominant sets. In *Proc. of ICMI*, 2011.
- [91] M. Ianeva, S. Faure, J. Theveniot, F. Ribeyron, C. Crossan, G. Cordon, and C. Garteriser. Pervasive technologies for smart workplaces: A workplace efficiency solution

- for office design and building management from an occupier's perspective. In *Human Work Interaction Design. Work Analysis and Interaction Design Methods for Pervasive and Smart Workplaces*, pages 73–82. Springer, 2015.
- [92] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *Journal of theoretical biology*, 271(1):166–180, 2011.
- [93] S. Jamil, A. Basalamah, and A. Lbath. Crowdsensing traces using bluetooth low energy (ble) proximity tags. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 71–74. ACM, 2014.
- [94] S. Jamil, A. Basalamah, A. Lbath, and M. Youssef. Hybrid participatory sensing for analyzing group dynamics in the largest annual religious gathering. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 547–558. ACM, 2015.
- [95] D. B. Jayagopi and D. Gatica-Perez. Discovering group nonverbal conversational patterns with topics. In *Proc. of International Conference on Multimodal Interfaces (ICMI-MLMI)*, 2009.
- [96] D. B. Jayagopi, H. Hung, C. Yeo, and D. Gatica-Perez. Modeling dominance in group conversations using nonverbal activity cues. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(3):501–513, March 2009.
- [97] K. Jayarajah, Y. Lee, A. Misra, and R. K. Balan. Need accurate user behaviour?: pay attention to groups! In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 855–866. ACM, 2015.
- [98] P. Jeffrey and A. McGrath. Sharing serendipity in the workplace. In *Proceedings of the Third International Conference on Collaborative Virtual Environments, CVE '00*, pages 173–179, New York, NY, USA, 2000. ACM.
- [99] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.
- [100] K. Katevas, K. Hänsel, R. Clegg, I. Leontiadis, H. Haddadi, and L. Tokarchuk. Finding dory in the crowd: Detecting social interactions using multi-modal mobile sensing. *arXiv preprint arXiv:1809.00947*, 2018.
- [101] S. Kauffeld and N. Lehmann-Willenbrock. Meetings matter: Effects of team meetings on team and organizational success. *Small Group Research*, 43(2):130–158, 2012.
- [102] S. Kauffeld and R. A. Meyers. Complaint and solution-oriented circles: Interaction patterns in work group discussions. *European Journal of Work and Organizational Psychology*, 18(3):267–294, 2009.

- [103] F. Kawsar, C. Min, A. Mathur, and A. Montanari. Earables for personal-scale behavior analytics. *IEEE Pervasive Computing*, 17(3):83–89, 2018.
- [104] M. Knapp, J. Hall, and T. Horgan. *Nonverbal communication in human interaction*. Cengage Learning, 2013.
- [105] T. Komine and M. Nakagawa. Fundamental analysis for visible-light communication system using LED lights. *IEEE transactions on Consumer Electronics*, 50(1):100–107, 2004.
- [106] V. Kostakos. Temporal graphs. *Physica A: Statistical Mechanics and its Applications*, 388(6):1007–1023, 2009.
- [107] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti. SpotFi: Decimeter Level Localization Using WiFi. In *Proc. of SIGCOMM*, 2015.
- [108] S. W. Kozlowski. Advancing research on team process dynamics theoretical, methodological, and measurement considerations. *Organizational Psychology Review*, 5(4):270–299, 2015.
- [109] D. Krackhardt and J. Hanson. Informal networks: The company behind the chart. *Harvard Business Review*, 71(4):104–111, jul/aug 1993.
- [110] D. Krackhardt and R. N. Stern. Informal networks and organizational crises: An experimental simulation. *Social psychology quarterly*, pages 123–140, 1988.
- [111] R. E. Kraut, R. S. Fish, R. W. Root, and B. L. Chalfonte. Informal communication in organizations: Form, function, and technology. In *Human reactions to technology: Claremont symposium on applied social psychology*, pages 145–199. Citeseer, 1990.
- [112] W. H. Kruskal and W. A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.
- [113] S. Kumar, S. Gil, D. Katabi, and D. Rus. Accurate indoor localization with zero start-up cost. In *Proc. of MobiCom*, 2014.
- [114] S. Kumar, E. Hamed, D. Katabi, and L. Erran Li. LTE Radio Analytics Made Easy and Accessible. In *Proc. of SIGCOMM*, 2014.
- [115] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta. Luxapose: Indoor positioning with mobile phones and visible light. In *Proc. of MobiCom*, pages 447–458. ACM, 2014.
- [116] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.
- [117] O. Lederman, D. Calacci, A. MacMullen, D. C. Fehder, F. E. Murray, and A. Pentland. Open badges: A low-cost toolkit for measuring team communication and dynamics. *arXiv preprint arXiv:1710.01842*, 2017.

- [118] O. Lederman, A. Mohan, D. Calacci, and A. S. Pentland. Rhythm: A unified measurement platform for human organizations. *IEEE MultiMedia*, 25(1):26–38, 2018.
- [119] Y. Lee, Y. Ju, C. Min, S. Kang, I. Hwang, and J. Song. CoMon: Cooperative ambience monitoring platform with continuity and benefit awareness. In *Proc. of MobiSys*, 2012.
- [120] Y. Lee, C. Min, C. Hwang, J. Lee, I. Hwang, Y. Ju, C. Yoo, M. Moon, U. Lee, and J. Song. Sociophone: Everyday face-to-face interaction monitoring platform using multi-phone sensor fusion. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2013.
- [121] Leesman. Activity Based Working - The rise and rise of ABW: Reshaping the physical, virtual and behavioural workspace, 2016. Last Accessed 03/08/2016”.
- [122] N. Lehmann-Willenbrock, J. A. Allen, and S. Kauffeld. A sequential analysis of procedural meeting communication: How teams facilitate their meetings. *Journal of Applied Communication Research*, 41(4):365–388, 2013.
- [123] N. Lehmann-Willenbrock, A. Grohmann, and S. Kauffeld. Task and relationship conflict at work. *European Journal of Psychological Assessment*, 2011.
- [124] N. Lehmann-Willenbrock, H. Hung, and J. Keyton. New frontiers in analyzing dynamic group interactions: Bridging social and computer science. *Small group research*, 48(5):519–531, 2017.
- [125] T. W. Leigh and J. O. Summers. An initial evaluation of industrial buyers’ impressions of salespersons’ nonverbal cues. *Journal of Personal Selling & Sales Management*, 22(1), 2002.
- [126] B. Lepri, J. Staiano, G. Rigato, K. Kalimeri, A. Finnerty, F. Pianesi, N. Sebe, and A. Pentland. The sociometric badges corpus: A multilevel behavioral dataset for social behavior in complex organizations. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*. IEEE, 2012.
- [127] L. Li, P. Hu, C. Peng, G. Shen, and F. Zhao. Epsilon: A visible light based positioning system. In *Proc. of NSDI*, pages 331–343, 2014.
- [128] K. Liu, X. Liu, and X. Li. Guoguo: Enabling fine-grained indoor localization via smartphone. In *Proc. of MobiSys*, 2013.
- [129] S. Liu, Y. Jiang, and A. Striegel. Face-to-Face Proximity Estimation Using Bluetooth On Smartphones. *IEEE Transactions on Mobile Computing*, 13(4), 2014.
- [130] J.-C. Lucet, C. Laouenan, G. Chelius, N. Veziris, D. Lepelletier, A. Friggeri, D. Abiteboul, E. Bouvet, F. Mentre, and E. Fleury. Electronic sensors for assessing interactions between healthcare workers and patients under airborne precautions. *PloS one*, 7(5):e37893, 2012.

- [131] Machine Learning Group at the University of Waikato. Weka 3: Data Mining Software in Java. <http://www.cs.waikato.ac.nz/ml/weka/>. Last Accessed 09/12/2015.
- [132] C. Martella, A. Miraglia, M. Cattani, and M. van Steen. Leveraging proximity sensing to mine the behavior of museum visitors. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–9. IEEE, 2016.
- [133] A. Mashhadi, U. G. Acer, A. Boran, P. M. Scholl, C. Forlivesi, G. Vanderhulst, and F. Kawsar. Exploring space syntax on entrepreneurial opportunities with wi-fi analytics. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, pages 658–669, New York, NY, USA, 2016. ACM.
- [134] M. S. Mast. On the importance of nonverbal communication in the physician–patient interaction. *Patient education and counseling*, 67(3):315–318, 2007.
- [135] R. Mastrandrea, J. Fournet, and A. Barrat. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PloS one*, 10(9):e0136497, 2015.
- [136] J. E. Mathieu, J. R. Hollenbeck, D. van Knippenberg, and D. R. Ilgen. A century of work teams in the journal of applied psychology. *Journal of Applied Psychology*, 102(3):452, 2017.
- [137] T. V. McGovern and H. E. Tinsley. Interviewer evaluations of interviewee nonverbal behavior. *Journal of Vocational Behavior*, 13(2):163–171, 1978.
- [138] J. E. McGrath and T. W. Altermatt. Observation and analysis of group interaction over time: Some methodological and strategic choicess. *Blackwell handbook of social psychology: Group processes*, pages 525–556, 2001.
- [139] A. Mehrabian. Some referents and measures of nonverbal behavior. *Behavior Research Methods*, 1(6), 1968.
- [140] E. M. Meijer, M. H. Frings-Dresen, and J. K. Sluiter. Effects of office innovation on office workers’ health and performance. *Ergonomics*, 52(9):1027–1038, 2009.
- [141] C. Min, A. Montanari, A. Mathur, S. Lee, and F. Kawsar. Cross-modal approach for conversational well-being monitoring with multi-sensory earables. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 706–709. ACM, 2018.
- [142] J.-K. Min, J. Wiese, J. I. Hong, and J. Zimmerman. Mining smartphone data to classify life-facets of social relationships. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 285–294. ACM, 2013.

- [143] M. Minami, Y. Fukuju, K. Hirasawa, S. Yokoyama, M. Mizumachi, H. Morikawa, and T. Aoyama. *DOLPHIN: A practical approach for implementing a fully distributed indoor ultrasonic positioning system*, pages 347–365. Springer Berlin Heidelberg, 2004.
- [144] A. Montanari. Multimodal Indoor Social Interaction Sensing and Real-time Feedback for Behavioural Intervention. In *Proceedings of the 7th Workshop on Wireless of the Students, by the Students, & for the Students*, pages 7–9. ACM, 2015.
- [145] A. Montanari, F. Kringberg, A. Valentini, C. Mascolo, and A. Prorok. Surveying Areas in Developing Regions Through Context Aware Drone Mobility. In *Proceedings of the 4th ACM Workshop on Micro Aerial Vehicle Networks, Systems, and Applications (DroNet18). Colocated with ACM MobiSys18*. ACM, 2018.
- [146] A. Montanari, C. Mascolo, K. Sailer, and S. Nawaz. Detecting emerging activity-based working traits through wearable technology. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, volume 1, pages 86:1–86:24, New York, NY, USA, Sept. 2017. ACM.
- [147] A. Montanari, A. Mashhadi, A. Mathur, and F. Kawsar. Understanding the privacy design space for personal connected objects. In *Proceedings of the 30th International BCS Human Computer Interaction Conference (British HCI '16)*, pages 18:1–18:13, Swindon, UK, 2016. BCS Learning & Development Ltd.
- [148] A. Montanari, S. Nawaz, C. Mascolo, and K. Sailer. Experience in deploying wearable devices for office analytics. In *Proceedings of the 1st International Workshop on Quantified Workplace. Co-located with CSCW 2016.*, 2016.
- [149] A. Montanari, S. Nawaz, C. Mascolo, and K. Sailer. A Study of Bluetooth Low Energy performance for human proximity detection in the workplace. In *Proceedings of the 15th IEEE International Conference on Pervasive Computing and Communications (Percom'17)*, pages 90–99. IEEE, 2017.
- [150] A. Montanari, Y. A. Pignolet, and E. Ferranti. Rest assured, we manage your microgrid. In *Proceedings of the 5th IEEE International Conference on Smart Grid Communications (SmartGridComm '14)*, pages 284–289, Nov 2014.
- [151] A. Montanari, Z. Tian, E. Francu, B. Lucas, B. Jones, X. Zhou, and C. Mascolo. Measuring interaction proxemics with wearable light tags. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, volume 2, pages 25:1–25:30, New York, NY, USA, Mar. 2018. ACM.
- [152] Nordic Semiconductor. SoftDevice S110 Specification v2.0. http://www.nordicsemi.com/eng/nordic/download_resource/20338/15/72793162, 2014. Last Accessed 02/04/2015.
- [153] D. O. Olgún, P. A. Gloor, and A. Pentland. Wearable sensors for pervasive health-

- care management. In *Proceedings of the 3rd International Conference on Pervasive Computing Technologies for Healthcare*. IEEE, 2009.
- [154] D. O. Olguín and A. S. Pentland. Social sensors for automatic data collection. *Proceedings of the 14th Americas Conference on Information Systems (AMCIS)*, 2008.
- [155] D. O. Olguín, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(1):43–55, 2009.
- [156] Osram Opto Semiconductors. High power infrared emitter (940 nm). http://www.osram-os.com/osram_os/en/products/product-catalog/infrared-emitters,-detectors-and-sensors/infrared-emitters/power-emitter-gt40mw/emitter-with-940-nm/sfh-4240/index.jsp. Last Accessed 02/09/2016.
- [157] Osram Opto Semiconductors. Silicon PIN photodiode with daylight blocking filter. http://www.osram-os.com/osram_os/en/products/product-catalog/infrared-emitters,-detectors-and-sensors/silicon-photodetectors/photodiodes/pin-photodiodes-in-through-hole-package/sfh-205-f/index.jsp. Last Accessed 02/09/2016.
- [158] S. Park, I. Locher, A. Savvides, M. B. Srivastava, A. Chen, R. Muntz, and S. Yuen. Design of a wearable sensor badge for smart kindergarten. In *Wearable Computers, 2002. (ISWC 2002). Proceedings. Sixth International Symposium on*, pages 231–238. IEEE, 2002.
- [159] A. Penn, J. Desyllas, and L. Vaughan. The space of innovation: interaction and communication in the work environment. *Environment and Planning B: Planning and Design*, 26(2):193–218, 1999.
- [160] A. Pentland. Social signal processing. *IEEE Signal Processing Magazine*, 24(4):108–111, 2007.
- [161] A. Pentland. The new science of building great teams. *Harvard Business Review*, 90(4):60–69, 2012.
- [162] B. T. Pentland, T. Hærem, and D. Hillison. Comparing organizational routines as recurrent patterns of action. *Organization Studies*, 31(7):917–940, 2010.
- [163] B. T. Pentland, A. P. Pentland, and R. J. Calantone. Bracketing off the actors: Towards an action-centric research agenda. *Information and Organization*, 27(3):137–143, 2017.
- [164] J. E. Perry-Smith and C. E. Shalley. The social side of creativity: A static and dynamic social network perspective. *Academy of management review*, 28(1):89–106, 2003.

- [165] F. Pianesi, M. Zancanaro, V. Falcon, and E. Not. Toward supporting group dynamics. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 302–311. Springer, 2006.
- [166] R. W. Picard. *Affective computing*, 1995.
- [167] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. The Cricket location-support system. In *Proc. of MobiCom*, 2000.
- [168] N. B. Priyantha, A. K. Miu, H. Balakrishnan, and S. Teller. The Cricket compass for context-aware mobile applications. In *Proc. of MobiCom*, 2001.
- [169] M. Radhakrishnan, A. Misra, R. K. Balan, and Y. Lee. Smartphones and ble services: Empirical insights. In *Mobile Ad Hoc and Sensor Systems (MASS), 2015 IEEE 12th International Conference on*, pages 226–234. IEEE, 2015.
- [170] R. Rawassizadeh, B. A. Price, and M. Petre. Wearables: Has the age of smartwatches finally arrived? *Commun. ACM*, 58(1):45–47, Dec. 2014.
- [171] E. Rea, J. Laffèche, S. Stalker, B. Guarda, H. Shapiro, I. Johnson, S. Bondy, R. Upshur, M. Russell, and M. Eliasziw. Duration and distance of exposure are important predictors of transmission among community contacts of ontario sars cases. *Epidemiology and infection*, 135(06):914–921, 2007.
- [172] M. S. Remland, T. S. Jones, and H. Brinkman. Interpersonal distance, body orientation, and touch: Effects of culture, gender, and age. *The Journal of social psychology*, 135(3), 1995.
- [173] P. Rong and M. L. Sichitiu. Angle of arrival localization for wireless sensor networks. In *Sensor and Ad Hoc Communications and Networks, 2006. SECON'06. 2006 3rd Annual IEEE Communications Society on*, volume 1, pages 374–382. IEEE, 2006.
- [174] I. Rossberg-Gempton and G. D. Poole. The effect of open and closed postures on pleasant and unpleasant emotions. *The Arts in psychotherapy*, 20(1):75–82, 1993.
- [175] K. Sailer. *The space-organisation relationship. On the shape of the relationship between spatial configuration and collective organisational behaviours*. PhD thesis, TU Dresden, Germany, 2010.
- [176] K. Sailer and I. McCulloh. Social networks and spatial configuration - how office layouts drive social interaction. *Social networks*, 34(1):47–58, 2012.
- [177] K. Sailer, R. Pachilova, and C. Brown. Human versus machine-testing validity and insights of manual and automated data gathering methods in complex buildings. In *Proceedings of the 9th International Space Syntax Symposium*. Sejong University Press, 2013.
- [178] K. Sailer and A. Penn. Spatiality and transpatiality in workplace environments.

- In *Proceedings of the 7th International Space Syntax Symposium*. Royal Institute of Technology (KTH), 2009.
- [179] K. Sailer, R. Pomeroy, and R. Haslem. Data-driven design - Using data on human behaviour and spatial configuration to inform better workplace design. *Corporate Real Estate Journal*, 4(3):249–262, 2015.
- [180] H. Salam, O. Celiktutan, I. Hupont, H. Gunes, and M. Chetouani. Fully automatic analysis of engagement and its relationship to personality in human-robot interactions. *IEEE Access*, 5:705–721, 2017.
- [181] H. Salamin, S. Favre, and A. Vinciarelli. Automatic role recognition in multiparty recordings: Using social affiliation networks for feature extraction. *IEEE Transactions on Multimedia*, 11(7):1373–1380, 2009.
- [182] H. Salamin, A. Vinciarelli, K. Truong, and G. Mohammadi. Automatic role recognition based on conversational and prosodic behaviour. In *Proceedings of the ACM International Conference on Multimedia*, 2010.
- [183] Samsung. Gear Live. http://www.samsung.com/global/microsite/gear/gearlive_design.html. Last Accessed 09/12/2015.
- [184] Samsung. Gear S2. <http://www.samsung.com/global/galaxy/gear-s2/>. Last Accessed 09/12/2015.
- [185] D. Sanchez-Cortes, O. Aran, M. S. Mast, and D. Gatica-Perez. A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Transactions on Multimedia*, 14(3):816–832, June 2012.
- [186] P. Schlyter. Radiometry and photometry in astronomy. <http://stjarnhimlen.se/comp/radfaq.html>, 2009. Last Accessed 02/09/2016.
- [187] R. Schmidt. Multiple emitter location and signal parameter estimation. *Antennas and Propagation, IEEE Transactions on*, 34(3):276–280, Mar. 1986.
- [188] D. J. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Pharmacodynamics*, 15(6), 1987.
- [189] N. Semiconductor. Concurrent MultiProtocol Library. <http://nordicsemiconductor.github.io/nRF51-multi-role-conn-observer-advertiser>. Last Accessed 28/11/2015.
- [190] M. Shoaib, S. Bosch, O. Incel, H. Scholten, and P. Havinga. A survey of online activity recognition using mobile phones. *Sensors*, 15(1):2059–2085, 2015.
- [191] T. Smieszek, V. C. Barclay, I. Seeni, J. J. Rainey, H. Gao, A. Uzicanin, and M. Salathé. How should social mixing be measured: comparing web-based survey and sensor-based methods. *BMC infectious diseases*, 14(1):1, 2014.

- [192] T. Smieszek, S. Castell, A. Barrat, C. Cattuto, P. J. White, and G. Krause. Contact diaries versus wearable proximity sensors in measuring contact patterns at a conference: method comparison and participants' attitudes. *BMC Infectious Diseases*, 16(1):1, 2016.
- [193] R. Sommer. Personal space in a digital age. *Handbook of environmental psychology*, pages 647–660, 2002.
- [194] A. Sorokowska, P. Sorokowski, P. Hilpert, K. Cantarero, T. Frackowiak, K. Ahmadi, A. M. Alghraibeh, R. Aryeetey, A. Bertoni, K. Bettache, et al. Preferred interpersonal distances: A global comparison. *Journal of Cross-Cultural Psychology*, 48(4):577–592, 2017.
- [195] J. B. Stryker and M. D. Santoro. Facilitating face-to-face communication in high-tech teams. *Research-Technology Management*, 55(1):51–56, 2012.
- [196] W.-T. Tan, M. Baker, B. Lee, and R. Samadani. The sound of silence. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2013.
- [197] Z. Tian, K. Wright, and X. Zhou. The Darklight Rises: Visible Light Communication in the Dark. In *Proc. of MobiCom*, 2016.
- [198] H. Tischler. *Introduction to Sociology*. Nelson Education, 1990.
- [199] Tizen Project, Linux Foundation. Tizen 2.3.1 Release Notes. <https://developer.tizen.org/development/tools/download/release-notes/2.3.1-sep-3-2015>. Last Accessed 28/11/2015.
- [200] A. Toshev and C. Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1653–1660, 2014.
- [201] S. Townsend, M. E. Larsen, T. W. Boonstra, and H. Christensen. Using bluetooth low energy in smartphones to map social networks. *arXiv preprint arXiv:1508.03938*, 2015.
- [202] J. M. Travaline, R. Ruchinkas, and G. E. D'Alonzo. Patient-physician communication: why and how. *J Am Osteopath Assoc*, 105(1):13–8, 2005.
- [203] A. Vinciarelli. Speakers role recognition in multiparty audio recordings using social network analysis and duration distribution modeling. *IEEE Transactions on Multimedia*, 9(6):1215–1226, 2007.
- [204] A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759, 2009.
- [205] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, and M. Schroeder. Bridging the gap between social animal and unsocial machine: A sur-

- vey of social signal processing. *IEEE Transactions on Affective Computing*, 3(1):69–87, 2012.
- [206] Vishay Semiconductors. Data formats for IR remote control. <https://www.vishay.com/docs/80071/dataform.pdf>. Last Accessed 02/09/2016.
- [207] Vishay Semiconductors. IR receiver modules for remote control systems. <https://www.vishay.com/docs/82491/tsop382.pdf>. Last Accessed 02/09/2016.
- [208] D. Vogel and R. Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*. ACM, 2004.
- [209] B. N. Waber, D. Olguin Olguin, T. Kim, and A. Pentland. Productivity through coffee breaks: Changing social networks by changing break structure. *Available at SSRN 1586375*, 2010.
- [210] S. Wang, C. Chen, and J. Ma. Accelerometer based transportation mode recognition on mobile phones. In *Wearable Computing Systems (APWCS), 2010 Asia-Pacific Conference on*, pages 44–46. IEEE, 2010.
- [211] W. F. Whyte. *Street corner society: The social structure of an Italian slum*. University of Chicago Press, 2012.
- [212] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6), 1945.
- [213] C. Wohlers and G. Hertel. Choosing where to work at work - towards a theoretical model of benefits and risks of activity-based flexible offices. *Ergonomics*, pages 1–20, 2016.
- [214] T.-W. Wong, C.-k. Lee, W. Tam, J. Lau, T. Yu, S. Lui, P. K. Chan, Y. Li, J. S. Bresee, J. J. Sung, et al. Cluster of sars among medical students exposed to single patient, hong kong. *Emerging infectious diseases*, 2004.
- [215] L. Wu, B. N. Waber, S. Aral, E. Brynjolfsson, and A. Pentland. Mining face-to-face interaction networks using sociometric badges: Predicting productivity in an it configuration task. *Available at SSRN 1130251*, 2008.
- [216] T. Wujec. The marshmallow challenge. <https://www.tomwujec.com/marshmallowchallenge>. Last Accessed 01/12/2018.
- [217] T. Wujec. *Imagine, design, create: How designers, engineers and architects are changing our world*, 2011.
- [218] J. Xiong and K. Jamieson. ArrayTrack: A Fine-grained Indoor Location System. In *Proc. of NSDI*, 2013.

-
- [219] E. Yoneki. Fluphone study: Virtual disease spread using hagggle. In *Proceedings of the 6th ACM workshop on Challenged networks*, pages 65–66. ACM, 2011.
 - [220] E. Yoneki and J. Crowcroft. Epimap: Towards quantifying contact networks for understanding epidemiology in developing countries. *Ad Hoc Networks*, 13:83–93, 2014.
 - [221] M. Zancanaro, B. Lepri, and F. Pianesi. Automatic detection of group functional roles in face to face interactions. In *Proc. of ICMI*, 2006.
 - [222] C. Zhang and X. Zhang. LiTell: Robust Indoor Localization Using Unmodified Light Fixtures. In *Proc. of MobiCom*, 2016.
 - [223] H. Zhang, W. Du, P. Zhou, M. Li, and P. Mohapatra. DopEnc: Acoustic-based encounter profiling using smartphones. In *Proc. of MobiCom*, 2016.
 - [224] Y. Zhang, J. Olenick, C.-H. Chang, S. W. Kozlowski, and H. Hung. Teamsense: Assessing personal affect and group cohesion in small teams through dyadic interaction and behavior analysis with wearable sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):150, 2018.
 - [225] N. Zhao. Full-Featured Pedometer Design Realized with 3-Axis Digital Accelerometer. Technical report, Analog Devices, 06 2010.