Opportunistic sensing platforms to interpret human behaviour

Niklas Palaghias

Submitted for the Degree of
Doctor of Philosophy
from the
University of Surrey

Institute for Communication Systems
Faculty of Engineering and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, U.K.

November 2016

© Niklas Palaghias 2016
Summary

Understanding human behaviour in an automatic but also non-intrusive manner, constitutes an important and emerging area for various fields. This requires collaboration of information technology with humanitarian sciences in order to transfer existing knowledge of human behaviour into self-acting tools to eliminate the human error. This work strives to shed some light in the area of Mobile Social Signal Processing by trying to understand if today’s mobile devices, given their advanced sensing and computational capabilities, are able to extract various aspects of human behaviour. Although one of the core aspects of human behaviour are social interactions, current tools do not provide an accurate, reliable and real-time solution for social interaction detection, which constitutes a significant barrier in automatic human behaviour understanding.

Towards filling the aforementioned gap in order to enable human behaviour understanding through mobile devices, particular contributions were made. Firstly, an interpersonal distance estimation technique is developed based upon a non-intrusive opportunistic mechanism that solely relies on sensors and communication capabilities of off-the-shelf smartphones. Secondly, based on user’s interpersonal distance and relative orientation, a pervasive and opportunistic approach based on off-the-shelf smartphones for social interaction detection system is presented. Leveraging information provided by psychology, analytical and error models are proposed to estimate the probability of people having social interactions. Then, to showcase the ability of mobile devices to infer human behaviour, a trust relationship quantification mechanism is developed based on users’ behavioural traits and psychological models. Finally, a prediction and compensation mechanism for the device displacement error that leverages human locomotion patterns to refine the device orientation is introduced.

The above contributions were evaluated through experimentation and hard data collected from real-world environments to prove their accuracy and reliability as well as showing the applicability of the proposed approaches in daily situations. This work showed that mobile devices are able to accurately detect social interactions and further social and trust relationships among people, despite the noise induced in real-world situations. Close collaboration between informatics and social sciences is imperative, to overcome the significant barrier in the development of human behaviour understanding. This work could constitute a fundamental building block, as the computational power and battery autonomy of mobile devices increases, for the development of novel techniques towards understanding human behaviour, by including multiple behavioural traits and enabling the creation of socially-aware information systems.

**Key words:** Social Signal Processing, Human Behaviour, Social Interactions, Machine Learning, Mobile Phones.
Email:  n.palaghias@surrey.ac.uk or palaghias@gmail.com
WWW:  http://www.surrey.ac.uk/5gic
Contents

1 Introduction .................................................. 1
   1.1 Motivation ................................................. 1
   1.2 Objectives .................................................. 3
      1.2.1 Interpersonal distance estimation with smartphones ... 3
      1.2.2 Modelling social interaction detection with smartphones in real-world ........................................ 4
      1.2.3 Quantifying trust relationships from social interactions ... 5
      1.2.4 A prediction and compensation mechanism for device displacement ........................................... 6
   1.3 Contributions ............................................... 7
   1.4 Outline .................................................... 8

Nomenclature ....................................................... 1

2 A Survey on Mobile Social Signal Processing .......... 9
   2.1 Introduction ............................................... 9
      2.1.1 A mobile and opportunistic point of view ............... 11
   2.2 An overview of Mobile Social Signal Processing .......... 13
   2.3 Sensing frameworks ........................................ 16
      2.3.1 Data collection ......................................... 17
      2.3.2 Inference engines ....................................... 19
      2.3.3 Framework comparison ................................... 21
   2.4 Social interaction detection ................................ 21
      2.4.1 Single modality ......................................... 22
      2.4.2 Multi-modality ........................................... 25
      2.4.3 Apposition of social interaction detection approaches ............................................................... 26
2.5 Behavioural cues extraction ..................................... 27
   2.5.1 Auditory .................................................. 28
   2.5.2 Physical activity .......................................... 34
   2.5.3 Gesture and posture .................................... 38
   2.5.4 Facial cues ................................................ 44
   2.5.5 Environment and space .................................. 48
   2.5.6 Device usage .............................................. 55
   2.5.7 Physiological ............................................ 56

2.6 From social signals to social behaviour inference ............ 57
   2.6.1 Stress ....................................................... 57
   2.6.2 Emotion ..................................................... 59
   2.6.3 Mood .......................................................... 62
   2.6.4 Personality traits ......................................... 63
   2.6.5 Dominance ............................................... 65
   2.6.6 Other social behaviours ................................. 66

2.7 Applications .................................................... 67
   2.7.1 Health-care ................................................. 68
   2.7.2 Organisational engineering ............................... 69
   2.7.3 Marketing .................................................. 70

2.8 Discussion ...................................................... 70

2.9 Challenges ...................................................... 74
   2.9.1 Context recognition ...................................... 74
   2.9.2 Multi-modal fusion ...................................... 75
   2.9.3 Interdisciplinary area ................................... 75
   2.9.4 Ground truth .............................................. 76
   2.9.5 Privacy ..................................................... 76
   2.9.6 Energy efficiency ....................................... 77

2.10 Conclusions of the literature review .......................... 78
3 Interpersonal Distance Estimation with Smartphones 81
   3.1 Introduction .................................................. 81
   3.2 Background .................................................... 83
      3.2.1 Coarse-grained distance estimation .................. 84
      3.2.2 Fine-grained distance estimation ................... 84
      3.2.3 Proposed approach .................................. 86
   3.3 Methodology .................................................. 87
      3.3.1 Overview .............................................. 88
      3.3.2 Training the models ................................ 89
   3.4 Experimental setup ......................................... 94
   3.5 Results ...................................................... 96
   3.6 Conclusion .................................................. 100

4 Modelling Social Interaction Detection with Smartphones in Real-world 103
   4.1 Introduction .................................................. 104
   4.2 Related work .................................................. 106
      4.2.1 Proximity detection ................................ 107
      4.2.2 Multi-modal detection ............................... 108
      4.2.3 The DARSIS approach ............................... 109
   4.3 DARSIS system ................................................. 110
      4.3.1 Interpersonal distance ............................... 111
      4.3.2 Relative orientation ................................ 112
      4.3.3 Collaborative sensing ................................ 114
   4.4 System model .................................................. 115
      4.4.1 Assumptions .......................................... 115
      4.4.2 Prediction models .................................. 116
      4.4.3 Error Models ......................................... 118
   4.5 Performance evaluation ...................................... 122
      4.5.1 Relative orientation evaluation .................. 122
      4.5.2 Error model evaluation ............................ 126
      4.5.3 Real-world social interactions .................... 129
      4.5.4 Power consumption .................................. 132
      4.5.5 Data communication burden ......................... 133
   4.6 Conclusion .................................................. 134
5 Quantifying Real-world Trust Relationships through Social Interactions 137

5.1 Introduction .................................................. 138
5.2 Background .................................................... 139
   5.2.1 On-line social networks ................................. 139
   5.2.2 Real-world social graphs ............................... 141
   5.2.3 Proposed approach ..................................... 142
5.3 Methodology .................................................. 143
   5.3.1 Background .............................................. 143
   5.3.2 Real-world social graph ............................... 143
   5.3.3 Deriving trust relationships ......................... 146
5.4 Experimental setup ........................................... 150
5.5 Results ....................................................... 151
   5.5.1 Real-world social graph ............................... 151
   5.5.2 Trust relationships .................................... 152
5.6 Potential applications ....................................... 154
5.7 Conclusion .................................................... 156

6 A Prediction and Compensation Mechanism for Device Displacement 159

6.1 Introduction .................................................. 159
6.2 Background .................................................... 162
6.3 Methodology .................................................. 164
   6.3.1 Assumptions ............................................. 164
   6.3.2 Approach ................................................. 166
   6.3.3 Kinematics .............................................. 166
6.4 System Design ................................................ 170
   6.4.1 Mechanism .............................................. 170
   6.4.2 Activity detection model ............................. 171
   6.4.3 On-body position detection model ................... 172
   6.4.4 Prediction .............................................. 172
   6.4.5 Error models ........................................... 175
   6.4.6 Kalman Filter .......................................... 176
## Contents

6.4.7 Correction ................................................................. 179

6.5 Experimental setup ....................................................... 179

6.5.1 Dataset ................................................................. 179

6.5.2 Performance metrics .................................................. 180

6.6 Results ................................................................. 181

6.6.1 Performance Evaluation .............................................. 181

6.6.2 Error Evaluation ...................................................... 190

6.6.3 Discussion ............................................................. 195

6.7 Conclusion ............................................................... 197

7 Conclusion ................................................................. 199

7.1 Future work .............................................................. 203

Appendix A Tables .......................................................... 207

A.1 Summary of Mobile SSP Literature ................................. 207

Bibliography ................................................................. 243
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Application architecture on Mobile Social Signal Processing</td>
<td>13</td>
</tr>
<tr>
<td>2.2</td>
<td>Application architecture of existing Sensing Frameworks</td>
<td>16</td>
</tr>
<tr>
<td>2.3</td>
<td>State-of-the-art social interaction detection approaches</td>
<td>22</td>
</tr>
<tr>
<td>2.4</td>
<td>Behavioural cues emitted during a social interaction</td>
<td>27</td>
</tr>
<tr>
<td>2.5</td>
<td>Process of understanding social behaviour through social signals</td>
<td>58</td>
</tr>
<tr>
<td>3.1</td>
<td>Effect of window size in distance estimation</td>
<td>83</td>
</tr>
<tr>
<td>3.2</td>
<td>Spatial arrangement of people including the concept of interaction zones</td>
<td>85</td>
</tr>
<tr>
<td>3.3</td>
<td>Performance of various propagation models based on Bluetooth RSSI</td>
<td>87</td>
</tr>
<tr>
<td>3.4</td>
<td>Features for 2-Layer DHC</td>
<td>90</td>
</tr>
<tr>
<td>3.5</td>
<td>ROC Diagrams for DARSIS and state-of-the-art approaches</td>
<td>98</td>
</tr>
<tr>
<td>4.1</td>
<td>Spatial arrangement of people in a social situation</td>
<td>108</td>
</tr>
<tr>
<td>4.2</td>
<td>The DARSIS system architecture</td>
<td>111</td>
</tr>
<tr>
<td>4.3</td>
<td>Relative spatial arrangement in a social situation</td>
<td>120</td>
</tr>
<tr>
<td>4.4</td>
<td>Error distribution of DARSIS and error model</td>
<td>128</td>
</tr>
<tr>
<td>4.5</td>
<td>Real-world experiment environment</td>
<td>130</td>
</tr>
<tr>
<td>4.6</td>
<td>DARSIS energy consumption</td>
<td>132</td>
</tr>
<tr>
<td>5.1</td>
<td>From social interactions to trust relationships</td>
<td>144</td>
</tr>
<tr>
<td>5.2</td>
<td>Information flow in the MobTrust</td>
<td>146</td>
</tr>
<tr>
<td>5.3</td>
<td>Real-world social graph from MobTrust</td>
<td>151</td>
</tr>
<tr>
<td>5.4</td>
<td>Trust relationship colour maps from MobTrust</td>
<td>153</td>
</tr>
<tr>
<td>6.1</td>
<td>Device displacement on real-world data</td>
<td>161</td>
</tr>
<tr>
<td>6.2</td>
<td>Displacement on uDirect</td>
<td>165</td>
</tr>
<tr>
<td>Figure Number</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>6.3</td>
<td>Device displacement compensation system.</td>
<td>166</td>
</tr>
<tr>
<td>6.4</td>
<td>Movement projection planes.</td>
<td>167</td>
</tr>
<tr>
<td>6.5</td>
<td>Device displacement detection points based on kinematics.</td>
<td>169</td>
</tr>
<tr>
<td>6.6</td>
<td>Complementary Kalman filter.</td>
<td>170</td>
</tr>
<tr>
<td>6.7</td>
<td>Displacement compensation in trousers front pocket.</td>
<td>182</td>
</tr>
<tr>
<td>6.8</td>
<td>Displacement compensation in trousers back pocket.</td>
<td>183</td>
</tr>
<tr>
<td>6.9</td>
<td>Displacement compensation when hand held.</td>
<td>184</td>
</tr>
<tr>
<td>6.10</td>
<td>Displacement compensation when hand held and using.</td>
<td>186</td>
</tr>
<tr>
<td>6.11</td>
<td>Displacement compensation in backpack.</td>
<td>187</td>
</tr>
<tr>
<td>6.12</td>
<td>Displacement compensation in handbag.</td>
<td>188</td>
</tr>
<tr>
<td>6.13</td>
<td>Displacement compensation in shirt pocket.</td>
<td>189</td>
</tr>
<tr>
<td>6.14</td>
<td>Error distribution based on on-body positions.</td>
<td>191</td>
</tr>
<tr>
<td>6.15</td>
<td>Error distribution based on orientation tracking methods.</td>
<td>192</td>
</tr>
<tr>
<td>6.16</td>
<td>Error distribution for all on-body positions and projection planes.</td>
<td>194</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Motivation

In the recent years the popularity and computational power of mobile devices such as smartphones, tablets and wearable devices has led to a new era where they are substituting computers and means of communication. Moreover, in order to facilitate user experience, mobile device production companies have integrated various sensors such as accelerometer, gyroscope, GPS, digital compass, microphone and camera. Furthermore, the release of application stores has given the opportunity to third party developers to implement their own applications. The combination of the embedded sensors and the application stores may introduce radical changes in fields such as healthcare, environment monitoring and human behaviour recognition.

The main challenge of this area is how to mine information about human behaviour derived from the device without endangering the user experience. The applications are categorised into participatory, where the user has active involvement in context recognition and into opportunistic where user does not affect the process [1]. Research in opportunistic sensing has focused mainly on extracting knowledge on backend servers. On the other hand, there is a need for development of new techniques for executing the above mentioned processes on mobile devices. By leveraging information derived from mobile devices, the following questions arise:
• What human behaviours are mobile devices capable of accurately recognising despite their limitations and constraints? Or more specifically to what extent can mobile devices contribute to a machine understanding of the social context of users?

• Is a mobile device capable of dealing with the demanding amount of energy, computational burden and memory constraints?

• What improvements must be applied to current techniques in order to accomplish knowledge extraction about human behaviour on mobile devices?

The execution of human behaviour recognition on mobile devices presents some major advantages [2]:

• It eliminates WiFi and Cellular communication with backend servers in order to infer user context, thus minimising the communication load and improving a mobile device’s battery life.

• Minimises financial cost due to cellular communication.

• Protection of user private information as behaviours are recognised locally and information leaving the phone can be better controlled.

• Allows user to be involved in the generation of labels concerning recognised context, allowing semi-supervised learning techniques to be leveraged through a crowd sourcing approach.

Based on the above advantages, it is hypothesised that mobile devices constitute the most qualified devices to infer social contexts of their users. The social interaction detection process is a very critical part of the context inference process. It constitutes the ground base for retrieving information regarding user’s social behaviour. Thus, a generic social interaction detection method for mobile devices will plausibly widen the current range of user context recognition applications, acting as a foundation of the area but also as a quantification mean of various aspects of user’s social daily life such as sociability [3], social role [4] and trust relationships among people.
1.2. Objectives

The key research problem that this study is striving to answer is to what extent could mobile devices contribute in understanding different aspects of social behaviour such as the trust relationships among people by detecting and combining several behavioural cues and social interaction information in an accurate, non-intrusive and privacy-preserving way.

1.2 Objectives

While mobile devices continue to provide more computation, memory, storage, sensing and communication bandwidth combined with their pervasive and ubiquitous character, they become the most qualified device for sensing and inference of social context [5] [6]. The aim of the present study is to develop techniques to extract knowledge about human behaviour on mobile devices. These advanced techniques will be essentially based on machine learning in order to recognise user social context and increase the inference accuracy which has been neglected due to device limitations. In the following subsections the main objectives of this study are outlined.

1.2.1 Interpersonal distance estimation with smartphones

Proxemics i.e. interpersonal distance among people has initially preoccupied research psychology, sociology etc. and later on the field of informatics. Initial works in psychology understood the importance of interpersonal distance among people by observing in the beginning the spatial arrangement of animals when they were in vicinity and later on followed similar approach for people. Hall [7] after conducting real-world experiment to understand the effect of interpersonal distance, concluded on creating different interaction zones based on the distance and mapping these zones into human social relations. This was the starting point towards understanding and measuring the interpersonal distance among people. Informatics strived to develop self-acting tools that would measure the interpersonal distance among people. Current approaches leverage obtrusive technologies in order to understand the interpersonal distance among people, to achieve high accuracy in the estimation. Also, in many cases they require from the
user to modify the firmware in order to deploy a particular configuration. Empirical and experimental evaluation (See Chapter 3) showed that state-of-the-art techniques have limited accuracy in estimating interpersonal distance as they are not able to cope with the signal fluctuation observed in real-world environment. Another major drawback of existing techniques for interpersonal distance estimation is the requirement for a large number of samples in order to perform inference. This induces a considerable amount of delay in the sensing and inference process and also may provide faulty estimations due to the usage of outdated samples.

In order to tackle the above drawbacks to infer accurately and in real-time the interpersonal distance estimation there is a need to develop a novel machine-learning technique that will consider a small number of signal-samples. Leveraging the Bluetooth received signal to estimate interpersonal does not induce any requirement about firmware modification and also provides the native capability for ad-hoc discovery among the devices. In order to tackle the fluctuation of the Bluetooth received signal, a hierarchical approach of machine-learning models combined with informative features and robust classifier could provide an accurate and reliable approach for estimating interpersonal distance. To select the most informative features there is a need to create a bank of features including a large number of samples that consider the human body absorption and then apply feature reduction and classifier wrapping techniques. To understand the applicability of the proposed interpersonal distance estimation technique, it will be evaluated against the state-of-the-art techniques in indoor environment.

1.2.2 Modelling social interaction detection with smartphones in real-world

A prerequisite for extracting social signals and in general human social behaviour, is the detection of social interaction among people with the above mentioned benefits of mobile devices. A first step towards mining information leading to possible social interactions is the identification of users’ spatial arrangement indicating that they are potentially interacting. Following in the rest of this report, the terms of social relation or social interaction will refer to information about the relative spatial arrangement
of the users during a social interaction. The detection of social interactions consists initially of two main parameters: a) relative orientation and b) interpersonal distance. The relative orientation of users requires modelling the knowledge of the direction of each of the participants. This cognition could be acquired by leveraging the capabilities of uDirect algorithm [5], which uses the walking locomotion of the user and only through two sequential steps is able to estimate his facing direction. Subsequently, proximity distance was leveraged in order to estimate if users’ interpersonal distance allows them to perform social interactions. For interpersonal distance estimation, the state-of-the-art techniques were assessed and concluded in developing two novel probabilistic models based on Bluetooth RSSI that detect: a) Interaction zone and b) Proximity. Interaction zone model provides a fine-grain estimation of the interpersonal distance of users. Proximity model, given the Bluetooth RSSI, infers if the users are in proximity to perform a social interaction. The developed system for detecting social interactions operates in a fully distributed manner, both the sensing and inference part. This is achieved by exchanging the user direction from uDirect algorithm and the Bluetooth RSSI logged. Thus, each device has retrieved the required information to carry out its independent inference. Further details about the system and the results of its evaluation are provided in Chapter [4].

1.2.3 Quantifying trust relationships from social interactions

Deriving trust relationships from real-world social interactions may contribute significant information towards social behaviour understanding. The level of trust among people constitutes a parameter for describing the social context but also an important measure for security and privacy in pervasive systems. Current works for deriving trust relationships either consider only on-line social networks to create trust networks or focus on users’ on-line social interactions. An opportunistic sensing system may allow the derivation and quantification trust relationships among people through smartphones based on the detected real-world social interactions. A real-world social graph may be derived from users’ daily social interactions by also considering snapshots of their social relation. A hybrid model is developed to quantify users’ trust relationships based on the extracted real-world social graph, the estimated social relations and the contex-
tual information provided by the detected social interactions. As a proof of concept, a real-world evaluation of the system is performed.

Trust has received many definitions depending on the context in which it is defined [8]. Among the most predominant definitions is the "generalised expectancy held by an individual that the word, promise, oral or written statement of another individual or group can be relied on" [9]. Quantifying such a complex notion constitutes a great challenge and requires particular care in the process followed [10]. In order to quantify this multidimensional notion there is a need to combine multiple modalities and behavioural cues to derive a coherent measurement of trust. This work will strive to provide an initial measurement of trust by leveraging information from social situations, given the importance of trust in social interactions [11]. In particular behavioural cues will be extracted from social interactions and social situations, which will be combined in a simplistic manner to try to derive an initial measurement of trust.

1.2.4 A prediction and compensation mechanism for device displacement

The pervasive computing paradigm has led to the current situation where a user’s environment is full of mobile and wearable computing and communication devices. Equipped with dozens of sensors, these devices allow to gather many different types of data produced by the user, and to infer contextual information and knowledge. Based on this various new services can be provided to the user. Such contextual knowledge has already been utilised in various types of applications including fitness, eHealth, behavioural, localisation etc. However, the devices are used in peoples’ daily routine, this means they usually operate in unconstrained and real-world situations. This leads to arbitrary movements that constitute a permanent source for errors for applications which work on the assumption that the device is following user’s body movement [5] [12] as it is considered attached to or implanted in user’s body. Common examples of these situations are, displacement of the mobile phone in users’ trousers’ pocket while they are walking, or users with sensor device implanted. This type of displacement can lead to arbitrary movements and erroneous conclusions based on the sensor data collected.
1.3. Contributions

There is a body of work that aimed to cope with the device displacement through filtering techniques or through analysis of the effect of device displacement in activity recognition \[13\] \[14\]. Further, occurrence of device displacement is detectable when user is standing, as there are no linear or angular accelerations applied on the device, excluding gravitational. However, there is no mechanism that tries to detect and compensate device displacement while the user is walking. The proposed approach here identifies particular reference points in the walking cycle and in an efficient manner it allows the estimation and compensation of device displacement.

1.3 Contributions

Due to the benefits of detecting human behaviour, the necessity of mining social signals in an automatic, consistent and non-intrusive manner, has been raised. This study intends to provide a set of tools that allow continuous monitoring, detection and inference of social behaviour of people through mobile devices.

The main contributions focus on:

- Development and evaluation of a novel technique for inferring interpersonal distance among people to detect when they are in physical proximity and in which interaction zone, by utilising machine learning methods.

- Modelling, development and evaluation of an opportunistic and collaborative system for real-world environments that leverages the proposed interpersonal distance estimation and the improvement of a state-of-the-art technique for estimating user’s direction to understand social interactions among people.

- Research on understanding and modelling of trust relationship among people based on social interactions and additional contextual information in real-world environments.

- Design, development and evaluation of a novel mechanism for mobile devices that detects and compensates the device displacement error induced due to arbitrary movements in users daily life.
1.4 Outline

The remainder of this report is structured as follows. Chapter 2 surveys the existing literature in Mobile Social Signal Processing including the state-of-the-art solutions for detecting social interactions among people. Chapter 3 presents the initiative approach of solving the challenge of estimating interpersonal distance among people through smartphones. The overall social interaction detection system is described in Chapter 4. Leveraging the social interaction detection system, a novel quantification mechanism for trust relationships is proposed in Chapter 5. Chapter 6 proposes a novel mechanism for predicting and compensating device displacement for mobile devices. Chapter 7 provides a conclusion and future work regarding this research.
Chapter 2

A Survey on Mobile Social Signal Processing

This chapter provides an overview of the area of Mobile Social Signal Processing (SSP) while striving to shed some light on the fundamental process of inferring social behaviour on mobile devices. It initially clarifies the core controversial terms of the field, continues by introducing the stages of mobile inference of human behaviour. Finally, it argues about potential applications in three main areas healthcare, corporations and marketing while describing the current shortcomings in literature regarding context recognition, multi-modal fusion, interdisciplinary character of the area and energy efficiency, which will be tackled in this research.

2.1 Introduction

Human behaviour understanding has received a great deal of interest since the beginning of the previous century. People initially conducted research on the way animals behave when they are surrounded by creatures of the same species. Acquiring basic underlying knowledge of animal relations led to extending this information to humans in order to understand social behaviour, social relations etc. Initial experiments were conducted by empirically observing people and retrieving feedback from them. These methods gave rise to well-established psychological approaches for understanding human behaviour,
such as surveys, questionnaires, camera recordings and human observers. Nevertheless, these methods introduce several limitations including various sources of error. Completing surveys and questionnaires induces partiality, unconcern etc. [15], human error [16], and additional restrictions in scalability of the experiments. Accumulating these research problems leads to a common challenge, the lack of automation in an unobtrusive manner.

An area that has focused on detecting social behaviour automatically and has received a great amount of attention is Social Signal Processing (SSP). The main target of the field is to model, analyse and synthesise human behaviour with limited user intervention. To achieve these targets, researchers presented three key terms which constitute different levels of abstraction in the process of educating social behaviour [17] [18] [19]. Behavioural cues include various characteristics of human behaviour that are extracted from a modality such as prosody of the voice and interlocutors spatial arrangement. The combination of these behavioural characteristics indicate a person’s current sentiment, understanding, attention, interest etc. which are social signals. Pentland [20] described social signals as non-verbal communication signals emitted when people are socially interacting. Merging these social signals in a longer temporal term leads to a person’s social behaviour. In recent literature the terms have been used in other areas such as social networks [21] to indicate every social related Internet activity of a user. However, this aspect is not considered in behaviour inference. Social networks may function as an enhancement of SSP to provide additional information regarding the context but in this work the two areas are considered distinct.

In [22] a generic procedure was proposed to detect social behaviour:

1. Data capture.

2. Person detection.

3. Extraction of audio and visual behavioural cues, and their mapping to social signals.

4. Incorporate context to detect social behaviour from social signals.
This procedure is focused on detecting the social behaviour of people through audio and visual data, from an external observer’s point of view. In order to achieve this observation, microphones and cameras are required to be deployed on the scene to monitor people. The major disadvantages of this approach are a) limited mobility of the system, where in case of the requirement for conducting an experiment in a different area there is a need for re-deploying and re-configuring the system to the specific environment, b) the confinement in scalability because the equipment is deployed at a certain environment and cannot follow the user’s mobility c) social signals are emitted during social interactions and when the detection process is based on audio and visual data, there is a need to perform person detection which is neglected, and finally d) establishing ground truth in audio and visual data requires labelling that is a time-consuming process and may induce human error.

2.1.1 A mobile and opportunistic point of view

The purpose of this chapter is to review state-of-the-art techniques for extracting social behaviour through mobile phones and also to introduce a discussion on the remaining challenges, existing gaps and potential extensions of existing solutions of the area. Understanding social behaviour in an automatic, non-intrusive, mobile, but also scalable manner constitutes a significant challenge with several potential applications. To address this challenge, close collaboration is required from the fields that accord two of the most important components of the field, information technology and psychology [23]. This collaboration will support the development of opportunistic non-intrusive self-acting tools for extracting human behaviour. These tools will expunge several sources of error introduced by current obtrusive and user engaging methods that incorporate human factor in the sensing process. In parallel, SSP focused on providing concrete solutions regarding modelling, analysis and synthesis of social behaviour. However, as mentioned some major gaps have been identified.

In order to fill these gaps the following objectives were determined, which will drive the research on extracting personalised social behaviour a step further.

- Utilising non-intrusive approaches.
- Capturing cues from user’s perspective, to produce personalised data.

- Leverage multiple modalities, to extract more robust and reliable behavioural information.

- Continuous sensing and inference process, without mobility and scalability restrictions.

- Elimination of external hardware requirement.

Smartphones have become a core feature of peoples’ daily lives. In recent years, popularity and computational power of mobile phones have led to a new era where they are substituting computers and other means of communication such as old feature phones, fixed line phones etc. Moreover, to facilitate a more rich user experience, mobile phone manufacturers have integrated various sensors such as an accelerometer, gyroscope, GPS, digital compass, microphone, camera etc. Furthermore, on-line application stores have given the opportunity to third party developers to implement their own applications utilising available integrated sensors seamlessly. Combining embedded sensors and application stores will introduce radical changes in fields such as healthcare, environment monitoring and human behaviour recognition by allowing easy, non-intrusive and wide deployment of mobile applications.

Given the pervasive and ubiquitous character of mobile devices and considering the built in sensing features, smartphones are considered as ideal devices for extracting social behaviour among people. To support this claim, Mobile Social Signal Processing (Mobile SSP) is introduced while proposing the main architecture of human behaviour inference for mobile applications. Further, each stage is analysed by providing state-of-the-art techniques capable of being executed on mobile devices. Also, potential application cases will help to familiarise the reader with areas that will benefit from the growth of Mobile SSP, followed by a discussion of research opportunities that may be leveraged for further contribution to the field.

In the remainder of this chapter a survey for Mobile SSP is provided. Section 2.2 describes the overall area of Mobile SSP while clarifying the core terms of the field. A brief description of existing sensing frameworks is shown in Section 2.3 and assists the
2.2 An overview of Mobile Social Signal Processing

Mobile devices and in particular smartphones are ubiquitous. Multi-modal sensing capabilities combined with increased computational power and available tools for mobile application development led to the view that smartphones are ideal devices for filling the gap of lack of automation in social behaviour understanding. Users can easily install an app from on-line application stores without any geographical restrictions and the device will automatically become a human behaviour aware smartphone. To discriminate and categorise more easily the types of applications two classes have been defined: participatory are the social behaviour detection applications that require the user’s participation in the sensing process and opportunistic where the user is not

reader in the selection criteria. State-of-the-art techniques utilised to detect social interactions among people on mobile phones are presented in Section 2.4. Behavioural cues extracted on smartphones whilst informing about their advantages and disadvantages are described in Section 2.5. Section 2.6 showcases methods of mining social signals and mapping them to social behaviours. Section 2.7 describes existing and potential applications of Mobile SSP. An overall discussion about methods presented for extracting social behaviour on mobile phones is argued in Section 2.8. Finally, challenges of the area are outlined in Section 2.9 and the chapter concludes with Section 2.10.

2.2 An overview of Mobile Social Signal Processing

Figure 2.1: Application architecture on Mobile Social Signal Processing.
involved in the process [1]. In order to minimise the obtrusiveness of the system and secure
the user’s spontaneous behaviour, the main attention of the work is focused on
opportunistic social behaviour detection applications.

Similar to [24] and based on the literature review it was concluded that the following
steps need to be taken for extracting social behaviour on mobile devices (See Figure 2.1).

1. Sensing.


3. Extraction of Behavioural Cues.

4. Understanding Social Behaviour by Inferring Social Signals.

Social behaviour inference on mobile devices is initiated by the Sensing process. During
daily life, users emit behavioural cues and social signals, which are captured by sensors
of the mobile device. These sensors may be integrated in the device or enclosed in
external hardware that communicates with the mobile device. Each sensor detects a
particular modality, then it converts the detected signal into a raw data signal. The
result is processed into a desired format or is directly forwarded to the next stage of
social behaviour inference. Researchers have developed various sensing frameworks to
allow developers to collect data in an abstract and uniform way, while in some cases
they also include an inference engine.

Pentland recognised the emission of social signals during a social interaction [25]. This
signifies the importance of recognising social interactions before initiating the process
of social behaviour understanding. After retrieving the appropriate data from mobile
device’s sensors, Social Interaction Detection may be performed as a preprocessing step
of social behaviour inference. Understanding social interactions provides important
contextual information that may be leveraged in the next steps of social behaviour
inference. The knowledge of on-going social interactions may also be utilised for filtering
data and allowing the development of adaptive sensing and inference techniques. In
applications focusing on extracting behavioural information not related to the social
aspect of a person, it is strongly encouraged to include this step as it provides important
contextual information.
Following the identification of on-going social interactions is the *extraction of behavioural cues*. Different modalities may be leveraged for the extraction of a behavioural cue, depending on the grammar defined in psychology. Each selected sensed modality is forwarded to behavioural cues extraction. Existing literature has been classified into seven categories based on the types of cues each work extracts (See Figure 2.1). The behavioural cues extraction is achieved through techniques such as decision models, statistical analysis etc.

The final stage of Mobile SSP is the transition from the understanding of *social signals* to *social behaviour inference*. Close collaboration with social sciences may provide the theoretical mapping among behavioural cues, social signals and social behaviours. Literature has been grouped based on the inferred social behaviour through mobile phones. The extracted behavioural cues are fed in decision making techniques to mine social signals and infer in long-term social behaviour.

To facilitate the reader’s understanding of the field, an outline of the main steps and requirements is provided for an integrated and real-world-enabled Mobile SSP:

- Define the context of the Mobile SSP application.
- Select the modalities required to infer a particular social behaviour.
- Define the grammar of behavioural cues and social signals that will lead to social behaviour inference.
- Evaluate and verify the reliability of the approach in a real-world environment based on ground truth.

In addition to the above requirements, researchers need to consider the intrusiveness, security and privacy of the system. Researchers need to take into account the computational burden and energy consumption which may endanger user experience. These parameters do not constitute a prerequisite for the realisation of Mobile SSP but will facilitate user experience and privacy.

In the following sections, each of the pre-defined stages will be analysed and state-of-the-art research are outlined. The works described in the next sections are summarised.
Figure 2.2: Application architecture of existing Sensing Frameworks.

in the Electronic Appendix, introducing the techniques developed in each stage of social behaviour inference.

2.3 Sensing frameworks

Sensing is the first stage in extracting human behaviour on mobile devices. In this stage, selection of appropriate modalities is performed. These will later on be processed and analysed to reveal information about user’s social behaviour. It constitutes the lowest level of the process, which collects raw data from sensors and other interfaces that can provide information relevant to the user (See Figure 2.2). After retrieving information from sensors either the raw data are forwarded to the next stages or lightweight and simplistic processing may be performed to minimise the complexity and computational burden at the upcoming stages. As shown in Figure 2.1, the next stages in social behaviour inference may be performed either on the device or at a backend server.

This section introduces and then compares existing sensing frameworks. Through this introduction, the reader should be able to understand the criteria based on which sensing framework should be selected for a desirable social behaviour application. An extensive analysis of existing sensing frameworks is outside the scope of this research.
2.3. Sensing frameworks

Table 2.1: Data collection frameworks for offline analysis

<table>
<thead>
<tr>
<th>Framework</th>
<th>Sensor Types</th>
<th>Energy Efficient</th>
<th>Privacy</th>
<th>License</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inertial</td>
<td>Ambient</td>
<td>Position</td>
<td>Virtual</td>
</tr>
<tr>
<td>MyExperience [26]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SeeMon [27]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>AnonymSense [28]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>OpenDataKit [29]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>PRISM [30]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LiveLab [31]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SystemSens [32]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Funf [33]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Medusa [34]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>METIS [35]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MSF [36]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

and the reader is referred to [1] and [2].

2.3.1 Data collection

This subsection focuses on sensing frameworks that perform only data collection and some minor pre-processing allowing the extraction of human behaviour off-line. An application is deployed on the device, which collects data from pre-configured sensors, and then either stores the information on device’s internal storage or performs uploading to a backend server. The retrieved data are merged and forwarded to the next stage in order to extract behavioural cues. Table 2.1 summarises existing frameworks for data collection by presenting a) the type of sensor data retrieved, b) the incorporation of a mechanism for energy efficiency such as adaptive sensing, c) the embodiment of a privacy preserving approach and finally d) the software license through which the authors released a particular framework. Urban sensing [37] is not included in this literature as it falls out of the scope of this research.

*MyExperience* [26] retrieves and logs contextual information from physical and virtual (e.g. phone usage patterns) sensors. *SeeMon* [27] investigates the context of the device and adapts the sensing process, by mining local sensors and installed applications in an energy efficient manner. To preserve privacy in crowd-sensing applications, [28] in-
troduced *Anonsense* that permitted, through a centralised sensing coordination, the collection and storage of anonymous smartphone sensed data in a collaborative manner. *OpenDataKit* [29] is a set of tools for sensing and aggregating data from mobile phones. Furthermore, [30] developed *PRISM* a platform for dynamic deployment of participatory and opportunistic sensing applications on mobile phones, while maintaining security through a sandboxed environment. Another approach that focuses on privacy-preservation but also on energy efficiency for long-range duration deployment is *LiveLab* [31], which stores inertial, phone usage, positioning and network-based data. *SystemSens* [32] is a data retrieval tool that enables researchers in capturing mobile data in large scale experiments. *Funf* [33] is a mobile data collection platform developed by MIT Media Labs, procuring easily configurable sensing and data retrieval capabilities. *Medusa* [34] achieved crowd-sensing and simultaneous coordination of multiple mobile devices. *METIS* [35] is a distributed system that decides, based on the device status and user context, to perform on-device or infrastructure-oriented sensing. *MSF* [36] is a recent data collection framework that complies to multi-pipeline architecture and targets in providing an abstraction regarding the sensing process.

Discussion. Examining the data collection frameworks, *MyExperience* [26] constitutes an event triggered approach that is energy efficient and does not require any polling process to identify state changes but does not utilise data from inertial sensors. *OpenDataKit* [29], *SystemSens* [32] and *Funf* [33] are three distinct configurable open-source data collection tools that allow off-line merging of data from different sensors. However, they do not perform adaptive sensing based on the context to improve energy efficiency. This is critical for this type of applications. *Medusa* [34] allows a coordinator to retrieve a certain type of sensor-data from a specific device. Furthermore, *METIS* [35] is the first work that lightens a mobile device by selectively perform sensing through the infrastructure but simultaneously narrows the mobility and increases the intrusiveness of the system. *MSF* [36] is focussing on easing the development of sensing applications. It requires the designer to implement the sensing functionality but handles tasks, power management and resource allocation. However, it does not constitute a ready to deploy solution and is suitable only for developers.
2.3. Sensing frameworks

2.3.2 Inference engines

As opposed to the previous subsection, this section includes frameworks that perform sensing and inference on the device or on a backend server. The sensor selection is predefined or configurable depending on the implementation. Data are retrieved from the sensors and forwarded to an inference pipeline. Based on the inference, one or more preprocessing stages could be performed. Then, the appropriate intelligence is applied to retrieve the requested knowledge. An optional post-processing phase, such as the consideration of historical inferences, outlier detection, smoothing etc., may be applied to remove results that deviate from normal. Table 2.2 presents the state-of-the-art inference engines and identifies a) the type of information that is extracted by each framework, b) the development of an energy efficient approach and c) the software license of the framework.

CenceMe [38] is a distributed platform that performs multi-modal sensing through mobile phones. A classification-based technique decides about inferring social context on the device or on a backend server. It also allows the user to publish the inference to social networks. BeTelGeuse [39] was one of the first tools that had the native capability of sensing and inferring about user’s context. Jigsaw [40] is a mobile platform that allows continuous data collection in an energy efficient way, through multiple pipelines (one for each modality) and adaptive sampling based on user behavioural patterns. EmotionSense [41] is a framework for inferring user emotion and incorporates an intelligent engine for adapting the sensing process. As an extension of it, SociableSense [3]...
measures sociability of people and introduces an adaptive inference mechanism (locally or distributed) based on reinforcement learning. AmbientDynamix \[42\] is an equally important framework that allows the deployment of custom inference modules in a sandboxed environment. Also Auditeur \[43\] is a context recognition framework that is focused only on the audio perspective, but provides a collection of inference mechanisms for the specific modality.

**Discussion.** Regarding state-of-the-art sensing frameworks that have a human behaviour inference, CenceMe \[38\] performs preliminary detection of activity and conversation. However only an application that publishes user context to social networks is publicly available. BeTelGeuse \[39\] focused mainly on the sensing process, enabling the integration of external Bluetooth-connected sensors. It also allowed the incorporation of inference through plug-ins while initially providing location and activity classification components. Jigsaw \[40\] limits its sensing capabilities to the accelerometer, microphone and GPS but provides integrated classification techniques for activity and voice recognition. Through a multi-threaded approach they try to limit the computational burden on the device due to the classification process. EmotionSense \[41\] and SociableSense \[3\] are based on the same framework, providing a quantification method for the user’s emotion and sociability whilst performing adaptive inference through learning techniques. It is available for developers but also for direct utilisation of the application for less technical people.

Furthermore, AmbientDynamix \[42\] allows the user to select existing or concrete components, integrate them in a main skeleton application and perform the desirable social behaviour detection. If the component exists, it constitutes an easy and reliable solution while if the module requires development it can be contributed to the community for further reuse. All the processes of sensing configuration, data logging, resource management, concurrent procedures are handled in a seamless manner by the skeleton application, which reduces the developer’s effort. If the targeting system is focused on mining social behaviour information through audio data, Auditeur \[43\] constitutes a reasonable solution that provides the appropriate mechanisms to extract audio features but also allows the configuration of the desired classifier. In addition, it includes state-of-the-art techniques for contextual sound recognition.
2.3.3 Framework comparison

A notable amount of works targeting sensing frameworks for mobile phones was briefly described in the two previous subsections. The literature was classified based on whether the framework enabled human behaviour inference or not.

Overall, the first step in the design of a social behaviour detection application is the decision about the sensing framework. Many researchers start by designing and developing the sensing process from scratch. However as shown, works on sensing frameworks have reached a certain maturity which allows component reuse. These frameworks provide off-the-shelf solutions for resource management, concurrency, data handling, energy efficiency and concrete structure of the application. This should be leveraged in order to reduce the development time cycle, human error and increase code reuse. Most of them are released with open licences, allowing clear understanding, editing but also contributing of the source code, from the research community. Selecting a data collection or inference engine framework is highly dependent on the targeting application and how sufficient the capabilities are of each framework with respect to the researchers’ envisioned outcome. Thus, it should be noted that selecting a certain framework does not lead to a right or wrong decision but in a tool that will provide more or less enabling capabilities for developing a social behaviour detection application.

2.4 Social interaction detection

The next stage of retrieving data from sensing the context continues is recognising ongoing social interactions. People are assumed to interact socially, when they are in close interpersonal distance, facing each other and participating in a conversation. Pentland definition of social signals [25], is that they are non-verbal communication signals that are conveyed when people are socially interacting. Thus, identifying possible social interactions accurately is an important stage of social behaviour understanding and requires tackling.

Researchers have developed several techniques to detect social interactions through smartphones. These techniques vary based on the level of accuracy and modalities
Chapter 2. A Survey on Mobile Social Signal Processing

(a) Single Modality  
(b) Multiple modalities

Figure 2.3: State-of-the-art social interaction detection approaches.

included in the inference process. Among these approaches are single modality that include coarse or fine grain distance estimation through Bluetooth and WiFi interfaces, audio-based distance or relative position estimation. There is also multi-modality where different modalities are combined in the inference process to indicate if people are interacting (See Figure 2.3).

2.4.1 Single modality

The majority of commercial off-the-shelf (COTS) smartphones today comprise wireless communication interfaces such as Bluetooth and WiFi. Due to their wide availability researchers often utilised them in order to detect when people are in proximity. By acquiring information about people being in proximity, researchers made a strong assumption of interpreting the knowledge of proximity into the existence of social interaction. An implicit assumption is that every person is required to carry a smartphone. For the determination of proximity, several techniques have been proposed.

The most common approach is to perform discovery through one of the two interfaces, log the Bluetooth IDs (BTIDs) or WiFi Service Set Identifier (SSID) and classify all the detected nearby devices as social interactions. This method was applied in miscellaneous works to estimate when people are interacting with each other. Some examples
of these works are Serendipity [44], CenceMe [38] and SoundSense [45], [46] aggregated the discovered smartphones based on BTIDs with static nodes. The Bluetooth ranges were overlapping to improve social interaction detection and provide some information about coarse-grain localisation. The accuracy of this method is limited to the range of the communication mean i.e. for Bluetooth the minimum nominal range is around 10m [47] and for WiFi the typical range is approximate to 35m for indoor environment\footnote{Bluetooth and WiFi ranges are highly dependent on the surrounding environment and chipsets characteristics of devices performing the discovery and the detected devices.}. Thus, every device - person detected is classified as being in a social interaction. It should be noted that these works do not provide error analysis of this social interaction detection approach.

The previous method introduces a noticeable amount of error. For that reason researchers focused on developing distance estimation techniques that would remove a percentage of error from the previous approach by limiting the communication mean range. So, for detecting social interactions [48] developed a machine learning based technique to estimate if users were in proximity or not, by retrieving WiFi Received Signal Strength Indicator (RSSI). They trained a model based on maximum and mean value of a 20-sample window of WiFi RSSI achieving a median error of 0.5m for social interaction detection. In particular, they detected ongoing social interactions with 86% accuracy and true negatives with 84%.

Another approach for detecting social interactions through Bluetooth RSSI-based distance estimation was presented in [49]. They developed a probabilistic model for indoor and outdoor environments. It utilised some predefined empirically calculated thresholds to compute the probability of being in proximity to socially interact, with a claimed error rate of 4.3%. In detail, they retrieved an RSSI measurement which was smoothed through exponential window moving average (EWMA) and a smoothing factor of 0.5. To compute the thresholds they retrieved Bluetooth RSSI measurement in different environments and distances. Bluetooth technology natively operates in a mode which allows the device to discover but also to be discoverable by other devices without any firmware modifications. A less complex approach was presented in [50], namely MAUC. This work focused on detecting social interactions through Bluetooth RSSI thresholds,
showing a detection rate over 90%. As opposed to previous approaches, it incorporated an adaptive discovery interval scheme, dependent on user’s activity movement.

**Discussion.** Scientists tackled the detection of social interactions on mobile phones with different granularity. The discovery of nearby devices (e.g., Serendipity [44], CenceMe [38] etc.) is easily implementable. However, it provides increased number of false positives due to inaccuracy in the interpersonal distance estimation, unawareness of spatial arrangement and conversation existence. As an example, Figure 2.3a shows four people in vicinity, where only (A) and (B) are interacting, although all four are in discovery range, thus classified as socially interacting. The WiFi interface on COTS smartphones cannot operate simultaneously in discovering and discoverable mode as opposed to WiFi Direct. For that reason, the authors [51] [52] were forced to modify the firmware of the devices to switch between the two modes. As an improvement, several works tried to estimate the interpersonal distance of users, to infer the existence of social interactions. For distance estimation based on RF (Bluetooth or WiFi) signals, RSSI has been leveraged in order to create empirical models, mainly due to simplicity in implementation. However, RSSI measurements received on mobile phones showcase great fluctuation which is affected by different environments, obstacles, human body absorption, reflections etc. Machine learning techniques constitute a notable effort to tackle the RSSI fluctuation [48]. Threshold-based approaches [49] [50] usually require adjustment of the algorithm’s boundaries based on the device and the environment. Other techniques such as Time Difference of Arrival (TDoA), Angle of Arrival (AoA) and Time of Arrival (ToA) showcase significant limitations such as firmware modification for device time synchronisation, lack of multiple antennas, the need for external hardware and therefore are not recommended for COTS mobile phones.

From simple discovery-based method, researchers have improved the social interaction detection through proximity. However, the assumption that when people are in proximity then are definitely interacting is strong. Hence, there is an imperative need to add other modalities in the inference process which will add new parameters such as spatial arrangement and conversation existence detection.
2.4.2 Multi-modality

To improve social interaction detection based on a single modality, researchers started to incorporate other modalities. These modalities enhanced social interaction detection by providing information about users’ relative orientations i.e. if they are facing each other or not and about the conversation existence i.e. if the users are really having a conversation and they are not two people having a spatial formation suitable for social interaction (See Figure 2.3b).

An important attempt to identify the relative spatial arrangement of the users combined with proximity detection was Virtual Compass [52]. This system utilises multiple RF interfaces such as Bluetooth and WiFi in order to estimate the interpersonal distance among users but also create a 2D localisation map based on users’ relative distances. For distance estimation given RSSI measurements from both interfaces, they computed the average distance and the uncertainty based on the difference of 90th and 10th distance percentile. Then, the authors applied regression on these features for distance estimation and achieved an error margin of 1.41m. For relative map localisation they propose a quick technique to estimate the initial coordinates of each user then they apply an iterative method [54] to refine the initially reckoned coordinates.

Matic [52] argued about a slightly different multi-modal approach that incorporated interpersonal distance estimation merged with relative orientation calculation and conversation detection. In this work, authors endeavoured to increase the accuracy of social interaction detection by taking into consideration users’ facing directions. The knowledge of user’s facing direction with respect to earth’s coordinates allowed them to compute the relative orientation of each pair of users, in order to understand if they had the appropriate spatial arrangement to interact. They estimated the interpersonal distance of the users through an initial calibration phase which led to a proximity detection model. During that period, they collected WiFi RSSI measurements at 1m distance and then based on an indoor path loss model (PLM), they created an artificial dataset for distances 0.5m, 1m . . . , 5m. Based on this dataset they computed the mean and maximum value of a 20-sample window and trained a Naïve Bayes with Kernel Density Estimator (KDE) model for proximity detection. An external accelerometer was at-
attached to a user’s chest to monitor his speech activity by performing spectral analysis of the signal it produced.

Discussion. In order to achieve higher accuracy and robustness in various environments, researchers combined various modalities. Virtual Compass [53] combined WiFi and Bluetooth to improve distance estimation and provide relative spatial arrangement detection. However, the simultaneous utilisation of these interfaces cannot be utilised for continuous sensing due to high energy consumption and lack of ad-hoc communication in current available WiFi on COTS mobile phones. For further improvement of social interaction detection on smartphones, [52] designed a work that provides a relatively accurate approach. As Figure 2.3b indicates, this method is able to identify correctly that (A) and (B) are socially interacting while (C) and (D) are not. They tracked user’s orientation and logged speech activity through an accelerometer attached to a user’s chest. Yet, the smartphone was in a fixed body position and the external accelerometer was intrusive as it was placed on user’s chest.

2.4.3 Apposition of social interaction detection approaches

In this section, detection methods for social interactions among people were surveyed by presenting different approaches that researchers have developed to tackle this problem. It is important to understand the strengths and the limitations of each method. As mentioned, the single modality works provide the benefit of low complexity, unobtrusiveness through limited constraints regarding the wearing position and the lack of external hardware. However, the approach is characterised by a large amount of false positive errors, which depending on the application could be neglected or require tackling. In case this error is not acceptable, a multi-modal approach is more suitable that incorporates the user’s facing direction and conversation detection. Although, these multiple modalities provide additional information to tackle the social interaction detection problem, researchers may consider the accumulated error introduced by each modality. Through efficient fusion of several modalities, the error of each individual modality may cancel each other out, driving the system to a less erroneous approach. Despite the numerous works on attempting to detect social interactions on mobile
phones, currently there is no tool that may be utilised in a real-world environment, without any constraints and with minimum intrusiveness. Ideally the framework may not require any firmware modification. Also, it would be able to be deployed on off-the-shelf smartphones providing a realistic and robust accuracy in a real-world environment, depending on the target application given only the integrated sensors of the device. Finally, both approaches are characterised by a trade-off among user friendliness, system complexity and accuracy [4] that should be considered by researchers depending on the needs of each application.

2.5 Behavioural cues extraction

Social interaction detection provides personalised information about who is interacting with whom. As Figure 2.4 shows, during a social interaction interlocutors emit cues such as spatial arrangement, posture and gestures indicating social signals such as intimacy, interest, mirroring etc. Acquiring this knowledge leads to the next stage of mining human behaviour on mobile devices, which is the extraction of behavioural cues. In this process, data obtained by sensing procedure are pre-processed or classified through a machine learning technique to retrieve some features that will be utilised in the next stage, the inference of social signals.
2.5.1 Auditory

In SSP, literature has focused mainly on extracting social signals from audio and visual data. This fact procures a reasonable indication regarding the importance of behavioural cues extracted from audio data. This section aims to provide a brief overview of the techniques applied on acoustic data recovered from ambient sensors and especially from a mobile device’s microphone. The techniques are categorised based on the type of features extracted. The same classification was performed in [22]. Thus, following and extending that taxonomy for techniques applied on mobile devices was considered as a reasonable continuity. It is noted that social signals refer to non-verbal communication signals emitted when a person is socially interacting [25]. Hence, natural language processing is excluded from Mobile SSP as it considered a separate field.

2.5.1.1 Prosody

This behavioural cue provides information about the characteristics of a person’s voice when socially interacting i.e. phrasing, stress, and intonation [55]. To infer about the prosody of a person’s voice on mobile devices, literature has focused on extracting certain features. Razak [56] extracted prosodic features such as energy, linear predictive coding (LPC) coefficients, duration, pitch and jitter of each recorded frame. VibeFones [4] focused on pitch, amplitude, mean energy, frequency of the fundamental format and spectral entropy. AMMON [57] calculated zero crossing rate (ZCR), root mean square, frame energy, pitch, harmonics-to-noise ratio and Mel-Frequency Cepstral Co-efficients (MFCC). StressSense [58] reported that pitch and its derivatives were the most informative features, followed by jitter, spectral centroid, high frequency ratio, speaking rate and MFCCs. In SoundSense [45] authors extracted ZCRs, low energy frame rates and other spectral features which were fed in a J48 decision tree [59]. The most recent work SocioPhone [60] calculated prosodic features through pitch, energy, loudness, rhythm and spectral features (formants\textsuperscript{1}, bandwidths, spectrum intensity).

Discussion. The optimal prosodic feature set varies based on the target application. [56] evaluated different prosodic feature sets. Those that included LPC coefficients had the

\footnote{Formant in a vocal signal is the accumulation of acoustic energy close to a certain frequency.}
best performance while the set with speech energy, duration, pitch and jitter performed worst. Authors did not apply any feature reduction technique on the training data, based on various factors such as information gain to retrieve the most informative feature set \cite{61}. *VibeFones* \cite{4} require a long-term analysis to derive standard deviations of the features and do not describe the process of concluding to the particular feature set. *AMMON* \cite{57} showed the performance improvements achieved when combining prosodic features with glottal timings\textsuperscript{1}. However, the system was evaluated off-line on datasets created in constrained environments while the performance in real-world situations was not provided. *StressSense* \cite{58} selected a particular feature set based on information gain. Authors did not provide any quantitative analysis of energy consumption of the system as they extract cumbersome features including MFCC. It should be noted that although MFCC improved the accuracy of the system, these features allow rebuilding of speech segments and further natural language processing. This fact induces some privacy issues as opposed to features such as pitch, speaking rate, jitter which do not allow derivation of speech segments. *SoundSense* \cite{45} focused on features that are not affected by the volume; in spectral features they removed DC components. To preserve users’ privacy they performed data processing on the device and then discarded any raw audio readings. *SocioPhone* \cite{60} is able to cope with ambient noise distributed uniformly to nearby devices but does not incorporate any on-body position detection mechanism of the device to discard positions such as bags that degrade the quality of raw sound signal.

In literature, many works inferred that features such as pitch and its statistics were considered as the most informative features. Additionally, features such as speaking rate, MFCCs, energy and spectral characteristics were also included in the process of detecting the vocal prosody of a user. The spectral-based features constitute a common ground in various works and especially formants, bandwidths and intensity. The disadvantage of these features is the requirement of transforming the time-based values into frequency based values before every inference, which induces additional computational burden. Only a specific set of the twenty MFCCs are mainly selected during the inference process depending on the classification target. Before including

\textsuperscript{1}Glottal timings refer to the air flow variations produced during speech.
MFCCs in the feature set, a designer may consider the accumulative burden of these coefficients due to a computational demanding extraction procedure.

2.5.1.2 Conversation vs silence

During a social interaction, speech and silence operate as regulators of a conversation emitting social signals such as consensus, rejection and reveal interlocutors’ social behaviour including their emotions [62]. One of the well-known and widely-used techniques to infer conversation existence was presented by Basu [63]. It specified a linked Hidden Markov Model (HMM) with three features: non-initial maximum of the normalized noisy autocorrelation, number of autocorrelation peaks, and normalised spectral entropy. The first layer of the model infers regarding voice existence and the second layer speech occurrence. This technique was adopted by [64], Vibefones [4], StressSense [58], MeetingMediator [65], [66]. Another technique widely used by systems such as SpeakerSense [67] and Auditeur [43] is to calculate the ZCR of an audio frame and then apply a classification method to infer if the segment contains speech [68].

Matic [52] inducted a privacy-oriented approach that incorporated an accelerometer on a user’s chest. By analysing the sum of the power spectral densities, computing integral and mean of the frames (80 - 256Hz), and feeding it to Naïve Bayes with KDE [69], they were able to detect when the user was speaking. CoenoFire [70], detects speech through Long-Term Signal Variability (LTSV) presented in [71]. Also, AutoSense [72] utilised a Respiratory Inductive Plethysmograph (RIP) in order to compute lung volume and breathing rate from which they detected conversation existence.

Discussion. For detecting conversation existence or absence through a mobile device, research focused on audio-based and accelerometer-based data. The conversation detection methods based on microphone data funnel on [63] or through some pre-processing steps they train a Gaussian Mixture Model (GMM) that identifies (non)conversation segments. [63] is a well-established approach for detecting speech in raw microphone data, which achieves less than 10% error estimations at 6.4m even with increased interpersonal distance. [43] brought together the most popular pre-processing steps in order to train a model for conversation and silence detection. A post-processing step may
be applied to add time dependence through an HMM. Among the pre-processing steps utilised, performing Fast Fourier Transform (FFT) and extracting MFCC features constitute the most energy consuming processes among the state-of-the-art techniques as opposed to ZCR which is a simple and robust feature. \cite{52} and AutoSense \cite{72} induced two privacy preserving approaches\footnote{\cite{52} utilised accelerometer data and AutoSense \cite{72} sensed lung volume and breathing rate, thus they are considered privacy preserving approaches as they do not focus on audio data that allow natural language processing.} for speech recognition through accelerometer (93% accuracy) and RIP data (over 87% accuracy), both evaluated in real-world situations. Although \cite{52} achieved around 10% error rate in real-world environments, it is prone to coughing and various mean of transportation that confers vibrations. Additionally, attaching an external sensor on user’s chest is considered intrusive. CoenoFire \cite{70} focused on LTSV, which is suitable for noisy environments but is not able to discriminate speech among various users. Finally, classifying conversation existence is a process that can be applied on mobile phones as shown in literature, nevertheless including energy consuming pre-processing steps will increase the computational burden.

\subsection{Turn-taking and vocal outbursts}

In linguistics, turn-taking refers to the process of exchanging speech turns during a conversation, including speech overlap, showing the willingness of a person to continue a conversation \cite{73}. The occurrence of non-linguistic vocalisations i.e. vocal outbursts, provide additional information about interest, boredom, willingness to continue a conversation etc. among the interlocutors \cite{74}. Given a set of features, turn-taking detection is identified mainly by training a GMM for each speaker through an Expectation Maximisation (EM) algorithm, to allow classification of the most probable speaker for a certain speech frame. This process is called speaker recognition and is harnessed to understand between which people turn-taking is occurring. \cite{75} developed an on-line speaker diarization system that in a distributed manner infers interlocutors turn-taking through the above described typical speaker recognition pipeline. In \cite{76}, authors perform speaker diarization\footnote{Speaker diarization refers to the process of speaker recognition followed by clustering with respect to each speaker.} and then they calculate three types of turn-taking features: a) independent (speaking length, number of speaking turns, turn duration statistics),
b) relational (interruptions, order, centrality), c) meeting (number of silent moments, overlapped speech). A similar approach was followed by *SocioPhone* [60]; instead of meeting features they incorporated interaction features that included the duration of speaking and non-speaking turn. For detecting vocal outbursts *VibePhones* [4] considered the distribution of utterance length i.e z-score.

**Discussion.** Understanding turn-taking in an audio data sequence, requires the execution of speaker recognition and then identification of the segments with speaker change or overlapping. The features selected are the same utilised in conversation detection. A great deal of attention should be paid in selecting the optimal dataset with respect to the application context in order to achieve high accuracy and robustness in the inference process. A certain model is trained for each speaker, inducing the requirement of a speaker models library if the system extracts information from all the interlocutors. In the case of adding a new user to the system, a model has to be trained especially for this user and incorporated to the library. Then a Maximum Likelihood algorithm is applied to identify the most likely model. The speaker diarization, in most cases, is executed off-line where all audio segments have been logged and categorised to each speaker such as [76]. For on-line execution, the process requires a connection with the centralised library in order to identify the speakers and then log information about turn-taking. *SocioPhone* [60] performs turn-taking detection on-line, it achieves its highest accuracy when the number of devices is equal to the interlocutors and the devices are placed on a meeting table. When the number of mobile devices is reduced, or the device is placed in a bag or in trousers pocket, the accuracy is degrading. In contrast, [75] was able to maintain similar accuracy even when the number of smartphones was reduced but the recognition accuracy was degraded in short-term turn-takings. Loading all speaker models on the device will increase the computational burden, energy consumption and will degrade user experience. In contrary, off-loading this computation burden to the cloud will induce a certain communication cost in terms of energy consumption and cost increase in users’ data plan. Having performed speaker diarization, statistical analysis such as speaking time, number of turns etc. can provide valuable information about the sociability and the overall social behaviour of the user. Regarding vocal outbursts *VibePhones* [4] utilised the z-score of utterance length which is an easy to extract fea-
ture but requires the knowledge of mean and deviation of the population to compute the particular feature.

### 2.5.1.4 Speech activity

Another auditory cue is speech activity which includes various derivatives of the time a person is talking in a conversation i.e. accumulated speech time, speaking frequency etc., indicating social behaviours such as sociability and dominance. Vibefones [4] detects user’s activity in voice, by performing initially speech recognition and then calculates the z-score of the time the user was speaking. In CoenoFire [70], having detected speech in audio data, they compute the speaking time for each user which indicates the speech activity of each participant.

**Discussion.** Speech activity is a feature that can be derived from previous auditory cues. Literature has extracted this cue by initially detecting voice and speech through the aforementioned techniques. Further, by inferring the speech segments that belong to the user, through a personalised speech detection model, an accumulation of the speech windows is performed through statistical analysis. This accumulation refers to the computation of speech frequency, overall speech time etc. Vibefones [4] considered z-score as measurement of speech activity. It provides the probability of a person speaking with respect to the rest of the speakers but requires all users’ speech segments to perform the computation. CoenoFire [70] performed a more light-weight speech detection method but considered only the overall speaking time. The initial process of personalised speech detection induces the computational burden, in contrast to speech activity inference which is mapped to statistical analysis on the speech segments of a specific user.

### 2.5.1.5 Auditory in essence

When Pentland [25] introduced social signals, he proved the applicability of the domain based on behavioural cues that were extracted from audio data. This induction may undermine the importance of these types of behavioural cues but also may indicate how informative they are. [77] refers to conversation as an *occasion of social interaction*
which highlights the correlation of the terms in context of human behaviour. For that reason a great deal of non-verbal communication may accompany the spoken linguistics. As shown, features related to the interlocutors’ voice can provide significant and informative features. However, some feature extraction processes introduce a considerable amount of computational burden and energy consumption but there are alternatives that can be utilised. A considerable amount of research has been conducted based on audio data. Thus, applying these techniques on mobile devices and in real-world environments will constitute a stable and robust solution for social behaviour inference, with a proportional cost in computation and energy.

2.5.2 Physical activity

After the incorporation of accelerometer sensor on mobile phones, initially by Nokia and later on by Apple which led to the evolution of smartphones, physical activity became a popular behavioural cue [78]. This provides additional contextual information allowing the reduction of false positives in situations such as stress detection. Furthermore, it may be utilised as an optimisation process e.g. to discard data under the assumption that when the user is running, he is not socially interacting. Most approaches focused on computing statistical features showing how active the user is but also what kind of activity the user was performing, such as standing, walking or running. In order to classify the activity the user is performing, researchers mainly train an activity detection model for each category that requires identification. In the next subsections state-of-the-art techniques are presented, that have been developed for measuring how active users are and in which activity state they are in.

2.5.2.1 Movement activity

One of the cues a person is conveying when socially interacting, is movement activity e.g. small but noticeable activity when being in standing position, which may reveal social behaviours such as stress. A low complexity and robust method was presented in Social fMRI [79]. The authors retrieved 3D accelerometer data from mobile phones, for 15 seconds every 2 minutes and then they computed a vector with the frame’s
magnitudes. They then calculated the variance for each second of the frame which was utilised as a ranking mechanism to classify the user’s activity as still, moderate or high based on a threshold.

In Jigsaw [70] authors extract movement activity by retrieving information from the accelerometer of the mobile phones. First, by calculating standard deviation of a moving window and then based on a threshold they classify the segment as active or non-active. In addition, they compute the movement intensity through the median magnitude of the linear absolute acceleration. It should be mentioned that these features are focused on fire-fighters, who work in an intense environment. Berke [80] utilised the accelerometer and barometer sensors to detect users’ activity and the time spent performing the specific activity given a certain weight. Based on the importance of the activity they provided an analogous weight to the time a user spent in performing a certain activity. Muaremi [81] computed activity movement of a user based on accelerometer and GPS data of the mobile device. From 3D accelerometer data they calculated the magnitude and further mean and variance values. From GPS data the amount of locations a user visited and the travelled distance, were computed.

Discussion. Movement activity detection has been tackled with several methods, mainly by computing statistics such as variance of a window over raw 3D accelerometer data and then applying an empirical threshold. Social fMRI [79] is one of these methods, which can be utilised easily through the Funf [33] open-source sensing framework that performs the procedure as a pre-processing step. A similar technique was implemented in Coefire [70], in [80] where authors included a time-dependent factor and [81] incorporated a location-change magnitude feature. Given the energy consumption of the aforementioned sensors, the most suitable approach for continuous sensing applications is the extraction of accelerometer statistics which is a lightweight and reliable procedure [33]. Although, the location incorporation provides an additional measurement and parameter, it is based on GPS; a high energy consuming sensor which cannot be used for long-term sensing applications or should perform sensing in an efficient manner. However, a less accurate approach may be utilised such as GSM localisation techniques to provide coarse grain location estimation. In essence, detection of activity movement for continuous sensing applications, may focus on less power “hungry” sensors such as
the accelerometer, and extract information through low complexity statistics suitable for the majority of social behaviour applications.

2.5.2.2 User’s activity state

Activity recognition provides important information for understanding the context in which a social interaction is taking place. Also, it allows researchers to create more accurate and reliable techniques for social behaviour understanding based on a specific activity state. There has been a considerable amount of research in order to detect the activity that a user is performing based on sensor data retrieved from mobile phones. In this part, only the most important works are mentioned analytically, to provide an overall understanding to the reader regarding the process of activity recognition. An extensive survey on activity recognition through body worn sensor is presented in [82].

Yang [83] presented an activity recognition approach that utilises orientation independent features for vertical and horizontal components of accelerometer data. He computed the mean, standard deviation, ZCR, interquartile range, 75% percentiles, spectral entropy and entropy of both components and their cross correlation. The features were forwarded to a decision tree [59] leading to 90% accuracy for sitting, standing, walking, running, driving and bicycling. Also, he proposes an approach to reduce data over-fitting, that combines K-Means clustering (a cluster for each activity) followed by an HMM-based Viterbi algorithm to leverage historical data.

A lightweight approach for detecting users’ state is presented in [84], that is based on the standard deviation of the magnitude of accelerometer data. It does not rely on the device orientation and achieves above 70% accuracy. Initially authors retrieve \( N \) (windows size) number of measures from the accelerometer and convert them to magnitude time series. Then from the produced signal, they extract entropy, power, value and amplitude of the highest magnitude frequency, weighted mean/variance of the top-\( t \) highest magnitude frequencies (weighted by amplitude). For classification they perform off-line supervised training for each user a C4.5 decision tree model (70% accuracy) by utilising accelerometer data which are labelled with respect to the activity that is performed. Feature extraction and classification is executed in real-time [85].
2.5. Behavioural cues extraction

*Jigsaw* [40] performed activity recognition and divided the process in four stages: a) calibration, b) pre-processing, c) feature extraction, d) activity classification. In the first stage, authors calculate the offset accelerometer parameters for the specific device through a linear least square estimator. The pre-processing stage includes outlier removal and projection of the accelerometer data to earth’s coordinates. Then, time and frequency domain features are extracted based on mean, variance, mean-crossing rate and spectral analysis. Finally, a decision tree is created followed by a sliding window smoothing method achieving 91.64% accuracy for cycling, vehicle, running, stationary and walking. Based on this method, they provided an extension which through a crowd-sensing technique creates a personalised model for detecting activity in a population of users [86].

Seiter [87] utilised mobile phones that incorporate accelerometer, barometer and GPS to understand the level of pain in a patient based on his activity. Authors concluded that based on 40% of pain relief, 10% degradation was detected. It should be mentioned that the study was conducted on only one person. Also, [88] presents an adaptive activity recognition method that leverages mean, variance, entropy and energy (FFT) of a frame, in order to compute the confidence of an activity and then apply the appropriate pair of sampling frequency and feature set. Furthermore, *PBN* [89] describes an approach with multiple sensors deployed on the user. For each sensor a unique classifier is trained and an overall classifier based on Adaboost [90] computes user’s state on the mobile phone.

**Discussion.** For estimating the user’s activity state, research focused mostly on extracting statistical and spectral features from a window of accelerometer samples and based on these features train a C4.5 decision tree. This constitutes a simple and straightforward approach, with lightweight feature extraction and classification model, ideal for mobile devices. Regarding the features, some exceptions are the processes of computing FFT and spectral entropy which add a considerable burden to the device. *Jigsaw* [40] sensing framework managed a comparable accuracy 91.64% on mobile phones. Both approaches could provide a realistic solution in order to tackle the certain problem. Achieving similar accuracy without the need of extracting the burdensome features, could constitute a significant challenge. In *PBN* [89] training a specific classifier for
each on-body position and combining those in a hierarchical model, provide an accurate outcome but limits the pervasiveness of the system in daily lives. In conclusion, incorporating users’ activity state recognition allows the creation of specific social behaviour inference models for different activities. These activity-dependent models led in reducing the error in social behaviour inference.

2.5.2.3 An outline of physical activity

Detection of users’ physical activity on mobile devices is an area that has triggered the interest of scientists mainly from the point of accelerometer integration in COTS smartphones. The works related to physical activity are highly correlated to the detection characteristics of user’s movement. This includes qualifying that a user is more active than another and classifying users’ current activity state such as standing, walking, running. Both fields provide important information about the user and his behaviour but also about the context in which he is. To summarise, movement activity detection is a relatively simple and lightweight process that is supplied by several sensing frameworks as a pre-processing step. In contrast, users’ activity state has a great spectrum of inference techniques that adds a notable amount of computational burden and demands model training for pre-known states. However, user’s activity state redounds significantly in detecting the context in which the social interaction is taking place and should be included in the inference process.

2.5.3 Gesture and posture

Gesture and posture are two means through which people emit signals during social interactions. A noticeable difference exists among several cultures but both cues convey important information about the social situation, attitude, relationship of the participants etc. In SSP, gesture and posture inference is performed through video recordings, in which researchers detect certain body parts of the participants. Then by tracking these points, they train classifiers that infer about various gestures and postures [22]. This section presents techniques developed to detect various gestures and user’s posture through mobile phones while in some cases with the incorporation of external sensors.
2.5.3.1 Gesture

There is lengthy research regarding gesture recognition through several means such as video and body-worn sensors. Literature focused on utilising smartphones’ integrated sensors but also incorporating external hardware. Regarding the inference, a popular approach is to train a Markov Model for each of the targeting gestures and based on the confidence that each model produces, the highest is selected.

PEYE [91] focused on detecting simple motion gestures on mobile phones by utilising the camera and recording small videos. These are split into sequential images from which they extract small rectangles that are tracked through an adaptive block matching approach in order to understand the device’s movement and further user’s hand gesture, with 12.86% minimum matching error. e-Gesture [92], proposed to train a generic HMM classifier for each gesture, which retrieved data from a hand-worn sensor and forwarded them to a smartphone. The accelerometer and gyroscope data are segmented in an adaptive manner based on gesture change through their magnitude. Then, they are forwarded to an adaptive (Maximum Likelihood Linear Regression for model update) or multi-situation HMM (one HMM for each situation: ride, stand, walk, run) for gesture recognition in four different situations, achieving 84.6% and 94.6% accuracy respectively.

Authors in [93] extracted pitch and roll of a user’s hand from body-worn sensors. By applying pre-processing mechanisms and an HMM they were able to identify user’s gestures with 97.7% accuracy. The pre-processing step was based on SWAB [94] that performs segmentation and approximation on time correlated data. These segments were grouped based on resemblance and the ones with the lowest similarity were selected. uWave [95] introduced Dynamic Time Warping (DTW) that performs adaptive gesture classification through sensed accelerometer data based on only one training sample and achieved 98.6% accuracy on 8 gestures. A pre-processing step is preceded, that performs quantization on raw accelerometer data to remove noise and reduce the size of the data.

As an improvement of DTW, [96] presented a frame-based descriptor and multi-class Support Vector Machine (SVM) that was able to detect 12 distinct gestures with 95.21%
accuracy. *Myo* [97] is a newly developed wearable armband that is able to perform gesture and motion control. It detects the muscle movement of the user’s arm (Electromyography) and transmits that information through Bluetooth to another device such as a mobile phone. It should be noted that *Myo* is a commercial product and its accuracy is not provided. For further information about the analytical works that have been done and proposed techniques in the area of gesture recognition the reader is referred to an extensive survey [98].

**Discussion.** In essence, *PEYE* [91] performed mainly device movement recognition through video recording. This may be replaced by orientation sensor readings due to lower power consumption of sensor and process. *e-Gesture* [92] proposed an adaptive method that continuously learns based on user-labelling but has increased computation due to learning and may not perform well due to data diversity. An adaptive model for each situation (e.g. standing, walking etc.) may achieve higher accuracy as it will create an activity-dependent classification. The multi-situation approach had the highest accuracy, but requires training of known situations. [93] focused mainly on detecting the type of activity the user was performing while gesture recognition was deficient due to lack of a garbage model; a model that infers if none of the target classes are detected. *uWave* [95] claimed high accuracy (98.6% for 8 gestures) through their adaptive approach. However, it should be noted that it is a user-dependent method which must be personalised to each user and there is a requirement for tracking the device orientation in case it is tilted. *Myo* [97] provides a pre-defined detectable gesture collection but a developer is able to add his own. In overall, both HMM and DTW methods achieved high accuracies. Because every person is different, a personalised approach will achieve highest accuracies. However, this constitutes a trade-off due to the requirement of additional training and user dependence. Also, in *e-Gesture* [92] the utilisation of a limited amount of training data, leads to the computation of non-optimised thresholds as a result the adaptive methods may achieve lower accuracy. Overall, approaches that utilise inertial data are more suitable than the video-based approach due to mobility and energy restrictions, however adaptive sensing and inference may be required.
2.5.3.2 Posture

A person’s posture is divided into head and body posture. Both produce non-verbal communication during a social interaction through the tilt of the certain body part. This is a brief description of existing literature for both classes in the following subsections.

2.5.3.2.1 Head posture  Being able to detect head posture through a mobile device can provide valuable information about a social interaction such as where the user is facing and if the head is tilted. For example, during a social interaction when people have a common interest or agree on a certain topic they tend to tilt their head to the same direction, i.e. mirroring [99]. Thus, head posture detection is another significant behavioural cue that could be utilised for social behaviour inference.

In SEPTIMU [100] an accelerometer and a gyroscope are integrated inside an earphone and utilised the microphone transmission of the headset, to communicate with a mobile phone which was used to infer about the head posture of the user. Smart Pose [101] and [102] employed the orientation sensor and the front-camera of the mobile phone to calculate the user’s neck angle with respect to the earth’s coordinates. Initially, the system performs face detection through Android API built-in functionality. It identifies if the user is holding the device in the hand by shake detection (threshold-based) on accelerometer data. Finally, based on the orientation sensor and user’s viewing angle with respect to the device, it computes the average neck tilt angle. Another technology that has been developed is Google Glass [103]; wearable glasses incorporating multiple sensors such as gyroscope, accelerometer and magnetometer allowing to be utilised as non intrusive technology.

Discussion. Research focused mainly on obtrusive approaches for head posture detection. SEPTIMU [100] claim of head tracking provides a simple but obtrusive solution because it requires the user to continuously wear an earphone. Smart Pose [101] constitutes a low complexity and multi-modal approach for head posture detection without any external hardware, it relies on off-the-shelf smartphone integrated sensors but requires the user to hold the device in the hand and also interact with it. Lately, wearable
devices constitute a viable solution for accurate and reliable head posture inference.

### 2.5.3.2.2 Body posture

An equally important class is body posture detection that convey social signals such as mirroring and intimacy during a social interaction. While people are interacting they tend to bend towards a person showing a certain level of intimacy while a body slope opposite to the interlocutor may indicate inconvenience. Thus, detecting this type of signals can provide underlying information about a social interaction.

In *imWell* [104], a sensor incorporating an accelerometer is placed under the left arm of the user and transmits logged data via Bluetooth to a mobile phone. To detect a different posture, they target identifying transition points. A pre-processing step is applied that computes the standard deviation of 1 second window of accelerometer samples, to remove minor movements. Then, the angle change with respect to vertical position is computed, which determines the upper body posture. Having the standing position of the user as a reference, they categorise the user’s body posture based on certain thresholds. *CONSORTS-S* [105] utilises the average of the accelerometer window of samples from a wireless on-body sensor and based on the device inclination it classifies through decision rules, about the posture of torsos (standing, facing up or down). [106] perform body posture recognition by retrieving measurements from mobile device’s orientation sensor and especially pitch, to classify if the user is sitting or standing, while the smartphone is in the user’s trousers pocket. They allow a margin of error of 20° around 180° or 0° of pitch to infer that the user is standing. With the same error margin around 90° or −90° they estimate if the user is sitting. Another approach that provides information about user’s torsos facing direction with respect to earth’s coordinates is *uDirect* [5] that utilises inertial sensors of off-the-shelf mobile phones.

**Discussion.** Literature has targeted mainly inertial sensors to estimate body posture. *imWell* [104] utilises a very simple technique to identify different body postures. However, it is considered as an intrusive methodology because it utilises an external sensor that is tied around the user’s torsos. *CONSORTS-S* [105] performs rule-based decision on average of accelerometer sample-window. It constitutes a lightweight process re-
2.5. Behavioural cues extraction

Regarding the feature extraction and the inference but is susceptible to on-body position changes of the device. So, in order to improve the accuracy of the approach it requires the creation of different rules for each on-body position. [106] described a threshold based approach on the orientation sensor’s pitch. It is easily implementable but is applicable only for the trousers pocket and requires the device to be in a vertical position, thus it does not provide a generic solution. Regarding uDirect [5], it assumes that the relative orientation between the device and user’s body is static. In unconstrained environments the devices are not fully attached to a user’s body and are able to move in a certain range. Also, a pre-processing step is required to identify the on-body position of the device. Overall, the extraction of body posture is mainly based on accelerometer data, a relatively low energy consumption sensor, but still requires an adaptive sensing mechanism. Also, the on-body position of the device may be considered as contextual information to target the inference on specific body parts and discard unrelated positions.

2.5.3.3 Revealing the methodology

Overall, gesture detection on mobile devices in current literature requires either the user to hold the device in his hand or the incorporation of an external sensor. The integrated accelerometer could be considered as the main source of data, through which gesture specific models can be trained. In posture detection, head tracking solutions are mostly intrusive (video or external sensor) however body posture detection could be implemented with COTS mobile phones. The increased popularity and close body attachment of wearable devices, which connect with smartphones, shows good potential in real-world situations for both gesture and posture detection. A gap identified in literature, is the lack of on-body position detection of the device before performing body posture inference. The on-body position of the device [108] [109] constitutes a necessity in order to accurately compute the posture in real-world applications.

\footnote{Most common wearing positions include trousers pockets, belt, hand, chest pocket, handbag, backpack [107].}
2.5.4 Facial cues

One of the most expressive parts of the human body that people used to externalise their interest, agreement, disagreement, surprise etc. is the face. This emission of social signals is mainly achieved through facial expressions and eye movement. Thus, providing a detection and quantification mechanism of behavioural cues vented from a person’s face is not negligible, while relying on mobile devices.

2.5.4.1 Facial expressions

People communicate verbally during a social interaction and in parallel emit social signals also through their faces. Several works in psychology showcase the importance of facial expression in recognising interlocutors’ emotions such as valence, arousal, disgust, embarrassment and amusement [110]. In addition, due to this high correlation they claim a high accuracy in detecting these emotions through facial expressions. For a comprehensive literature review of face recognition, the reader is referred to [111] [112]. State-of-the-art techniques for detecting facial expressions through mobile phones are presented in the following paragraphs.

Detection of facial expression of a user through a mobile phone is presented in Visage [113]. The approach is based on information retrieved from the camera and motion sensors of the device, for face and head pose recognition respectively. Through the data, authors perform face detection [114] (Adaboost-based object detection) inferring rectangle features allowing face tracking. The knowledge of head and face reference points combined with Active Appearance Models [115] merged texture and shape of the face and allowed them to detect different facial expressions. Performing real-time training and recognition of facial features is presented in [116] based on novel non-orthogonal local random basis. According to the authors, this method provides a robust but energy efficient solution for extracting facial characteristics. The features are forwarded to a neural network which performs classification and updates the decision thresholds. They evaluated against six well-known face databases and benchmarked against Principal Component Analysis (PCA) approach [117].
Another proposed method in order to detect facial expressions on mobile phones is [118]. Initially, the authors propose the utilisation of two SVMs, a micro and a macro component. The first layer computes the score of the input image with respect to pre-trained (non-)facial classifier and then a second layer SVM calculates the fiducial points: eye, nose, mouth. The acquisition of fiducial points leads to the extraction of Local Gabor wavelet features through Gini selection method. In [119] authors performed face recognition on mobile phones based on eigenfaces [117], a well-known and established approach. By retrieving an image from the smartphone’s integrated camera, they detected the user’s eyes. Utilising the eyes’ position they were able to mine several facial feature points. By tracking these facial features, a classifier can be leveraged that enables the identification of various facial expressions.

Discussion. Most of the works in facial expression detection were designed, implemented and evaluated in constrained environments, mainly to showcase the applicability of such cumbersome techniques. One of these works was Visage [113] that was able to detect facial expressions on mobile phones through a well-established face recognition method [114]. However, the system realisation in real-world environments including the energy consumption was not evaluated, implying its limited applicability in realistic situations. Applying NN approach for classification on mobile phones in [116], was proven insufficient. It is worth noting that although PCA with SVM outperforms conventional random basis, it requires training. [118] tracked users eyes, mouth and nose, which provided some concrete identification points for facial expression detection. During real-time classification they claim an inference cycle of five seconds by the Boosting Naïve Bayesian (BNB) algorithm, with an overall accuracy 75% for four expressions. [119] applied eigenfaces [117], distance projection and computation, classification on mobile phones. The inference process duration was about 1.2 seconds, which constitutes a reasonable delay for face recognition. However, they have not included any analysis regarding the computational burden and the energy consumption of the approach, which will limit the applicability of the method in real-world situations. Even though facial expression detection constitutes a cumbersome process, the value of this behavioural cue is important.
2.5.4.2 Eye tracking

During a social interaction a person conveys non-verbal communication signals also from the eyes. Social signals indicating intimacy, interest, personal relation but also conversation regulation are some of the information emitted through eye contact [120]. Furthermore, [121] induces the ability of predicting interlocutors’ attention during a group conversation based on the eye movement. Thus, as eye movement is considered an informative behavioural cue in social behaviour inference, state-of-the-art techniques for eye tracking are reviewed in the following paragraphs.

One of the first works for detecting eye movement was eyeLook [122], where authors attached an Eye Contact Sensor (ECS) [123] surrounded by LEDs, on a mobile phone. Through the flashing of LEDs, their reflection was displayed on the user’s pupils, logged by the attached camera and then transmitted for off-line analysis. LEDs’ reflection was displayed near user’s pupils which allowed detection and tracking of his eye. By detecting the movement of the participant’s eye, they were able to identify turn-taking among users.

For detecting facial movement and especially eye motion [124] developed custom wearable goggles. These were constituted by dry electrodes, light sensor, accelerometer and were connected through a wire to a digital signal processor (DSP) and a data storage. Based on the application they proposed alternative methods, such as electrooculography (EOG) or camera recordings, to detect eye movement. EOG is performed through dry electrodes of the goggles that are attached to participant’s face. Authors depict six main feature categories from which they compute various statistics and signal characteristics; a) saccades, b) fixations, c) blinks, d) microsaccades, e) vestibulocular reflex, f) smooth pursuit movements. It should be noted that the authors utilised only saccade, sequence, blink and fixation features. By triggering the interest of a participant wearing the goggles, these features are extracted and a model is trained which is applied later on to identify certain eye movements.

In order to create an eye-controlled mobile phone, [125] utilised the front camera of a smartphone for achieving eye tracking. The system takes sequential pictures from the user. Then, it performs a Haar classification that identifies features in a rectangular
space through summing the intensity of the pixels. The method detects two similar spaces of the picture and classifies it as the eyes. Having detected the user’s eyes, they utilise the CAMSHIFT \cite{126} algorithm to keep track of their position in upcoming images. To transform the detected eye position to the device’s display coordinates, they compute the centroids of the rectangles and then they apply the Starburst algorithm \cite{127} for tracking.

**Discussion.** A common practice for extracting facial cues is defined by detecting initially fiducial points \cite{117} \cite{128}. Eye positions are some points which may be leveraged for tracking a person’s eye movement. Regarding the aforementioned works, eyeLook \cite{122} is based on an obtrusive mechanism that requires detecting the flash reflection near the eye pupil which is additionally prone to daylight reflections. \cite{124} tried to deviate from the main visual-based approaches by utilising a EOG attached to muscles surrounding the eye. This is a less computational consuming approach that provides a raw signal indicating the muscle movement, allowing detection of eye activity. Although, the method requires specific glasses with integrated dry electrodes, the reduction of the complexity is noticeable regarding the video based approach. Furthermore, \cite{125} also focused on a video-based eye tracking method. By applying a modified Haar feature extraction and classification they were able to achieve a speedy inference of eye points in an efficient manner. The classification process was only initiated when a certain movement threshold was overcome. Finally, performing eye tracking requires tackling some key challenges including a) the high computational power required for retrieving information from visual data and b) the difficulty in detecting the eye pupil in a sunny outdoor environment that is characterised by brightness fluctuations and saturation \cite{129} c) the utilisation of ubiquitous and non-intrusive sensors for retrieving data that will allow the inference.

**2.5.4.3 Facial cues at a glance**

Face is a very descriptive part of the human body during social interaction in terms of social signal conveyance. However, the capability of detecting these behavioural cues on mobile devices constitutes a great challenge. Along with the computationally complex
process of detecting facial cues other provocations arise including the intrusiveness of the system, the applied training data but also the real-time on-device classification. Detecting facial cues is stemmed by identifying several fiducial points of a person’s face, including the mouth, nose and eye. In many cases the methods include a pre-processing step of detecting these parts of the face and then utilise these points for classifying facial expressions and monitoring eye movement.

Overall, identifying facial cues i.e. facial expressions and eye tracking is a burdensome process, especially for constrained devices such as smartphones. As shown, there are works that have managed to execute these computationally demanding processes on mobile phones. However, they do not provide a concise energy consumption and computation burden analysis which will indicate the applicability in continuous sensing and inference applications. SociableSense [3] is a system that decides based on computational requirement to perform the inference on-device or at a cloud infrastructure. Applying a distributed inference adaptation model such as SociableSense, combined with an adaptive sensing technique based on the context is a viable solution for performing such burdensome processes on smartphones. Finally, wearable technologies such as Google Glass [103] constitute a promising approach for real-world applications. However, energy consumption is still a great challenge in continuous sensing systems.

2.5.5 Environment and space

Equally important behavioural cues are space and the environment in which a social interaction is taking place. According to psychologists, the interpersonal distance and the spatial arrangement of the interlocutors provide a large amount of information about their social relation, their intimacy but also the probability of people are interacting in multi-personal interactions [130] [7].

2.5.5.1 Interpersonal distance

In psychology, proxemics is an area that has been exploited for many years, starting from the work of Edward T. Hall [7]. In this work, Hall following the social behaviour among animals, defined some imaginary concentric circles around each person during a
social interaction, which indicate the type of relationship among the people. Interpersonal distance is a significant element of social interactions, not only to detect if people are interacting but also to estimate their relationship.

2.5.5.1.1 Sound The most promising approach for distance estimation through sound is BeepBeep [131]. It is based on ToA without the requirement of time clock synchronisation among the devices. Each of the devices sent out an audible Beep sound and logs its own sample and the remote sound. The device continues to record until it receives the remote Beep. Then, they exchange the standby time and compute the interpersonal distance from the number of samples recorded and the time required to receive them. An extension of BeepBeep is [132] in which the authors develop a transmission scheme and apply an adaptive ToA mechanism to improve the accuracy of the system. Also, Whistle [133] is an approach akin to BeepBeep but relies on TDoA by recording the sound from multiple devices and performs the computation at a centralised point. A recent work called RF-Beef [134] combined the methodology of BeepBeep with RF interface to apply TDoA by sending initially a Beep sound followed by an RF beacon. A ToA-based scheme is introduced in [135] that uses a speaker and a mobile phone’s microphone to perform distance estimation. [136] exploited and developed a mechanism for estimating the distance based on TDoA among devices, by producing ultrasound through COTS mobile phones.

Discussion. A different modality for distance estimation, the sound, was considered in some approaches in order to be able to apply techniques that are difficult to deploy on mobile phones such as ToA, TDoA and AoA. BeepBeep was the first work that was able to leverage these types of techniques (ToA) on mobile phones based on sound. By exchanging the time duration, there was no need to perform clock synchronisation among the devices. The technique was applied between two devices only. Following this approach, different works used other techniques such as TDoA or combined them with RF signals. As claimed, they are able to achieve a fine-grained distance estimation among the devices. However, the sound based methods are prone to relative orientation of device and user with respect to the interlocutor. The majority of these approaches utilise audible beacons that are not suitable for ubiquitous usage. [136] claimed to
have achieved the transmission of ultra-sound through COTS mobile phones. For the appliance of TDoA, there is a requirement for speaker array deployment at the environment in order to calculate the time difference between arrival of the beacon at the two speakers. This increases the intrusiveness of the system.

2.5.5.1.2 RF interfaces  Interpersonal distance estimation through RF-based technologies (e.g. RF, Bluetooth, WiFi) constitutes a common approach due to its easiness in development and implementation. Researchers have developed various techniques to estimate interpersonal distance among users in a coarse-grained and fine-grained manner.

**Coarse-grain.** A commonly used method to detect if people are in vicinity is to utilise the Bluetooth interface. This is available in the majority of today's mobile devices. By performing an enquiry scan process, a smartphone retrieves discoverable nearby devices. This approach takes into consideration every device that is in the range of Bluetooth radius (∼10m). It is not affiliated with any intelligence to mine more specific information about interpersonal distance; only details such as the identifier and timestamp are logged. One of these works was Serendipity [44] in which the author developed BlueAware framework for mobile phones to log the Bluetooth identifier and current timestamp. [46] deployed Bluetooth dongles inside a building and through Bluetooth discoverable mobile phones they were able to detect if users were in vicinity. CenceMe [38], Friends and Family [79] [137] [138] [139] [140] are other examples of works where they utilised simple Bluetooth discovery to infer if users were in vicinity. Unlike previous approaches, PeopleTones [141] leveraged cell tower readings to estimate if the users are nearby in a larger scale, claiming an error around 322m.

**Discussion.** As noted before, the most common approach for distance estimation on mobile device is through RF-based technologies. In coarse-grain distance estimation based on the targeting device, researchers focused on Bluetooth discovery or GSM localisation. On one hand, due to the popularity and implementation simplicity of Bluetooth discovery, it constitutes a widely used method when conducting research into social behaviour. It comprises only of the discovery of nearby devices and logging their BTIDs including the timestamps. There is no processing or inference required,
thus if the induced error in distance estimation is acceptable for a certain type of application, this method may be preferred. On the other hand, a large amount of people set their devices on non-discoverable mode or disable the Bluetooth interface of their smartphones, making the coarse-grain distance estimation non-applicable. Nevertheless, the range of Bluetooth introduces a large amount of error, e.g. two people may be in different rooms, but in through this method they are considered close enough to interact. These interpersonal distance estimation techniques rely on the assumption that when devices are in vicinity then their users are as well. However, in real-world situations this assumption is not always valid, thus there is a need to incorporate a mechanism to detect when a user is not carrying the device such as [142].

**Fine-grain.** Alternative and more advanced techniques have been proposed to achieve a more accurate result in estimating if people are in vicinity. These techniques are mainly based on ToA, TDoA, AoA and RSSI. Due to ease of implementation on mobile phones, most approaches focus on retrieving the Bluetooth/WiFi RSSI and then through a PLM, threshold-based classification or machine learning technique they try to estimate the interpersonal distance of the users.

An initial approach to estimate interpersonal distance through Bluetooth/WiFi RSSI, is the development of a PLM. The most simple method is *Free space* PLM that considers an ideal environment without reflections and obstacles. It requires a reference $\text{RSSI}_\text{ref}$ measurement at a specific distance. Given the RSSI reference the model estimates the distance between the two devices. An improvement of this model is *Office* PLM [143] that modifies Free space PLM. In particular, it adds the impact of the indoor environment and especially of a normal office while assuming line-of-sight between the devices. Regarding the environmental parameter for indoor environments, there are predefined values for certain types of rooms that can be utilised. Alternatively, by retrieving RSSI samples at different distances and through an optimisation technique, researchers may compute their own parameters. Based on these generic PLMs, several variations have been proposed which add more parameters in order to consider other factors. One of these variations is *BlueEye* [144] that strengthens the office PLM by incorporating two environmental constants and the relative orientation of the two devices; one of the factors which affects the RSSI is the directionality of device antennas. The output of the
improved PLM was forwarded to k-means clustering to estimate users’ interpersonal distance.

Stankovic [145] applies a PLM with computed parameter for indoor and outdoor environments to detect when people are in vicinity; the interpersonal distance boundary utilised is 3m. Regarding the WiFi interface, Matic [52] created an artificial dataset through an indoor PLM by leveraging WiFi RSSI measurements at 1m distance. Then, he trained a Naïve Bayes with a KDE classifier to detect if people were at a distance to socially interact. Features utilised to train the classifier were average and maximum values of a 10-sample window. Finally, Comm2Sense [51], followed the same process for training a classifier on 20-sample window that determines in which interaction zone people socialise.

Discussion. Researchers managed to achieve an improved accuracy in distance estimation. In order to achieve this, techniques such as RSSI, ToA, TDoA, AoA were utilised. For the implementation of some of these methods on smartphones, there are particular requirements such as firmware modification, multiple antennas etc. Thus research has focused mainly on leveraging RSSI provided from the core API of the majority of COTS mobile devices. Based on RSSI, various PLMs have been proposed for environments such as free space, indoor and office, which require certain parameters for the specific environment. Even given the environmental parameters, RSSI is prone to antenna type and orientation, human body absorption, reflections and obstacles. Authors in [53] and [52] strive to tackle this through machine learning techniques. They incorporated uncertainty in distance measurements and utilised a 20-sample window on which certain statistics were computed. However, they performed only small-scale experiments while viability, reliability and robustness of such a solution in the real-world environment is not proven. Additionally, the number of samples (window) required should be taken into consideration. As the number of samples increases, depending on the sampling frequency of the RF interface, the waiting time for an inference may increase. Also, when using a large window of samples (e.g. 20 samples [51]), the data may be outdated leading to erroneous inference results.
2.5.5.2 Spatial arrangement

Kendon [77] introduced F-Formation referring to the spatial formation created by the participants during a social interaction. In more detail, an F-Formation can include various configurations such as face-to-face, side-by-side, rectangular, circular, semi-circular and L-Formation\footnote{In L-Formation users’ torsos draw a right angle, similar to letter L.}. So, depending on the formation that participants frame, different information about their social relationship is conveyed. This signifies the importance of a user’s spatial arrangement. Researchers in *Virtual Compass* [53] by considering the interpersonal distance among users in vicinity, they created a virtual map through computing the Euclidean distance of the users. [52] utilised off-the-shelf mobile phones to detect social interactions. Each participant carried the device on a static body position. While knowing the position of the mobile phone relative to a user’s body, they used the orientation of the mobile phone in order to detect forward direction of torsos, hence to estimate users’ spatial formations.

Discussion. *Virtual Compass* [53] calculated users’ relative spatial arrangement. Due to the lack of users’ facing directions and absolute locations they were not able to estimate the absolute spatial arrangement of the users. As the approach incorporated both WiFi and Bluetooth RSSI to perform the computations, unless an energy efficient mechanism is added, this work is not suitable for continuous sensing applications. Also, RSSI is highly dependent on the environment and prone to human body absorption. In contrary, [52] used the orientation sensor to keep track of the user’s facing direction. However, the orientation sensor is based on a fusion mechanism of accelerometer and magnetometer that is affected by accelerometer bias and magnetic disturbance. A fusion mechanism that incorporates gyroscope with a drift compensation approach could prove to be a more reliable solution. Researchers in this work increased the intrusiveness of the system by limiting the smartphone’s wearing position to the user’s belt. A less restrictive approach regarding user’s wearing position would improve the pervasiveness of the system.
2.5.5.3 A disclosure of environmental and spatial cues

The environment and space in which a social interaction is taking place conveys information. A brief comparison of state-of-the-art techniques was presented to understand and provide quantification mechanisms to allow the extraction of these types of information.

Interpersonal distance estimation is an explored field with several proposed approaches. The classification of these works was based on the modality utilised to perform distance estimation. Sound-based distance estimation is the most recent approach where scientists have shown interesting results. BeepBeep [131] was able to tackle the device synchronisation problem required in ToA-based methods. Audible beacons constitute an issue which could be tackled through ultra-sound beacons, however they are still in an immature phase regarding mobile phones. In RF-based approaches, there is high dependence between the accuracy and system complexity required. Techniques such as ToA, TDoA and AoA are mainly contingent on external stationary or mobile hardware which introduces a certain level of intrusiveness and also mobility issues. RSSI is a popular solution for estimating distance but is highly dependent on the environment and is characterised by large fluctuations. Overall, these methods are prone to the environment and to human body absorption which both introduce a considerable amount of error. Preliminary results have shown that ultra-sound methods could achieve accurate distance estimation. However, there is no evaluation in unconstrained real-world environments.

Regarding spatial arrangement of the users, its importance has been indicated in psychology [77] however there are not considerable amount of works. Researchers focused on detecting the relative spatial arrangement of the users. Furthermore, relative spatial arrangement induces error, as the absolute position is not known and through various parameters researchers focus on reducing the location uncertainty. There is no analytical work in order to quantify the error induced by this approach. Absolute positioning systems may reduce the error introduced by estimating relatively the spatial arrangement of the users. This could be achieved by inertial tracking systems that are built upon these types of information bearing in mind the requirement for energy efficiency.
2.5. Behavioural cues extraction

due to continuous sensing.

2.5.6 Device usage

The term behavioural cues mainly refers to non-verbal signals that are conveyed from a person during a social interaction. This constraint does not only refer to physical presence, but also to a social interaction in different physical places. For example, during an SMS text conversation people emit social signals such as response time, call frequency, punctuation, emoticons etc. These are all a small part of features that could be extracted from the usage of a mobile device.

SenseMs [146], was one of the first works that argued about non-verbal signals in SMS messaging. [147] logged user’s interaction with the device in order to understand the effect on the network and the energy consumption. The data utilised in this work could be forwarded to a human behaviour understanding mechanism to extract contextual information. [139] monitored calling and SMS text behaviour on the mobile phone of a person and categorised it to different social groups. [145] utilised GPS and Calendar to understand the context of a social interaction, while logging call records to list the interlocutors. Altshuler [148] introduced six categories of features based on user’s patterns, that could be retrieved from a mobile phone: a) Internet usage, b) Calls, c) SMS messages, d) Phone applications, e) Alarm clock, f) Location. BeWell [149] also monitored smartphone usage such as device charging, screen lock, power off etc. Apart from the previous works describing the features that could be extracted from mobile phones, Olivier [150] created dataset from 17300 Blackberry devices in which he logged data representative of the user’s interaction with the device. These datasets could prove to be a useful mean for predicting user’s context.

Discussion. The most important advantage of these types of signals, is that they are collected from virtual sensors. This type of information is stored locally on the device while a person uses it and can be retrieved at user’s discretion. Researchers can collect these types of data through the device’s API or a sensing framework. Then, they can extract behavioural information with negligible energy consumption due to lack harnessing any of the burdensome physical sensors. These types of cues can be
employed for long-term behavioural analysis of a user by inferring social characteristic patterns, but also to acquire contextual information.

2.5.7 Physiological

Extracting physiological characteristics of people during social interactions provides precious intelligence of natural state of the body. During a social interaction, based on the user’s mental state, feelings, stress etc. the physiological body states are changing such as heart rate, skin temperature and humidity. For example people are interacting and due to the conversation context they feel stressed, which increases their heart rate and skin temperature. To detect these types of signals, researchers have focused on Galvanic Skin Response (GSR), Respiratory Inductance Plethysmography (RIP), Electrocardiography (ECG), Electroencephalography (EEG) sensors.

*AutoSense* [72] is a system composed of physiological sensors such as GSR, RIP, ECG, a mobile phone and a software component called FieldStream. Through external sensors (RIP and ECG), FieldStream performs a windowing pre-process, producing information such as a window of R-peak locations, followed by feature extraction computing mean, variance, heart rate and respiration rate. *NeuroPhone* [151] was the first work that incorporated mobile phones with a wireless electroencephalography (EEG) headset in order to perform actions on the mobile device emitted directly from a person’s brain. The headset transmits data to the mobile phone, on which an initial averaging is performed followed by the appliance of a bandpass filter for noise reduction. Then, they utilise weighted classifiers, multivariate equal-prior Bayesian and decision stump classifiers. This approach could be applied in order to detect other brain signals which will lead to other social signal detection. *imWell* [104] connected a smartphone with a physiological sensor called Zephyr BioHarness 3 [152] through Bluetooth interface. The mobile phone was monitoring and storing information about user’s heart activity and later uploaded the data to a mHealth backend server for off-line processing. In [153] authors employed physiological sensors in order to extract user’s heart rate and variability. They deployed a feature extraction framework [154] for filtering noise from ECG raw data in a robust and lightweight manner. *SEPTIMU* [100] utilised an
earphone in which they incorporated a microphone in order to detect user’s heart rate.

Discussion. Literature has mainly focused on detecting a person’s heart rate and skin temperature. This is performed through off-the-shelf sensors transmitting through wire(less) communication to mobile phones which conduct the inference. Off-the-shelf sensors have incorporated mechanisms of noise reduction, thus provide accurate estimations and usually do not need any pre-processing step. However, current approaches introduce a certain level of intrusiveness which should be considered during the design of Mobile SSP applications.

2.6 From social signals to social behaviour inference

Extraction of behavioural cues constitutes an abstraction layer, in which some preliminary knowledge is retrieved from raw sensor data. Combining these different types of information leads to the process of mining social signals. These signals convey significant information that characterise a person’s feelings, mental state, interest and boredom during a social interaction. As the duration of the social signals is limited, a long-term analysis of the information they provide, will infer a person’s social behaviour. This section will outline different social behaviours that can be extracted from long-term analysis of certain social signals with respect to the behavioural cues analysed in the previous section (See Figure 2.5).

2.6.1 Stress

A social behaviour that has attracted a noticeable amount of interest among researchers is stress. Stress detection is mainly based on vocal, physical and physiological activity cues that are forwarded to a machine learning technique responsible for providing an estimation. As claimed by researchers, state-of-the-art techniques are able to achieve an acceptable accuracy over 80% in most cases.

AMMON [57] extracted prosodic features and utterances which were fed in a linear SVM and performed stress classification with 84.4% accuracy and 93.6% for stress increase-decrease. StressSense [58] exploited three different approaches to train two
GMMs for stressed and neutral voice. They developed a universal model for all participants (71.3% indoor accuracy), an adaptive model that starts from the universal model and through Maximum A Posteriori the model fits to a specific user (81.3% supervised, 77.8% unsupervised indoor accuracy), and finally a personalised model trained especially for each participant (82.9% indoor accuracy). *AutoSense* [72] requires physiological measures such as cardiovascular and respiratory data to infer about user’s stress levels with 90% accuracy.

In [153], authors combined activity, posture and physiological features through neural networks and a fuzzy logic algorithm in order to detect if a person is stressed. [155] performed stress classification in three different activities. Authors utilised physiological (ECG and GSR) and activities (e.g. sitting, walking, standing) features to determine if a person is stressed by applying J48 decision tree [59], Bayesian Networks [156] and an SVM [157] achieving the corresponding accuracies 92.4%, 85% and 84%. [81] integrated physical activity, auditory, phone usage and heart rate variability features and achieved 61% accuracy for stress detection through multinomial logistic regression.

**Discussion.** As shown, existing literature has focused on inferring stress through auditory, activity and physiological cues. *AMMON* [57] was able to manage 84.4% accuracy through prosody including glottal features and utterances given the trade-off of computational burden introduced by eigenvalues solving and other glottal features. In
StressSense \cite{58} as expected the personalised classifier achieved the highest accuracy. But for each user there is a need to train a separate model, followed by a supervised, an unsupervised adaptation model and last a generic classifier managing the worst accuracy. It is worth noting that external equipment was required in order to be able to perform speaker segmentation i.e an indoor array of microphones and outdoor a second smartphone.

AutoSense \cite{72} and \cite{153} require additional physiological equipment. This introduces a certain amount of intrusiveness but includes supplementary features such as heart-rate achieving multi-modal inference. As opposed to AutoSense that utilises a J48 classifier which is prone to over-fitting, \cite{155} applies fuzzy logic-based rules that insert softer boundaries in the classification process. \cite{155} with similar modalities achieved a relatively robust approach, without auditory cues, as for different types of classifiers there is a small variation in the overall claimed accuracy. Muaremi \cite{81} utilised lightweight and easy to extract features but achieved the lowest accuracy for stress detection in the literature that was reviewed.

In essence, the approaches for stress detection are concentrated either on auditory cues or on a combination of physiological, activity and auditory cues. Literature indicates that the most significant cues are auditory and physiological for detecting stress. In detail, researchers were able to detect stress accurately (over 80%) by utilising only auditory data and extracting the aforementioned cues in contrast to physiological cues that were combined with additional modalities. Another important point that should be taken into consideration is the identification of the activity that the user is performing before executing the stress classification. Depending on the activity, the approach may be prone to false positives when carrying out intense activities. In conclusion, stress detection is a promising area and with the incorporation of the field of psychology will become mature, multi-modal and coherent.

2.6.2 Emotion

After analysing existing techniques for stress detection in Mobile SSP, this subsection focuses on emotion detection. To detect emotion in a preliminary stage, researchers
perform some simplification by focusing on the identification of major emotions such as happiness, anger, neutral, sadness etc. or just classifying if the user has positive or negative emotions. For emotion detection, scientists utilised audio datasets targeting different emotions and used them as training sets for machine learning techniques. Next, state-of-the-art techniques researchers utilised will be presented and a brief discussion about them will be provided.

At first, AMMON [57] extracted prosodic and spectral features from Belfast Naturalistic Database [158], and trained an SVM [157] classifier with 75% accuracy for emotion recognition i.e. positive or negative. An important work is EmotionSense [41] which used Speech and Transcripts library [159] to train an emotion recognition model and succeeded in 71% accuracy for 5 emotions based on prosodic features. Visage [113] detected users’ emotion on mobile phones through facial expression detection [128]. To evaluate their approach, they applied it on the JAFFE dataset [160] achieving the corresponding accuracies: a) anger 82.16%, b) disgust 79.68%, c) fear 83.57%, d) happiness 90.30%, e) neutral 89.93%, f) sadness 73.24% and g) surprise 87.52%. In [118] they apply facial expression classification to detect a user’s emotion and discriminate among four different emotions: a) neutral, b) joy, c) sad, d) surprise.

Discussion. As mentioned above, AMMON focused only on extracting information regarding the users having a positive or negative emotion, which induces some generalisation. By performing classification with several feature sets, they achieved acceptable accuracy given the trade-off of computational load when including glottal timings in the feature set. Formant tracking including Newton-Raphson method is a high work-load process, while in case the eigensolver fails additional burden is created by the construction of Toepliz matrices. FFT is another technique that is computationally expensive and should be considered before being applied on a mobile device intended for continuous inference.

EmotionSense includes components for adapting the sensing process based on the context. It showcases the effects in computation, communication cost and energy for performing the computations on the device or on a backend server. Authors trained the emotion detection model on a state-of-the-art library. However, there is a need
2.6. From social signals to social behaviour inference

to evaluate this model not only based on the trained library but also in a real-world environment to understand the robustness of the model. They performed speaker recognition on samples retrieved from 10 users. But there is no indication in what type of environment the data were collected from i.e. indoor, outdoor, with(out) ambient noise etc. Furthermore, adding Brownian noise is not sufficient to prove that the detection model is able to tolerate noise introduced by real-world environments. Similarly, the emotion recognition model was only evaluated on data from the training library. In essence, providing an evaluation of each of the components (speaker, emotion recognition) individually and as a holistic approach on real-world data, would indicate the robustness of the system in daily life monitoring. This necessitates the conduction of a larger-scale experiment for further analysis.

Visage utilises a well-established, robust and accurate method for face recognition combined with the device’s orientation. However, this approach requires the user to hold the device in a position so as the mobile phone’s camera is targeting the user’s face. The face recognition approach through Fisherfaces [128] provides tolerance in variations of lightning and expressions in comparison to other techniques such as Eigenfaces [117]. Also, it should be noted that the system operates in a supervised manner. Thus, it requires from the user to provide predefined facial expressions to construct a personalised model that classifies the seven distinct emotions.

In [118] authors were able to achieve a reasonable emotion recognition accuracy (70-80%) for four emotions. They utilised a boosted Naïve Bayes for classification which introduces a certain computation load in the training process due to the creation of domain specific classifiers. Likewise, this approach is prone to the creation of domain specific classifiers for possible outliers, inducing over-fitting. The system requires pre-loaded images in the device and does not support real-time recognition of user’s emotion through facial expressions.

Based on the above techniques certain parameters should be considered. The highest accuracy was achieved through facial expression recognition in Visage. However, it induces intrusiveness due to the requirement that the device’s camera should target user’s face. Also, the computational burden induced by face recognition and facial
point tracking must be considered. *EmotionSense* managed an acceptable accuracy in an energy efficient manner, without requiring a specific on-body position of the device or any external hardware. *AMMON* provided only a preliminary classification result regarding the user’s emotion but based on the application could be utilised. Regarding [118], the restricted inference context of the application indicates it as a less qualified system with respect to the others, for continuous sensing and inference.

### 2.6.3 Mood

In contrast to emotion, mood constitutes a generic emotional state difficult to describe and infer due to its multidimensionality. For that reason, researchers tried to approximate this emotional state through detecting several social signals based on their extraction complexity and significance with respect to mood. In order to detect the emitted social signals, researchers employed physiological sensors connected to mobile phones, on which they performed the inference. The most common social signals to infer mood in literature were arousal and valence while activeness and pleasure were also leveraged. Next, the techniques developed in literature will be outline and finally a brief comparison will be provided.

One of the first pieces of research, in which the authors interpreted user’s mood was *eMoto* [161]; through a sensor that was measuring pressure and arbitrary movement (gestures) the user was applying on it. They decoded valence, effort, pleasure and arousal. Another work of mood inference on mobile phones was [162]. Authors extracted physiological features from the user and through a certain threshold they were able to detect the level of arousal of the user i.e activated and relaxed. *MoodScope* [163] is a mobile application that takes advantage of a user’s phone usage patterns. Through a two-month training they were able to estimate mood i.e. activeness and pleasure, with 93% accuracy through multi-linear regression. Authors in *eyeLook* [122] leveraged the ECS eye tracking tool to extract social signals such as attention though fixations and arousal through eye contact.

**Discussion.** The majority of works have concentrated on detecting mood, especially valence and arousal based on physiological features. Sensors measuring this type of
features provide valuable information about a user’s physiological state but require intrusive equipment that reduces the ubiquitous character of Mobile SSP. eMoto obliges the user to hold a stylus and does not consider the cultural background of the user e.g. people around Mediterranean sea tend to utilise many gestures during their conversation in comparison to people in Scandinavian that seldom perform gestures while discussing. [162] considered only the level of user’s arousal, while the threshold-based classification approach is prone to misclassification when applied to people of different cultures.

As a continuation, MoodScope performed long-term analysis and included also phone usage data. The model’s training required a considerable amount of time. Initial models had a poor performance (60-70%) and only a personalised model was able to achieve high accuracy (93%) in mood inference. Another disadvantage was that the mood detection model needed to be stored at a cloud infrastructure, requiring continuous internet communication and adding a noticeable burden on battery consumption. It is worth pointing out that the inference is based only on phone usage data, inducing minimal sensing energy consumption. eyeLook detected eye pupils and when they are dilated, which is considered as an indication of arousal. However, the eye-tracking mechanism is quite intrusive and is prone to false positives (eye detection) in an outdoor environment when it is sunny.

The emotional state of mood is not fully described by valence and arousal, indicating there is a requirement for incorporating other social signals to provide a more holistic approach. A missing parameter is the collaboration with psychologists, who will indicate the grammar of several modalities. This will provide the area with an understanding of the appropriate combination of social signals for inferring mood accurately given its multidimensionality.

2.6.4 Personality traits

Following the emotional state of mood, a more static approach in terms of time is the characterisation of a person’s personality traits [164]. These are mainly parts of a person’s character where a long-term analysis is required to identify them. Due to
a broad spectrum of personality traits, the majority of researchers have focused their works on the so called Big-Five in psychology: a) extraversion, b) emotional stability, c) conscientiousness, d) agreeableness and e) openness [165]. Although some works substituted emotional stability with neurotism, the overall concept of Big Five was the same.

Thus, in [166] authors monitored proximity among people and smartphone usage. For each trait a distinct set of features was fed into an SVM [157] and C4.5 [59] classifiers to designate the Big-Five with accuracies in the scale 69-75.9%. [167] initially examined the correlation of auditory cues with personality traits and then showed that laughter and backchannel influence significantly increased the perception of social attractiveness. [168] utilised data retrieved from mobile phone usage (calls, SMSs) and proximity (Bluetooth) to classify the happiness of the user with accuracy 80.81% through a Random Forest classifier.

Discussion. Given the afore analysed trade-off of proximity-based detection of social interactions (See Section 2.4) authors in [166] and [168] utilised simple Bluetooth discovery in order to measure the social interactions in which a user participated. Although this method is easily implementable it introduces a noticeable amount of false positives that should be taken into account. A supplementary social interaction feature is the number of remote communications that existed among the users i.e. call and SMS logs. These features assume that the owner is the only user of the device and therefore there is a need to immunise it. Overall the achieved accuracy in both works is acceptable. However, there is a lack of incorporating several informative cues such auditory, activity-based etc., which would provide a significant amount of information about the personality traits of the user.

In contrast, [167] concentrates on auditory cues and shows the correlation between them and the Big-Five. However, they do not make any attempt to classify personality traits given these specific behavioural cues. The audio data are retrieved from recorded calls and do not include any data from real face-to-face situations. Furthermore, they extracted a large amount of features, some of them are computational demanding, rising issues regarding the applicability of such continuous feature extraction on mobile
2.6. From social signals to social behaviour inference

phones. Social attractiveness inference is based on laughter and backchannel which were proven reliable cues according to the authors. Other cues could provide additional information such as physical appearance, eye contact, mimicry in speech and movement etc.

In conclusion, inferring personality traits requires mining several social signals. The Big-Five is a first step for identifying the most important social signals related to one’s personality. Nevertheless, literature includes works for distinct cues, thus an initiation of incorporating these different cues will gather a large amount of information and may provide a more holistic characterisation of a user’s personality.

2.6.5 Dominance

After the analysis of inferring a user’s personality traits, another characteristic of social behaviour is dominance. During a social interaction, a dominant person has higher social status in contrast to other submissive people. Dominance detection is a popular topic in SSP, which triggered the research on mobile devices as well. In Mobile SSP researchers mainly inferred dominance through auditory features by applying various distinct sets. In the following paragraphs existing literature of dominance inference on mobile phones will be described and analysed.

[76] is not based on mobile phones, but the methodology according to the authors is applicable to smartphones. In detail, they propose several approaches including simple rule-based inference. To introduce multi-modality, they perform feature-fusion based on the rank or a score and then utilise a rule-based classification. The features are extracted from audio (prosodic and turn-taking) and visual data. MeetingMediator [65] also detects dominant persons by computing turn-taking (speaking time, average speech segment length), prosodic features (variation in speech energy) and physical activity. A recent approach was introduced on SocioPhone [60] in which they extracted prosodic and turn-taking features, and fed them in a supervised SVM [157].

Discussion. Regarding [76], the approach utilised in this work constitutes a lightweight and simple method, however it utilises only one feature. For that reason the authors decided to perform multi-modal fusion. In detail, they applied fusion techniques based
on rank or score to generate a unique feature that incorporates a series of multi-modal features. In *MeetingMediator* the only inference they perform is to compute the correlation of each person with respect to dominance, without developing a dominance detection model. Their conclusion about important features for dominance detection showcases high significance in the speaking time and speech energy variation. Thus indicates that a possible dominance detection model should include the aforementioned features. Finally, *SocioPhone* created an SVM-based dominance detection model but did not perform any evaluation to quantitatively understand the accuracy of the model. It should be noted that they were able to achieve high accuracy in the extraction of prosodic and turn-taking features in different environments and on-body positions of the device. Thus, a real-world evaluation would provide significant information about the applicability of such a model.

2.6.6 Other social behaviours

Previous subsections analysed various social signals that contribute to the inference of some major social behaviour characteristics such as stress, emotion, mood, personality traits and dominance. Based on the literature, these are the main social characteristics that have driven researchers’ interest. However, in parallel with social behaviours inference in these works, other social signals were mined which could trigger the interest of researchers to focus on other social behaviours or even invigorate existing inferences.

Other social behaviours were predicted in [169], such as diversity (69% accuracy), loyalty (69% accuracy) and overspending (71% accuracy) through phone usage information based on calls, SMSs and calendar. In [80] authors calculated the sociability of a person based on the time speaking during his participation in a conversation. In *SocioPhone* except from training a dominance detection model, they focused on estimating characteristics such as interactivity through the number of turns-takings per minute, sparseness based on the number of silences with duration at least three seconds and skewness based on standard deviation of turn-takings.

Referring to a previous analysis about [169], they were able to achieve a medium accuracy based on survey and receipt/credit data combined with proximity and phone
usage data. The features were calculated on data collected over 1 year. Each social behaviour considered multiple modalities except overspending that utilised only proximity data fused with survey data. This method includes survey and receipts/credit data which induces human error. However, the integration of NFC technology allowing payments through mobile devices combined with incorporation of a connection of the system with user’s bank account, will eliminate the human factor and create an opportunistic sensing system with higher accuracy.

[80] estimated sociability through auditory data in comparison to SociableSense [3] that combined speaking time with proximity data. The utilisation of multiple modalities allows the inference of a larger amount of information, such as co-location. In the case of an adaptive sensing system, proximity data can be utilised as a mean that triggers the conversation detection module. Thus, there is no need for continuous speech detection while avoiding missing events. In addition, other modalities could be incorporated for sociability inference such as calls, SMSs and instant messaging services. Finally, SocioPhone extracted with high accuracy prosodic and turn-taking features but similarly to dominance inference, they did not evaluate their models for interactivity, sparseness and skewness in real-world situations, in order to understand their applicability.

2.7 Applications

In previous sections state-of-the-art techniques were analysed that may be used to infer social behaviour on mobile phones. Currently, the leading application areas will be showcased in which Mobile SSP can contribute or has already been utilised, indicating the importance and applicability of the field. Among a wide variety of applications where Mobile SSP can be leveraged, the main identified areas are health-care, organisational engineering and marketing.
2.7.1 Health-care

*Health-care* constitutes one of the most significant applications of Mobile SSP. A mobile device, through the large variety of internal and external sensors, allows constant monitoring of a patient in an unobtrusive way by simultaneously minimising the error introduced by human observer. They are able to detect minor and unnoticeable changes or anomalies in behaviour which may lead to diagnosing a disease even in the preliminary stages. Social behaviour-aware mobile devices, are capable of benefiting from the diagnosis and prevention of both physical and mental diseases [170].

A notable amount of applications focus on the physical illness aspect of health-care, diagnosis, prevention and even prediction of various physical diseases. In detail, through continuous monitoring a minor behavioural variation that may not be noticeable to a human observer or even the patient himself, may be identified by anomaly detection in a patient’s social behaviour. As an example [87] observed the pain relief of a patient resulting from surgery, by detecting behavioural cues such as activity and posture. An application focused more on prevention was presented in [79], where a user’s activity was inferred and combined with a reward system to engage users in a more healthy way of life. Also, there are situations where the patient requires long-term monitoring of physiological cues such as heart-rate, skin temperature etc. These may provide more detailed information about the overall health of the user and predicting common diseases such as obesity, high blood pressure and others including multiple sclerosis, Parkinson etc.

Apart from diagnosing physical diseases, Mobile SSP has applications in monitoring mental health as well. This area is described by changes and abnormalities outlined in patients’ behaviour which can be identified through continuous monitoring of user behaviour. A common application of Mobile SSP is the quantification of user’s stress levels in pursuance of limiting the effects of long-term high stress levels. This constitutes an application that requires short range monitoring. However, there are other mental diseases that require long-term monitoring. An example of this is the detection of evidence referring to the possibility of a person being depressed or bipolar [171] by collecting information such as mobility patterns, sociability etc.
Physical and mental diseases require a continuous, pervasive and ubiquitous monitoring tool that will provide significant information about anomalies or routines in a user’s social behaviour. This will provide unbiased information to medical experts, enabling them to perform an initial diagnosis which will be verified by them. Also, there are some works that concentrate on the way diseases spread [172] and which mechanisms are suitable to prevent these occasions.

2.7.2 Organisational engineering

Another important application field is an automatic manner to quantify and analyse several aspects of organisational engineering i.e. employees’ sociability, stress, job satisfaction [30] including information flow, workload efficiency etc. These are all significant parameters that contribute to a healthier environment with respect to the employee and the organisation itself but also in increasing the efficiency and productivity of the organisation.

Mobile SSP will fabricate a new era in understanding, modelling and predicting the behaviour of organisations while introducing the importance of the social aspect. Social behaviour of an organisation’s employees is an important parameter that is neglected today. So providing a quantification method for employees’ sociability, stress levels etc. will indicate the job satisfaction employees are feeling, and accordingly perform the appropriate adjustments. As it has been shown in preliminary research [173] it is achievable to comprehend the overall work-flow at an organisation by spotting lack of communication among different departments. This may lead to the identification of any existing or future eruption. Organisation are keen on being knowledgeable about the relationship among people, to reduce customer churn (e.g. churn prediction [174] [175] [176]), to minimise any gap in the functional process of the corporation [173] or to procure a suitable working team [177].

Organisations are dependent on their employees. This indicates the importance of being aware of their healthy social behaviour [178] to cope with early identified issues such as lack of intercommunication among various teams. Additionally, Mobile SSP may also identify possible unsatisfied customers and further assist in a correctly structured
organisation. In conclusion, Mobile SSP has the potential to provide various enablers in the field of organisational engineering.

2.7.3 Marketing

Finally an area that several applications of Mobile SSP is foreseen that will emerge, is marketing. Social sciences have become essential in marketing, due to the comprehension of human behaviour required to fulfil the appropriate needs [179]. The knowledge of the user’s general but also present social behaviour constitutes a new parameter in the area of marketing [180] [181].

One of the benefits of Mobile SSP is the ability to provide a personalisation aspect in today’s generalised marketing campaigns. This allows the identification of certain perspectives of user’s behaviour. Following, it will enable marketers to target their campaigns to a specific audience that is keen on or open to the promotional target [182].

As an extension, modelling user’s social behaviour through a mobile device may guide marketing to a new era, in which the environment will adjust automatically based on a user’s predicted preference and mood [183]. Another application that would provide benefits is the identification of potential customers [184]. An example proposed by Pentland in [25] was through leveraging only characteristics of voice, they were able to predict negotiation outcomes. This achievement would constitute an enabler for telephone-based marketers. In particular, they will recognise in short-term customers willing to accept an offer, reducing the time and effort spent for customers unwilling to be convinced.

In overall, as the field of marketing is largely correlated with the area of psychology, there is a large amount of applications that may benefit from Mobile SSP, in order to improve and facilitate customer understanding and personalised marketing.

2.8 Discussion

Mobile SSP is an important domain that has started to gain a great deal of interest due to its wide applicability. Not only psychology, but as presented there are several
fields that will potentially benefit from the growth of this area. As described, research has not concluded in the terms and the taxonomy of SSP. Thus, researchers are need to agree and finalise in the terminology of the field so a concrete area is created. This will directly affect the development of Mobile SSP, while also enhancing the modelling of social signals. Having modelled social signals will provide a more clear understanding and classification of which behavioural cues can lead to certain social signals. By analysing these signals, an explicit guideline will tutor researchers in mining social behaviour in the long-term.

As shown in Section 2.3, there are numerous works released in order to provide the appropriate abstraction for retrieving information from mobile device sensors. Some frameworks have reached a certain maturity. This enables the utilisation of these tools in the design and development of mobile social behaviour applications without the need for handling low-level procedures required for sensing, processing, storing and retrieving information. The majority of sensing frameworks are built based on modern software design patterns to ensure robustness, security, extendibility but also openness. The latter two characteristics are highly correlated through the common ground of allowing third-parties to develop their own applications upon these frameworks but also contribute custom modules to extend and improve them. In addition, selecting a sensing framework is ostensibly a complicated process. But the designer should understand that concluding on a certain framework will only constitute a (less) significant enabler in the application and will not limit its capabilities. Last but not least, the intelligence that some frameworks provide regarding energy-efficiency may prove to be an additional succour.

Detecting social interactions through mobile devices is a topic that has drawn researchers’ attention. Several approaches have been proposed by leveraging COTS mobile phones. Researchers have focused on detecting social interactions by utilising a single or multiple modalities. Each method has its own advantages and disadvantages. Most works have performed simple discovery due to the pervasive and robust character of the approach, given a large amount of error. However, this method does not limit the user on a specific wearing position, with a very low design and development complexity. It is less sensitive to environmental factors in comparison to other approaches, because
of the large spatial range it covers. The biggest disadvantage of the approach is the large amount of false positives it provides, especially in crowded places. Researchers tried to tackle this error through distance estimation based on Bluetooth, WiFi or audio signals. These approaches are highly dependent on the environment while human body absorption constitutes a significant obstacle. Voice and conversation detection have been incorporated in social interaction inference to increase accuracy. It should be noted that conversation detection constitutes a great challenge which requires tackling. It is also highly dependent on the environment and the on-body position of the device. Thus, depending on the accuracy required by the social interaction detection system, a less or more complex approach could be utilised while also considering the development effort for each methodology.

In the development of custom mobile devices, the designer decides about the components required based on the application. For that reason, there is an advantage to selecting a robust and accurate solution (sensor) that will constitute the appropriate denouement. Although this approach may provide a reliable and robust solution, the designer must put a lot of effort in limiting its intrusiveness. To this point there is no robust and reliable off-the-shelf solution for detecting social interactions on mobile phones in real-world environments.

At the moment, research has focused mainly in extracting behavioural cues because of immediacy among the device and the cue. This stage is based on the engineering part of Mobile SSP and does not necessarily require the collaboration with psychologists. Researchers have been mining several types of behavioural cues. Among them are the auditory for which, although a lot of research was conducted already from SSP, researchers applied various of these techniques on mobile devices. Although the majority of them were successfully adjusted to smartphones, there are some techniques that increase the computational burden and the energy consumption, thus this should be taken into consideration. Physical activity detection is also a topic that has gained researchers attention from the point of accelerometer incorporation on COTS mobile phones. As described, this is not a burdensome process and can be executed on mobile phones with high accuracy. Gesture recognition is an arguable cue, that to this point required either complex video processing or the user to hold the mobile phone in the
2.8. Discussion

hand. This raises questions regarding its real-world applicability. Posture detection is also mainly based on accelerometer data, with(out) external hardware, which can reliably be inferred. However the on-body position of the device should be included in the process. Facial cues extraction are based primarily on burdensome video processing and object identification. This may not be ideal for continuous sensing applications despite the psychological importance of the cues. Environmental cues provide a significant view of the behavioural cues regarding the context. A reliable and robust solution for detecting interpersonal distance and spatial arrangement from COTS mobile phones in real-world environments is still not available. The device usage based cues may not provide information about face-to-face interaction, but it constitutes a lightweight and unobtrusive that can indicate reliable contextual knowledge. Physiological cues have been extracted through specific external sensors that limit the ubiquitous character of the area. Nevertheless the type of cues they detect convey information with high significance.

As literature indicated, as opposed to extracting behavioural cues, mining social signals and social behaviours on mobile phones is still immature. This occurs due to some important reasons that researchers need to take into consideration. Social signals and social behaviours include a noticeable amount of psychological knowledge. They require systematic collaboration with psychologists which will indicate behavioural cues postulated for mining a certain type of social behaviour. There has not been any tutorial providing a clear guideline of state-of-the-art techniques utilised in each of the steps detecting social signals. A tutorial will provide a definite understanding of the area and the methodology of mining social signals and social behaviour. A popular social behaviour is stress which can be detected robustly through auditory, physiological and physical activity cues. Emotion and mood detection was also mainly performed by auditory, facial and physiological cues with over 70% accuracy. Different personality traits were primarily detected by auditory, proximity and phone usage cues indicating the need for incorporating additional cues in the inference process. Finally, dominance and social role of a person was focused on auditory cues as in SSP neglecting information such as spatial arrangement.

Overall, Mobile SSP is a multidisciplinary area that acquires a considerable amount of
knowledge from adjacent fields, indicating the importance of active collaboration. These will drive researchers to incorporate multiple modalities in each of the inference stages. Each of these modalities introduce a certain level of error, intrusiveness, computational burden and energy consumption that should be considered, as the area targets mobile phones characterised by autonomy issues.

2.9 Challenges

In previous section a discussion about the overall area of Mobile SSP and its main components was presented, identifying the key outcomes of the literature review. This research concluded in some of the most significant challenges of the area that require tackling. These challenges constitute potential opportunities for research regarding the overall area of Mobile SSP which will provide a significant stride in the development and evolution of the area. In the following subsections each of the challenges is described, while in some cases initial steps are outlined in order to fill these gaps and to provide a further reference to the reader.

2.9.1 Context recognition

Context is one of the most important factors in affective [185] and context-aware computing [186] [187], anticipatory sensing [188] and in Mobile SSP. As described earlier, SSP delves to interpret social behaviour, that requires detection of interactions among people, intertwined with the context in which it is taking place. Acquiring the knowledge of context in a more efficacious way of monitoring and understanding social behaviour, is looming. Due to the broad meaning of the term, one proposed solution for context recognition is to limit the scope of an application in order to focus on certain aspects of a specific context (e.g. monitor productivity in organisations [173]). However, comprehending and construing context is a great challenge, which requires attentive and systematic research to depict a more holistic view. An example of context recognition is to detect accurately social interactions among people which will function as a significant enabler of social signal recognition through mobile phones. An important
2.9. Challenges

step to understand context, is to combine different modalities in a seamless manner to infer social behaviour.

2.9.2 Multi-modal fusion

At this point in time, research has mainly been focussing on extracting various behavioural cues by utilising different modalities. A limited part of them has tried to infer social behaviour, either through individual or by combining a few behavioural cues in a simplistic manner. Due to their continuously increasing computational power, mobile devices allow incessant sensing of various modalities without compromising the user’s experience. In order to infer accurately social behaviour, merging information from physical and virtual sensors is an indispensable need. Novel fusion techniques may be developed to perform this data amalgamation, precluding information redundancy, increasing the classification accuracy and mining contingent additional social signals. Targeting the incorporation of multiple modalities through novel fusion techniques, researchers must be able to model the area with help of psychology to understand which combination of modalities will lead in the identification of certain social behaviour.

2.9.3 Interdisciplinary area

Mobile SSP is an area that requires coordination of different fields, in each of the stages for mining social behaviour. Starting from the sensing layer, experts in different modalities need to cooperate to leverage the most from every modality by providing appropriate pre-processing, fusion and post-processing mechanisms. These stages include expertise mainly from Electrical and Computer Engineering such as signal processing. A rife approach to extract behavioural cues and social signals is by utilising machine learning techniques. Understanding the type of modalities required to extract a certain form of social behaviour, indicates that the most important collaboration is between Engineering and Psychology \[189\]. Psychologists have great experience in social behaviour and could provide the guideline on how to infer different aspects of human behaviour. This will supply them with an automatic and concise way to monitor and
understand social behaviour. In addition, a common challenge among the areas is the issue of acquiring the knowledge of ground truth.

### 2.9.4 Ground truth

Another important challenge in Mobile SSP is the fact of establishing ground truth opportunistically in real-world experiments. In state-of-the-art methods including Mobile SSP and SSP, scientists have acquired ground truth through human observer, camera recordings or user data labelling. As mentioned before, all three methods are time-consuming and prone to human error. Establishing ground truth by asking the user to label the data, induces subjectivity from the user’s perspective and eliminates the opportunistic character that is a core idea of the field. It relies on the user’s willingness to provide the experiment’s baseline. Another approach adopted by researchers is to perform experiments in a small scale and controlled environment such as a room, in order to estimate the accuracy while understanding the method’s limitations. Knowing the limitations of the approach and achieving an acceptable accuracy for a particular application, leads to a concrete solution. This enabler is then deployed in a large scale environment to extract a higher level knowledge of a population with the accuracy that was established in the initial experiments. Although, this method has been evaluated in a controlled environment and achieved a particular accuracy, scaling the approach will introduce new sources of error that may need to be tackled. An alternative approach that will be utilised potentially as Mobile SSP is evolving, is considering as ground truth the outcome of state-of-the-art techniques. However, this method limits an enabler’s results to the state-of-the-art technique’s accuracy. Thus providing a viable methodology for establishing ground truth in social sciences and especially Mobile SSP while preserving user’s privacy, is an imperative need.

### 2.9.5 Privacy

Every application that is directly or indirectly related to humans, is also correlated to privacy [190]. For that reason a very important trade-off to be made during the design and implementation of a Mobile SSP application, is usability against privacy [191].
Regarding usability, in this context the opportunistic and non-intrusive character of Mobile SSP is considered. The target of a Mobile SSP application is to extract a certain type of behavioural information from the user. However, this target should be achieved with respect to the user’s privacy. Some solutions have been proposed to minimise the impact on the user’s privacy in crowd-sensing application, where the data are first anonymised and then retrieved from the device [28]. Privacy could be preserved by performing sensing and inference of social behaviour on the user’s device. Thus, the collected data are not transmitted to a third party application while the user has the ability to delete unwanted or sensitive information. In some cases, on-line inference is not applicable due to device resource limitations. In that sense, the designer should introduce a privacy preserving mechanism that protects users’ anonymity but also allows them to manage and expose only the desirable information in an energy efficient manner [192].

2.9.6 Energy efficiency

Today’s mobile devices have evolved significantly in terms of sensing and computation during the last decade. But a remaining issue that is challenging researchers in the field of Mobile SSP is battery consumption. To tackle this challenge, scientists may adopt alternative techniques to continuous sensing and inference. One promising approach is to apply adaptive mechanisms (e.g. reinforcement learning) in both sensing and inference regarding the context in which user/device is in. Another proposed solution is to perform the computations with subtlety either on the device or on a backend server, in an adjusted manner based on user’s preference and the device’s status. In order to allow devices to cope with the continuous computational and energetic demand, applications should be able to adapt based on user’s context e.g. to apply a conservative policy in situations were user’s social behaviour is insignificant. Regardless the existing solutions, it has been identified that there is a great deal of research that has yet to be conducted and requires exploitation in each of the stages during the inference process.


Chapter 2. A Survey on Mobile Social Signal Processing

2.10 Conclusions of the literature review

After Pentland’s introduction of *Honest Signals* [20], the research community focused on modelling, analysing and synthesising human behaviour in an automatic manner. This interest was raised mainly due to the novel point of view introduced by incorporating the social, spontaneous and native aspect of human behaviour. Capturing this type of physical signals is a challenge, but mobile devices with the pervasive, ubiquitous and unobtrusive characteristics are a candidate solution. Mobile devices are a personalised tool, that is able through intelligent learning techniques to adopt to its user’s preferences. Additionally, it eliminates the person detection process of SSP and thus provide more accurate results though less computational demanding processes. Mobile SSP is a promising area but requires a great deal of effort to overcome its main challenges. The scientific community has to finalise the core term-definition in order to establish a common ground. There have been noticeable works at lower layers of extracting social behaviour on mobile devices, e.g. open source sensing and context recognition frameworks that provide an important abstraction enhancement. Currently, there is no concrete framework for detecting and measuring social interactions on mobile phones in contrast to wearable devices that are able to accurately identify social interactions. Also, context recognition based works have to be leveraged and combined with the theoretical knowledge from the field of psychology. This will lead to modelling and analysing an additional sizeable amount of various behavioural cues in an energy efficient way. However, mining social signals and combining them to infer a user’s social behaviour is still an area in which limited research has been conducted due to lack of coordination with the field of psychology. By tackling the challenges of Mobile SSP a new realm will emerge with applications in several fields and providing numerous benefits to areas such as health-care, organisational engineering and marketing.

The contribution of this literature review that analysed in a critical manner existing works on social behaviour inference based on smartphones, provided a taxonomy of the state-of-the-art techniques and identified the key challenges and application areas in which future works may focus on, led to the identification of four significant gaps:

- The lack of an accurate and reliable mechanism targeting smartphones, in order
2.10. Conclusions of the literature review

to estimate the interpersonal distance among people in real-world situations.

- The requirement of social behaviour inference for an accurate and non-intrusive
  social interaction detection mechanism that does not induce any mobility or ge-
  ographical deployment restrictions.

- The absence of considering real-world social interactions and contextual informa-
  tion in order to quantify trust relationships among people.

- The need to compensate the mobile device displacement error that is induced due
  to arbitrary movement in daily life through a generic solution.

The aforementioned gaps of the literature review will drive the research of this the-
sis, which will tackle each of the above significant gaps. Four distinct solutions are
presented in the next four chapter, tackling each of the identified gaps. Chapter 3
proposes a novel interpersonal distance estimation technique focused on smartphones,
which does not rely on any additional hardware or require any firmware modifications.

Then, an opportunistic and collaborative sensing system that detects social interac-
tions among people based on interpersonal distance and relative orientation is designed
and developed in Chapter 4. Leveraging the social interaction detection system, a
real-world social graph is derived and combined with contextual information from the
social interaction, trust relationships among users are quantified in Chapter 5. Finally,
Chapter 6 presents a generic compensation mechanism for device displacement error.
Chapter 3

Interpersonal Distance Estimation with Smartphones

This chapter presents a novel machine-learning model that leverages Bluetooth RSSI to detect interpersonal distance among people through off-the-shelf smartphones. This technique does not induce any mobility restrictions as it is independent on any external hardware and does not require any firmware modifications allowing large-scale deployment through on-line application stores. The model classifies the interaction zone in which the users are and if they are in proximity or not. A hierarchical approach combined with the informative features and a robust classifier lead to the development of accurate models for interaction zone and proximity detection. The model is evaluated against the state-of-the-art techniques and the results show that the proposed approach outperforms the prior works.

3.1 Introduction

Interpersonal distance among people is a significant behavioural cue for human behaviour understanding. It constitutes an informative behavioural cue that allows the inference of social signals and higher level of social behaviour. Researchers have leveraged interpersonal distance to understand if people are close enough to interact [46]. In psychology interpersonal distance has been also used to recognise the social relation of
people, as interlocutors tend to interact in distinct distances depending on their social relation [193]. Context-aware and localisation systems tend to include the distance between users or even between a user and a device. The above applications constitute only a certain part of the wide application areas where distance estimation is required.

The literature for interpersonal distance estimation is extensive. For distance estimation, several types of sensing data were utilised such as Bluetooth, WiFi, audio signal and inertial data. The later requires continuous tracking of the absolute position of the user and is prone to large error deviations due to gyroscope drift, accelerometer bias and magnetic disturbance. Audio signals are combined with ToA, AoA and TDoA for distance estimation, which could not be used in pervasive computing. For WiFi and Bluetooth researchers have developed various techniques including ToA, AoA, TDoA and RSSI but the accuracy is limited.

The initiation point for the research is the question, whether smartphones are able to accurately estimate interpersonal distance among users in a real-world environment. This environment is characterised by many factors that may affect the accuracy of the distance estimation technique. Hence, there is a need for an accurate technique that will be able to cope with the environmental factor and provide an accurate estimation in real-time.

This chapter presents a novel interpersonal distance estimation based on Bluetooth RSSI. The technique requires only 6 RSSI samples\(^1\) in order to infer if the users are in proximity and also to detect what type of social relation users have based on [193].

For the social relation detection, a hierarchical approach is introduced that trains a specific model for each social relation and then an overall model infers about the social relation of two persons. Both models are based on machine-learning and in particular on a boosting technique called MultiBoostAB [194]. The models are trained based on an extensive dataset i.e. 48000 Bluetooth RSSI samples and a large bank of extracted features. The models are considering the fluctuation of the Bluetooth RSSI in the

---

\(^1\)Empirical evaluation was conducted in order to conclude on the 6-sample window. The evaluation included window sizes from 2 to 20 samples and resulted in selecting 6 due to the trade-off between accuracy and number of samples required for inference (See Figure 3.1. It should be noted that as the number of samples increases the first inference delay increases theoretically by at least 20 seconds per additional sample. In practice, evaluation showed that the 6 sample window is filled in around 1-2 minutes.
3.2 Background

This section discusses about the related works for interpersonal distance estimation. As shown in previous works custom devices managed to achieve high accuracy. However, they have some major drawbacks including: a) requirement for specific on-body position of the device, b) need for additional hardware such as RFID Reader that must be

indoor environment and also take into account the human body absorption. It is worth noting, that both models infer about the interpersonal distance with only 6 Bluetooth RSSI, allowing the interpersonal distance inference in short-term contacts.

The rest of this chapter is structured as follows. In Section 3.2 the related works to interpersonal distance estimation are presented. Section 3.3 describes and analyses the proposed methodology for interpersonal distance estimation. The setup of the evaluation experiment is outlined in Section 3.4. The results of the experiment are discussed in Section 3.5 followed by the conclusion of the chapter in Section 3.6

Figure 3.1: Empirical evaluation was conducted for different window size 2-20 samples to understand the effect of the number of samples in the window towards the inference process.
deployed in the experimental environment, which limits the mobility and the scale of
the deployment, and c) users must carry non-pervasive devices, which affect their spontaneous behaviour that is critical in this type of applications. To tackle these challenges the research target only commercial off-the-shelf mobile phones for the detection of social interactions. For a more extensive analysis of the interpersonal distance estimation techniques the reader is refereed to Chapter 2.

3.2.1 Coarse-grained distance estimation

Literature initially focused on detecting social interactions through classifying nearby devices as interacting. Prior works like [46], [195] CenceMe [38], Serendipity [196] and SoundSense [45] performed discovery and the devices that were detected in vicinity were considered as in proximity. For the discovery process, researchers leveraged technologies such as Bluetooth, GSM and WiFi. Figure 3.2 depicts the users (A), (B), (C) and (D) in a social situation. Users (A), (B), (C) and (D) are detected as being in discovery range. It should be note that only users (A) and (B) are interacting in the particular context. Also, when the discovery range of a technology increases then the number of discovered devices may potentially increase in a multi-person situation when people are in a certain distance. Thus the number of false positives regarding the people socially interacting increases.

3.2.2 Fine-grained distance estimation

Estimating distance through only using the discovery range of a particular interface, introduces a large amount of error. To improve upon this approach, researchers focused on reducing the discovery range of the interface. Literature avoided to modify the transmission strength of the signal as this would require firmware modification because the mobile software development kits do not provide such functionality. Thus, to limit the range of the discovery, researchers induced various interpersonal distance estimation techniques. Among these techniques are the ToA, TDoA, AoA and RSSI. The first three techniques suffer from some significant drawbacks including a) the need for firmware modification, b) the requirement for external hardware, c) the need for
3.2. Background

Figure 3.2: This figure shows the spatial arrangement of people including the concept of interaction zones [193]. In particular, the figure highlights the importance of understanding the interpersonal distance among people when socially interacting.

time synchronisation among the devices and d) the high development complexity. The last technique constitutes a more pervasive approach as RSSI is provided by the mobile software development kits, it does not fall in any of the above mentioned drawbacks and it has a very low development complexity.

Literature has focused on developing several path loss models based on Bluetooth RSSI such as Free Space PLM, Office PLM [143] and BlueEye [197]. Figure 3.3 shows the performance of the PLMs for interpersonal distance estimation based on the Bluetooth RSSI. This initial evaluation of the PLMs indicates the lack of ability to estimate interpersonal distance in real-world situations. This is because the PLMs are general analytical models and require specific parameters for a particular environment, but even in that case there is large deviation from the actual values. To further improve the interpersonal distance estimation, researchers developed a machine-learning technique called Comm2Sense [51] to perform proximity detection. They extracted the mean and the maximum of a 20-sample window of WiFi RSSI samples. The features were provided into a Naïve Bayes classifier with a Kernel Density Estimator. As the approach operates in an ad-hoc mode, there was a need to switch between discovering and hot
spot mode, so the devices would be able to discover each other. Switching between the two mode is only possible through modifying the mobile phones’ firmware and it also induced a certain amount of delay in the discovering process. *PhoneMonitor* [49] is probabilistic proximity estimation technique that used Bluetooth RSSI to estimate how probable it is that the two people are in proximity to perform a social interaction. In *MAUC* [50] researchers developed a threshold-based technique to detect if two people are in proximity to interact considering also if they are standing or moving. The majority of these techniques are not able to cope with the RSSI fluctuation in real-world environments and do not consider human body absorption which is a critical factor for social situations. Also they perform the analysis off-line and not in real-time, introducing several privacy issues by transferring the data to a third party component and not performing the analysis on the users’ devices providing the with full control over their personal data.

In order to improve the accuracy of the distance estimation, researchers introduce a multi-modal approach *Virtual Compass* [53]. This method combines the WiFi and Bluetooth RSSI through developing regression models. The major shortcomings of this approach is the complexity of incorporating both modalities but also the high power consumption induced by simultaneous operation of both Bluetooth and WiFi. Both drawbacks indicate the lack of applicability of the approach in real-world social situations.

State-of-the-art techniques fall short mainly in terms of accuracy but also applicability in real-world environments. To address these challenges, in the next subsection an initial introduction of the *DARSIS* system is provided, an attempt to detect interpersonal distance through an accurate and reliable manner in real-world situations.

### 3.2.3 Proposed approach

In order to overcome the gaps identified in the state-of-the-art techniques *DARSIS* interpersonal distance estimation technique is designed, developed and evaluated. This technique estimates interpersonal distance in a fine-grained manner and detects the interaction zone in which users are and if they are in proximity or not. It performs the
Figure 3.3: This figure shows the performance of different propagation models for estimating distances 0.5, 1, ..., 4.0m based on Bluetooth RSSI.

detection opportunistically when a user is detected in vicinity to the smartphone and does not require any additional hardware or users’ involvement in the sensing or the inference process. This is a novel technique that requires only 6 Bluetooth RSSI samples to provide a fine-grained interpersonal distance estimation. Machine-learning models are trained for interaction zone and proximity considering the human body absorption, and show that they are able to tolerate real-world fluctuation of the RSSI signal. The approach does not require any firmware modification and is able to operate real-time on an off-the-shelf smartphone, as it leverages the native ability of Bluetooth for ad-hoc discovery, allowing large-scale deployment through app stores. It is a privacy-preserving approach as the sensing and the inference are performed on-line on the smartphone and the data are not transmitted to any third party components.

3.3 Methodology

Psychology has show that proxemics is the an important component of social behaviour and social interaction. Hall [193] first introduced the correlation between the interpersonal distance in which people interact and the social relation they have. In particular, he mapped the interpersonal space among people into four interaction zones, depending on the social relation of the people:
• **Public.** People that have interpersonal distance larger than 3.5m do not participate in a social interaction and do not consider to have a social relation.

• **Social.** In this interaction zone people have interpersonal distance in the range of 1.5-3.5m and have an impersonal social relation.

• **Personal.** In this interaction zone people have interpersonal distance in the range of 0.5-1.5m and have social relation such as friends and family.

• **Intimate.** People who have an intimate social relation tend to have interpersonal distance less than 0.5m.

For the detection of these interaction zone, there is a need to develop an accurate interpersonal distance estimation technique. In Section 3.2 it is argued that in order to remove any requirements of firmware modifications or external hardware this research focuses on RSSI-based interpersonal distance estimation techniques. Current methods are not accurate as they are not able to cope with the fluctuation of the RSSI. Also they depend on a large number of RSSI samples such as in [51] where authors require 20 samples to perform inference. The requirement of large number of samples lead to an increase in the sensing period but also in many cases will include outdated samples. To overcome the above mentioned problems, in the following subsection a novel technique is presented for fine-grain interpersonal distance estimation.

### 3.3.1 Overview

Through empirical evaluation, RSSI-based state-of-the-art techniques showed a lack of accuracy in interpersonal distance estimation. A novel machine-learning based technique is proposed that leverages only a 6-sample RSSI window. The main idea is to develop a classification of interaction zones through a hierarchical fashion in order to achieve higher accuracy and in the meantime maintaining a low number of window-samples, to allow the detection short-time interactions. The DARSIS Hierarchical Classifier (DHC) is based on two layers of machine-learning models. On the first layer, the classifiers are trained for a particular region, for example public, social and personal zone. These constitute a set of domain expert classifiers. The confidence of the
classification from this layer constitute a fuzzy membership for each window of RSSI samples. On the second layer, the classifier tries to identify the optimum thresholds for classifying the membership values into the correct interaction zone. It should be noted that all classifiers that were evaluated faced important difficulties in identifying the correct class between intimate and personal zone. Thus, the intimate and the personal zones were merged and the classification provided results among public, social and personal zones.

Also, the DARSIS Proximity Classifier (DPC) was developed to infer if a user is in proximity or not. In order to improve the accuracy of the estimations faced in state-of-the-art approach, special attention was paid in the feature extraction and training model processes.

### 3.3.2 Training the models

The interpersonal distance estimation technique relies on initially generating a generic training set and then extracting a bank of features from this training set. A feature selection process is followed in order to choose the most informative and less redundant feature set. Each classifier in each level of the hierarchy is evaluated in order to find the most appropriate choice.

A data collection campaign was performed in an indoor office environment through HTC One S smartphones, in order to construct the training set for the classifier. The Bluetooth interface on one of the smartphones was configured as discoverable and the other device was performing the discovery process. After the end of the experiment, Bluetooth RSSI data were collected from eight different distances, three different device relative orientations. These device relative orientation were a) Screen-to-screen b) Screen-to-Back c) Back-to-Back. Empirical evaluation showed that these vertical relative orientations constitute representative for the effect of the facing direction variation. A large number of Bluetooth RSSI samples was collected i.e. 2000 samples for each different distance and orientation combination, resulting in a dataset of 48000 Bluetooth RSSI samples, for reasons of statistical significance. As the data collection process was extremely lengthy, the humans were replaced with water-filled cylinders...
Figure 3.4: Features for 2-Layer DHC.

to which devices were attached, in order to simulate to human body absorption [198].
The devices with the bottles were placed at 0.8m height from the floor to simulate the
most common wearing position (i.e. trousers pocket) [107].

For each different distance and different orientation, 2000 samples were collected. This
resulted in a large dataset of 48000 samples. As Figure 3.2 indicates, the interaction
zones were considered as Public, Social and Personal+Intimate. So the training data
were split into three interaction zones according to the corresponding distances. In this
way the Public and Personal+Intimate zones resulted with 12000 samples each and
the Social zone had 36000 samples. This occurred because the Social zone takes into
account samples at 2m, 2.5m and 3m. For that reason in the Social zone, the number
of samples for each distance and orientation was reduced in order to have 12000. This
was done in order to avoid any bias in the training process towards the target class of
the final classifier.

Literature has mainly focused on extracting at most 3 features in order to develop a
machine-learning model that will be able to perform the mapping between the RSSI
of either Bluetooth or WiFi signal towards the distance between the emitter and the
receiver device. In Comm2Sense [51] authors selected only the maximum and the av-
average value of a 20-sample window. Based on these two features authors training a
machine-learning model. To further improve on the state-of-the-art distance estima-
tion technique, there was a need to create a large bank of features, from which the most informative features would be selected. As Comm2Sense [51] selected only the maximum and the minimum value as feature set, the proposed process included all the basic statistics such as min, max, average, standard deviation etc. In addition, similar to the approach followed in K-means algorithm, the distance of the window statistics with the basic statistics of the target class were also included. Also, the deviation and z-score used to derive social signals [25] in literature, were also included in the feature bank.

A large feature set of 3050 features including several statistics was extracted from this dataset, considering a maximum window of 6 Bluetooth RSSI samples. Table 3.2 shows the basic features that were extracted and refer to various statistics. Table 3.3 shows the relative features which are produced by combining basic features with various statistics of the target class. Table 3.4 are features that combine the basic and relative features through a statistical metric such as deviation and z-score. Such large number of features was generated in order to be confident that the feature reduction techniques will produce the most informative features. Given the level of consistency [199] of each feature based on the target class, a subset of features is chosen. A wrapper subset evaluation [200] followed to retrieve an optimised feature set for the given classifiers.

To conclude on the appropriate feature selection technique, the methods presented in Table 3.1 were evaluated on the dataset. Each method was evaluated on the dataset with a particular ranker to understand the importance of the selected feature set that the approach concluded on. Only the feature selection techniques based on the information gain and the level of consistency were able to provide a high ranked feature set. Having concluded on these two feature sets, they were used in order to train two classifiers with the same classification technique (J48 [201]). The technique that evaluates the worth of a subset of attributes by the level of consistency in the class values when the training instances are projected onto the subset of attributes, achieved the highest accuracy and for that reason the particular subset was selected.

The next step is the creation of the machine-learning based model for understanding the interpersonal distance among the people. Towards the fulfilment of that goal,
Table 3.1: Feature selection techniques

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>202</td>
<td>Evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy between them.</td>
</tr>
<tr>
<td>Information Gain Ratio</td>
<td>Evaluates the worth of a feature by measuring the gain ratio with respect to the class.</td>
</tr>
<tr>
<td>Correlation</td>
<td>Evaluates the worth of a feature by measuring the correlation (Pearson’s) between it and the class.</td>
</tr>
<tr>
<td>Information Gain</td>
<td>Evaluates the worth of a feature by measuring the information gain with respect to the class.</td>
</tr>
<tr>
<td>Chi squared</td>
<td>Evaluates the worth of a feature by computing the value of the chi-squared statistic with respect to the class.</td>
</tr>
</tbody>
</table>

Table 3.2: Basic Features

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mean</td>
</tr>
<tr>
<td>2</td>
<td>Median</td>
</tr>
<tr>
<td>3</td>
<td>Min</td>
</tr>
<tr>
<td>4</td>
<td>Max</td>
</tr>
<tr>
<td>5</td>
<td>MidRange</td>
</tr>
<tr>
<td>6</td>
<td>MinMode</td>
</tr>
<tr>
<td>7</td>
<td>MaxMode</td>
</tr>
<tr>
<td>8</td>
<td>Percentile 25th, 75th and 90th percentile</td>
</tr>
<tr>
<td>9</td>
<td>IQR</td>
</tr>
<tr>
<td>10</td>
<td>MAD</td>
</tr>
<tr>
<td>11</td>
<td>STD</td>
</tr>
<tr>
<td>12</td>
<td>Kurtosis</td>
</tr>
<tr>
<td>13</td>
<td>Skewness</td>
</tr>
</tbody>
</table>

Table 3.3: Relative Features

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Interaction Zone</td>
<td>The difference between the basic feature of a window of samples and each of the interaction zones (Public, Social, Personal+Intimate).</td>
</tr>
<tr>
<td>2 Distance</td>
<td>The difference between the basic feature of a window of samples and each of the distances (0.5m, 1m, 1.5m, . . . , 4m).</td>
</tr>
<tr>
<td>3 Orientation</td>
<td>The difference between the basic feature of a window of samples and each of the distances and orientations (0.5m Screen-to-Screen, 0.5m Back-to-Screen, 0.5m Back-to-Back, etc.).</td>
</tr>
</tbody>
</table>

Table 3.4: Combined Features

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 MovDev</td>
<td>The deviation of a specific window of samples with respect to the mean of a class (interaction zone, distance, distance &amp; orientation).</td>
</tr>
<tr>
<td>2 ZScore</td>
<td>The z-score of a specific window of samples with respect to the mean and standard deviation of a target class (interaction zone, distance, distance &amp; orientation).</td>
</tr>
</tbody>
</table>
various evaluations were performed. Machine-learning models were trained using the feature set described in the previous paragraph and defined in Table 3.5. Classification algorithms such as decision trees, naive bayes, ada boost etc. were used to train machine-learning models. The models were evaluated based on the training set with 10-fold cross validation and 25% split. The algorithms that achieved the highest accuracy were the decision tree and the MultiBoostAB [203] with decision tree J48 [201]. The MultiBoostAB approach achieved even higher accuracy than the decision tree, thus it was selected. The model was then evaluated in small 10-minute experiments in indoor office environments at different distances, where two users were placed at the centres of different interaction zones. The inference of the model was logged to understand the accuracy of the model. Also, the evaluation presented in PhoneMonitor [49], where two users walked for a particular distance in indoor environment next to each other, in order to understand if the people are interacting. In all experiments, the model achieve accuracy higher than 80%.

It should be noted that the machine-learning models were trained and evaluated on the same dataset. The algorithms were initially trained and evaluated based on 10-fold cross validation and 25% split on the different machine-learning models. Once the final machine-learning algorithm was selected, based on which the model will be trained, the model was trained based on all the dataset. This process improved the accuracy of the algorithm but tuned the algorithm towards the particular dataset. In the evaluation Section 3.4 it is detailed that the state-of-the-art machine-learning models that the approach was evaluated against, were also tuned with the same process. The PLMs were configured for the particular indoor environment, based on the RSSI measurements of the training dataset. This process reduces the generality of the model, however as mentioned in previous paragraph additional evaluation was performed that still achieved over 80% in different indoor settings e.g. office environment, standing, walking etc.

Initially a DARSIS Single Classifier (DSC) is trained for all three interaction zones based on several algorithms. Various evaluations showed that MultiBoostAB [203] with decision tree J48 [201] performed best combined with the features showed in Table 3.5. In detail, it managed the highest accuracy in a robustly manner because of its native
capability for variance and bias reduction. To further improve the performance, the hierarchical classifier DHC depicted in Figure 3.4 was introduced. The accuracy and robustness of MultiBoostAB in the inference process, led to the development of the models of each layer of the DHC based on the same algorithm. This led the approach in achieving an even higher accuracy than the DSC.

Table 3.5: DARSIS Single Classifier

<table>
<thead>
<tr>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>MovMedian6</td>
</tr>
<tr>
<td>MovDev6_Mean3m</td>
</tr>
<tr>
<td>MovDev6_Mean1mF2F</td>
</tr>
<tr>
<td>Kurtosis6 - Kurtosis4mB2F</td>
</tr>
<tr>
<td>MovMax3</td>
</tr>
<tr>
<td>MovDev6_Mean4mF2F</td>
</tr>
<tr>
<td>MovDev4_Mean4mB2F</td>
</tr>
<tr>
<td>STD6</td>
</tr>
</tbody>
</table>

Based on psychology and also as depicted in Figure 3.2, people are interacting when they are in social or personal zone. Following that, a DPC was developed that detects if users are in proximity or not. Given that DHC infers in which interaction zone the users are, DHC was used to infer also if users are in proximity by considering social and personal zone as proximity and public as no-proximity.

3.4 Experimental setup

In order to benchmark the DARSIS interpersonal distance estimation, some commonly used evaluation techniques were applied on the training data of the classifiers. The most common state-of-the-art techniques were developed and evaluated on the same dataset. The benchmarking was based on data collected from an indoor office environment and considered a large amount of data including different distances (0.5m, 1m, \ldots, 4m) and device orientations (screen-to-screen, back-to-screen, back-to-back). The machine-learning approaches were evaluated based on the 10-fold cross-validation technique, which determined the accuracy of the model.

Before the evaluation of the state-of-the-art techniques each approach was configured. The machine-learning technique was trained based on the same dataset introduced in Section 3.3. The features used were the same as the ones used in the initially proposed technique. The PLMs are mathematical models that require some parameters according to the environment in which they are focused. There is no training involved. So, these
3.4. Experimental setup

parameters for each of the PLM, were extracted from the same dataset introduced in Section 3.3. For example the $RSSI_{Ref}$ was calculated from the average RSSI signal at 1m distance at the dataset. In this way the PLMs were configured for the particular environment that they will evaluated. Also, each machine-learning technique received a 6-sample window to perform inference. The same approach was followed with the PLMs, where the 6-sample window was averaged and then the average value was fed into the PLM.

From the state-of-the-art techniques used in the evaluation process only Comm2Sense [51] was based on machine-learning. Both Comm2Sense and [52] used the same machine-learning model, thus by referring to Comm2Sense both approaches are considered. The rest of the techniques used in the evaluation process were PLMs including BlueEye [197], Free Space PLM, Office PLM [143].

For PLMs given equation (3.1), the following parameters were computed for the specific environment $RSSI_{Ref} = -56.7977$ at 1m distance, for Free Space $n_{free} = 2$ and for Office $n_{office} = 3.134$; also the parameter $X_\sigma = 0$ was considered for line-of-sight.

$$RSSI = RSSI_{Ref} - 10 \times n \times \log_{10}(d) + X_\sigma$$  \hspace{1cm} (3.1)

As mentioned, Figure 3.3 depicts the estimations of PLMs in the indoor environment with empirical measurements. When the RSSI is close to the mean then the PLM are able to estimate the correct interpersonal distance. As in real-world environments Bluetooth RSSI has a large fluctuation, this indicates that PLMs are not able to follow that variation. Especially, BlueEye [197] performs close to the Free Space PLM. Based on the evaluation process, BlueEye was not able to solve the equation required in order to perform the inference. When the equation was solved, then the accuracy of the approach was still very low. Thus, from this point the BlueEye method in not considered in the evaluation process.
Table 3.6: Confusion Matrices for evaluation of Interaction Zone Detection against state-of-the-art in percentages (%).

<table>
<thead>
<tr>
<th></th>
<th>Public</th>
<th>Social</th>
<th>Personal + Intimate</th>
<th>Public</th>
<th>Social</th>
<th>Personal + Intimate</th>
<th>Public</th>
<th>Social</th>
<th>Personal + Intimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>88.05</td>
<td>10.23</td>
<td>1.72</td>
<td>72.88</td>
<td>24.63</td>
<td>2.49</td>
<td>35.58</td>
<td>60.12</td>
<td>4.3</td>
</tr>
<tr>
<td>Social</td>
<td>4.93</td>
<td>94.13</td>
<td>0.94</td>
<td>9.16</td>
<td>89.38</td>
<td>1.46</td>
<td>5.36</td>
<td>92.31</td>
<td>2.33</td>
</tr>
<tr>
<td>Personal</td>
<td>1.16</td>
<td>1.1</td>
<td>97.74</td>
<td>2.03</td>
<td>2.55</td>
<td>95.42</td>
<td>5.02</td>
<td>14.76</td>
<td>80.22</td>
</tr>
<tr>
<td>+Intimate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) DHC  
(b) DSC  
(c) Comm2Sense

<table>
<thead>
<tr>
<th></th>
<th>Public</th>
<th>Social</th>
<th>Personal + Intimate</th>
<th>Public</th>
<th>Social</th>
<th>Personal + Intimate</th>
<th>Public</th>
<th>Social</th>
<th>Personal + Intimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>24.15</td>
<td>34.5</td>
<td>41.35</td>
<td>40.12</td>
<td>34.52</td>
<td>25.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>15</td>
<td>51.8</td>
<td>33.2</td>
<td>45.95</td>
<td>37.04</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal</td>
<td>0.02</td>
<td>0.23</td>
<td>99.75</td>
<td>0.03</td>
<td>12.34</td>
<td>87.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Intimate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(d) Office PLM  
(e) Free Space PLM

3.4.0.1 Performance metrics

For the evaluation process, one of the performance metric utilised is the accuracy of the approach with respect to the target classes. In particular the performance is calculated through confusion matrices, Receiver Operating Characteristic (ROC) curves and the overall accuracy. The confusion matrices is a performance metric to understand the misclassification of the model. The ROC curves shows the trade-off between the true positives (sensitivity) and the false positives (1 - specificity) of a particular classifier and a particular class i.e. interaction zones and proximity. In addition, ROC curves provide useful information in cost sensitive classifications. The diagonal (y=x) in ROC plot shows the random classification. The perfect classifier would plot the line in the upper left corner of the ROC curve with 100% sensitivity and specificity. The confusion matrices in Table 3.6 show an overview of each techniques’ average accuracy, then the ROC curves indicate the rate of correct classification.

3.5 Results

Table 3.6 presents the confusion matrices for detecting interaction zones. The DHC achieved the best accuracy among the evaluated approaches, having a small misclassification error between public and social zones. The DSC performed also very well having the main misclassification error between public and social zones. Regarding the
3.5. Results

Table 3.7: Confusion Matrices for evaluation of Proximity Detection against state-of-the-art in percentages (%)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) DPC</td>
<td>77.36</td>
<td>22.64</td>
<td>(b) Comm2Sense</td>
<td>49.94</td>
<td>52.06</td>
<td>(c) Office PLM</td>
<td>40.13</td>
<td>59.87</td>
<td>(d) Free Space</td>
</tr>
<tr>
<td>Not</td>
<td>5.94</td>
<td>94.06</td>
<td>Prox.</td>
<td>8.78</td>
<td>91.22</td>
<td>Not</td>
<td>30.64</td>
<td>69.36</td>
<td>Not</td>
</tr>
</tbody>
</table>

state-of-the-art, Comm2Sense \[51\] detected accurately the social and personal zones, however there was a very high misclassification percentage error for the detection of the public zone. The Office PLM achieved high accuracy for the detection of the personal zone, yet a high misclassification error was observed when detecting the social and public zones. The Free Space PLM managed satisfactory accuracy for personal zone as opposed to social and public zones where the approach could not detect these zones. In overall the DHC method achieves the highest and the most stable accuracy over the three interaction zones, and also has the highest rate for social zone. It should be mentioned that Office PLM achieved the highest accuracy for personal zone detection, however public zone is detected with a very low accuracy. Thus, the total accuracy of Office PLM is low. The low performance of the PLMs for recognising the public and social zones as opposed to personal, is occurring because in short distances the fading effect is negligible.

Next, Table 3.7 presents the confusion matrices for proximity detection. The DPC achieves high accuracy for proximity classification. There is a small misclassification error with the false negatives when users are not in proximity. The Comm2Sense managed to detect proximity, however the is misclassification error for the detection of non-proximity. Both the Office PLM and the Free Space PLM are characterised by considerable misclassification error. The Free Space PLM advanced in comparison to Office PLM, but both accuracies are not applicable in real-world environments. In overall, the DPC achieved the highest accuracy for detecting if users are in proximity comparing to the state-of-the-art techniques.

Although Office and Free Space PLMs are able to detect with high accuracy the personal zone, there is misclassification error when estimating social-public zone and detecting proximity or not. This confusion error is produced due to reflections, which are forcing
the RSSI values to overlap at specific distances (See Figure 3.3). As proved, this fluctuation of the RSSI values is tackled by the machine-learning techniques through the time dependence, in contrast to PLM that estimate based on the current value. The DHC takes into account the hierarchical structure, informative features and the robust classifier, allowing the approach to tackle the confusion between public and social zones, which the state-of-the-art techniques were not able to cope with. When users are in close distances such as personal zone, the state-of-the-art techniques are able to classify correctly the personal zone, as the effect of environmental factors is minimal. Thus, in close distances all approaches were able to achieve high classification accuracy.

Figure 3.5: ROC Diagrams for DARSIS and state-of-the-art approaches.

Figure 3.5 presents the ROC curves of the state-of-the-art techniques and the proposed approaches for classifying interaction zones and proximity. The Figure 3.5a shows that the DHC performs better than all the approaches used in the evaluation. In particular,
Table 3.8: Overall accuracy for Interaction zone and Proximity detection in percentages (%).

<table>
<thead>
<tr>
<th>Zones</th>
<th>DHC</th>
<th>DSC</th>
<th>DPC</th>
<th>Comm2Sense</th>
<th>Office PLM</th>
<th>Free Space PLM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>93.52</td>
<td>86.76</td>
<td>-</td>
<td>75.10</td>
<td>56.87</td>
<td>50.46</td>
</tr>
<tr>
<td>Proximity</td>
<td>88.50</td>
<td>89.88</td>
<td>88.50</td>
<td>80.39</td>
<td>62.05</td>
<td>54.03</td>
</tr>
</tbody>
</table>

more than 90% of the true positives are produced, while having less than 10% of false positives. The DSC managed close to 80% of true positives while having produced 10% of the false positives. Then, the ROC curve made a shallow slope and significantly increased the rate of true positives and also increased the rate of false positives. For the Comm2Sense, the model presents a high rate of false positives from the beginning. But when almost managing 50% of false positives, it achieves the perfect true positive rate. For the Free Space PLM and the Office PLM, the predictions are close to that of a random classifier. This means that the performance is close to a line where true positive rate is equal to the false positive rate.

Figure 3.5b presents the results for detecting the social zone. The DHC technique outperformed the rest of the methods, followed by the two machine-learning techniques, the DSC and the Comm2Sense. Likewise to the public zone detection, Comm2Sense manages 100% of true positives while achieving 50% of false positives. The Free Space and Office PLMs have worse performance, but they indicate improvement in their performance in comparison to the public zone. For the personal zone, the results are presented in Figure 3.5c. Related to the public and social zones, the majority of the techniques for personal zone detection have managed to perform better in terms of the ROC curves. The DHC outperforms all the other techniques for personal zone detection and is able to almost achieve the perfect performance in the top left corner. Finally, the Figure 3.5d presents the performance in terms of ROC curves for the proximity classifiers. The proposed approach managed the best performance regarding the state-of-the-art techniques. The Office and Free Space PLMs performed worst with respect to the machine-learning techniques, having performed close to a random classifier.

Table 3.8 presents the overall accuracy for each of the evaluated approaches, with respect to the interaction zone and proximity detection. The proposed approaches
DHC, DSC and DPC performed best with respect to the rest of the evaluated methods. This is because of the utilisation of informative and consistent features, which were combined with a very powerful classifier. As the table shows, the Comm2Sense managed to perform well using only two features for classifying interaction zones and proximity. Regarding the PLMs, Free Space and Office, the accuracy was quite limited showing that these methods are not applicable to real-world environments due to to the generality of the models. The proposed DHC model achieved the highest accuracy that proved the effectiveness of the hierarchical structure. Finally, the DHC achieved an improvement of at least 8% in the detection accuracy in comparison to the rest of the evaluated techniques.

3.6 Conclusion

This chapter presented and evaluated a novel methodology to perform interpersonal distance estimation in an accurate, reliable and low complex in terms of implementation technique. The technique is based on machine-learning, were an extensive data collection process was performed that considered the fluctuation of the indoor environment but also the absorption of the human body. Using only a 6-sample window of Bluetooth RSSI samples the technique was able to accurately recognise the interaction zone in which people are and if they are in proximity or not. Informative features and a robust classifier led to the development of accurate models for interaction zone and proximity detection. In addition, for interaction zone detection an hierarchical approach was introduced that trained domain expert classifier to allow the creation a accurate and reliable classifier. The technique does not require any firmware modification or external hardware. It utilises the native capability of all the modern software development kits for smartphones to retrieve the Bluetooth RSSI. The devices are able to discover and be discovered by other devices as the technique is based on Bluetooth’s native capability for ad-hoc discovery. The proposed approaches were evaluated against the state-of-the-art techniques based on a 48000-sample dataset with different distances and device orientations. The evaluation showed that the proposed models were able to outperform the state-of-the-art techniques. As potential future work, it is identified
that deep learning approaches such as Recursive Neural Networks [204] could be lever-aged to explicitly learn the features of the model and to replace the ad-hoc approach proposed in this work.

The development of this interpersonal distance estimation technique constitutes the first step towards the recognition of social interactions and further trust relationships based on off-the-shelf mobile phones. In social situations utilising only the interpersonal distance in order to understand social interactions is not enough. In the next chapters, to improve the accuracy of social interaction detection, the relative orientation of users is incorporated in the process. For the relative orientation computation there is a need to estimate the users’ facing direction and then perform the appropriate computation. A state-of-the art technique, uDirect [5], provides the ability to recognise users’ facing direction based on users’ walking locomotion, independent of the on-body position of the device. In the next chapter, the complete DARSIS system is presented that detects social interactions through off-the-shelf mobile phones. In addition, error analysis is performed by developing analytical models and combining them into a social interaction mechanism as well as providing error models for each of the components and an overall error estimation model.
Chapter 4

Modelling Social Interaction Detection with Smartphones in Real-world

Understanding real-world social interactions can provide valuable insights into human behaviour. In particular, quantifying social interactions provides a strong measure for describing social contacts and real-world engagements of humans in their daily lives. Current tools to capture social interactions are either limited in accuracy or require additional hardware, which in turn makes them intrusive for long-term observations in unconstrained environments. This chapter presents the complete DARSIS system, an opportunistic sensing system that solely relies on sensors and communication capabilities of off-the-shelf smartphones to detect social interactions. DARSIS exploits a novel hierarchical model to estimate the interpersonal distance of humans based on Bluetooth Received Signal Strength Indication (RSSI). Furthermore, it combines the derived knowledge, with estimations of users' direction, which are retrieved from a collaborative sensing process between mobile phones\footnote{The term collaborative sensing refers to dyadic pairing of mobile devices, that allows the exchange of information among two devices.} to detect social interactions. Through a series of real-world experiments, the accuracy of the proposed mechanisms is evaluated and benchmark them against other state-of-the-art systems. The evaluation
results confirm the applicability of the proposed system as non-intrusive solution for social interaction detection and show that our approach outperforms the state-of-the-art solutions in terms of accuracy and reliability.

4.1 Introduction

Social interactions provide important information about people’s behaviour in their daily live. While participating in a social interaction, humans communicate through verbal and non-verbal signals. Humans are interacting when they are proximate to each other, they are facing each other and they exchange verbal signals. As presented in [205], when people are interacting they also transmit social signal to each other, in order to convey different messages. Also, a correlation between human networks and contextual information was presented by Pentland [206] including the location, interpersonal distance and social interaction.

Literature tried to strive social interaction detection through surveys, questionnaires, human observers and camera recordings. These techniques are not automated and are prone to human errors as they involve the human factor in the sensing and inference process. By performing the sensing and the inference process automatically i.e. without involving the human factor, a reduction the error is anticipated. One of the first attempts that tried to create self-acting tools, was the development of custom mobile device like wearables. Some of these custom mobile devices such as [207] [208] managed to accurately understand real-world social interactions among people. It was shown that interpersonal distance and relative orientations among people is sufficient in order to understand real-world social interactions. Even though these works were able to accurately understand social interactions, they used intrusive hardware in order to achieve that result. In particular, users have to wear intrusive mobile devices on a specific on-body position that are not common in their daily life, while also there is a need to deploy external hardware that increases even more the intrusiveness of the systems.

Recent advancements in the area of smartphones have created an emerging opportunity to overcome the drawbacks of current mobile devices [2]. The advances in sensing and
computational capabilities and the use of smartphones in people’s daily life have led to consideration that smartphones are pervasive devices and ideal to observe peoples’ social behaviour and social interactions.

The question that drives this research is whether smartphones could operate as suitable platforms for understanding social interactions in real-world situations. Prior works have some major drawbacks: a) interpersonal distance estimation techniques suffer from inaccuracy in real-world environments, b) they do not consider users’ relative orientation in a realistic way, c) an ad-hoc communication channel for transmitting users’ facing directions, and d) the social interaction detection is prone to delay in sensing process.

This chapter introduces the overall DARSIS system, the realisation of a pervasive social interaction system based on smartphones that address the above drawbacks. DARSIS is an opportunistic and collaborative sensing system that leverages the sensing and computational capabilities of off-the-shelf smartphones to understand social interactions among people. The system incorporates the novel interpersonal distance estimation technique for interaction zone and proximity detection presented in Chapter 3. It computes users’ relative orientation based on the estimated facing direction each device calculated through an improvement of a state-of-the-art technique for facing direction estimation. DARSIS is operating in a distributed manner through performing the sensing and inference process locally on the device and thus preserving users’ privacy. As a result, the system is does not require any additional hardware or infrastructure and is not affected by any mobility restrictions.

The contributions of this chapter are briefly described below:

- An improvement of a state-of-the-art technique for facing direction direction is presented, which considers the error distribution of the approach to remove outliers and cluster the correct estimations but also the device displacement (See Chapter 6) occurring in real-world situations.

- A collaborative sensing technique is designed and developed to enable the exchange of facing directions and Bluetooth RSSIs, required to compute relative orientation estimation and to speed up the sensing and inference process.
Chapter 4. Modelling Social Interaction Detection with Smartphones in Real-world

- An opportunistic and collaborative system for social interaction detection is presented that incorporated a novel interpersonal distance estimation (See Chapter 3) and the proposed relative orientation computation technique.

- An analytical model in order to calculate the probability of two people participating in a social interaction based on their interpersonal distance and relative orientation in a certain context. Based on this analytical model, a probabilistic error model is derived that calculates the expected error of the approach in a certain context.

- The proposed system is evaluated in real-world experiments to showcase the importance of incorporating the relative orientation, to benchmark against an RFID-approach but also to understand the error of the proposed system and analytical model.

The rest of this chapter is structured as follows. In Section 4.2 the related works to social interaction detection are presented. Section 4.3 presents the DARSIS system and analyses the proposed methodology for social interaction detection. An analytical model is designed in Section 4.4 from which an error model is derived. Section 4.5 presents the evaluation of the DARSIS system and the error model in different real-world environments to understand the effect of incorporating the relative orientation, the similarity of the DARSIS system with the analytical model but also the system’s performance in comparison to a state-of-the-art technique based on RFID-approach.

4.2 Related work

This section provides a discussion about related works for understanding social interactions. Prior works for detecting social interactions focused on designing and developing custom mobile devices. These devices leveraged technologies such as Infrared [207] [209] and RFID [208] to understand if the users are in proximity to socially interact. In order to incorporate the relative orientation of the two users, researchers leverage the different properties of each modality. In the case of Infrared [207] [209], there is a need to keep line-of-sight contact to transmit an appropriate signal, which means that by placing
4.2. Related work

the mobile device on users’ chest and when the a signal is received from a device, the two people are facing each other i.e. they are interacting. In the case of RFID [208], to achieve the directionality of the RF-signal, they adjusted the transmission power of the signal so as it will not be able to penetrate the human body due to its absorption, leading to a iconic transmission range at the front plane of the torsos of each user. In that way, these technologies managed to induce the relative orientation factor in the social interaction detection process.

Although these technologies were able to achieve high accuracy for social interaction detection, they have some major shortcomings. In particular, they have the following requirements that increase the intrusiveness of the systems and limit their mobility: a) constant on-body position of the device, b) additional hardware needed to be deployed in the monitored environment and c) participants have to carry obtrusive devices that affect their spontaneous behaviour which is critical in social behaviour inference applications. Given these major drawback of these technologies, the research is focused on detecting social interactions based on smartphones.

4.2.1 Proximity detection

For detecting social interactions through smartphones, researchers focused on detecting proximity among users and considering that they are interacting if they are close enough, independent of their relative orientations. For detecting proximity literature considering in the beginning the devices that were detected after the discovery process of Bluetooth or WiFi such as [46]. The range of discovery depends on the means utilised, for example Bluetooth average discovery range is 10m radius [210]. However, according to psychology [211] people interact at most until 3m interpersonal distance, thus the discovery range of Bluetooth or WiFi constitutes a technique that induces false positive error.

To overcome this error researchers strived to perform interpersonal distance estimation to limit the discovery range between 2 to 3m radius. The techniques used were ToA, TDoA, AoA and RSSI, but this work focuses only on RSSI-based techniques to discard the firmware modification and external hardware requirement that the first three tech-
Chapter 4. Modelling Social Interaction Detection with Smartphones in Real-world

Figure 4.1: This figure shows the spatial arrangement of people. In particular, the figure highlights the importance of incorporating both interpersonal distance and relative orientation in the process of social interaction detection. In this situation only user (A) performs a real-world social interaction with user (B). Users (C) and (D) have interpersonal distance and relative orientation with respect to user (A), not suitable to perform a social interaction [212].

Techniques are dependent on. To estimate proximity through RSSI, literature developed PLMs such as Free Space PLM and Office PLM [143], probabilistic models such as [49], threshold-based such as [50] and based on machine-learning such as [51]. Empirical and experimental evaluation in Chapter 3, these approaches are not able to deal with the fluctuation observed in real-world environment, meaning they have limited accuracy. For extensive discussion of the interpersonal distance estimation techniques, the reader is referred to Chapter 2 and 3.

4.2.2 Multi-modal detection

To improve the accuracy of detecting social interactions, literature tried to incorporate multiple modalities in the inference process. An initial approach was to combine Bluetooth and WiFi in the interpersonal distance estimation but also provide the relative spatial arrangement of the people through Euclidean geometry. Although the approach claims an error of 1.45m in interpersonal distance estimation, the technique requires firmware modification to adjust the signal transmission power in Bluetooth and WiFi.
but also the combination of both interfaces increases the power consumption of the device, which makes it unsuitable for continuous daily monitoring. It should be noted that the approach allows the inference of only the relative spatial arrangement and not the relative orientation of the users which is significant for social interaction detection.

Figure 4.1 shows an example of a spatial arrangement in a social situation. The techniques that perform discovery of nearby devices would detect all participants as socially interacting. However, participants (A), (C) and (D) are not participating in a social interaction. By detecting proximity, the participant (D) would be correctly discarded from the social interaction but participant (C) would be considered as participant due to the interpersonal distance. This indicates that there is a need to incorporate the relative orientation of the participants. Authors in [52] leveraged a state-of-the-art machine-learning technique for interpersonal distance estimation i.e. [51] and combined it with the relative orientation of the users to improve the accuracy of interaction detection. This technique falls short in the requirement for firmware modification due to the need for transmission signal adjustment but also switching between Hotspot and discovering mode of the WiFi. Also, authors fixed the position of the device on users’ belt which induces the intrusiveness of the system and does not allow pervasive monitoring of the participants in their daily life.

State-of-the-art techniques have some major drawbacks related to the accuracy of the approach in real-world environment but also the requirement for firmware modification or need of external hardware. In order to address these challenges a novel social interaction detection approach for real-world situations will be proposed in the next section.

4.2.3 The DARSIS approach

To overcome the drawbacks of the prior works for social interaction detection DARSIS is proposed. DARSIS is an opportunistic and collaborative sensing system that leverages the sensing and computational capabilities of smartphones to understand social interactions. It does not require any external hardware or firmware modifications while does not rely on the user in order to perform the detection. The inference of social
interactions is based on the estimated interpersonal distance through the technique developed in Chapter 3 and the relative orientation of the users. For the relative orientation of the users, a state-of-the-art technique [5] for facing direction estimation that is independent of the wearing position of the device, is advanced through a proposed clustering technique to eliminate outlier and based on the Gaussian error distribution provide a more reliable outcome.

A collaborative sensing technique is proposed to provide an abstraction layer for information exchange, required for the relative orientation estimation and to speed up the Bluetooth RSSI sensing process in real-time. The sensing and inference processes are performed on the device and data are not transmitted to other third-parties, which preserves users’ privacy and do not rely on any other network. The DARSIS system does not rely on any other network, additional hardware or firmware modifications, thus allows the large-scale deployment through online application stores without mobility restrictions. Furthermore, an analytical model is proposed to understand social interactions based on interpersonal distance and relative orientations. An error model is derived from the analytical approach to compute the expected error of the approach. Evaluations are performed in different indoor environments to understand the accuracy of the approach, the effect of incorporating the relative orientation, the error of the DARSIS system and the analytical model, and finally the accuracy of the DARSIS system against a state-of-the-art RFID-approach.

4.3 DARSIS system

This section presents the design and implementation of the DARSIS system. The most important components of the DARSIS system are depicted in Figure 4.2. Initially, a brief description of the novel interpersonal distance estimation technique proposed in Chapter 3 is provided. Then, the proposed relative orientation estimation and the collaborative sensing component are described, which all together constitute the DARSIS system.
4.3. DARSIS system

Figure 4.2: This figure shows the major components of the DARSIS application. Each device performs initially Detection of nearby devices and facing direction of the user. The Collaborative sensing component exchanges users’ facing directions and mutual Bluetooth RSSI measurements. The Inference component estimates the interpersonal distance and relative orientation of the users. Finally, it infers about on-going social interaction.

4.3.1 Interpersonal distance

A brief description of the novel interpersonal distance estimation technique will be provided, which was proposed, analysed and evaluated in Chapter 3. The interpersonal distance estimation models focus on inferring in which interpersonal zone users are and if they are in proximity or not. To improve the accuracy for interpersonal distance estimation machine-learning models were developed that are based on Bluetooth RSSI. The developed models receive a window of only 6-RSSI samples in contrast to the state-of-the-art techniques that require from 10-20 samples. This reduces the time required to sense the appropriate number of samples needed for the inference process and also provides an up-to-date estimation, as while the number of RSSI sample increases it is possible that these samples are no more valid.

The machine-learning models were trained based on a 48000 sample dataset, collected from indoor environment at different orientations and distances. Water-filled bottles were used to simulate the human body absorption. An extensive feature set was extracted that was minimised through feature reduction techniques and was also wrapped
on MultiBoostAB algorithm [203]. For the interaction zone detection, in order to further improve the accuracy, a hierarchical machine-learning model was developed that infers first about the confidence of estimation regarding each interaction zone and then it combined these confidences to provide an fuzzy membership in the overall inference. The models are lightweight enough to operate in real-time on a off-the-shelf mobile phone.

4.3.2 Relative orientation

As proved by authors in [211] and also depicted in Figure 4.1, relative orientation is a substantive component in the process of understanding social interactions in a real-world environment. Figure 4.1 depicts a social situation in which even though participants (A), (B) and (C) are in a proximity, only participants (A) and (B) are socially interacting. The participant (C) does not have the appropriate relative orientation in order to interact with participants (A) and (B). The relative orientation estimation implicitly is extracted from the facing direction of each participant and is defined as the angle required for a user to turn in order for both users are facing each other directly [212]. An implicit assumption is made, that the participants’ direction is the same as their facing direction [211]. One of the drawbacks of a prior work [52] was the requirement for fixed on body position, thus there is a need to remove the dependency of the device on-body position to limit the intrusiveness of the system and make it suitable for real-world situations.

To tackle the aforementioned drawback, a state-of-the-art technique for direction estimation was advanced through developing an outlier removal algorithm that considers the error distribution of the approach [5]. The estimation algorithm does not depend on the wearing position of the device. The algorithm is based on two phases. In the first phase considered as calibration phase, the algorithm computes the relative orientation of the device with respect to the earth’s coordinates. In the second phase, considered as direction estimation, the algorithm leveraging on the walking locomotion of the user, estimates the relative orientation of the device with respect to the user. Then, a given the calibration and the direction estimation phase, the device orientation
is transformed to the user’s orientation with respect to the earth’s coordinates. This allows the orientation tracking of the user’s facing direction. To smooth the results of direction estimations, a low pass filter with smoothing factor $a = 0.5$ was introduced, which increased the robustness of the facing direction estimations.

Given the Gaussian error distribution of uDirect [5], a novel outlier removal technique was introduced to discard erroneous facing direction estimations. The algorithm is a categorisation technique that clusters the direction estimations and selects the most popular cluster (See Algorithm 1). Through a voting approach the most popular cluster is selected based on the number of estimation contained in each cluster. The values in the most popular cluster are averaged based on the Equation (4.1). The computed average value of the most popular cluster is considered as the direction estimation. The average of $n$ facing direction estimations is calculated through Equation 4.1.

$$\text{Mean}(\theta) = \arctan \frac{\sum_{i=0}^{n} \sin \theta_i}{\sum_{i=0}^{n} \cos \theta_i} \quad (4.1)$$

The decision about averaging the most popular cluster was taken due to the Gaussian error distribution that characterises uDirect [5]. The above process is performed iteratively until the algorithm converges to a satisfactory facing direction. The satisfactory clause is defined by two sequential facing direction estimations that do not differ more than $10^\circ$ i.e. the distance between the two estimations. When then criteria are satisfied, the two final facing directions are averaged and produce the final facing direction. The $10^\circ$ distance between two following sequential facing directions was chosen to reduce the error of the magnetometer sensor.

After the above process, the device orientation with respect to the users coordinates has been estimated. Given the initial calibration phase and the above process, the users’ facing direction is being tracked. When a user in proximity is detected, then the current facing directions are exchange among the devices. This allows the computation of the relative orientation among the users. At the front part of the users’ torsos an imaginary cone of $90^\circ$ is considered as the appropriate relative orientation in order to perform a social interaction. For two users to participate in a social interaction these two imaginary cones need to overlap (See Figure 4.1) and also that the two users are
in proximity to interact.

**ALGORITHM 1:** Outlier Removal Algorithm

**Data:** List of directions;  
**Result:** The final direction;  
assign each direction to a cluster;  
**while** not at end of direction list **do**  
read current direction;  
**while** not at end of cluster list **do**  
if distance(direction, cluster) \( \leq \) THRSLD then  
add direction to cluster;  
end  
end  
max cluster = SelectMostPopularCluster(clusters);  
direction = Average(max cluster);

### 4.3.3 Collaborative sensing

To compute the relative orientation there is a need to acquire the facing directions from the users in vicinity. As DARSIS is an opportunistic and collaborative system that does not rely on any additional hardware, third party components or other networks, there is a need to develop a layer to enable ad-hoc communication among the devices. The devices that are encountered in vicinity will exchange their users’ facing directions and the mutual Bluetooth RSSI to speed up the inference process. This will decrease the time required to fill the 6-sample window of Bluetooth RSSIs and will further reduce the time to perform the inference of social interaction detections. One of the targets of this system is to preserve users’ privacy. This is ensured through performing the social interaction detection on the device without transmitting the data to any third party components.

The collaborative sensing component targets in overcoming the aforementioned challenges identified in prior works. This will enable the devices to opportunistically exchange the facing directions in order to calculate the relative orientations of the users and the mutual Bluetooth RSSI samples in order to speed up the interpersonal distance detection process. DARSIS is an opportunistic system and automatically initiates ad-hoc communication channels when the devices of the system are in vicinity. This allows the devices to speed up the sensing and inference process. Following, the appropriate
steps are described, required to exchange information among the device through the collaborative sensing component:

1. **Discovery.** In this step the devices strive to discover other devices of the system in vicinity through Bluetooth discovery. Each discovery circle has a duration around 12 seconds.

2. **Pick Over.** The devices in vicinity identify each other to understand if they belong to the social interaction detection system. The ones that do not belong are discarded.

3. **Role Assignment.** The devices that belong to the network negotiate in pairs to assign a master/slave role to the dyadic relation. As master is assigned the device with the highest ID number in the dyadic comparison.

4. **Exchange Information.** Having established the role in their master/slave relation, the master device initiates the connection with the slave and then a communication channel has been established, with both devices ready to transmit the appropriate information.

### 4.4 System model

This section proposes a generic analytical model for the detection of social interactions having as input the interpersonal distance and the relative orientation of the users. The error for the interpersonal distance estimation technique and the relative orientation computation is quantified. Following to the analytical model for social interaction detection, an error model is derived and the expected error is computed for a certain context.

#### 4.4.1 Assumptions

This subsection presents the assumptions that were taken in the process of developing the analytical and error models for social interaction detection based on the interpersonal distance and the relative orientation between two users.
Chapter 4. Modelling Social Interaction Detection with Smartphones in Real-world

1. Social interaction is a dependent variable while interpersonal distance and relative orientation are independent [212].

2. Peoples’ torsos orientation is considered as the facing direction in order to compute their relative orientation [211].

3. The relative orientation between two users is the angle one of the users is required to turn to achieve direct contact i.e. face-to-face (0°).

4. The relative orientation of the torso of two people interacting with respect to earth’s coordinates may vary from 0° (face-to-face) to 180° (side-by-side) [77] [213].

5. Accuracy for interpersonal distance and relative orientation is the distance of the estimated from the actual value [214] [215].

4.4.2 Prediction models

This subsection introduces the analytical models for calculating the probability of estimating a social interaction based on the interpersonal distance and the relative orientation of the users. The analytical models that are combined for the overall social interaction detection model are derived from the estimation techniques for interpersonal distance and relative orientation, and from the field of psychology.

4.4.2.1 Interpersonal distance model

The probability of the users being in distance x when the estimated interpersonal distance is \( \hat{x} \) is defined as \( P(x|\hat{x}) \). In this case, as the distance estimation technique used is the one introduced in Chapter 3 the \( P(x|\hat{x}) \) (See Equation 4.2) is calculated from the confusion matrices of the method; meaning this is the probability of the interpersonal distance estimation technique to estimate correctly the distance among the users.

The probability of two people interacting at a distance x provided by psychology is defined as \( P(I|x) \). This probability is modelled as a Gaussian distribution [212] [213]
4.4. System model

\begin{equation}
\begin{aligned}
P(x|\hat{x}) &= \begin{cases} 
P_{\text{Personal}}, & \text{is the probability of correctly estimating personal zone.} \\
P_{\text{Social}}, & \text{is the probability of correctly estimating social zone.} \\
P_{\text{Public}}, & \text{is the probability of correctly estimating public zone.} 
\end{cases}
\end{aligned}
\end{equation}

and its PDF is shown in Equation (4.3),

\begin{equation}
D(I|x) \sim N(\mu_D, \sigma_D^2)
\end{equation}

where $\mu_D = 0.7473$ is the mean interpersonal distance in which two people socially interact and $\sigma_D = 0.7453$ is the standard deviation [212] 213. The probability of two people taking part in a social interaction is defined in Equation (4.4), given the interpersonal distance between the two users.

\begin{equation}
P_{\text{Dist}}(I|\hat{x}) = \sum_{i=1}^{N} P(x_i|\hat{x}) \cdot P(I|x_i)
\end{equation}

4.4.2.2 Relative orientation model

The probability of two people actually having a relative orientation $\theta$ when the estimated relative orientation is $\hat{\theta}$, is defined as $P(\theta|\hat{\theta})$. This probability is computed from the relative orientation computation methods leveraged in the approach. In this case, the relative orientation computation method consists of two estimations of uDirect [5], from which the $P(\theta|\hat{\theta})$ is calculated (See Equation 4.5); meaning the probability of both uDirect estimations are correct.

\begin{equation}
P(\theta|\hat{\theta}) = P_{\text{uDirect}} \ast P_{\text{uDirect}}
\end{equation}

where $P_{\text{uDirect}}$ is the probability of uDirect to correctly estimate the facing direction of one person, derived from the error model of uDirect.

More direct facing direction tend to facilitate social interactions [212]. The probability of two people socially interacting at a relative orientation $\theta$ is provided by psychology.
and is defined as $P(I|\theta)$. This probability is modelled as a Gaussian distribution \cite{213} and its PDF is given in Equation (4.6),

$$ \Phi(I|\theta) \sim N(\mu_\Phi, \sigma_\Phi^2) $$ (4.6)

where $\mu_\Phi = 0$ is the mean relative orientation in which two people socially interact and $\sigma_\Phi = 7.3314$ is the standard deviation \cite{213}. Equation (4.7) computes the probability of two people performing a social interaction based on their relative orientation. The limits of the integral are bounded between $[-\pi, +\pi]$ based on the assumption that the relative orientation is the angle between their facing directions.

$$ P_{\text{Orient}}(I|\hat{\theta}) = \int_{-\pi}^{+\pi} P(\theta|\hat{\theta}) \cdot P(I|\theta) d\theta $$ (4.7)

### 4.4.2.3 Social interaction model

Mehrabian considered interpersonal distance and relative orientation of the users as two independent variables. The social interaction model computes the probability of two people taking part in a social interaction based on a particular interpersonal distance and relative orientation, defined in Equation 4.8,

$$ P_{\text{Inter}}(I|\hat{x}, \hat{\theta}) = P_{\text{Dist}}(I|\hat{x}) \cdot P_{\text{Orient}}(I|\hat{\theta}) $$ (4.8)

where $P_{\text{Dist}}(I|\hat{x})$ is the probability of performing a social interaction at an estimated interpersonal distance $\hat{x}$ and $P_{\text{Orient}}(I|\hat{\theta})$ is the probability of a social interaction taking place given the estimated relative orientation of two users $\hat{\theta}$.

### 4.4.3 Error Models

Following the developments of the analytical models for detecting real-world social interactions given the interpersonal distance and the relative orientation of the users,
this section derives the corresponding error models. These error models describe the error introduced by each product of the analytical model and computes the overall social interaction detection error for real-world situations given the interpersonal distance and the relative orientation among the users.

### 4.4.3.1 Interpersonal distance error model

For the interpersonal distance estimation the error model is derived from the evaluation of the developed technique. Evaluation techniques such as k-fold cross validation provide the misclassification error. Equation (??) computes the probability of interaction based on an estimated interpersonal distance $\hat{x}$. DARSIS classifies among three distinct targets, thus the model is converted from continuous to discrete in Equation (4.4).

where $x_i$ with $i = 1, ..., N$ are the different interactions zones the users may be standing on and $\hat{x}$ is the estimated interaction zone. The confusion matrices and the distance of the erroneous classification from the correct value quantify the error introduced when classifying the three different interaction zones. Section 3.5 shows the confusion matrices for the interpersonal distance model. The estimation error for classifying the interpersonal distance is derived from these confusion matrices and is described in Equation (4.9).

$$\Delta x = E[x - \hat{x}] = \sum_{j=1}^{M} \sum_{i=1}^{N} P(x_i|\hat{x}) \cdot (x_i - \hat{x}) \cdot P_j$$  \hspace{1cm} (4.9)$$

where $\hat{x}$ is the estimated interpersonal distance between the two users and $x$ is the actual interpersonal distance of the two users. The probability of the users being in interaction zone $x_i$ when the estimated interpersonal distance of two users is $\hat{x}$ is defined as $P(x_i|\hat{x})$. The distance error between the estimated and the actual value is given by $(x_i - \hat{x})$. The probability of the users being in the geographical plane (front or back) where they are facing each other, or is at the back of the other (See Figure 4.3) is defined as $P_j$ with $j = 1, ..., M$. The DARSIS system assumes that users are either
Figure 4.3: This figure shows the detection of social interaction through interpersonal distance and relative orientation estimation. Given only the interpersonal distance and relative orientation, the detected user may be in positions A, B and C. In psychology \[213\], people may interact $\pm 90^\circ$ from their facing direction. Thus, the xy plane is divided into two sub-planes, the front (white) plane including A,B positions and the back (grey) plane including C position.

in the front or the back plane, thus the $M = 2$. For the error computation a user is considered to be in the front and back plane with equal probability.

The probability of two people being at distance $x$ from each other in a particular context is defined as $P(x|C)$. In psychology this probability in a normal office environment is modelled as a Gaussian distribution and its PDF is described in Equation (4.10).

$$L(x|C) \sim N(\mu_L, \sigma_L^2) \quad (4.10)$$

where $\mu_L$ is the mean interpersonal distance in which two people are when they are placed in a normal office environment and $\sigma_L$ is the standard deviation. Following, the expected error for the interpersonal distance estimation is provided by Equation (4.11).

$$E[\Delta x] = \sum_{i=1}^{N} \Delta x \cdot P(x_i|C) \quad (4.11)$$

where $\Delta x$ is the error in estimating the interpersonal distance between two users and $P(x|C)$ is the probability of two people having an interpersonal distance $x_i$ with $i = 1, ..., N$ i.e. being in each of the interaction zones at a particular context, provided by psychology.
4.4.3.2 Relative orientation error model

The evaluation of the direction estimation method provides the error of calculating the relative orientation of the users. The error calculation for the relative orientation includes also the processing error $Q_\epsilon$. Authors of the direction estimation technique [5] quantified the error of the approach and proved that it follows the Gaussian distribution. The amount of error is double as the approach due to the fact that two users are involved in the inference process and is described in Equation (4.12).

$$\Delta \theta = E[\theta - \hat{\theta}] = \Delta \theta_{uDirect} \cdot 2 + Q_\epsilon \quad (4.12)$$

where $\theta$ is the actual relative orientation of the two users, $\hat{\theta}$ is the estimated relative orientation of the two users, $\Delta \theta_{uDirect}$ is the error distribution of $uDirect$ presented in [5] and $Q_\epsilon$ is the processing error for computing the relative orientation. However, $Q_\epsilon$ error factor is discarded due to its very small effect in comparison to the facing direction estimation error.

To compute the expected error of the relative orientation model, the occurrence of how often $\Delta \theta$ error is taking place. The probability of two people having a relative orientation $\theta$ from each other in a particular context is defined as $P(\theta|C)$. This probability in a normal office environment is modelled as a Gaussian distribution in Equation (4.13) and is also derived from psychology [213],

$$R(\theta|C) \sim N(\mu_R, \sigma_R^2) \quad (4.13)$$

where $\mu_R$ is the mean relative orientation two people have when they are placed in a particular context and $\sigma_R$ is the standard deviation. Then, the expected error for the relative orientation of two users is described in Equation (4.14),

$$E[\Delta \theta] = \Delta \theta \int_{-\pi}^{\pi} P(\theta|C) d\theta \quad (4.14)$$

where $\Delta \theta$ is the error in computing the relative orientation between two users and
$P(\theta|C)$ is the distribution of users' relative orientation in a particular context, provided by psychology.

### 4.4.3.3 Social interaction error model

The social interaction model error is presented in the following equation, to compute the error of the approach:

$$
\Delta P_{\text{Inter}}(\hat{x}, \hat{\theta}) = \frac{\partial P_{\text{Inter}}(\hat{x}, \hat{\theta})}{\partial \hat{x}} \Delta x + \frac{\partial P_{\text{Inter}}(\hat{x}, \hat{\theta})}{\partial \hat{\theta}} \Delta \theta
$$

(4.15)

where $P_{\text{Inter}}(x, \theta)$ is the probability of a social interaction taking place between two users given the interpersonal distance $x$ and their relative orientation $\theta$ (See Section 4.4.2.3). $\Delta x$ is the error in estimating the interpersonal distance provided by Equation (4.9) and $\Delta \theta$ is the error in estimating the relative orientation of the users given by Equation (4.12).

### 4.5 Performance evaluation

The evaluation of the DARSIS and the analytical model are presented in this section. Three experiments are conducted in real-world environments to understand the effect of incorporating the relative orientation in the social interaction detection process. The DARSIS error and the analytical model error are evaluated and compared to benchmark the two approaches against each other in different real-world environments. The DARSIS system is benchmarked in a real-world environment against a state-of-the-art technique based on active RFID tags [208]. Finally, the DARSIS system is evaluated in terms of power consumption and data transmission complexity.

#### 4.5.1 Relative orientation evaluation

The DARSIS system is evaluated in several environments to understand the effect of incorporating the relative orientation effect in the social interaction detection process.
Table 4.1: Confusion Matrices for the evaluation of the effect of Relative Orientation estimation for DARSIS system in percentages (%)

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
<th>Positive</th>
<th>Negative</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>DARSIS</td>
<td>Pos. 83</td>
<td>2.29</td>
<td>Pos. 60</td>
<td>40</td>
<td>Pos. 77</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>Neg. 17</td>
<td>97.71</td>
<td>Neg. 40</td>
<td>8.14</td>
<td>Neg. 23</td>
<td>98.72</td>
</tr>
<tr>
<td>Proximity</td>
<td>Pos. 83</td>
<td>36</td>
<td>Pos. 65</td>
<td>60.57</td>
<td>Pos. 83</td>
<td>39.43</td>
</tr>
<tr>
<td></td>
<td>Neg. 17</td>
<td>64</td>
<td>Neg. 35</td>
<td>35</td>
<td>Neg. 35</td>
<td>56.43</td>
</tr>
</tbody>
</table>

(a) Meeting room - 1 (b) Meeting room - 2 (c) Corridor

The system is evaluated in two situations: a) the overall DARSIS system and b) only the interpersonal distance estimation technique that DARSIS incorporates. The section starts with the evaluation methodology that describes the experimental setup and the performance metrics followed by the results of the evaluation.

4.5.1.1 Evaluation methodology

Section 4.2 highlighted the importance of incorporating the relative orientation is the detection of social interaction. State-of-the-art techniques do not consider users' relative orientation, apart from [52] that obliges the users to restrict the wearing position of the device at a specific point, inducing a particular level of intrusiveness. An experimental analysis is performed to understand the effect of incorporating the relative orientation calculation in the social interaction detection process.

Several experiments were conducted in real-world indoor environments to evaluate the incorporation of relative orientation calculation in the social interaction detection process. Five participants in the age range of 25-30 and height range 1.65m-1.96m were involved while each participant received an off-the-shelf mobile (HTC One S) and was placed in one of the participant’s trousers pocket. The environments in which the experiments took place were meeting rooms and corridor. The participants were split in two groups of people, into two and three participant-interactions. Regarding the experimental process, the participants placed the smartphones in their trousers pockets. The smartphones had pre-installed the DARSIS application and it was operating as a background service on the phone. After placing the devices in their pockets, participants started walking for some meters in order to calibrate the relative orientation of
Chapter 4. Modelling Social Interaction Detection with Smartphones in Real-world

the device with respect to the user. Then, after the direction estimation has converged, the participants entered the environment in which the experiment would take place. Once the participants entered the environment in which the experiment would take place, they started to interact in two groups as mentioned before. In order to establish ground truth, a human observer was employed that logged down the on-going social interactions. The groups in which the participants were interacting were changed after 10 minutes.

It should be noted that all participants were placed in close proximity, to force the social interaction detection to engage the relative orientation calculation mechanism. This will allow the understanding and quantification of the effect of the relative orientation calculation in the social interaction detection process. If the participants were not in proximity, then the interpersonal distance estimation technique would directly discard the detected participant from the social interaction detection process and classify them as non-interacting. At the end of all experiments, the produced data were analysed in order to infer the existence of social interaction first only based on the interpersonal distance and then including also the relative orientation.

4.5.1.2 Evaluation results

The initial performance metric used in the evaluation is the confusion matrices, which were calculated for both approaches, the overall DARSIS system and only the interpersonal distance estimation for interaction detection. The confusion matrices are presented in Table 4.1. The aim of incorporating the relative orientation calculation in inference process is to reduce the rate of false positives, which are the rate of social interactions that were detected but did not actually take place. Following the reduction of the rate of false positives would increase the overall accuracy of the system through the increase of the true negatives rate.

The initial evaluation showed that incorporating relative orientation estimation produced high accuracy for the social interaction detection. The system managed 93.3% classification accuracy in the first scenario in the meeting room in contrast to the standalone interpersonal distance technique that managed only 69.7% of accuracy for
4.5. Performance evaluation

Table 4.1 shows the improvement in the false positives rate because of the incorporation of the relative orientation. It should be noted that the interpersonal distance estimation technique is the same in both approaches. In the second scenario the system achieved 82.3% accuracy as opposed to the standalone interpersonal distance estimation technique that managed only 47.1% accuracy. This shows that an improvement of 33% in the overall accuracy was added by incorporating the users’ relative orientation estimation. In the corridor scenario, the DARSIS system was able to detect correctly 85.2% of the social interactions and the standalone interpersonal distance estimation technique achieved only 64.2% accuracy. This provides another indication about the significance of incorporating the relative orientation calculation in the social interaction detection process. The incorporation of the relative orientation calculation in the social interaction detection process showed an improvement of at least 20% in the overall accuracy in comparison to the standalone interpersonal distance estimation.

The classification rates of each approach are presented in Table 4.1 to provide a better understanding of the effect of the relative orientation calculation in the social interaction detection process. As shown, the true positives rates are the same in both approaches and did not receive any improvement. This is because in the social interaction detection process, the interpersonal distance estimation technique has the dominant role. It should be noted that the incorporation of the relative orientation calculation mechanism induces a small amount of error. This error is due to processing and computation error of the algorithm but also due to magnetic disturbance in the environment. Tables 4.1b and 4.1c quantify this error through the reduction of the true positives rates. The largest error around 5% is observed in Table 4.1c due to the relative orientation algorithm. However, while observing also the 42% decrease in the false positive rate followed by similar increase in the true negative rate, the 5% error in the true positives is a trade-off that is worth taken. Due to the dominant factor of interpersonal distance estimation, the false negative rate can only be reduced by the relative orientation detection. The false negative rate was not affected by the relative orientation calculation according to Table 4.1a. Tables 4.1b and 4.1c show a minor increase of at most 6% in the false negative rate due the relative orientation estimation, including the facing
direction algorithm and the magnetic interference of the environment.

The aim of this evaluation was to show the improvement in the accuracy of social interaction detection through incorporating the relative orientation computation but also the robustness of the system in various environments. Table 4.1 shows that the aim of reducing the rate of false positives was achieved from 33% to 42% improvement in the accuracy, while also led to the increase of true negative through 33% and a 97.7% accuracy for the true negatives rate due to the incorporation of the relative orientation calculation mechanism.

### 4.5.2 Error model evaluation

This section presents and benchmarks the error of DARSIS system against the theoretical error derived from the analytical model that was described in Section 4.4. The purpose of this evaluation is to compare the error of the two approaches, show the validity of the error model and provide a better understanding of the error distribution of the DARSIS system.

#### 4.5.2.1 Methodology

There was a need to perform a comparison between the error derived from the theoretical model with the DARSIS error, to evaluate the error model and prove the applicability of in real-world environment based on the error. Understanding the error of the DARSIS system and the error model will evaluate both approaches as coherent systems that include the interpersonal distance and relative orientation estimations.

The evaluation of the two approaches was performed based on the datasets acquired in the previous experiments. The DARSIS and the theoretical error were evaluated based on the datasets from the three different indoor environments, involving five participants in real-world social interactions (See Section 4.5.1.1). The outcome of this evaluation will provide the error for both approaches in real-world situations. To measure the error the evaluation in both approaches was performed by considering both as coherent system that include both the interpersonal distance and relative orientation estimation.
4.5.2.2 Performance metrics

The key performance metric of this evaluation is the error of each approach. The percentage of faulty estimation of each approach with respect to the existence/absence of a social interaction given the ground truth is defined as error rate. The error is calculated for each user in each experiment.

To evaluate the error between the DARSIS system and the theoretical model, various performance metrics were utilised. Table 4.2 presents the mean error of each model and the standard deviation (SD) of the error to measure the variation of the error. The standard error of the mean (SEM) statistic provides a metric for the standard deviation of the distribution of the mean. Figure 4.4 depicts box-and-whisker plot with the error of the DARSIS system and the error model. This box plot constitutes a coherent representation of the error distribution of the two approaches through five statistics including minimum, 1st quartile, median, 3rd quartile and maximum. In the middle of the box plot, the red horizontal line describes the median value. The blue horizontal edges of the box describe the 1st quartile and the 3rd quartile. The horizontal black line that is connected through the dotted line with the box describes the minimum and the maximum error values of the error distribution. The red crosses at the top of the box plot are represent the outliers of the error distribution and are 1.5 times bigger than the 3rd quartile.

4.5.2.3 Evaluation results

The results of the evaluation of the DARSIS system and the error model based on the aforementioned performance metrics are shown in Table 4.2 and Figure 4.4. The target of this evaluation is show that the error model has similar performance with the DARSIS system and to analyse the error distribution of both approaches.

Table 4.2 shows a comparison between the DARSIS and the error model based on the mean, STD and SEM error of each approach. As shown, both models perform similar by having similar error statistics including mean 13-16% and SD of 18-19.5%. A small decrease in the mean error of DARSIS is observed, around 1.84% in comparison to
Table 4.2: Comparison between DARSIS and the error model regarding the percentage (%) of error introduced in real-world environments.

<table>
<thead>
<tr>
<th></th>
<th>Mean Error</th>
<th>StdDev Error</th>
<th>Standard Error of Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>DARSIS</td>
<td>13.83</td>
<td>18.92</td>
<td>3.45</td>
</tr>
<tr>
<td>Error Model</td>
<td>15.67</td>
<td>19.43</td>
<td>3.55</td>
</tr>
</tbody>
</table>

Figure 4.4: Comparison of error distribution of DARSIS and Error model through box plots.

the error model. An even smaller reduction of 0.51% is observed in the SD of the error, which shows a decrease in the error variation of DARSIS with respect to the error model. The smallest difference of 0.1% is observed in the SEM between the two approaches. In all three statistics, the difference of the two approaches does not exceed the error of 2%.

Figure 4.4 shows the diagram that includes the two box plots of the error distribution of DARSIS and the error model. In both cases the error distributions are quite similar.

A difference of 3% is observed by comparing the medians of the two box plots. The minimum with 1\textsuperscript{st} quartile of the error plot are overlapping in both cases. The two plots indicate that the 1\textsuperscript{st} quartile i.e. 25% of the lowest error cases the models do not produce any error. The 3\textsuperscript{rd} quartile i.e 75% of the lowest error cases for both models produces 25% of error. Regarding the maximum percentage of error introduced
in the experiments, DARSIS and the error model produce 42% of error. As shown in the box plots, the error model produced one outlier with 68% of error as opposed to DARSIS which generated two outliers at similar levels. Although the median values have a difference around 3%, both plots present similar error distributions considering the other four statistics provided by the box plots.

Overall, DARSIS and the error model performed similarly regarding the error distributions. The difference in the mean, SD and SEM error did not exceed 2% of error between the two models. The box plots revealed a difference of around 3% between the median error of the two models. The minimum, 1\textsuperscript{st} quartile, 3\textsuperscript{rd} quartile and maximum percentage of error of both error distributions performed very close to each other. In the error statistics and the box plots, a small increase in the error is observed due to errors induced by the models derived from psychology. The outliers are sources of decalibration of the facing direction mechanism in one of the devices, which affects the relative orientation computation. Finally, both models managed to achieve less than 16% of mean error with SD less than 19.5% having consistent mean error with SD of 3.5%.

4.5.3 Real-world social interactions

This section presents the evaluation the DARSIS system against a state-of-the-art approach that is based on active RFID \[208\] in a real-world environment. The objective of this real-world experiment is to prove the viability and robustness of the DARSIS approach to detect social interactions in real-time. In addition, there is a need to understand if the DARSIS approach is capable of achieving similar accuracy to an RFID-based method. Both systems operate in real-time and perform the social interaction detection online.

Regarding the participants of these experiments, eight PhD students were recruited between the age of 25-30 years. As shown in Figure 4.5, an indoor office environment typically furnished was selected as the experimental environment. The evaluation consists of three sets of experiments. During each experiment, five randomly selected participants were placed in the office room to interact with each other. The partici-
pants were interacting in two different groups. Both the DARSIS and the RFID-based approaches were simultaneously evaluated during the experiments. To achieve this, each participant was provided with an HTC One S smartphone running the DARSIS application and an active RFID tag [208]. Because the number of users was not too large, the ground truth was established from a human observer.

At the beginning of the experiment, users placed the provided smartphone in a trouser pocket with an arbitrary device orientation. To estimate the user’s facing direction, uDirect [5] requires an initial calibration phase. Hence, the users were advised to walk around in arbitrary directions for a few seconds until the algorithm converged. After the completion of this calibration phase, the users entered the office environment for the initiation of the experiment. Users were given RFID-tags that were placed on their chest. To log the detected interactions of the RFID-approach, an RFID-reader and a laptop were deployed in the office room.

During the empirical evaluations 756 social interaction inferences were logged from DARSIS and 40000 from the RFID-approach. The RFID-approach does not provide any information about the people in vicinity that do not interact and logs only the detected social interaction. The DARSIS approach detects all users in vicinity and infers individually for each one whether there is an on-going social interaction or not.
4.5. Performance evaluation

The evaluation of the DARSIS system indicated that it is able to detect correctly 81.4% of the on-going social interactions with only 6 Bluetooth RSSI samples. The RFID-approach detected correctly 67.9% of the social interactions. As DARSIS introduces and incorporates collaborative sensing it is able to receive Bluetooth RSSI samples faster than other mobile-phone-based systems. For that reason, DARSIS is able to also accurately identify short-term social interactions that were missed by state-of-the-art solutions.

By observing the number of inferences collected, the RFID-approach provides more frequent detections in contrast to DARSIS. This improves the granularity of the detected social interactions. On the one hand, an important outcome of the experiments is the effect of incorporating the facing direction in the inference on the reduction of false positive errors. In particular, the ratio of false positives and true positives in DARSIS is reduced by more than 9% in comparison to the RFID-approach. On the other hand, due to facing direction estimation errors, magnetic disturbance and hard thresholds in relative orientation computation, there is an increase in the number of false negative errors.

Overall, given the above observations DARSIS constitutes a more conservative approach than the RFID-based system. DARSIS performs the inference by considering more concrete evidence such as the relative orientation of the users and their interpersonal distance. This results in a decrease of the false positive error but also in a small increase in the false negative error. The RFID-approach constitutes a more liberal inference method for detecting social interactions leading to an increased number of false positives. It should be noted that both systems performed instant inference while providing snapshots of detected social interactions. This means that in social situations where longer (in terms of duration) social interactions take place, multiple inference snapshots will be provided by both systems and could bias third-party systems that leverage the information received from the two enablers. In order to tackle this issue, there is a need for a post-processing step to aggregate the inference snapshots into social interactions with longer duration.
Chapter 4. Modelling Social Interaction Detection with Smartphones in Real-world

Figure 4.6: DARSIS power consumption (W) of a device when different number of devices are in vicinity.

4.5.4 Power consumption

To understand the applicability of DARSIS in real-world environments the power consumption of the application was evaluated. During the experiments the current and the voltage were measured in order to compute the power consumption in Watts (W). The power consumption of DARSIS as a coherent system was evaluated when different numbers of users are in the vicinity. It was benchmarked with different numbers of devices in the vicinity to understand the effect of the induced communication burden. The results for 0-4 users in vicinity derived during the previous experiments are provided. Each device was running only the DARSIS application, while having the device’s screen switched off.

Figure 4.6 presents box plots of the power consumption of the DARSIS system deployed on different number of devices (See Section 4.5.2.2). The power consumption shown in the plot refers to an individual DARSIS device. The results show that there is a slight increase in the power consumption as the number of devices increases. The minimum value and 25% of the lowest power consumption are stable in both statistics, having the corresponding values of 0.44W and 0.52W. From zero to two devices in the vicinity, the median power consumption is around 0.54W and the maximum value between 0.7-0.75W. When two devices are in the vicinity 75% of the lowest power consumption increases relatively to the two others close to 0.61W. When three devices are in the
vicinity, a notable increase in the 3rd and maximum value of the power consumption is observed i.e. 0.64W and 0.82W. Then, four devices in the vicinity increases the median power consumption slightly over 0.6W, the 3rd shows a value close to 0.7W and the maximum around 0.91W. In each case there are also statistical outliers indicating an instantaneous increase in the power consumption.

Overall, there is a slight increase in the power consumption as the number of users in the vicinity increases. This occurs due to the collaborative sensing. The device needs to exchange the facing direction and mutual Bluetooth RSSI measurements with each device in the vicinity. Also, there is a need to compute the interpersonal distance with each user, which also adds some computational burden. In each case the maximum power consumption is below 1W for at least 0.09W and the median below 0.62W, while having at most four users in the vicinity. Adaptive sensing and inference techniques may further improve the power consumption of the application e.g. [50].

4.5.5 Data communication burden

This section provides an evaluation of the communication burden induced from the messages that the devices exchange. DARSIS is an opportunistic system that performs collaborative sensing to compute the relative orientation of the users and speed up the sensing process. The devices exchange the mutual Bluetooth RSSI measurements and the facing directions of their users.

Each message the devices exchange induces a certain communication burden. In particular, the size of each message is 186 bytes. The number of messages sent by a device depends on the number of devices in the vicinity. When there are $n$ devices in vicinity, a device will transmit at most $n \times (message\_size + header)$ bytes of data during one inference period. Thus, as both the header and the message size are constant, the message complexity of our system is $O(n \log n)$ for each interval of sensing and inference.
Chapter 4. Modelling Social Interaction Detection with Smartphones in Real-world

4.6 Conclusion

This chapter presented the design, analysis and evaluation of the DARSIS system that detects social interactions through smartphones in an opportunistic manner. Towards the realisation of the DARSIS system several contributions were achieved. First, an accurate and reliable interpersonal distance estimation technique for interaction zone and proximity detection that was introduced in Chapter 3 was incorporated. Secondly, a relative orientation computation method was developed and incorporated in the DARSIS system that addresses the absence of facing direction in previous works and the dependence of on-body position. Thirdly, a collaborative sensing component was introduced to allow information exchange such as Bluetooth RSSI samples and users’ facing directions in order to speed up the sensing and inference process. Fourthly, the development of a generic analytical model was introduced to estimate the existence of social interaction based on interpersonal distance and relative orientation with a certain probability. From this analytical model and the probability of occurrence, an error model was derived for the DARSIS system and further its expected error.

DARSIS system and the proposed analytical model were evaluated in three different indoor environments. The number of participants in the study was quite small (5 participants) and the standard deviation of the error is large relative to the separation of the mean error. In order to empower the evaluation results, as future work it is proposed to perform evaluation of the error of the DARSIS system and the proposed analytical model with a larger number of participants to minimise the standard deviation of the error relative to the mean error.

DARSIS is a privacy-preserving system as it performs the inference online and does not transmit the data to third-party components. It does not depend on any external hardware, therefore there are no mobility restriction for the operation of the system. Understanding social interactions and quantifying the related context provide valuable information for inferring different aspect of social behaviour. The DARSIS system is an initial step towards ubiquitous and pervasive observation of real-world social networks with applications in healthcare, the Internet of Things, epidemiology, marketing and others. One of the aspects of social behaviour that are related to social interactions
is the trust relationships among users. Trust is a fundamental notion based on which society has been build, from the rules and law of a society to the financial transactions among people. Understanding these trust relationships among user would improve the sustainability of future smart cities, create relationships among people without any prior knowledge and advance the trustworthiness of the social environment. An initial attempt towards understanding trust relationships among people is presented in the next chapter, were a real-world social graph is extracted by leveraging the social interactions detected by the DARSIS system. Furthermore, based on the social relations that the real-world social graph provides and combined with the contextual information from the daily social interaction, people trust relationships are inferred.
Chapter 4. Modelling Social Interaction Detection with Smartphones in Real-world
Chapter 5

Quantifying Real-world Trust Relationships through Social Interactions

This chapter introduces a novel approach to quantify trust relationships among people based on detected social interaction through off-the-shelf mobile phones. Previous chapter presented and evaluated an opportunistic and collaborative sensing system to detect real-world social interactions based on smartphones. In this chapter, initially the information provided by the social interaction detection system are leveraged to extracted a novel real-world social graph that considers the social relation among people. Following the extraction of a real-world social graph, the derived social relation among people is combined with contextual information from the social interaction detection in order to quantify the trust relationships among people. A proof-of-concept evaluation was performed, where people were placed in an indoor environment and started to interact in a real-world situation, providing some initial insights regarding the applicability of the proposed approach.
Chapter 5. Quantifying Real-world Trust Relationships through Social Interactions

5.1 Introduction

Trust is an important factor that plays a significant role in the structure of our society today. Fields such as psychology, organisational engineering, marketing and informatics have focused on understanding and measuring trust relationships. Psychological and emotional well-being have been correlated with trust, where a trusted person tends to be happier, more open to new relations and less neurotic in social situations [216].

Researchers focused on understanding and quantifying trust in various contexts. Internet applications [217], on-line social networks [218], on-line service provisioning systems [219], Internet of Things [220] [221] and many others constitute a significant background for understanding trust.

Literature initially strived to measure trust relationships among people through less automated methods including questionnaires and surveys. These methods induce a considerable amount of error due to the involvement of the human factor [16]. Several techniques focused on understanding trust relationships among people in on-line social networks. These techniques consider information retrieved from users’ social accounts. Research has indicated that on-line and real-world social networks may be different, as in on-line social networks users tend to have a large amount of false positive relations [222]. Until now, there is no prior work that was able to quantify trust relationships among people through smartphones based on contextual information derived from real-world social interactions and social relations.

The starting point of the research is whether smartphones are able to provide appropriate contextual information extracted from daily interaction to create a real-world social graph and derive trust relationships among people. To initiate this research an assumption is taken, while people participate in a social interaction the level of trust and trustworthiness among them increases [223].

This chapter presents MobTrust, the first work towards quantifying trust relationships among people based on real-world interactions through smartphones. To support this work, a social interaction detection mechanism based on off-the-shelf smartphones is leveraged [224]. A real-world social graph is developed, leveraging the social interaction detection information. The edges of the real-world social graph are weighted with the
5.2 Background

Important research has been conducted on the development of trust models for the identification of trusted peers in various networks including EigenTrust [225], TrustMe [226], PeerTrust [227] and PowerTrust [228]. However, the research in this work considers the trust quantification in real-world social situations. Prior works for quantifying social trust in on-line social networks and real-world social networks. A more extensive analysis of the prior works for trust computation is provided in [218].

5.2.1 On-line social networks

Literature has focused on developing trust models based on graphs extracted from the on-line social network, the on-line interactions among the users of the social network and the combination of both as a hybrid solution.

5.2.1.1 Graph

These approaches leverage the on-line social network and extract a social graph and in some cases consider also the information flow among the users. Initially, literature
focused on developing a trust model based on the weights of the social graph [229]. As an improvement researchers also added a user feedback mechanism related to peoples’ levels of trust [230] [231]. In [232], authors present a dynamic approach to compute the trust distance among users, which varies in favourable and unfavourable situations. A trust graph was developed in [233] and was further improved in [234] by incorporating a recommendation system based on the similarities of the trust network. Prior works have based their trust models mainly on the structure of the social network and the way users are linked with each other. In addition the knowledge provided by these techniques is related to the information and trust flow among the users of the network. Researchers have also extracted social graphs from images [235]. A major drawback of these techniques is that they do not consider the contextual information from the interactions of the users and also require knowledge of the overall network.

5.2.1.2 Interaction

For trust quantification, researchers also developed models based on contextual information including social behaviour, related to the interactions among the users of the network. Authors in [236] developed a trust model that builds upon information related to the users’ interactions and their actions in the system including commenting and rating. In [237] authors leveraged information from users’ interactions such as frequency and duration of interaction. Also, they incorporated the propagation of trust; when some type of information is transferred from one person to another a certain level of trust is implied at both ends. In [238] authors classified the type of trust based on features such as the popularity and the engagement. The first feature measures how popular a person is and the second feature shows how active the person is in the community. [239] developed a trust computation mechanism using interaction data and fuzzy logic. These trust models are characterised by some advantages and some disadvantages. Considering each individual user may could provide interaction-based information without prior knowledge of the overall social network. A drawback of these approaches is that they consider only contextual information from users’ interactions. Thus, they are not able to derive any knowledge about the overall structure of the network.
5.2. Background

5.2.1.3 Hybrid

A hybrid solution was proposed in [240] that considered both the overall social network structure but also from the interactions among the users. Two trust models were developed. The first model is called explicit trust model where users actively provide their friend-list to others. The second model is called implicit trust model in which an algorithm opportunistically calculates the level of trust given the duration and frequency of the interactions.

These works consider only on-line social networks and do not take into account the social graph derived from peoples’ daily interactions. A real-world social graph constitutes a more realistic representation of the social and trust relationships of people. In addition, existing literature does not take into account the need to differentiate the level of trust based on the context [241]. For example a car mechanic is a trusted person in the context of repairing cars but may not be in the IT-sector.

5.2.2 Real-world social graphs

A limited amount of works strived to extract social graphs from real-world situations. Understanding the social links among people including friendships and social relations, constitute a vital requirement for developing a real-world social graph. Pervasive and self-acting tools are required to ensure the accurate derivation of peoples’ social links. In the beginning, researchers developed obtrusive mobile device to monitor users’ social interactions [207] and extracted a real-world social graph. To reduce the obtrusiveness of the system, researchers focused on mobile phones and stationary devices through. Using the Bluetooth interface and triangulation techniques they inferred about proximity [46]. This technique did not consider the existence of social interactions and required additional hardware to be deployed in the environment. Authors in [208], developed an active-RFID based approach to measure social interactions among people and extracted a real-world social graph. The drawback of this approach is the requirement for deploying additional hardware in the environment. This increases the intrusiveness of the system and forbids large-scale deployment. Cityware [242] is a platform for combining on-line social network data and mobility traces captured through Bluetooth scanning.
However, this approach calculates only the encounters and not the social interactions among the users. People could encounter each other in their daily life in a radius e.g. 35m (Bluetooth average range in outdoor environments), but this does not necessarily mean that the users have a social relation.

In overall, the major shortcomings of the prior techniques are intrusiveness and the lack of the ability to perform a large-scale deployment. None of the previous works considered the social relation to develop a real-world social graph and derive trust relationships. Although some of the works detected social interactions, they did not consider the extracted contextual information to quantify trust relationships.

5.2.3 Proposed approach

In this chapter, an opportunistic system is proposed based on smartphones that extracts a real-world social graph based on users’ social relations and derives their trust relationships through contextual information sensed from real-world situations. The work on social interaction detection through smartphones is leveraged, presented in Chapter 4. A model is proposed to extract real-world social graph based on the detected social interactions and social relations. In particular, based on the detected social relation snapshots and the contexts, an analytical model is proposed for extracting a novel real-world social graph. The graph’s edges are weighted with the estimated social relation and the confidence of the estimation. Given the contextual information from users’ social interactions and the extracted real-world social graph, a hybrid probabilistic model is build to derive trust relationships. There is no prior work that leverages users’ social interaction patterns and a real-world social graph to derive trust relationships among people. The proposed models are based on probabilities and provide results through simple and lightweight computations. This allows the system to sense and infer continuously on a mobile device. The system does not introduce any restrictions spatial restrictions and allows large-scale deployment of the system, as it does not depend on external hardware or firmware modifications.
5.3 Methodology

This section presents the methodology for extracting a real-world social graph based on social relation snapshots and for deriving trust relationships based on a hybrid probabilistic model that considers contextual information from social interactions and from the real-world social graph.

5.3.1 Background

The mechanism for quantifying trust relationships has its roots on the social interaction and social relation detection work presented in Chapter 4 (See Figure 4.1). In particular, an opportunistic sensing and inference system based on smartphones was proposed, independent of external hardware or firmware modifications. Upon this, Hall [193] introduced the term interaction zones that mapped the interpersonal distance with the social relation among users, while interact­ing. Considering these finding in psychology, an interpersonal distance estimation technique was developed to automatically perform this mapping and infer the social relation among people. To understand the direction­ality of people in vicinity, a relative orientation computation technique was developed. Furthermore, a collaborative sensing component was introduced allowing the exchange among the devices of users’ facing directions and the mutual Bluetooth RSSI samples to speed up the sensing process. Following, leveraging the social interaction and social relation detection mechanism two models are developed to extract a real-world social graph and derive the trust relationships among people.

5.3.2 Real-world social graph

Social interactions may produce important knowledge about the social behaviour of people. Researchers has mainly extracted social graphs from on-line social networks. However, as pointed out in [222], a large amount of false positive links among people are identified in on-line social networks. Monitoring peoples’ daily social interactions may provide a more accurate estimate of the users’ actual social graph.
Chapter 5. Quantifying Real-world Trust Relationships through Social Interactions

Figure 5.1: Social interactions provide snapshots of human social relation allowing the extraction of a real-world social graph. Given the contextual information from peoples’ social interactions and their social relation from the real-world social graph, the trust relationships among people are derived.

5.3.2.1 Approach

A dynamic social graph is proposed that is able to adjust based on the real-world situations. As people interact in daily basis, they form certain social relations. The interpersonal distance estimation provides the snapshots of peoples’ social relation, which are leveraged in this model to extract a real-world social graph. The social relation estimation and the confidence of the estimation are provided as weights of the edges of the social graph. For each context, the model computes the confidence of the social relation estimates and selects the most confident estimate in each context. The confidences from the various contexts are merged through probabilistic weights of each context given by psychology. The most confident social relation is selected. A more detailed description of the social relation model is provided in the next subsection.
5.3. Methodology

5.3.2.2 Social relation model

Initially, a set \( S \) is defined that includes all the possible social relations that two people may have with respect to the MobTrust system.

\[
S := \{ r : r \in \{\text{Public, Social, Personal}\}\}
\]  \hspace{1cm} (5.1)

Equation (5.2) provides the confidence of each social relation between two users in a certain context.

\[
P(r) = \frac{Q(r)}{N}, \text{ where } r \in S \text{ and } R(r), N \in \mathbb{N}^+ \]  \hspace{1cm} (5.2)

where \( Q(r) \) is the number of inferences that are related to social relation \( r \) and \( N \) is the total number of social relation inferences. The most confident social relation between two people in a certain context is computed by Equation (5.3).

\[
f(r) = \operatorname{argmax}_{r \in S} \left\{ \frac{Q(r)}{N} \right\} \]  \hspace{1cm} (5.3)

Equation (5.3) calculates the confidence of the social relation between two people in a certain context. In order to include multiple contexts in the model, the Equation (5.4) includes the weighting probability of how important is the estimated social relation \( r \) in a particular context \( c_j \).

\[
R(r) = \operatorname{argmax}_{r \in S} \left\{ \frac{1}{C} \sum_{j=1}^{C} P(r \cap c_j) \cdot w_j \right\} \]  \hspace{1cm} (5.4)

\[
= \operatorname{argmax}_{r \in S} \left\{ \frac{1}{C} \sum_{j=1}^{C} P(c_j) \cdot P(r|c_j) \cdot w_j \right\}
\]

where \( P(c_j) \) is the probability of the users being in the context \( c_j \), \( P(r|c_j) \) is the probability of the two people having a social relation \( r \) given that they are in context
Chapter 5. Quantifying Real-world Trust Relationships through Social Interactions

Figure 5.2: This figure shows the information flow among the different component of the system. The social interaction detection component measures contextual information about users’ interactions and provides snapshots of their social relation. The social relation snapshots are merged through a context-related probabilistic model in order to extract a real-world social graph. Contextual information from users’ interactions and the social relation from the real-world social graph are combined through a probabilistic model to derive their trust relationship.

$c_j$ and $w_j$ is a probabilistic weight of a particular context $c_j$ with respect to a social relation.

5.3.3 Deriving trust relationships

Following the extraction of a novel real-world social graph that weights its edges by the social relation, this subsections presents the proposed model for deriving trust relationships among people.

5.3.3.1 Assumptions

In this subsection the assumptions taken in order to develop the trust relationship model, are presented. The goal of this work is to create an initial and simplified model to derive a measurement of trust based on information extracted in daily life and in particular during social interactions. To achieve this goal, there is a need to make some assumptions that will enable the creation of the simplified trust model. Among these assumptions are parameters used to create the trust model, are the information derived from the social interaction detection system. In future work, these parameters will be expanded by including information from on-line social networks and other behavioural data. An another assumption that was made, is that each of the four parameters is independent variable. This constitutes a logical assumption and allows the creation of a simplified trust model, enabling the integration of the various parameters into one trust score.
5.3. Methodology

- Users’ social relation, relative orientation, frequency of interactions and duration of interactions are considered for trust relationship estimation.

- The above four parameters leveraged to derive trust relationships are independent variables.

- Users have equal opportunity to interact with each other. [243].

- As relative orientation is considered the angle a user has to turn in order to face directly another user. [211].

5.3.3.2 Approach

The hybrid probabilistic model for trust relationship estimation is based on probabilistic models provided by psychology and the real-world social graph. The model is developed based on the assumption described in the previous subsection. The model receives as input contextual knowledge such as users’ social relation, relative orientation, frequency of interactions and duration of interactions. The probability density function of each of the above parameters with respect to the trust relationships among people, is provided by psychology. The trust relationship estimation model combines the probabilistic models provided by psychology and calculates an overall trust score. This probability of trust is related to a particular context, while the trust scores from different context are combined into one overall trust score.

5.3.3.3 Trust model

In the case of two people participating an unfriendly argument, they tend to face each other directly with high probability. This reduces the level of trust between them. The parameter of relative orientation between two people is modelled as a Gaussian distribution [244].

$$\Theta(t|\theta) \sim N(\mu_\Theta, \sigma_\Theta^2)$$ (5.5)
where $\theta$ is the relative orientation between the two users in degrees, $\mu_\Theta$ is the mean relative orientation between two users having a conversation that increases the trust among them and $\sigma_\Theta$ is the standard deviation.

Based on the social relation between two people, there may have different levels of trust \[211\]. People that have a personal social relation are socially interacting in a close distance in contrast to people that have an impersonal relation who tend to interact in a further distance. The parameter of social relation between two people is extracted from the real-world social graph and is modelled as a discrete distribution \[211\].

$$R(t|r) = \sum_{k=0}^{n} P(r_k)$$ \hspace{1cm} (5.6)

where $r \in \mathbb{S}$ is the social relation between two users extracted from the real-world social graph introduced in previous section, $n$ is the discrete number of social relations and $P(r_k)$ is the probability of trust while having a social relation $r_k$.

When two people interact more frequently, then the level of trust between them increases \[243\]. The parameter of frequency of social interactions is modelled by the cumulative density function of a normal distribution \[245\].

$$F(t|f) = \frac{1}{\sigma_F \sqrt{2\pi}} \int_{-\infty}^{f} e^{-\frac{(t-\mu_F)^2}{2\sigma_F^2}} dt$$ \hspace{1cm} (5.7)

where $f$ is the frequency of interaction between two people in a particular context, $\mu_F$ is the mean frequency of interactions and $\sigma_F$ is the standard deviation.

When two people interact for a longer period of time, then the level of trust between them increases. The parameter of duration of social interactions is modelled by the cumulative density function of a normal distribution \[243\].

$$D(t|d) = \frac{1}{\sigma_D \sqrt{2\pi}} \int_{-\infty}^{d} e^{-\frac{(t-\mu_D)^2}{2\sigma_D^2}} dt$$ \hspace{1cm} (5.8)

where $d$ is the duration of interaction between two people in a particular context, $\mu_D$ is the mean frequency of interactions and $\sigma_D$ is the standard deviation.
The probabilistic model estimates the trust relationship between two people based on the above models and a particular context. The goal of this model is to provide an initial technique to measure trust based on information extracted from users’ social interactions. In order to achieve this, Section 5.3.3.1 defined four assumptions that would allow the creation of the trust model, among which a hard assumption regarding the independence of the variables. The later statement constitutes a logical assumption as the variables do not clearly depend on each other. Meaning that although there may be cases where people have a social relation and interact frequency, but it does not constitute a clear dependency between the two variables as there are cases that have social relation and do not interact frequently. It should be noted that the model assumes that each of the parameter contributes equally to the trust computation. Adding weights to these parameter constitutes part of the future work that is envision and described in the conclusion of the chapter. So the probabilistic trust level between two people is provided by the following equation.

\[
P(t|c) = \prod_{i \in \{\theta, r, f, d\}} P(t|i)
\]  

(5.9)

where \(i\) includes each of the parameters of the model i.e. relative orientation, social relation, frequency of interaction and duration of interactions.

The estimations of trust in different context are combined in order to calculate the overall trust score in the equation below.

\[
T(t) = \arg\max_{0 \leq t \leq 1} \left\{ \frac{1}{C} \sum_{j=1}^{C} P(c_j) \cdot P(t|c_j) \cdot w_j \right\}
\]  

(5.10)

where \(P(c_j)\) is the probability of the users being in the context \(c_j\), \(P(t|c_j)\) is the probability of the two people having a trust relationship \(t\) given that they are in context \(c_j\) and \(w_j\) is a probabilistic weight of a particular context \(c_j\) with respect to the trust.
5.4 Experimental setup

In this section the experimental setup is described. The experiment presented in this section has the sole purpose to evaluate the proposed real-world social graph model and the trust relationship model. During the experiment, participants were placed in an indoor room and socially interacted to measure the establishment of trust relationships among them.

Five participants took part in the experiment. When people meet each other for the first time, they establish an initial trust relationship [246]. Thus, the five participants did not have any prior knowledge about each other. This allows the monitoring of the trust relationship establishment in a short time frame. Hence, the proposed system will be evaluated in small-scale experiment, as a proof-of-concept.

Surveys were provided to the participants including questions about their social and trust relationships before and after the experiment, to establish the ground truth. In that way, it will be possible to observe the increase of trust relationship among the participant in short duration, while verifying the results through the survey. The survey was provided to the participants at the beginning of the experiment to ensure that they do not have any prior knowledge about each other. The survey provided to the participants before and after the experiment, were identical.

Some of the questions provided to the participants through the survey were:

- What is your social relation with each of the participants?

- What is your trust relation with each of the participants?

The experiment took place in a common indoor environment, a conference room. Before the beginning of the experiment, every participant answered the survey and the was given an HTC One S smartphone with the application deployed. Participants were asked to place the mobile phone in one of their trouser pocket in an arbitrary orientation. Then, they walked for some meters in order for the application to estimate users’ facing direction. After the calibration phase, the participants entered the room and started interacting in groups. After the experiment, the participants filled in again the
5.5 Results

This section discusses the results of the real-world experiment performed given the described experimental setup, with respect to the real-world social graph and the trust relationships among the participants.

5.5.1 Real-world social graph

The real-world social graph inferred from MobTrust and from the surveys given to the participants are presented in Figure 5.3. The graph extracted from MobTrust includes also the confidence of the estimated social relation as weights on the edges of the graph. The graph that shows the ground truth of the experiment depicts only the relation among the participants of the experiment without the inclusion of any weights.

Figure 5.3 shows that MobTrust was able to correctly recognise all social relation among the participants. As participants did not have any prior knowledge about each other and two pairs of them did not interact, they did not establish a social relation. The
relations (1, 3) and (2, 4) do not exist in both social graphs, indicating the correct recognition of the lack of relation by MobTrust. All detected social relations were classified as Social, which was verified by the answers participants provided in the surveys.

MobTrust provides also the confidence of estimation regarding the identified social relations. The social relation inference does not include any temporal factors in the estimation. But in cases where people interact for a longer time, the system receives a larger amount of social interaction and social relation snapshot samples. This allows it perform a more confident estimation of the social relation and the trust relationship between the two persons. From the results of the experiments, it is shown that (1, 2) and (3, 4) constitute the most confident pairs with respect to the estimated social relation. This is because these pairs interacted for the longest time (over 20 minutes), enable MobTrust to include a larger number of samples in the inference.

Regarding (1, 5), (2, 5), (3, 5) and (4, 5), MobTrust inferred their social relation with confidences between 0.35 and 0.57. The participants interacted between 10 and 15 minutes. The pairs (1, 4) and (2, 3) have the least confident estimations of social relations. This is because the two pairs interacted for less than 5 minutes, resulting in very few samples of social interaction inferences and social relation snapshots. In overall, Mobtrust managed to recognise correctly the social relations of the participants and provided the corresponding confidences of estimations.

5.5.2 Trust relationships

The trust relationships inferred from MobTrust and extracted from the ground truth provided by participants are depicted through two colou maps in Figure 5.4. The first colour map depicts the ground truth regarding the trust relationships the participants provided. The second colour map represents the output of the trust relationship inference of the MobTrust system.

In addition to the surveys, a human observer was recording the ground truth of the features used to infer trust relationships. Initially, the features extracted in MobTrust
5.5. Results

Figure 5.4: This figure shows the trust relationships derived from MobTrust and the surveys provided by the participants during the experiment (See Section 5.4). The numbers on x and y axis represent the ID of each user, while each cell indicates the level of trust between the corresponding IDs \((x_i,y_j)\). The white colour indicates the highest level of trust and the black colour shows the lowest level of trust i.e. not trusted person.

were analysed individually to provide a more detailed justification of the system’s inference result. In every pair of participants that took part in a social interaction, the relative orientation varied between the angles 40-90°. Regarding the social relations provided by the real-world social graph, every pair of participants was correlated with a confidence of estimation. During the experiment participants were interacting in different groups, thus the frequency and duration of interactions with different people varies.

MobTrust detected accurately the trust relationships among the participants of the experiments with respect to the ground truth they provided, as shown in Figure 5.4 and in Table 5.1 The Table 5.1 provides a more detailed analysis of the error of MobTrust with respect to the ground truth. The minus sign indicates that MobTrust predicted a lower value than the ground truth provided by the users. It should be noted that users input rates 1-5 in contrast to MobTrust that computes the trust rating in a continuous form. The pairs (1, 2), (3, 4) and (2, 5) showed the highest level of trust among all the participants of the experiment, managing over 0.5 confidence in at least
Chapter 5. Quantifying Real-world Trust Relationships through Social Interactions

Table 5.1: Error in estimating the trust relationship based on ground truth ($-1 \leq error \leq 1$). The

<table>
<thead>
<tr>
<th></th>
<th>User 1</th>
<th>User 2</th>
<th>User 3</th>
<th>User 4</th>
<th>User 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>0</td>
<td>-0.0429</td>
<td>0</td>
<td>-0.1300</td>
<td>0.0792</td>
</tr>
<tr>
<td>User 2</td>
<td>-0.0429</td>
<td>0</td>
<td>-0.1300</td>
<td>0</td>
<td>-0.0849</td>
</tr>
<tr>
<td>User 3</td>
<td>0</td>
<td>-0.1300</td>
<td>0</td>
<td>-0.1037</td>
<td>-0.1001</td>
</tr>
<tr>
<td>User 4</td>
<td>-0.1300</td>
<td>0</td>
<td>-0.1037</td>
<td>0</td>
<td>-0.0404</td>
</tr>
<tr>
<td>User 5</td>
<td>0.0792</td>
<td>-0.0849</td>
<td>-0.1001</td>
<td>-0.0404</td>
<td>0</td>
</tr>
</tbody>
</table>

15 minutes. In addition, the pair (2, 5) achieved a higher frequency of interaction in comparison to the other two pairs, which is verified by the human observer. This pair of participants interacted initially for 10 minutes and then interacted again towards the end of the experiment for 5 more minutes. Regards the pairs (1, 5) and (4, 5), MobTrust inferred correctly the trust relationships between the participants of each pair, given the ground truth. The pairs (1, 4) and (2, 3) interacted for a very short time, which is less than 5 minutes. Thus, their calculated trust score was quite low showing that it does not constitute a confident the estimation. In this experiment, participants did not have any prior knowledge about each other. Thus, a very short interaction and further a very small number of social interaction and social relation inferences constitute a challenge to infer absolutely accurately the level of trust among the participants.

In overall, MobTrust was able to provide an accurate estimation of the level of trust among the participants given the ground truth they provided. The system was able to differentiate between the less and the more trusted persons. In one case there was a discontinuity in the social interaction detection result that led in one faulty increase of the frequency of interactions for the particular user. Concluding, the system was able to estimate the probability of the trust relationship among the participants and proved the applicability of the system.

5.6 Potential applications

Several application areas may benefit from deriving trust relationships from real-world social interactions. A significant application, is to leverage the trust relationship among
people in order to add another important parameter into the definition of the multi-dimensional notion of context, leading the way to a large number of applications.

iKaaS is a project aiming to create a distributed multi-cloud environment based on the key discrimination between Global and Local Clouds, where the Global Cloud can be viewed as a legacy cloud computing paradigm whereas Local Clouds can be formed on demand [1] to extend cloud coverage so as to reduce latency and facilitate real-time provisioning. Local Clouds can range from small home installations to area/regional clouds based on the specific deployment scenario each time. End-devices are expected to play a key role in iKaaS not only for the data themselves they can provide for the optimization of the iKaaS provided services (i.e. GPS data for location based services, accelerometer /luminance data for enabling monitoring of the posture and home environment conditions of elderly citizens), but also for the data they can provide with respect to establishing social relationships.

Establishing the social relationships based on data coming from end-user devices in the context of iKaaS, can help in defining suitable end-points (both at granularity of end-devices and local clouds themselves) in a service provisioning chain. For example if two people are deemed as being in a trusted relationship, then in the case of a fall or accident of the first person, an alert can be sent only to the other trusted person to go assist. Once trusted people are identified in a service provisioning chain, then their devices and local cloud resources (if owned) can then be leveraged upon for service provisioning. E.g. if a person goes to a trusted person’s home, then these home local cloud resources can be utilized and have the notion of the follow me as I move service.

Once these end-points (end-devices and/or local clouds) are fixed, based on social relationships, this can be also taken into account by service provisioning functionalities, which can instantiate/migrate other processing functionalities needed in appropriate locations in the cloud to meet latency/response time requirements and cloud platform optimization objectives (e.g. to reduce the length of network paths that traffic will need to traverse in a cloud-based service provisioning chain and minimize as such the number of network links that may be overloaded due to that traffic).

As mentioned, trust relationships may be leveraged to understand the trustworthiness
of a particular environment. Users that are carrying their smartphones during the day, could be utilised to understand if the user is in a trusted environment or not. Given the understanding of the trustworthiness of the surroundings, different security and data access policies could be applied. In addition, in crowd-sensing applications users could focus the sensing process on particular trusted communities.

5.7 Conclusion

In this chapter, a novel real-world social graph model and a trust relationship estimation model were presented. The system leverages social interaction detection and social relation information provided by the work presented in Chapter 4. The real-world social graph advances against state-of-the-art, by creating a probabilistic model that takes into consideration snapshots of users’ social relation based on their interpersonal distance. The edges of the real-world social graph are weighted with the estimated social relation and the confidence of the estimation. Furthermore, a hybrid probabilistic model for deriving trust relationships based on contextual information from users’ social interactions and social relations, was presented. The model consider probabilistic information from psychology in order to merged the confidence of trust based on the social interaction and social relation information. As a proof-of-concept, a real-world experiment was conducted to showcase the applicability and accuracy of the approach in a small-scale experiment. Based on the ground truth, the results shown that MobTrust was able to understand the social relations among the participants, and provide a confidence of the estimation related to the number of samples the enabler collected in the time-frame participants interacted. MobTrust was able to accurately estimate the trust relationships among people of the experiment. In very short interactions, meaning less than 5 minutes MobTrust provided a low level of trust indicating the lack of confidence in the estimation. In overall, the system was able to understand accurately the social relations and the trust relationships among people. Finally, considering trust in the definition of context may advance a large amount of applications including service provisioning chain, health-care, data sharing and security.

MobTrust quantifies trust by considering the various features extracted from social
interactions as independent and identical distributed variables. Also the model assumes that each feature is equally weighted. As future work, research could be conducted in order to understand the weight of each of the parameter in the model, in order to design and develop a weighted model considering the effect of each of the parameters in the trust model.

As the DARSIS and MobTrust systems were evaluated in various conditions, a certain error was observed introduced by arbitrary movement induced when the users are walking. uDirect, the facing direction technique used in both systems, performs an implicit assumption that the relative orientation between the user and the device is constant. However, in daily life situations due to diverse movements of the user, the relative orientation of the device with respect to the user coordinate system changes (i.e. device displacement). As both DARSIS and MobTrust are performing inference based on the relative orientation of the users, the displacement error propagates in both systems. In order to tackle this source of error and improve the overall error distribution of the above system, the next chapter presents a generic solution for detecting and compensating device displacement for mobile devices.
Chapter 5. Quantifying Real-world Trust Relationships through Social Interactions
Chapter 6

A Prediction and Compensation Mechanism for Device Displacement

Context-awareness constitutes a notable characteristic for the development of mobile device applications. Device orientation is a feature utilised in various mobile applications. However, human movement cause the displacement of the initial reference of the device orientation. This chapter introduces a prediction and compensation mechanism for the device displacement error that leverages human locomotion patterns to refine the device orientation. The mechanism introduces prediction models for the device orientation based on the sensor and the human walking patterns. Error propagation models are developed that enable the prediction of the device displacement errors through a Kalman filter, as well as the correction of the device orientation. Finally, the proposed approach is evaluated through datasets provided by prior works.

6.1 Introduction

Over the last decade mobile devices have shown an large increase in popularity among people. Advanced computational and sensing capabilities allowed researchers to lever-
Chapter 6. A Prediction and Compensation Mechanism for Device Displacement

age mobile devices in order to extract contextual information [1]. This type of knowledge may be leveraged to create novel applications in various areas including Internet of Things, health-care, multimedia etc.

A derivative of contextual information is the extraction of personalised data produced by a device that belongs to a user. Mobile devices are able to sense data related to the user such as movement, voice etc. depending on the integrated sensors [2]. In particular, movement related sensing considers the data produced when the user is performing some activity and requires a device with an integrated Inertial Measurement Unit (IMU) i.e. at least an accelerometer and a gyroscope. Then, the data retrieved from the IMU are passed to the computational logic depending on the application, in order to infer about different types of contextual information such as user’s current location [254] and facing direction [5].

However, to infer personalised information related to movement, in a great amount of situations an implicit assumption is made. The movement performed by a particular body part is fully mapped to the movement of mobile device [255]. This assumption is valid to some extent, if the mobile devices are attached to user’s body and maintain the initial on-body position. As the user is engaged in various real-world situations, arbitrary movements also may affect the device position [12]. Even though the mobile device may be placed at a specific on-body position due to particular movements, the device may be displaced with respect to the initial device orientation with respect to the user (See Fig. 6.1). Thus, the initial reference of device orientation with respect to user’s body is lost, causing a considerable amount of error. As most of the applications target real-world situations, a detection and self-correction mechanism of device displacement is required.

In this chapter, a robust mechanism to detect and compensate on-body device displacement based on user’s walking locomotion, is designed, implemented and evaluated. The approach focuses on devices with an integrated IMU, which are placed on user’s body. An initial calibration phase when the user is in standing position computes the device orientation with respect to earth’s coordinates. Then, when the user initiates a walking cycle, the system detects potential device displacement through some reference points.
during the walking locomotion. A Kalman filter \cite{256} based on error modelling is incorporated to compensate for the detected device displacement and refine the device orientation with respect to the earth’s coordinates i.e. the relative orientation of the device with respect to the user.

![Azimuth Displacement](image1.png)

![Pitch Displacement](image2.png)

![Roll Displacement](image3.png)

Figure 6.1: This figure shows the displacement of the device occurring at each user in the dataset provided in \cite{257}. For the data collection process, users placed the mobile phones in their trousers’ pocket and walked on a straight line. Each user initially stood, then walked for some distance and then stood again. The difference in the orientation angle between the two standing positions was considered as displacement.

The main contributions of this chapter are summarised in the following bullet points:

- Development of analytical models for the device orientation error and the refinement mechanism that leverages users’ kinematics.

- Design and development of a device displacement compensation mechanism that considers users kinematics to refine the device orientation. The mechanism consists of a complementary Kalman filter that predicts and then compensates the device displacement error through the above analytical models.

- Evaluation of device displacement prediction and compensation mechanism.

The remainder of this chapter is structured as follows. Section 6.2 provides the background about the existing works related to the device displacement. The methodology
is explained in Section 6.3 to introduces the reader to the overall approach of the proposed device displacement compensation mechanism. Section 6.4 describes the system design by presenting the analytical error models and the complementary Kalman filter developed for device displacement compensation. The experimental set up including inertial dataset provided by prior works, is described in Section 6.5. Finally, Section 6.6 provide the evaluation of proposed mechanism based on different contextual parameters.

6.2 Background

Researchers have tried to tackle device displacement in various ways. The most common approach about this problem in literature is to assume that the device is fully attached to the user’s body. This means that device displacement does not occur and the relative orientation of the device with respect to the user does not change. In theory this constitutes a valid assumption, however in real-world situations where the environment is not controlled, arbitrary movements may cause device displacement [12]. Losing the initial reference of device orientation with respect to the user will induce a large amount of error which will propagate into the technique that is based on this assumption.

Another approach followed in literature is the process of applying a low pass filter or a moving average on raw accelerometer data in order to minimise the effect of loose attachment of device with respect to the user’s body [258] [259]. A filtering mechanism’s purpose is to reduce the fluctuation observed on the raw accelerometer data while the user is walking and the device is partially following user’s body part movement. In case of device displacement this approach is not able to cope with the orientation change and the user is required to place the device in the initial position.

A third methodology is to retrieve the device orientation when the user is in standing position as a calibration phase [5]. In this case an initial orientation of the device with respect to earth’s coordinate system is retrieved. After the induced arbitrary movement while the user is walking, a potential device displacement may occur (See Fig. 6.2). Thus, when the user enters the standing position, it is possible to perform again the calibration phase to detect and refine device displacement while comparing
the two orientations from the standing positions. However, this approach is limited by the requirement of the user being in the standing position. There is no approach that is able to detect and correct the device displacement during daily life situations.

A few works focused on understanding and trying to minimise the effect that the device displacement induces in the process of activity recognition. Kunze [260] analysed the effect of on-body sensor displacement and provided only some suggestions on how to reduce the induced error in systems performing activity recognition. In [261] authors presented an activity recognition approach that compensates the displacement by adapting the model based on the on-body position through extreme-learning machines. It should be noted that this method requires retraining of the activity model when device displacement occurs. Also, the model was tested only by changing the on-body positions, not on displacement originating from arbitrary movements. Banos et al. [14] developed activity recognition models through various machine-learning techniques. The models were evaluated in the following situations: a) ideal on-body position, b) users placed the devices on their body and c) device displacement was manually introduced. However, the above approaches focused only on the effect of device displacement on activity recognition models and did not present a generic mechanism on detecting and eliminating device displacement error.

The proposed approach detects and compensates device displacement for on-body mobile devices, which is based on user’s walking locomotion and is caused by arbitrary movements. In particular, it is able to refine the initial reference orientation of the device with respect to earth’s coordinates not only when users are standing but also when they are walking. The method is based on Kalman filter and as the prediction models used in the development filter are linear, the filter maintains its ability to provide an optimal estimate of displacement [256]. Furthermore, the system is not focused only on activity recognition, it constitutes a generic solution that can be leveraged by various applications. Because the device displacement compensation mechanism is not based on machine-learning it does not require any training or adaptation. In terms of computational and memory requirements, the system is capable on running on commercial off-the-shelf mobile phones and also may be deployed on any mobile device with an integrated IMU.
6.3 Methodology

This section provides the overall methodology developed in order to estimate device displacement and correct it. Initially, the assumptions are outlined based on which the methodology of the system was built. Then, the approach is presented to provide an overall understanding of the system’s functionality. Next, the kinematics utilised in the approach are analysed, followed by the displacement refinement models utilised during walking cycle.

6.3.1 Assumptions

In this subsection the assumptions are provided, based on which the on-body device displacement detection and compensation mechanism was developed.

- A mobile device orientation is represented with respect to earth’s coordinates. They constitute a reference system utilised to understand the relative orientation of other elements (e.g. devices, human body etc.) with respect to the mobile device.

- A mobile device has an integrated IMU with at least a 3D accelerometer and a 3D gyroscope, while a sensor fusion mechanism provides the instantaneous orientation of the device with respect to earth’s coordinates.

- In an IMU, a 3D accelerometer produces measurements of acceleration; it incorporates linear, gravitational and Gaussian white noise. A 3D gyroscope produces measurements of angular velocity; this includes angular velocity, an offset and Gaussian white noise.

- A person that carries a mobile device is able to perform a human walking cycle.

- A person’s human body is modelled as rigid body parts.
6.3. Methodology

Figure 6.2: This figure shows an example of the effect of device displacement on uDirect estimations, when computing user’s facing direction. An initial calibration phase is performed while user is standing and then user’s facing direction is estimated while walking on a straight line. This initial reference of device orientation with respect to earth’s coordinates was rotated to showcase the effect of device displacement on facing direction estimation approach.

Figure 6.3: This figure presents the information flow in the device displacement detection and compensation system. The IMU provides the required data to the preprocessing steps, where the activity and on-body position detection are taking place. The orientation and reference model predict the orientation based on the sensor and walking model as well as the displacement model that predicts the device displacement. From these predictions, the corresponding error values are generated and fed into the Kalman Filter. The refined error values are combined with the predicted device orientation to compensate potential device displacement.
6.3.2 Approach

Strapdown integration was selected for device orientation tracking as opposed to Gimbal system. This decision was taken to avoid the Gimbal lock occurring in particular rotations, as the device may be in arbitrary orientation. Rotation tracking can be performed through Euler angles, rotation matrices and quaternions. The latter was selected as they are not prone to Gimbal lock as Euler angles and are more computational efficient in comparison to rotation matrices.

6.3.3 Kinematics

Humans’ walking cycle constitutes a periodic movement. During the walking cycle each rigid body part is performing a particular movement, following a certain pattern. The movements performed during the walking cycle may be mapped to sagittal, coronal and transverse planes (See Fig. 6.4). This mapping allows tracking and analysis of the movement of particular body segments in a three dimensional space. Depending on the on-body position different linear, rotational and gravitational accelerations are applied while a person is walking.

The main idea is to take advantage of the walking cycle pattern to compensate the device displacement, while a person is walking. Each cycle includes some specific moments when the orientation of a body part is similar to the standing position of the user. These points may operate as reference instances to detect and correct any potential device displacement with respect to the human body.

6.3.3.1 Trousers pocket position

Trousers pocket constitutes one of the most common wearing positions of mobile devices [262]. Combined with the complexity of the human kinematics for the particular on-body position, this chapter will analyse the proposed mechanism for trousers pocket. Similar methodology may be applied on other on-body positions to detect and compensate on-body device displacement.

1When two out of three axis of rotation are in parallel in Euler angles, the system is not able to provide reliable orientation measurements, known as Gimbal lock.
Figure 6.4: This figure shows the three planes through which every movement of human body may be described in the following planes: (S) sagittal, (C) coronal and (T) transverse. In particular, the angular rotations performed during a walking cycle may be represented by rotations on these three planes.
Chapter 6. A Prediction and Compensation Mechanism for Device Displacement

Figure 6.5: Device displacement detection points based on kinematics. This figure shows the applicability of the approach on trousers’ pocket on-body position. The two reference points in walking cycle where the device orientation is similar to the standing position, are crossing the zero point [264].

The on-body position of trousers pocket is described as a system including pelvis, thigh and a joint that connects these two body segments. The angular rotations of these three rigid body parts are projected on sagittal, transverse and coronal planes (See Fig. 6.4). During a person’s walking cycle, angular movements in each plane are described in orthopaedic terminology [263] as: (S) Flexion and extension, (C) Adduction and abduction and (T) Internal and external. These projections lead to the creation of angular rotation models which provide information about the relative position of each segment with respect to the standing position.

Fig. 6.5 showcases the phases of gait cycle and at which percentage of the cycle they occur for the thigh movement. There is a direct mapping of each of the phases of human walking cycle with the thigh’s angle with respect to the standing position. This angle model is retrieved by projecting to the sagittal (S) plane, representing the angular motion between flexion and extension.
6.3.3.2 Other on-body positions

As shown in the literature [262] users place their mobile devices at other on-body positions also such as the belt, in the hand, in the backpack and in the shirt pocket. These constitute some of the common wearing positions of a mobile device. While a human is walking, each of these on-body positions perform different movement as well as different forces [265] and accelerations [266] are applied on each position. For example, the ground reaction force is applied to the pelvis joint from one leg as opposed to the chest where there are weaker ground reaction forces applied from both legs. Thus, an initial step to differentiate among these positions is to understand where the device is placed and then perform the analysis based on that on-body position.

Understanding and refining the device displacement based on the walking locomotion of the user, is performed by modelling and predicting the device movement based on each of these on-body positions. The relative movement of the particular body part is modelled and provides the reference model (See Section 6.4). The proposed system detects the on-body position of the device and based on that applies the appropriate walking locomotion pattern model to the compensation mechanism.

6.4 System Design

In this section the system design is presented for detecting and compensating on-body device displacement. Initially, the overall mechanism is described in an abstract overview by specifying major components. Then, these major components are analysed based on the workflow presented in Fig. 6.3 until the final device orientation is corrected.

6.4.1 Mechanism

The displacement compensation mechanism is depicted in Fig. 6.3. An IMU is considered as the basic sensing component that provides the data for the overall mechanism. Two preprocessing steps are introduced to discard any irrelevant data that could provide faulty estimations [257]: a) the activity recognition and b) the on-body position
6.4.2 Activity detection model

As an initial pre-processing step before the displacement compensation, there is a need to perform activity recognition. Literature [267] has stressed that in order to perform accurate on-body position detection, it is important to perform activity detection as a pre-processing step. Furthermore, activity detection is required as the device displacement detection and compensation mechanism leverages the walking patterns of the user in order to understand when the device has changed position related to user’s body. As the mechanism is based on users’ walking locomotion, the system requires the knowledge of when the user is walking. This knowledge is acquired through an activity detection model. This component will operate as a filter, by removing any data retrieved from the sensors when the user is not walking and performs some other activity, which could provide erroneous data to the system.
Various techniques have been proposed for the detection of the users’ activity. A simple and lightweight activity detection model was selected, which has proven its robustness in prior works such as [257]. This consists of the simple process of calculating the magnitude of the 3D accelerometer presented in Equation (6.1) and applying a threshold-based filtering mechanism. This model allows the system to understand if the user is standing or walking in a lightweight manner. As this model is based only on the magnitude of the accelerometer data, it discards the dependency of the accelerometer signal with respect to the device orientation.

\[
y_{m,t} = \sqrt{x_t^2 + y_t^2 + z_t^2}
\]  

(6.1)

As previously mentioned, there are plethora of activity recognition approaches that could provide higher robustness but with the penalty of the implementation complexity and higher computational burden. It is a trade-off between efficiency and computational burden. Modern mobile operating system platforms depending on the devices’ hardware, offer a native activity detection component that also could be utilised.

### 6.4.3 On-body position detection model

As explained in Section 6.3 the knowledge of the on-body position of the device is required. This system utilises different types of inertial and positioning data that differ based on the on-body position of the device. For example, when the device is placed in the trousers pocket then a certain pattern is produced in the accelerometer data. Once the device is placed in the shirt pocket, the amplitude of the signal of the pattern is reduced and also the frequency of the pattern is increased. This is because the device detects vibrations from both legs instead of one in the case of the trousers pocket. If all on-body positions would be processed as one, then a large number of faulty estimations would be induced in the system.

To reduce the erroneous input data to the system, an on-body position detection mechanism was introduced in [268]. This approach utilises inertial and positioning sensed data retrieved from sensors such as accelerometer, magnetic field and gyroscope. From
the sensed data, there are several features extracted in the frequency and time domain. A machine learning model infers based on the provided features about the particular on-body position. A garbage class is also included in the model for the cases where the model was not able to accurately infer one of the pre-defined on-body positions. Other techniques for on-body position detection could be leveraged such as [269], which is mainly based on data related to angular velocity and radius.

6.4.4 Prediction

This subsection describes the component that predicts the device displacement based on the current device orientation (See Fig. 6.6).

6.4.4.1 Orientation model

Initially, an orientation model predicts the current device orientation based on strap-down integration combined with the previously refined orientation. The orientation of the device is computed with respect to the earth’s coordinates. The device orientation is modelled as a first order Markov process where Gaussian white noise is applied [270]:

\[ y_{s,t} = q_{s,t} + p_{s,t} + \nu_{s,t} \]  \hspace{1cm} (6.2)

where \( y_{s,t} \) is the device orientation retrieved from the sensor output, \( q_{s,t} \) is the actual orientation of the device and \( p_{s,t} \) is the offset incorporate by the sensor while computing the device orientation and \( \nu_{s,t} \) is the Gaussian white noise. The actual orientation of the device \( q_{s,t} \) is modelled as a low pass filter with a smoothing factor \( 0 < c_s < 1 \):

\[ q_{s,t} = c_s \cdot q_{s,t-1} + w_{s,t} \]  \hspace{1cm} (6.3)

where \( q_{s,t} \) is current device orientation represented as a quaternion, \( q_{s,t-1} \) is previous device orientation also described as a quaternion and \( w_{s,t} \) is the Gaussian white noise.
The offset in the device orientation computation is modelled as a first order Markov process followed by white Gaussian noise.

\[ p_{s,t} = p_{s,t-1} + w_{p,t} \] (6.4)

where \( p_{s,t-1} \) is the offset computed in the previous orientation measurement and \( w_{p,t} \) is the white Gaussian noise.

### 6.4.4.2 Reference model

To refine potential displacement through the complementary filter, a reference model was developed that is based on human body kinematics. It is assumed that the device is placed at a particular on-body position. This model is based on user’s walking locomotion. During the walking cycle there are particular points where the device has similar orientation to the standing point. These points operate as ground truth to understand the existence of device displacement and compensate it.

Equation (6.5) describes the walking model utilised as the reference model, in order to understand which is the theoretical orientation of the device. Depending on the on-body position of the device, the parameters of the walking model are modified in order to adapt the model to the specific on-body position. The reference model is provided by kinematics [271]. A graphical representation of the model for the trousers’ pocket is depicted in Fig. 6.5.

\[ q_{R,t} = a_0 + a_1 \cdot \cos(t \cdot w) + b_1 \cdot \sin(t \cdot w) \]
\[ + a_2 \cdot \cos(2 \cdot t \cdot w) + b_2 \cdot \sin(2 \cdot t \cdot w) \] (6.5)

The model in Equation (6.5) provides the values of the device orientation during the human walking cycle. Leveraging the prior reference device orientation, the model is able to predict the different device orientations during the walking cycle. As shown in Fig. 6.5 during the walking cycle there are two points where the device orientation is
similar to the initial reference orientation. The reference model for the walking model is modelled also as a first order Markov process and described in Equation (6.6).

\[ y_{w,t} = q_{w,t} + b_{w,t} + \nu_{w,t} \]  

(6.6)

where \( y_{w,t} \) is the device orientation retrieved from the walking model output, \( q_{w,t} \) is the actual orientation of the device, \( b_{w,t} \) is the offset of the reference model with respect to the actual orientation and \( \nu_{w,t} \) is the Gaussian white noise. The actual orientation of the device \( q_{w,t} \) is modelled as a low pass filter with a smoothing factor \( 0 < c_w < 1 \):

\[ q_{w,t} = c_w \cdot q_{w,t-1} + w_{w,t} \]  

(6.7)

where \( q_{w,t} \) is current reference orientation represented as a quaternion, \( q_{w,t-1} \) is previous reference orientation also described as a quaternion and \( w_{w,t} \) is the Gaussian white noise. The offset of the reference model with respect to the actual reference orientation is modelled as a first order Markov process followed by white Gaussian noise.

\[ b_{w,t} = b_{w,t-1} + w_{b,t} \]  

(6.8)

where \( b_{w,t-1} \) is the offset computed in the previous time point and \( w_{b,t} \) is Gaussian white noise.

6.4.4.3 Displacement model

The displacement model produces the device orientation that constitutes the reference to understand the existence of device displacement. The displacement model compares the sensor orientation model and the kinematics walking model to infer if the device has been displaced and which is the displacement angle. This will allow the detection as well as the compensation of the actual device displacement.

\[ y_{d,t} = q_{s,t} - q_{w,t} + \nu_{d,t} \]  

(6.9)
where $q_{s,t}$ is the predicted orientation retrieved from the device sensor model at time $t$, $q_{w,t}$ is the predicted orientation retrieved from the kinematics walking mode at time $t$ and $\nu_{d,t}$ is Gaussian white noise.

### 6.4.5 Error models

The device orientation tracking is based on strapdown integration introduced in [272]. This allows the computation of the device orientation and also showcases the error propagation in process of computing the device orientation. The error propagation is described in Equation (6.10) as introduced in [273]. As shown in the following error propagation model, there is a dependency on the time $T$.

$$ q_{e,sensor,t}^- = q_{e,sensor,t-1}^+ - T \cdot c_s \cdot p_{e,t-1}^+ + T \cdot \nu_{s,t} $$ (6.10)

where $q_{e,sensor,t-1}^+$ is the estimated orientation error in the previous time point, $p_{e,t-1}$ is the estimated error in the offset of the orientation and $\nu_{s,t}$ is the noise term. This error propagation formula is used for both the sensor orientation error and the walking orientation error model.

Similarly, the error for estimating the device orientation based on human kinematics was also modelled as the first approximation of the strapdown integration described in Equation (6.11). The error model applies to $T$ human walking cycles.

$$ q_{e,walking,t}^- = q_{e,walking,t-1}^+ - T \cdot c_w \cdot b_{e,t-1}^+ + T \cdot \nu_{w,t} $$ (6.11)

where $q_{e,walking,t-1}^+$ is the estimated orientation error in the previous time point, $b_{e,t-1}$ is the estimated error in the offset of the orientation and $\nu_{w,t}$ is the noise term.

The propagation of the displacement error is modelled as a low pass filter with smoothing factor $c_d$ and is described by Equation (6.12).

$$ q_{e,displacement,t}^- = c_d \cdot q_{e,displacement,t-1}^+ + T \cdot \nu_{d,t} $$ (6.12)

where $q_{e,displacement,t-1}^+$ is the estimated displacement error in the previous time point.
and $\nu_{d,t}$ is the noise term.

### 6.4.6 Kalman Filter

Having modelled the different information sources of the system and the error of each source, a Kalman filter is developed. The purpose of the Kalman filter is to compensate the error in the device displacement and provide at the end an estimation of that error. The Kalman error state and the error measurement equations are described in Equation (6.13).

\[
\begin{align*}
    x_{\epsilon,t} &= A \cdot x_{\epsilon,t-1} + w_{x,t} \\
    z_{\epsilon,t} &= C \cdot x_{\epsilon,t} + \nu_{z,t}
\end{align*}
\]  

(6.13)

Following, the Kalman error state incorporates the errors in the offset of the orientation sensor, the offset of the walking model and the error in the displacement of the device. The Kalman error state is described in Equation (6.14).

\[
    x_{\epsilon,t} = \begin{bmatrix} p_{\epsilon,t} & b_{\epsilon,t} & d_{\epsilon,t} \end{bmatrix}^T
\]  

(6.14)

where $p_{\epsilon,t}$ is the error of the orientation in the sensor fusion mechanism, $b_{\epsilon,t}$ is the error in the offset of the walking locomotion model and $d_{\epsilon,t}$ is the error in the device displacement.

For the description of the Kalman filter there is a need to compute the uncertainty of each estimation $P_t$. As defined, the Kalman state is based on the various sources of error introduced in device displacement. The Kalman filter predicts the amount of error. Each estimation of the filter provides the instantaneous error. This means that each error estimation does not depend on the previous estimation. Thus, the matrix $A$ of Kalman state equation (6.13) is equal to zero matrix. Considering that $A$ is the equal to the zero matrix, the augend is removed in Equation (6.15).

\[
    P_{t+1} = A(I - K_t C)P_t A^T + Q_{w,t+1}
\]

\[
    = Q_{w,t+1}
\]  

(6.15)
The prediction of the next Kalman error state is provided by Equation (6.16).

\[
\hat{x}_{\epsilon,t}^+ = \hat{x}_{\epsilon,t}^- + K_t \cdot (z_{\epsilon,t} - C \hat{x}_{\epsilon,t}^-)
\] (6.16)

where \( \hat{x}_{\epsilon,t}^- \) is predicted error state based on the prediction models previously defined, \( K_t \) is the Kalman gain computed at each step of the filter and \( C \) is the matrix describing the conversion of the error state to the error measurement. Following, the Kalman Gain is described by Equation (6.17) where \( Q_{\nu,t} \) is the measurement noise of the filter.

\[
K_t = \frac{P_t C^T}{C P_t C^T + Q_{\nu,t}}
\] (6.17)

The measurement of the Kalman filter is described by Equation (6.18) and includes measurements from the error of the orientation sensor, the walking model and the potential displacement of the device with respect to the human body.

\[
z_{\epsilon,t} = \begin{bmatrix} z_{\epsilon,\text{sensor},t} & z_{\epsilon,\text{walking},t} & z_{\epsilon,\text{displacement},t} \end{bmatrix}^T
\] (6.18)

In order to calculate the error for the orientation sensor, the estimated device orientation \( q_{\theta,t}^- \) based on the sensor model at the particular time point \( t \) is subtracted from the measured value of the device orientation \( y_{\theta,t} \) at the same time point \( t \). For the error of the walking model from kinematics, the estimated device orientation \( q_{\theta,t}^- \) is subtracted from the measurement value of the device orientation based on the walking mode \( y_{w,t} \) at the same time point \( t \). The error for the displacement of the device is computed through subtracting the estimated displacement of the device \( q_{d,t}^- \) from the measured device displacement \( y_{d,t} \) at the same time point \( t \). In order to compute the matrix \( C \) in Equation (6.13), the measured kalman state is subtracted from the predicted state.

### 6.4.6.1 Covariance matrices

In order to complete the Kalman filter there is a need to calculate the covariance matrices for the noise terms in the Kalman equations presented in Equation (6.13), meaning \( w_{z,t} \) the system noise and \( \nu_{z,t} \) the measurement noise. As the previous errors are in-
Chapter 6. A Prediction and Compensation Mechanism for Device Displacement

tegrated in the current estimation of the Kalman state, the matrix $A$ that represents the transition between the two error states is equal to the zero matrix. Thus in order to calculate the covariance matrices for the noise terms $w_{z,t}$ and $v_{z,t}$, there is a need to compute the variance of the equation indicating the propagation of the error.

The system noise is calculated through the matrix in Equation (6.19). This is because the offset in the orientation error, the offset in the walking model and the displacement of the device are independent of each other. So the system noise consists of a matrix with the covariances of each of the error source on the diagonal of the matrix $Q_{w,t}$.

$$Q_{w,t} = E(x_c x_c^T) = \begin{bmatrix} Q_{p,t} & 0 & 0 \\ 0 & Q_{b,t} & 0 \\ 0 & 0 & Q_{d,t} \end{bmatrix}$$  \quad (6.19)

Having defined the covariance matrix of the system noise, the individual covariance matrices of each source of error are derived from the error propagation equation.

### 6.4.7 Correction

This section describes the correction of the orientation by considering the previously estimated error device displacement and further the actual device displacement. The correction process includes the refinement of the device displacement by considering the estimated error. Following, taking into account the corrected device displacement, the device orientation is recalibrated to the correct orientation. The device displacement is computed by considering the predicted device displacement and the estimated error in the prediction. Considering the estimated device orientation reference, the potential device displacement, and the previous device orientation reference, the new device orientation is computed.

### 6.5 Experimental setup

This section presents the experimental setup utilised for the evaluation of the proposed system for detection and compensation of on-body device displacement. Initially
the evaluation datasets are introduced, followed by the performance parameters and metrics.

6.5.1 Dataset

This section provides a brief description of the dataset included in the evaluation process. The dataset collection process is presented for the reader to understand the conditions under which the data were collected.

Authors in [257] collected a dataset using commercial off-the-shelf mobile phones from 27 users. For the data collection process, they utilised the smartphone Galaxy Nexus GT-19250 with Android 4.1.1, having integrated the Bosch BMA220 accelerometer with sampling frequency 100Hz. The data collected were extracted from sensors such as accelerometer, magnetic field and gyroscope. For the collection of gyroscope data, authors leveraged the calibrated sensor provided by Android OS [275]. The calibration scheme strives to tackle the drift (bias) and noise introduced by the mechanical part of the sensor. Regarding the experiment, all users walked in a straight line for 70-90 steps. The direction of the users was not compulsory, so participants may have walked towards both directions on the predefined line.

Initially, the participants stood for a few seconds and then started to walk for a again a few seconds in a normal pace. Following, the walking pace increased for some time and then decreased to less than the normal walk pace. After 70-90 steps the participants stood for a few seconds. Authors collected data for various on-body positions including front and back trousers pocket, hand held, hand held using, backpack, hand bag and shirt pocket. The smartphone device was placed in an arbitrary orientation in these seven different on-body positions. The participants’ were both female and male, in the age range of 15-29 years old and their height ranged from 150-189cm. For more information about the statistics of the participants, the reader is referred to Table 6.1.

6.5.2 Performance metrics

This section describes the performance metrics used in order to evaluate the proposed detection and compensation mechanism for device displacement. The first metric used
for the evaluation is the accuracy of the mechanism (See Section 6.6.1); the remaining 3D displacement after the proposed mechanism has been applied. The results are shown in radians, zero is the ideal case where the device displacement has been detected and compensated fully. The sign in the displacement represents the directionality of the displacement. The second metric used for the evaluation of the proposed mechanism is the error of the approach represented in percentage (%). Boxplots are used to showcase the error distribution of the approach in different on-body positions (See Section 6.6.2.1), with the use of different orientation tracking methods (See Section 6.6.2.2) and with respect to different rotation planes (See Section 6.6.2.3). The upper and lowest horizontal lines are connected through a vertical line with the boxplot and represent the maximum and minimum values of the error distribution. At the boxplot the upper and the lowest horizontal sides represent the 3rd and the 1st quartile of the error distribution. The horizontal line in the middle of the boxplot indicate the median error rate of the approach. Based on these performance metrics, the evaluation of the proposed mechanism provided the results discussed in the next section.

### 6.6 Results

This section provides discussions about the results of the evaluation of the detection and compensation mechanism for device displacement. The section is divided into two subsections that analyse the performance of the mechanism in terms of accuracy and the error distribution of the approach.
6.6.1 Performance Evaluation

The following sections discuss the results for the performance evaluation of the compensation mechanism in the seven different on-body positions, which were included in the evaluation dataset (See Section 6.5). Each of the related figures presents the remaining displacement after the compensation mechanism has been applied. These boxplots enable the presentation of the displacement distribution after the compensation mechanism. The figures showcase the displacement through the use of Euler Angles that constitutes an easy and readable way for the reader to understand the performance of the proposed mechanism. The rotation angles are measured in radians $[-\pi, \pi]$ to be compliant with the modern smartphone operating systems.

6.6.1.1 Trousers front pocket

Fig. 6.7 presents the results of applying the proposed compensation mechanism to refine the device displacement occurring while the user is walking. The mechanism has in the majority of the cases, compensated the device displacement. As shown in the figure, the azimuth and roll displacement distributions while user is walking are in best cases $[-0.1, 0.1]$ and in worst cases between $[-1.1]$ radians. The pitch of the device displacement is placed mainly between $[-0.2, 0.2]$ radians. This difference is reasonable based on kinematics, as the rotational movements performed on the traverse and the sagittal planes are much larger than the rotational movements conducted on the coronal plane.

The displacement distribution for the pitch is very small and 75% of the samples for all the users, are close to 0 displacement. For azimuth and roll, it could be observed that for around half of the users, the compensation mechanism is able to refine the device displacement, so the distribution of the displacement is very small. In the other half, the displacement distribution indicates a higher variation. This is occurring because of the combination of large rotational movement at the particular on-body position and the generality of the walking model used to compensate the device displacement. As the rotational movements at this on-body position are larger, any potential displacement is proportionally large also. In addition, according to kinematics the walking pattern
Figure 6.7: This figure depicts the displacement that occurred while the users were walking and had the smartphone in their front trousers pocket.

of a person is different depending on several factors including height. Thus, as the walking model used for the detection and compensation of the device displacement is generic, there would be cases where the model does not fit on particular users. An adaptive model that could be calibrated based on a person’s walking pattern, could lead to a more personalised approach for compensating the device displacement and would improve the performance of the proposed mechanism.

6.6.1.2 Trousers back pocket

Fig. 6.8 depicts the boxplots for the device displacement at the trousers back pocket. The results across the users showcase similar distribution with very small variations, except from user #21. The maximum device displacement observed in Fig. 6.8 at the majority of the users is between \([-0.2, 0.2]\) radians for azimuth, pitch and roll. Given the large rotational movements occurring in the trousers back pocket, the compensation mechanism has managed to refine over 0.8 radians for azimuth and roll. The median device displacement for azimuth and roll is close to 0 radians, while a small deviation in observed in the pitch, where in average 0.02 radians displacement is taking place.

As indicated by the Fig. 6.8 there is inconsistency for the user #21, where large remaining displacement is observed, which the proposed mechanism was not able to refine
Figure 6.8: This figure depicts the displacement that occurred while the users were walking and had the smartphone in their back trousers pocket.

accurately. This error occurred because the data of that users showcased significant difference to the normal walking patterns of the particular on-body position. So the theoretical walking model that was responsible for compensating the displacement, could not fit properly on the sensor data in contrast to the remaining 23 users. However, as shown in the figure the mechanism performed very well at 23 out of 24 users and was able to reduce the displacement by over 80%.

6.6.1.3 Hand held

The results of evaluating the detection and compensation mechanism for the hand held on-body position are presented in Fig. 6.9. The compensation mechanism managed the highest performance when the users are holding the devices in their hands. This is occurring due to the goodness of fitting of the walking model, for the particular on-body position. While observing the displacement at the hand held on-body position, the distribution of displacement is similar across the three rotation axis. Across the different users, the performance is similar and does not showcase any variation larger than ±0.05 radians. The displacement values for azimuth, pitch and roll are between $[-0.1, 0.1]$ radians.

The maximum displacement in average based on the boxplots in Fig. 6.9 are between
Figure 6.9: This figure depicts the displacement that occurred while the users were walking and were holding the smartphone in the hands.

$[-0.05, 0.05]$ radians with a slight increase in the azimuth around $[-0.07, 0.07]$. As observed, the variation in pitch and roll is smaller than in the azimuth. The median displacement across all three rotation axis is close to 0. For pitch and roll, the 1st quantile of the displacement is around -0.01 and the 3rd quantile of the displacement is between $[0.01, 0.02]$ radians. For azimuth, a slightly larger variation of the boxplots is observed between $[-0.02, 0.02]$ radians that in some cases reaches $[-0.02, 0.04]$ radians.

The variations observed with the particular margins are normal to be present, as the users walk using different models and for that reason the compensation mechanism utilises the complementary Kalman filter to reduce the effect of that noise. The proposed mechanism achieved its best performance at the hand held on-body position, in comparison to the other on-body positions.

6.6.1.4 Hand held using

Fig. 6.10 showcases the displacement distribution for each user. As observed the deviation from 0 for the different users, is small at most $[-0.1, 0.1]$ radians. This indicates that at the on-body position where the user is holding the device in their hands and using it, the proposed mechanism has managed to successfully compensate the device displacement introduced while the users are walking. It is observed that the average
6.6. Results

As the users walk, hold the devices in their hands and use them, the rotations that take place are small and the displacement that occurs is for a longer period i.e. the user may decide to change the distance between the head and the device. In this case, the compensation mechanism is able to refine the displacement with greater accuracy. Another parameter that facilitates the compensation mechanism is the reduction of the accelerations applied on the smartphone, while the users’ hands operate as a smoothing factor when using the device. For example, at the trousers front pocket the rotation movements are larger and introduce a larger amount of error to the proposed mechanism.

At the majority of the users, 75% of the samples for azimuth, pitch and roll indicate no displacement. The median displacement for the different users is also around -0.01 radians, indicating the compensation capabilities of the approach. Regarding the maximum displacement, it can be observed that users #8 and #12 reached 0.1 radians as maximum displacement for azimuth, pitch and roll. Observing the rest of the users, the maximum displacement for azimuth, pitch and roll is around ±0.05 radians. The best performance is observed at users #11, #14 and #23 in all three rotational axis, where the 3rd quantile of the displacement distribution is 0, meaning that there is no

Figure 6.10: This figure depicts the displacement that occurred while the users were walking, were holding the smartphone in the hands and using it.
6.6.1.5 Backpack

This section presents the results from the evaluation of the compensation mechanism when the smartphone is placed in the user’s backpack. Fig. 6.11 depicts the results of the evaluation based on the displacement distribution through boxplots for each user. The device in the backpack is affected by the movements introduced during the walking cycle of the user. The compensation mechanism managed a good performance as indicated by the figure, in terms of refining the device displacement, as the median and 3\textsuperscript{rd} quantile of the displacement are close to 0.

About the azimuth, across all users the displacement distribution is centred around 0, with the maximum values being between \([-0.2, 0.2]\) radians. As mentioned the median and 3\textsuperscript{rd} quantile of the displacement distribution are at most \(\pm 0.05\) radians from 0, indicating a very small remaining displacement. User #7 showcases a certain variation around \(\pm 0.05\) radians for median and \(\pm 0.16\) radians for 3\textsuperscript{rd} quantile. The maximum values for user #7 show a greater variation, reaching almost \(\pm 0.6\) radians.

For pitch, the compensation mechanism managed to reduce the displacement to the
maximum values of less than ±0.1 radians, except from user #14 that indicated a
greater variation with maximum values ±0.18 radians. The median values are close to
0, while the 3rd quantile is at most ±0.04 radians. For roll, median and 3rd quantile
show some slight displacement around ±0.05 while the maximum values vary between
±0.2. User #7 show some larger variation where the median displacement is 0.1 and
the 3rd quantile is 0.2. The maximum values for the particular user are between [±0.8]
radians. Considering the overall result, the compensation mechanism managed to refine
the device orientation by compensating the device displacement in all three rotation
axis.

6.6.1.6 Handbag

This section discusses the performance results of evaluating the proposed mechanism
for the handbag on-body position. Fig. 6.12 depicts the boxplots for the displacement
distribution for each user when the device is placed in a handbag and the compensa-
tion mechanism is applied on the orientation data. It should be noted that the dataset
utilised for the evaluation of compensation mechanism, included sensor data from the
handbag on-body position from only 6 users, and for that reason, Fig. 6.12 presents
only six boxplots. The compensation mechanism provides consistent displacement dis-
tributions for azimuth, pitch and roll among the six different users.

The compensation mechanism achieved removing the displacement for azimuth, pitch
and roll towards 50% of the samples. The 3rd quantile of the samples for each user
indicated a small displacement around 0.01-0.04 radians for azimuth, pitch and roll.
Regarding the maximum displacement values, 5 out of 6 users had maximum displace-
ment less than ±0.15 radians, and one user showed 0.4 radians displacement in azimuth.
For pitch, all users had at most ±0.13 radians displacement. For roll, the maximum
displacement was at most ±0.2 radians. The compensation mechanism managed sim-
ilar performance at users #1-6. A very small variation around ±0.05 radians could
be observed at the pitch of user #5. Also, user #2 azimuth displacement distribution
showcases a small variation [−0.16, 0.1] radians for the 1st and 3rd quantiles. Apart
from that, the displacement distribution among these users is similar.
Chapter 6. A Prediction and Compensation Mechanism for Device Displacement

6.6.1.7 Shirt pocket

This section presents the evaluation of the compensation mechanism with respect to the shirt pocket on-body position. The results are depicted in Fig. 6.13; the evaluation dataset included data for only two users, for that reason the figure includes only two boxplots for each rotation axis. In comparison to previous on-body positions, the smartphone is placed loosely in the particular on-body position allowing it to perform small arbitrary movements during the walking cycle. The margins of the boxplots indicate that the compensation mechanism was able to refine the device displacement in azimuth, pitch and roll. Different variations are observed in the three rotation axis, due to the kinematics of the particular on-body position and the loose attachment of the device.

Starting with azimuth, the 3rd quantile of the result samples does not include any displacement, while the median displacement is approximately -0.04 radians. The maximum displacement values are included between [-0.2, 0.17]. For the pitch, the 3rd quantile of the result samples is between [0.007, 0.018] and the median varies between [-0.05, 0] radians, where 0 radians is considered as fully refined displacement. The maximum values showcase some variation, as for user #1 the maximum values are ±0.02 radians and for user #2 the maximum values are ±0.04 radians which are still con-

Figure 6.12: This figure depicts the displacement that occurred while the users were walking and having the smartphone in the handbag.
Figure 6.13: This figure depicts the displacement that occurred while the users were walking and had the smartphone in their shirt pocket.

sidered very small variations. For roll, the rotational movements are a little bit larger and for both users the 3rd quantile of the result samples is around -0.02 radians, while the median is -0.06 radians. The maximum values in the displacement distribution for both users are between $[-0.2, 0.1]$. In overall, the compensation mechanism was able to reduce the device displacement and in over 50% of the cases to fully compensate the device displacement, while the user is walking.

6.6.2 Error Evaluation

This section presents results of the evaluation of the error of the displacement detection and compensation mechanism. The error of the approach is evaluated with respect to the different on-body positions of the device, the utilisation of different orientation tracking method and also the distribution of the error on the different projection planes (See Fig. 6.4).

6.6.2.1 On-body positions

This section presents the results from the evaluation of the device compensation mechanism with respect to different on-body positions. Fig. 6.14 depicts the error distribu-
Chapter 6. A Prediction and Compensation Mechanism for Device Displacement

Figure 6.14: Error of the detection and compensation mechanism for device displacement on different on-body positions. The proposed approach was evaluated on 7 different on-body positions: a) trousers back pocket, b) trousers front pocket, c) hand held, d) hand held while using, e) backpack, f) handbag and g) shirt pocket.

As shown in Fig. 6.14 the compensation mechanism managed less than 10% maximum error values in different on-body positions except from the front trousers pocket. The smallest error was achieved when the users were holding the smartphone in their hands, with maximum error 2.6%, 3rd quantile error of 1% and median error of 0.6%. The largest error was identified in the front trousers pocket, where the maximum error value is 25.4%, the 3rd quantile was 11.8% and the median 4.6%. Next, the trousers back pocket had 8% maximum error value, 3rd quantile 4.3% and 3% median error. The backpack, handbag and the shirt pocket has similar error distributions where the maximum value was close to 5%, the 3rd quantile was around 3% and the median error around 2%. The hand held on-body position when the person is using the smartphone, had similar error distribution with the hand held with a slight increase in the maximum and the 3rd quantile values.

The compensation mechanism indicates robustness by showing that 75% of the samples
have less than 5% error, except the front trousers pocket. As discussed in Section 6.6.1, this on-body position is affected a lot by large rotational and linear movements during the walking cycle. Also, depending on the natural characteristics of a person such as age, gender, height etc., the walking patterns change from person to person. As shown in previous section the walking model that was utilised, fitted well in half of the participants. Thus, for the trousers pocket, where the rotational movements are larger and the displacement errors increase proportionally, there is a need to develop an adaptive walking model that will be able to cope with the different walking characteristics of people. The adaptive walking model could benefit also the trousers back pocket, which also showed a small increase in the error, but still had maximum error less than 9%. In the rest of the cases, where the rotational movements are smaller the compensation mechanism operated very well and managed less than 5% maximum error.

6.6.2.2 Orientation tracking method

This section presents the results from the evaluation of the device compensation mechanism with respect to different orientation tracking methods. Fig. 6.15 depicts the accumulated error distribution of the device displacement with respect to different orientation tracking methods and the proposed compensation mechanism from all the on-body positions. It should be noted that the state-of-the-art techniques do not take into account the patterns in the kinematics of the human walking cycle, to refine the device displacement.

The proposed mechanism for device displacement compensation outperformed the state-of-the-art technique for orientation tracking in terms of error distribution with maximum error 5%, 3rd quantile at 2% and median at 1%. The largest error was observed in the orientation tracking based on the accelerometer and magnetic field sensor with maximum error 66%, 3rd quantile at 28% error rate, and median error at 8%. This orientation tracking approach was highly affected from the magnetic disturbance that is present in real-world environments. In addition, it utilised raw sensor measurements that added a certain amount of noise and combined with the users’ movement, resulted in the highest error. The orientation tracking through strapdown integration of gyro-
Figure 6.15: Error of the detection and compensation mechanism for device displacement with respect to different orientation tracking methods; a) accelerometer and magnetic field sensor readings, b) strap-down integration using angular velocity from gyroscope, c) complementary filter sensor fusion of accelerometer, magnetic field and gyroscope, d) Kalman filter based sensor fusion of accelerometer, magnetic field and gyroscope and e) the proposed device displacement compensation mechanism.

The complementary filter tackled the drift and achieved 49.6% maximum error, 21.7% for the 3rd quantile and 6.8% median error. Kalman filter improved even more the error distribution for device displacement by managing 41% maximum error, 28% 3rd quantile and 6% median error. The further improvement is observed as the Kalman filter considers also the error distribution of the sensor signal as Gaussian white noise. The median and 1st quantile error in all state-of-the-art technique are relatively similar. The difference is observed in the 3rd quantile of the error distribution. The proposed compensation mechanism managed to reduce the 1st quantile, the median and the 3rd quantile of the error distribution by considering the walking effect in the orientation signal.

Overall, the displacement compensation mechanism managed to improve the error dis-
tribution of device displacement while the user was walking. In particular, the maximum error value was improved by more than 50%, the 3rd quantile by at least 16% and the median error by 4%. The orientation tracking through accelerometer and magnetometer had the worst error distribution, as the magnetic disturbance or real-world environments combined with the users’ movement, added a considerable amount of error. Due to gyroscope drift and users’ movement, similar error distribution was observed in the strapdown integration of the gyroscope. The complementary and the Kalman filters reduced the errors introduced by magnetic disturbance and gyroscope drift, but still the device displacement is observed due to users’ walking movement. Only, the proposed compensation considered the walking patterns of the users, to refine the device displacement that occurs while the users are walking.

6.6.2.3 Projection planes

This section presents the results from the evaluation of the device compensation mechanism with respect to the different projection planes at the device orientation. The results of this evaluation are depicted in Fig. 6.16 where the error distribution of the device displacement is presented with respect to the different project planes and to the different on-body positions.

The overall error with respect to the three projection planes, observed in all on-body positions is less than 5% (See Fig. 6.14). An increase is observed in the trousers front pocket, at the azimuth and roll. The error in the pitch at the trousers front pocket is below 7%. The median error of the mechanism in all on-body position is 9% for azimuth and roll, while for pitch it is at 3%. In the traverse plane, the maximum error value is detected at the trousers front pocket with 30% even though the median is at 9%. In the coronal plane, the maximum error observed, is in both front and back trousers pocket between 6-7%. In the sagittal plane, the maximum error is at 30% at the front trousers pocket but the rest of the on-body positions have a maximum error of 10%.

Evaluation showed that the proposed mechanism managed to reduce the displacement error over 40% for azimuth, 5% for pitch and 20% for roll. The largest displacement
error was observed when the accelerometer and magnetic field sensors were used to
track the device displacement, with median error 20% and 3rd quantile at 50% in
azimuth, 4% median error and 3rd quantile at 10% in pitch, and 9% median error
and 3rd quantile at 30% in roll. Similar displacement error was observed in gyroscope
strapdown integration. The complementary filter tackled the drift and reduced the 3rd
quantile of azimuth, pitch and roll by 1-11%. The Kalman filter managed a further
improvement and achieved a reduction of 3rd quantile of pitch (3%) and roll (8%).

The evaluation shows that the error distribution of the proposed compensation mecha-
nism, with respect to the projection planes is highly dependent on the on-body position
of the device. In some on-body positions, where the rotational movements are larger
such as the trousers front pocket, higher amount of error is introduced. With respect to
the projection planes, the error is accumulated mainly in azimuth and roll and less in
pitch. This can be justified through kinematics, where the larger rotational movements
are taking place on the traverse and sagittal planes (See Fig. 6.4). At some on-body
positions such as hand held while using the smartphone, the rotational movements of
the device across all the projection planes are smaller, and for that reason the proposed
mechanism is able to refine it almost perfectly.

6.6.3 Discussion

As shown in the above evaluation, the compensation mechanism is able to refine the device displacement at different on-body positions. It is important to note that this is the first work that proposes such a compensation mechanism and evaluates it in such a diverse dataset. The evaluation included 27 users, with different ages, genders and heights in order to introduce diversity in the evaluation process. In addition, the evaluation showed that the compensation mechanism is able to cope with different walking speeds, as the dataset included normal, fast and slow walking speed. These characteristics that were introduced through the selection of such a diverse dataset, add the robustness to the compensation mechanism indicating its capability to cope with different people, on-body positions and walking speeds, making it applicable for real-world environments. There is no restriction in the evaluation that would limit the applicability of the approach, except from the assumption that the person is able to conduct normal walking cycles. The user is able to perform other activities, and for that reason the approach includes also an activity recognition mechanism. While the user performs other activities, the system discard the particular data, until it detects that the user starts to walk, so it is able to perform the procedure, using correct data.

As shown in the evaluation, at on-body position such as hand held, hand held using, backpack, handbag and shirt pocket, the approach was able to eliminate the median displacement and achieve less than 0.05 for the 75% of the samples. At the trousers pocket and especially the front pocket, where there are large rotational movements taking place, the approach managed to cope very well with the displacement and achieved similar performance to the rest of the on-body positions. There were some cases where the walking model for the front trousers pocket did not fit well with the walking patterns of some users, and for that reason the performance was degraded. As proposed, an adaptive walking model that will be able, to adapt based on the walking characteristics of a person, will improve the accuracy for the remaining users.

An error evaluation was also performed to understand the error distribution with re-
Chapter 6. A Prediction and Compensation Mechanism for Device Displacement

spect to different on-body positions, orientation tracking methods and also the projection planes. The evaluation showed that the compensation mechanism has median error less than 5% for all seven on-body positions that were evaluated. All on-body positions managed less than 5% error for the 3rd quantile, two of them less than 2% error and the trousers front pocket 12%. For the different orientation tracking methods, the maximum error value was improved by more than 50%, the 3rd quantile by at least 16% and the median error by 4% with respect to the state-of-the-art techniques. Regarding the projection planes, in most cases azimuth and roll showcased an increase in the variation in comparison to pitch. This is because the movements that are taking place during the walking cycle are mainly projected on the traverse and sagittal plane. Overall, the approach was able to accurately tackle the device displacement and provide a robust compensation solution.

6.7 Conclusion

In this chapter a novel approach for predicting and compensating a potential device displacement was designed and evaluated with respect to user’s body induced by arbitrary movement. The proposed mechanism leverages the walking locomotion of the user in order to detect and compensate the device displacement. State-of-the-art activity recognition and on-body position detection components are applied as pre-processing steps, in order to filter out erroneous data. Analytical models for orientation sensor, walking locomotion and device displacement were developed, followed by the corresponding error models. A complementary filter merged information provided by the analytical and error models. In particular a Kalman filter estimated the orientation, walking locomotion and displacement errors. The estimated errors were used to correct device orientation by considering the device displacement. The proposed mechanism was evaluated on a state-of-the-art dataset from 27 users on seven distinct device on-body positions and with three different walking speeds.

The device displacement compensation approach will constitute an important enabler for applications that require the knowledge of the device orientation reference with respect to global coordinate system. This method managed to provide an updated
6.7. Conclusion

and self-calibrated estimation of the device orientation reference not only while the user is standing, but also while the user is walking. This will benefit a wide variety of applications that require the initial reference of the device orientation with respect to the earth’s coordinates. An important application is localisation based on inertial sensors, that requires the initial reference of the device orientation to calculate the user’s facing direction [5] and further track the user’s location. Gaming is an area that utilise inertial sensors and device orientation to improve the interactiveness of the app. The device orientation compensation mechanism could filter out the walking locomotion effect from the device orientation data and allow the users to walk while they are playing. This will introduce a new era of mobile gaming. Other example application for the device displacement compensation mechanism could constitute the ability to record video or take panorama image while the user is walking. The method will filter out the walking movement, resulting in a constant and without horizontal and vertical movement video or image.
Chapter 7

Conclusion

The main purpose of this research was to understand through scientific methods if mobile devices, given their advance in computational power and sensing capabilities, are capable of operating as self-acting tools for detecting human behaviour. This work was based on the assumption that mobile devices are pervasive, utilised by people in their daily lives. For that reason, mobile devices while providing various functionalities such as phone calls, texting, Internet browsing etc., may passively monitor the users and infer various aspects of their human behaviour. Due to their computational and battery limitations, efficient and intelligent techniques may be applied that will allow continuous and unobtrusive inference of human behaviour. Understanding of human behaviour in an automatic but also non-intrusive manner, constitutes an important and emerging area for various fields. This requires collaboration of information technology with humanitarian sciences in order to transfer existing knowledge of human behaviour into self-acting tools. These tools will eliminate human error that is introduced by current obtrusive methods such as questionnaires. To achieve unobtrusiveness, the focus is directed towards exploiting the pervasive and ubiquitous character of mobile devices.

In this research, a survey of existing techniques for extracting social behaviour through mobile devices was provided. Initially the terminology of the area was discussed, followed by the introduction of a concrete architecture of social signal processing applications on mobile phones, constituted by sensing, social interaction detection, behavioural cues extraction, social signal inference and social behaviour understanding. State-of-
the-art techniques applied on each of the distinct stages of the process were presented. Furthermore, potential applications were shown while arguing about the main challenges of the area. Following to the main challenges of the area of social behaviour extraction, the main gap that was identified in literature, is the detection of real-world social interactions in an accurate and unobtrusive manner. Even though a noticeable number of works were included in this survey, the area of Mobile SSP is still at the initial stages of development.

Close collaboration among different areas such as engineering and psychology is still an open issue. Most of the works have focused on extracting various behavioural cues, by developing various processing and inference techniques on sensor data. However, the lack of the aforementioned collaboration is becoming obvious in the inference of social behaviours. Engineers do not have the knowledge regarding the grammar of different social signals, which need to be incorporated. In parallel, understanding the context in which the social behaviour occurs is important. Regarding the inference process, researchers may start to combine multiple modalities to provide more coherent and accurate social behaviour detection. Also, labelling the data and establishing ground truth is a burdensome problem for researchers while developing inference mechanisms. As social behaviour detection applications are based on sensing personal data, there is a need for developing privacy-preserving frameworks that would protect users’ personal data. Last, this research focused on mobile devices and smartphones that people use to improve their daily lives. Social behaviour detection may take into consideration the energy consumption of these approaches, as they may affect the users’ experience by draining the battery of the mobile devices. In overall, the area of Mobile SSP has shown great advance but still has open challenges that require tackling.

This work tried to provide concrete solutions on some of the challenges identified in the area of Mobile SSP. One of the core gaps of the area, was the lack of considering the existence of social interaction while detecting social behaviour. In order to detect social interactions there is a need for an accurate, reliable and real-time recognition, of users’ interpersonal distance. Towards fulfilling this requirement, research was conducted for an interpersonal distance estimation technique based upon a non-intrusive opportunistic mechanism that solely relies on sensors and communication capabilities
of off-the-shelf smartphones. A set of novel hierarchical classifiers was developed for interpersonal distance estimation, produced by a training set of Bluetooth RSSIs and a feature selection process. The proposed interpersonal distance estimation models outperformed state-of-the-art solutions and achieved up to 93.52% accuracy. The hierarchical approach that was introduced managed to outperform the other solutions, which showed the effectiveness of the structure of the classifier. The large number of features (3050) combined with the large dataset (48000 RSSIs) enabled the Multi-BoostAB classifiers to find the appropriate balance between variance and bias, to cope with the fluctuation of the RSSI signals. Using a more advanced communication interface such as WiFi-Direct will allow the retrieval of larger number of RSSI samples in a shorter time, meaning that the classifier may utilise a larger window of RSSI samples. This may potentially increase even more the accuracy and robustness of the model.

In overall, the interpersonal distance estimation constitutes an accurate and reliable approach for estimating the distance between two devices.

The realisation of social interaction detection requires accurate, reliable and real-time inference of interpersonal distance and relative orientation. A social interaction detection system was the outcome of the research in order to address these challenges through a pervasive and opportunistic approach based on off-the-shelf smartphones. The novel machine-learning model for interpersonal distance estimation previously developed was utilised. A relative orientation computation technique was developed that lightens the on-body position restriction of prior works. A collaboration sensing approach was introduced to allow real-time transfer of sensed data, required for the inference process. To understand the accuracy of the approach, a generic analytical model was provided that predicts the probability of people interacting given the interpersonal distance and relative orientation. From the analytical model the error model of the approach was derived and from psychology the error probability to compute the expected error of the approach was extracted. The importance of incorporating the relative orientation in social interaction detection was showcased in three real-world environments. The overall system was benchmarked as a coherent system against an RFID-approach in a real-world environment. The system showed robustness in different types of indoor environments. DARSIS constitutes a coherent solution that leverages the relative spatial
arrangement of users to understand when they are socially interacting. As the system focuses on being opportunistic and non-intrusive, the speech recognition element was not included. This decision was made to discard any privacy issues that are introduced when the microphone is used as well as the fact that speech recognition adds another layer of complexity to the system.

Deriving trust relationships from real-world social interactions may contribute significant information towards social behaviour understanding. The level of trust among people constitutes an important parameter for describing the social context but also for security and privacy in pervasive systems. An opportunistic sensing system that derives and quantifies trust relationships among people through smartphones based on the detected social interactions in a laboratory environment was presented. A real-world social graph was derived from users’ daily social interactions by also considering snapshots of their social relation. A hybrid model was developed to quantify users’ trust relationships in a laboratory environment based on the contextual information provided by the detected social interactions. As a proof of concept, an evaluation of the system in a laboratory environment provided some initial insights regarding quantifying trust relationships. Further evaluation in a larger experiment would showcase the accuracy of the system, prove the ability of the system to cope in various contexts and showcase the saturation that is observed in trust measurements after a certain period of time.

Context-awareness constitutes an important characteristic for the development of mobile device applications. Device orientation is a feature utilised in various mobile applications, including the detection of users facing direction and relative orientation. However, human movement causes the displacement of the initial reference of the device orientation. A prediction and compensation mechanism for the device displacement error that leverages human locomotion patterns to refine the device orientation was introduced. The mechanism introduced prediction models for the device orientation based on the sensor and the human walking patterns. Error propagation models were developed that enable the prediction of the device displacement errors through a Kalman filter, as well as the correction of the device orientation. The proposed approach was evaluated through datasets provided by prior works in different on-body positions, dif-
7.1 Future work

This research strived to provide a taxonomy of the literature for detecting human behaviour in an automatic way through mobile devices. Having identified the challenges and gaps of the literature, certain key aspects were chosen and were tackled through this research. However, there are still remaining gaps of human behaviour inference that required tackling in order to allow the area to become more mature. This section focuses on providing some insight on the remaining gaps and how this research may be exploited to further advance the area of Mobile SSP.

The first step is to improve the social interaction detection mechanism and the trust relationship computation system with the device displacement detection and compensa-
Chapter 7. Conclusion

tion method. As the literature indicated [12], in real-world situations various arbitrary movements introduce device displacement. uDirect, the facing direction estimation algorithm is utilised in the social interaction detection mechanism and the trust relationship computation system. However, uDirect is prone to error introduced by device displacement, as the reference orientation of the device with respect to the earth’s coordinates is affected. So the next step for enhancing the two systems, is to incorporate the device displacement mechanism that will discard any displacement of the device introduced by arbitrary movements, which are generated in real-world environments.

The second step is to evaluate the trust relationship computation mechanism in a large scale experiments. This research constituted a proof-of-concept for quantifying trust relationships through social interactions in a laboratory environment. Thus, the evaluation was conducted in a small-scale experiment, where five participants were involved for two hours in the experiment. A large-scale experiment and with larger duration should be conducted to understand the accuracy and the error distribution of the approach. The large-scale experiment will also introduce various types of contexts that would exploit fully the capabilities of the system, as the small-scale experiment considered only one particular context.

The trust relationship computation mechanism may provide valuable input to the area of crowd-sourcing. Mobile devices produce a large amount of data through their sensors. Crowd-sourcing platforms collect and utilised the data produced by mobile devices and other data sources. However, at this point there is no mechanism that may verify if the produced data are trustworthy. Anomaly detection mechanisms have been developed to tackled this is issue, but these are based on the patterns observed in the sensed data. Using the level of trust of the user, crowd-sourcing platforms may have a prior knowledge of the trustworthiness of the user and handle the data produced by the particular user in a corresponding manner. In addition, various data correlations may be empowered based on the users’ social and trust relationships, and provide new streams of information.

Also, the trust relationship computation mechanism may provide real-time service provisioning. Currently, on-line services utilise various types of authentication to secure
the end-to-end communication channel. However, apart from the authentication credentials, there is no other way to ensure that both ends are legitimate. For example, there is no way for an on-line platform to understand if the user/device is trustworthy enough in order to perform a particular transaction. Quantifying trust relationships provides real-time and continuous evaluation of the level of trust of a user. Services may leverage the trust measurement provided by the enabler, to ensure that the users are trusted enough, to perform certain transactions. Different types of services may be provided to users/devices with different levels of trust, which ensures a more personalised service. Users may be able to share information and data with particular communities and ”circles of trust” based the social and trust relationships. The devices themselves may adapt their security policies depending on the trustworthiness of the environment. In health-care, when an accident happens, only trusted people nearby may operate as first responders. Trust constitutes a fundamental aspect in every transaction, either in real-world or on-line, so a quantification mechanism may provide an automatic way to ensure that trust among the peers.

Finally, this research focused on showing that mobile devices are able to detect human behaviour. This was performed by developing a set of enablers for understanding different levels of human behaviour e.g. interpersonal distance, social interactions and trust relationships. These self-acting tools may be utilised as a whole or as individual components to detect other aspects of human behaviour, and also in various applications. As services become more personalised, social behaviour inference may start to play an even more important role in advancing today’s applications and services. This will create the need for developing more advanced and complex social behaviour tools. In overall, social behaviour inference constitutes a well-promising area and may provide great insight in today’s and future applications.
Appendix A

Tables

A.1 Summary of Mobile SSP Literature

This section summarises in a table, all works analysed in literature review of Mobile SSP. Table A.1 presents each work and classifies its components into the different stages of social behaviour inference. The initial taxonomy includes the distinct social behaviour inference stages of sensing, social interaction detection, behavioural cues extraction and from social signals to social behaviour inference. In sensing column we outline the type of sensor data utilised by a particular work. Then, social interaction column describes the methodology developed by researchers in order to estimate ongoing social interactions. Following, the column of behavioural cues refers to the approach induced by researchers to extract behavioural cues. In order to clarify this process, the column is divided into extracted features, method developed to perform decision making, and classification target i.e. the result of inference. Similarly, the social behaviour column includes the research conducted by each work in terms of understanding social behaviour. As social behaviour inference is performed through a decision mechanism, this column is also divided into the extracted features, the decision method and the classification target. In overall, Table A.1 constitutes a summary and brief categorisation into Mobile SSP inference stages, of the works discussed in the literature review. This article categorised related works based on the developed methodology. To enhance
reader’s understanding of the various methodologies, Table A.1 provides an alternative view by outlining each work and the corresponding methods for every stage of social behaviour inference.
### Table A.1: Literature on Mobile Social Signal Processing.

<table>
<thead>
<tr>
<th>System</th>
<th>Sensing</th>
<th>Social Interaction</th>
<th>Behavioural Cues Features</th>
<th>Method</th>
<th>Target</th>
<th>Social Behaviour Features</th>
<th>Method</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sociometer</em></td>
<td>Microphone</td>
<td>Infrared</td>
<td>non-initial maximum of the autocorrelation, number of autocorrelation peaks, normalized spectral entropy</td>
<td>2-layer HMM</td>
<td>Conversation, Turn-taking</td>
<td>conversation duration, frequency, ratio of interaction, centrality scores</td>
<td></td>
<td>social network structure</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>influence, embeddedness in community</td>
</tr>
<tr>
<td>[276][173]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GPS, WiFi, Bluetooth, Calls, SMS</td>
<td>Bluetooth Discovery</td>
<td></td>
<td>number and time at locations, social interactions, calls, SMS</td>
<td></td>
<td>deviation, standard deviation, confidence interval</td>
<td></td>
<td>Stress</td>
</tr>
<tr>
<td>[139]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>NeuroPhone</em></td>
<td>EEG</td>
<td>Person identification from images through P300 brain signals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[151]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing Interaction</td>
<td>Social Behaviour</td>
<td>Target</td>
<td>Features</td>
<td>Method</td>
<td>Target</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------</td>
<td>------------------</td>
<td>--------</td>
<td>-------------------------------</td>
<td>--------------</td>
<td>--------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AutoSense</td>
<td>ECG, RIP, GSR, Skin Thermometer, Ambient Temperature Sensor, Accelerometer</td>
<td></td>
<td></td>
<td>Performing windowing on raw data, preliminary features (virtual sensors)</td>
<td>Analysis of HRV</td>
<td>Conversation, Activity, Posture</td>
<td>mean, variance, heart rate, respiration rate</td>
<td>Stress 90%</td>
</tr>
<tr>
<td>Surround Sense</td>
<td>WiFi, Camera, GSM, Microphone, Accelerometer</td>
<td></td>
<td></td>
<td>raw ECG data</td>
<td>QRS detection</td>
<td>motion activity, posture, heart rate</td>
<td>activity, heart rate</td>
<td>Neural Network, Fuzzy Logic algorithms</td>
</tr>
<tr>
<td>Virtual Compass</td>
<td>Fusion Bluetooth and WiFi RSSI</td>
<td></td>
<td></td>
<td>Sound, Light, Color, WiFi, Accelerometer fingerprints</td>
<td>colour clustering, light extraction, feature selection</td>
<td>social context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[118]</td>
<td>Camera</td>
<td></td>
<td>Image</td>
<td>Fiducial Points</td>
<td>Boosting Naive</td>
<td>Emotion (neutral 76.3%, joy 78.3%, sad 74.7%, surprise 78.7%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[My Experience]</td>
<td>Location, Bluetooth, User interaction, Device state</td>
<td>Device charging, SMS, Cellular info</td>
<td>Data aggregation</td>
<td>Social Context</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[PEYE 91]</td>
<td>Video recording</td>
<td>partitioned image in 4 equal regions</td>
<td>block matching in 16x16 image through three step search, four step, diamond, hexagon, and the adaptive multiple-mode search</td>
<td>Gesture recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e-Gesture [92]</td>
<td>Accelerometer, Gyroscope</td>
<td>raw, delta, integral data for each axis</td>
<td>HMM, Viterbi algorithm maximum likelihood</td>
<td>Gesture recognition (94.6%, in 4 different contexts)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------------------</td>
<td>------------------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Features</td>
<td>Method</td>
<td>Target</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[93]</td>
<td>Orientation sensor</td>
<td>Pitch and Roll</td>
<td>SWAB [94], HMM</td>
<td>Gesture recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uWave [95]</td>
<td>Accelerometer</td>
<td>Quantized accelerometer data</td>
<td>Dynamic time warping</td>
<td>Gesture recognition (8 distinct gestures, 98.6% accuracy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myo [97]</td>
<td>Electromyographic sensor</td>
<td>electrical activity of skeletal muscles</td>
<td></td>
<td>Gesture recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEPTIMU [100]</td>
<td>Accelerometer, Gyroscope, Microphone in earphones</td>
<td>Raw data</td>
<td></td>
<td>Head Posture, Physiological (Heart rate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smart Pose [101]</td>
<td>Accelerometer, Orientation Sensor, Camera</td>
<td>Raw data</td>
<td>Face detection, Device shaking detection, average tilt of device</td>
<td>Head Posture (User’s neck tilt angle)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>onWell [104]</td>
<td>Zephyr Bio-Harness 3</td>
<td>Heart and Physical Activity</td>
<td>Standard deviation and Threshold-based</td>
<td>Body posture</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------</td>
<td>----------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CONSORTS-S</strong></td>
<td>Accelerometer, Electrocardiograph, Thermometer</td>
<td>last value(thermometer), maximum value (electrocardiograph), average (accelerometer), variance and primary spectrum frequency from DFT</td>
<td>Decision Rules</td>
<td>Body posture (Standing, still, facing downwards and upwards), Physical Activity (staying, walking, running), Physiological (Heart rate, skin temperature)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[105]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>uDirect</strong></td>
<td>Orientation Sensor</td>
<td>Pitch</td>
<td>Threshold</td>
<td>Body posture (sitting, standing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>uDirect</strong></td>
<td>Accelerometer, Magnetometer</td>
<td>Raw data</td>
<td>Device calibration, Detect relative orientation between user and device</td>
<td>Body posture (facing direction)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>------------------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SenseMs</td>
<td>Virtual Sensor (SMS)</td>
<td></td>
<td>SMS</td>
<td>Phone Usage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Content Interpretation (Facial expressions, avatars, colours, size, location)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SeeMon</td>
<td>BVP, GSR, Light, Temperature, Humidity, 2-axis accelerometer, GPS</td>
<td>Raw data, skin conductance</td>
<td>Physical activity (Strain, movement), Physiological (Heart rate)</td>
<td>BVP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OpenDataKit</td>
<td>All available physical and virtual sensors of device</td>
<td></td>
<td></td>
<td>Stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SystemSens</td>
<td>All available physical and virtual sensors of device</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------------------------------------</td>
<td>-----------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Funf</strong></td>
<td>All available physical and virtual sensors of device</td>
<td>Bluetooth Discovery, GSM, GPS</td>
<td>Accelerometer</td>
<td>Physical activity (high, moderate, low)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Medusa</strong></td>
<td>All available physical and virtual sensors of device</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>METIS</strong></td>
<td>Accelerometer, Bluetooth, discovery</td>
<td>Bluetooth discovery</td>
<td>Vibrations on desks, location, noise level, presence duration</td>
<td>Social context</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MSF</strong></td>
<td>All available physical and virtual sensors of device</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Anonysense</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LiveLab</strong></td>
<td>Phone usage, Network usage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>------------------</td>
<td>------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PRISM</strong> [30]</td>
<td>microphone, camera, GPS, external accelerometer</td>
<td>GPS location</td>
<td>GPS, microphone</td>
<td>Social context</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CenceMe</strong> [38]</td>
<td>Accelerometer, Microphone, Camera, GPS, Bluetooth</td>
<td>Bluetooth Discovery</td>
<td>Accelerometer (mean, std, peaks number), Audio (mean and std of DFT), Bluetooth MAC</td>
<td>Decision tree, Rule-based, Physical activity (sitting, standing, walking, running), Auditory (Conversation), Social context</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BeTelGeuse</strong> [39]</td>
<td>Integrated Sensors, Camera, GSM, Phone Usage data, GPS, Acceleration, Temperature, Heart Rate</td>
<td>Accelerometer, Heart Rate, GSM, GPS</td>
<td>Physical Activity, Social Context</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
<td>Method</td>
<td>Target</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>---------------------------------------</td>
<td>------------------</td>
<td>----------------</td>
<td>------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jigsaw [40]</td>
<td>Accelerometer, Microphone, GPS</td>
<td></td>
<td>mean, variance, mean crossing rate, spectrum peak, sub-band energy, sub-band energy ratio, spectral entropy, MFCC</td>
<td></td>
<td>J48, GMM</td>
<td>Physical Activity, Posture (sitting, standing), Conversation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EmotionSense [41]</td>
<td>Accelerometer, Bluetooth, Location, Microphone</td>
<td>Bluetooth Discovery, Conversation Detection</td>
<td>Audio, Accelerometer data</td>
<td>HTK - GMM - Maximum A Posteriori</td>
<td></td>
<td>Conversation (Speaker Recognition), Physical Activity (movement)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambient Dyna-</td>
<td>All available sensors of device and external</td>
<td></td>
<td>Accelerometer, Zephyr Hx</td>
<td></td>
<td></td>
<td>Colocation, interaction patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>namiz [42]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Compute relation strength</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Emotion (happiness, sad, fear, anger, neutral)
<table>
<thead>
<tr>
<th>System</th>
<th>Sensing</th>
<th>Social Interaction</th>
<th>Behavioural Cues</th>
<th>Social Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditeur</td>
<td>Microphone</td>
<td>FFT, ZCR, RMS, 13-MFCCs, Low Energy (Weak), Frame Rate, Spectral (Entropy, Energy, Flux, Roll-off, Centroid), Bandwidth, Phase Deviation, Pitch, and statistics of these</td>
<td>Naïve Bayes, Decision Tree, GMM, MLP, SVM, kNN, HMM</td>
<td>Auditory (Prosody, Turn-taking, Vocal Outbursts, Conversation)</td>
</tr>
<tr>
<td>Visage</td>
<td>Camera</td>
<td>Shape, Texture</td>
<td>Fisher Linear Discriminant Analysis (Fisherface)</td>
<td>Facial cues (Facial Expression)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Geometric, Appearance</td>
</tr>
<tr>
<td></td>
<td>Camera</td>
<td>Local face features through local random bases</td>
<td>Sequential Neural Network</td>
<td>Facial cues (Facial Expression)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Emotion (angry, disgust, fear, happy, neutral, sad, surprise)</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Features</td>
<td>Method</td>
</tr>
<tr>
<td>[118]</td>
<td>Camera</td>
<td></td>
<td>Fiducial Points</td>
<td>Boosting Naïve</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>with Local Gabor and Gini</td>
<td>Bayesian</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Facial cues</td>
<td>Facial Expression</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Facial Express-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ion)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[119]</td>
<td>Camera</td>
<td></td>
<td>Image</td>
<td>Eigenface de-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>composition,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Distance pro-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>jection and</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>computation,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SVM</td>
</tr>
<tr>
<td>eyeLook [122]</td>
<td>Camera, Eye</td>
<td></td>
<td>Eye contact</td>
<td>Facial cues</td>
</tr>
<tr>
<td></td>
<td>contact sensors</td>
<td></td>
<td>interaction</td>
<td>(Eye tracking -</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sensor readings</td>
<td>turn taking)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>-------------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>WiFi, Bluetooth, cell tower, GPS, accelerometer, apps Info, calls, SMS, contacts, phone/network info, power/screen state, alarm clock</td>
<td>Bluetooth discovery</td>
<td>Accelerometer data</td>
<td>variance of magnitude</td>
<td>Activity level</td>
</tr>
<tr>
<td>Internet usage, call, SMS, apps, alarm clock, GPS, Bluetooth</td>
<td>Bluetooth discovery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bluetooth</td>
<td>Bluetooth discovery with static and mobile nodes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing Interaction</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>[80]</td>
<td>[accelerometer, microphone, barometric pressure, temperature, humidity, visible light, infrared light, battery]</td>
<td>[Conversation detection]</td>
<td>[maximum autocorrelation peak, total number of autocorrelation peaks, and relative spectral entropy of sound detected]</td>
<td>[Conversation (speech or not), Activity level]</td>
</tr>
<tr>
<td>[168]</td>
<td>[Calls, SMS, Bluetooth discovery]</td>
<td>[Bluetooth discovery]</td>
<td>[Device usage, Proximity]</td>
<td>[general phone usage, diversity, active behaviours, regularity]</td>
</tr>
<tr>
<td>[124]</td>
<td>[Accelerometer, Bluetooth, light sensor, dry electrodes]</td>
<td>[face movement]</td>
<td>[Wearable electrooculography]</td>
<td>[Facial cues (Eye tracking)]</td>
</tr>
<tr>
<td>[154]</td>
<td>[ECG]</td>
<td>[ECG data]</td>
<td>[Kalman filter, Pan Tompkins, R Detection]</td>
<td>[Stress]</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------</td>
<td>--------------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td><strong>Comm2Sense</strong></td>
<td>WiFi</td>
<td>Interpersonal distance based on WiFi RSSI</td>
<td>maximum and mean of 20-sample window of WiFi RSSI</td>
<td>Naïve Bayes with KDE</td>
</tr>
<tr>
<td><strong>AMMON</strong></td>
<td>GPS, Calendar, Microphone</td>
<td></td>
<td></td>
<td>linear SVM</td>
</tr>
<tr>
<td><strong>Application usage, Bluetooth, SMS, calls</strong></td>
<td>Bluetooth discovery</td>
<td>number of occurrences, mean, median</td>
<td></td>
<td>Extraversion, Agreeableness, Conscientiousness, Emotional Stability, Openness to Experience</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------</td>
<td>--------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>The Mobile Sensing Platform</td>
<td>microphone, light phototransistor, accelerometer, barometer, thermometer, IR, humidity, compass</td>
<td></td>
<td>linear and log-scale frequency coefficients, cepstral coefficients, spectral entropy, band-pass filter coefficients, correlations, integrals, means, variances</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HMM, CRF</td>
<td>Activity Recognition</td>
</tr>
<tr>
<td>Social Serendipity</td>
<td>Bluetooth</td>
<td>Bluetooth discovery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[44]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[129]</td>
<td>Camera</td>
<td>Video recording</td>
<td>ERICA</td>
<td>Facial cues (Eye-tracking)</td>
</tr>
<tr>
<td>[140]</td>
<td>Static and mobile Bluetooth nodes</td>
<td>Bluetooth discovery</td>
<td></td>
<td>Social Context</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------------</td>
<td>--------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td>Microphone, Accelerometer, Barometer, Bluetooth</td>
<td>Bluetooth discovery</td>
<td>Microphone, accelerometer, pressure data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>standard deviation of the acceleration magnitude, median of the absolute linear acceleration magnitude, inter-quartile-range of the absolute linear acceleration magnitude, long-term-spectral-variability</td>
<td></td>
</tr>
<tr>
<td>Coenofire</td>
<td>[10]</td>
<td></td>
<td>movement activity, movement intensity, movement variability, speech activity</td>
<td></td>
</tr>
<tr>
<td>BlueEye</td>
<td>Microphone, speakers, accelerometer, magnetometer</td>
<td>Microphone data, accelerometer, magnetometer</td>
<td>DSP filters, trilateration, ultrasound frequencies</td>
<td>Interpersonal distance</td>
</tr>
<tr>
<td></td>
<td>Bluetooth</td>
<td>Bluetooth RSSI</td>
<td>Modified PLM incl. devices’ relative orientation</td>
<td>Interpersonal distance (proximity)</td>
</tr>
<tr>
<td>System</td>
<td>Sensing Interaction</td>
<td>Social Interaction</td>
<td>Behavioural Cues Features</td>
<td>Method</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>---------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>[162]</td>
<td>Biomonitoring sensors</td>
<td></td>
<td>avg heart beats per min, root mean square of std, % of differences between adjacent beat-to-beat intervals, avg breaths per min, EDA turning points, EDA percentage of increase, EDA responses, avg slope of EDA response, EMG number of contractions, EMG % of activity, mean temperature, gradient of linear regression of temp</td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>[209]</td>
<td>IR cameras, IR beacons</td>
<td>users spatial formation through IR beacons</td>
<td>IR cameras and beacons</td>
<td>interpersonal distance and spatial arrangement</td>
</tr>
<tr>
<td>PBN [89]</td>
<td>accelerometer, GPS, WiFi, external(2-axis accelerometer, microphone, light, and temperature)</td>
<td>On-body sensor data</td>
<td>Pearson correlation coefficient, Adaboost, Kullback-Leibler divergence</td>
<td>Activity classification</td>
</tr>
<tr>
<td>Meeting Mediator [65]</td>
<td>Microphone, Bluetooth</td>
<td>Conversation detection, Bluetooth discovery</td>
<td>Total speaking time, Overlap speaking time, Turn taking per sec, avg length of speech, avg speaking energy, avg speaking speed</td>
<td>Auditory cues (conversation, silence, turn-taking)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>speaking time, avg speech length, variation speech energy, variation in movement, questionnaire data</td>
<td>Dominance</td>
</tr>
<tr>
<td>System</td>
<td>Sensing Interaction</td>
<td>Behavioural Cues Features</td>
<td>Method</td>
<td>Target</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>BeWell [19]</td>
<td>GPS, accelerometer and microphone</td>
<td>accelerometer, GPS, microphone data, frequency and duration phone charging, time stationary or silent sound environment</td>
<td>[15] [10]</td>
<td>Conversation, physical activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[86]</td>
<td>GPS, accelerometer, microphone</td>
<td>accelerometer, microphone data, GPS trajectory, stay points, location interest etc.</td>
<td>[40] naive Bayes, smoothing Markov model</td>
<td>Activity detection, Social context</td>
</tr>
<tr>
<td>PeopleTones</td>
<td>GSM cell tower based</td>
<td>Cell tower based</td>
<td>Cell tower based</td>
<td>Cellular data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PhoneMonitor</td>
<td>Bluetooth, light sensor, WiFi, GPS, cellular data, device info</td>
<td>Bluetooth proximity, Bluetooth RSSI</td>
<td>Probabilistic model</td>
<td>Interpersonal distance (Proximity)</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>[107]</td>
<td>Camera, accelerometer, orientation sensor</td>
<td>User’s image, accelerometer, orientation data</td>
<td>face detection, threshold based hand detection, avg device tilt</td>
<td>Head posture</td>
</tr>
<tr>
<td><em>SocioPhone</em> [60]</td>
<td>Microphone</td>
<td>Conversation and turn-taking detection</td>
<td>300 ms-frames: power, average of the square, decibel given sound pressure level, ZCR, RMS</td>
<td>multi-class SVM, two consecutive window frames to same cluster</td>
</tr>
<tr>
<td>[145]</td>
<td>Microphone, GPS, Bluetooth, calls, calendar, BSN</td>
<td>Bluetooth distance estimation</td>
<td>Bluetooth RSSI, Amplitude modulation, spectral profile, harmonicity, accelerometer variation and raw data, calendar, location</td>
<td>PLM, tree classifier, threshold,</td>
</tr>
</tbody>
</table>

Appendix A. Tables
<table>
<thead>
<tr>
<th>System</th>
<th>Sensing Interaction</th>
<th>Social Interaction</th>
<th>Behavioural Cues</th>
<th>Social Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Moodscope</strong></td>
<td>Device usage, SMS, calls, location</td>
<td>Device usage, SMS, calls, location</td>
<td>Orientation sensor data</td>
<td>Orientation sensor data, threshold for different orientations</td>
</tr>
<tr>
<td><strong>[163]</strong></td>
<td><strong>[106]</strong></td>
<td><strong>[95]</strong></td>
<td><strong>[132]</strong></td>
<td><strong>[67]</strong></td>
</tr>
<tr>
<td><strong>Orientation sensor</strong></td>
<td><strong>Accelerometer</strong></td>
<td><strong>Microphone, speakers</strong></td>
<td><strong>ToA</strong></td>
<td><strong>Interpersonal distance (distance estimation)</strong></td>
</tr>
<tr>
<td><strong>uWave</strong></td>
<td><strong>[95]</strong></td>
<td><strong>[132]</strong></td>
<td><strong>[67]</strong></td>
<td><strong>[132]</strong></td>
</tr>
<tr>
<td><strong>Guoguo</strong></td>
<td><strong>[95]</strong></td>
<td><strong>[132]</strong></td>
<td><strong>[67]</strong></td>
<td><strong>[132]</strong></td>
</tr>
<tr>
<td><strong>SpeakerSense</strong></td>
<td><strong>[95]</strong></td>
<td><strong>[132]</strong></td>
<td><strong>[67]</strong></td>
<td><strong>[132]</strong></td>
</tr>
<tr>
<td><strong>[106]</strong></td>
<td><strong>[132]</strong></td>
<td><strong>[67]</strong></td>
<td><strong>[132]</strong></td>
<td><strong>[67]</strong></td>
</tr>
</tbody>
</table>

**Method**

- histogams, multi-linear regression
- Mood (average 73%, generic 66%, personal 93%)

**Features**

- emails, SMS, calls, website domains, apps, locations, user input

**Target**

- Mood (average 73%, generic 66%, personal 93%)

**Features**

- Moodscope: Device usage, SMS, calls, location
- Orientation: Orientation sensor data
- uWave: Accelerometer
- Guoguo: Microphone, speakers
- SpeakerSense: External and internal microphones
<table>
<thead>
<tr>
<th>System</th>
<th>Sensing</th>
<th>Social Interaction</th>
<th>Behavioural Cues</th>
<th>Social Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>StressSense</strong> [58]</td>
<td>Microphone (microcone indoors)</td>
<td>Conversation</td>
<td>statistics pitch, spectral centroid, high frequency ratio, speaking rate, MFCCs, TEO-CB-AutoEnv</td>
<td>Auditory (Conversation)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GMM with diagonal covariance matrix, EM, MAP (adaptation)</td>
<td>Stress</td>
</tr>
<tr>
<td><strong>SoundSense</strong> [45]</td>
<td>Microphone</td>
<td>ZCR, Low energy frame rate, spectral flux, spectral rolloff, spectral centroid, bandwidth, normalised weighted phase deviation, relative spectral entropy, MFCC</td>
<td>Decision tree, Markov models, smoothing, Bayes, HMM smoothing</td>
<td>Auditory (Conversation), Social Context</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
<td>--------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Auditory (Conversation, turn-taking, utterances, influence), Physical activity</td>
<td>z-scored percentage of speaking time, z-scored influence on turn-taking (HMM), z-score of pitch and amplitude variation, z-score of short utterances</td>
</tr>
<tr>
<td>[52]</td>
<td>WiFi, Magnetometer, external accelerometer</td>
<td>Interpersonal distance, relative orientation, conversation</td>
<td>10-sample window WiFi RSSI, device azimuth, power spectral density of mean, maximal, minimal, and integral of 10 sec audio</td>
<td>Naïve Bayes with KDE, raw azimuth, noise cancellation, Naïve Bayes with KDE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Interpersonal distance (Proximity), Relative spatial arrangement, Auditory (conversation)</td>
<td>Stress, activity level, engagement, emphasis, mirroring</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phone calls</td>
<td>mean acceleration variation, sleep duration, Mean RR, std RR, RMS RR, 50ms difference RR intervals, HRV index, triangular interpolation, approximate entropy, coefficients of Poincari, LF, HF, LF/HF</td>
<td>Physical activity, Physiological (heart rate)</td>
</tr>
<tr>
<td>Microphone, Accelerometer, GPS, phone calls, address book, calendar, battery, Biomonitoring sensor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BeepBeep [131]</td>
<td>Microphone, speaker</td>
<td>peaks, sharpness of a peak, maximum peak</td>
<td>ToA combined with threshold</td>
<td>Interpersonal distance</td>
</tr>
<tr>
<td>Accelerometer, calls, ringer, GPS, microphone</td>
<td>call status, ringer status, ambient sound, location, accelerometer data</td>
<td>threshold based categorisation</td>
<td>Social context, Physical activity</td>
<td>Social context, Physical activity</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------</td>
<td>--------------------</td>
<td>---------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td><strong>[125]</strong></td>
<td>Accelerometer, camera</td>
<td></td>
<td>Haar-like features, rectangles, sum of pixel intensities, difference among adjacent rectangles</td>
<td>CAMSHIFT, Starburst algorithm, Haar Eye Detection</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Facial (eye-tracking)</td>
<td></td>
</tr>
<tr>
<td><strong>[56]</strong></td>
<td>Microphone</td>
<td></td>
<td>Energy, pitch, LPC coefficients, duration, pitch, jitter</td>
<td>Auditory</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Auditory</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Neural Networks, Fuzzy networks</td>
</tr>
<tr>
<td><strong>CONSORTS-S</strong></td>
<td>ECG, accelerometer, thermometer, hygrometers, microphone array</td>
<td></td>
<td>latest, maximum, minimum, average</td>
<td>Physiological (heart rate, skin temperature), Posture, Physical activity</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Features</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>Accelerometer, GPS, barometer</td>
<td>location clusters, transition clusters, rectified signal, intensity, cadence, number of steps, number of instance in climbing stairs</td>
<td>t-test</td>
<td>Physical activity, Social Context</td>
</tr>
<tr>
<td>135</td>
<td>Bluetooth, calls, SMS, questionnaire</td>
<td>Bluetooth discovery</td>
<td>interaction events, different contacts, interaction diversity, number of total purchases, colocation</td>
<td>Naïve Bayes</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Features</td>
<td>Target</td>
</tr>
<tr>
<td>[85]</td>
<td>Accelerometer</td>
<td></td>
<td>the highest magnitude frequency, magnitude of highest magnitude frequency, weighted mean of top-5 highest magnitude frequencies, weighted variance of the top-5 highest magnitude frequencies</td>
<td>C4.5 decision tree, with confidence threshold, with pre-classification</td>
</tr>
<tr>
<td>[137]</td>
<td>Bluetooth, calls, survey</td>
<td>Bluetooth discovery</td>
<td>centrality, efficiency, transitivity, triadic measures</td>
<td>Random forests</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Features</td>
<td>Target</td>
</tr>
<tr>
<td></td>
<td>RFID</td>
<td>Interpersonal</td>
<td>RF signal</td>
<td>Interpersonal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>distance with</td>
<td>interpretation</td>
<td>distance, partial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>partial relative</td>
<td></td>
<td>relative</td>
</tr>
<tr>
<td></td>
<td></td>
<td>orientation</td>
<td></td>
<td>spatial</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>arrangement</td>
</tr>
<tr>
<td>208</td>
<td></td>
<td></td>
<td>Mean RR, Std RR,</td>
<td>SVM, Bayesian</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean HR, Std HR</td>
<td>network,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSSD, pNN50,</td>
<td>decision tree</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LF, HF, LF/HF</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ratio, Mean</td>
<td>Physiological,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SCL, Std SCL,</td>
<td>Activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total magnitude,</td>
<td>Physiological,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Duration, and</td>
<td>activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Number of startle</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>responses, Mean,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>std, energy of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>XYZ axis,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Correlation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>coefficient of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>XY, YZ, and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ZX</td>
<td></td>
</tr>
<tr>
<td>155</td>
<td>Accelerometer, ECG, GSR</td>
<td>Mean RR, Std RR, Mean HR, Std HR RMSSD, pNN50, LF, HF, LF/HF ratio, Mean SCL, Std SCL, Total magnitude, Duration, and Number of startle responses, Mean, std, energy of XYZ axis, Correlation coefficient of XY, YZ, and ZX</td>
<td>SVM, Bayesian network, decision tree</td>
<td>Physiological, activity</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------</td>
<td>--------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>eMoto [161]</td>
<td>external stylus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pressure, shaking of stylus</td>
<td>Laboratory notation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Physical activity, gestures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Activity, gestures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Emotion (arousal, valence)</td>
</tr>
<tr>
<td>TagMobile [280]</td>
<td>RFID</td>
<td>Interpersonal distance and spatial arrangement</td>
<td>RF signal strength</td>
<td>Triangulation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Interpersonal distance and spatial arrangement</td>
</tr>
<tr>
<td>UbiSense [281]</td>
<td>Ultra-wideband</td>
<td>Interpersonal distance and spatial arrangement</td>
<td>Ultra-wideband signal strength</td>
<td>TDoA, ToA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Interpersonal distance and spatial arrangement</td>
</tr>
<tr>
<td>RF-Beep [123]</td>
<td>Microphone, speakers, WiFi</td>
<td>WiFi beacon frame, acoustic beacon signal</td>
<td>TDoA</td>
<td>Interpersonal distance</td>
</tr>
<tr>
<td></td>
<td>EyeContact sensor, microphone</td>
<td>Eye contact and speech detection</td>
<td>pupil detected from sensor, energy of the voice</td>
<td>Auditory (conversation, silence), Facial (eye-tracking)</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------------------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>[167]</td>
<td>Microphone</td>
<td></td>
<td>syllables, harmonicity, spectral centroid, skewness, kurtosis, shimmer etc. and statistics</td>
<td>Auditory (turn-taking, silences, laughter, utterances)</td>
</tr>
<tr>
<td>[84]</td>
<td>GPS, accelerometer, microphone, WiFi</td>
<td></td>
<td>energy, FFT, peaks</td>
<td>Social context, activity</td>
</tr>
<tr>
<td>[143]</td>
<td>Bluetooth</td>
<td>Bluetooth RSSI</td>
<td>triangulation</td>
<td>Interpersonal distance, spatial arrangement</td>
</tr>
<tr>
<td>[86]</td>
<td>Accelerometer</td>
<td>mean, energy, entropy, std of amplitude and correlation</td>
<td>Frame-based Descriptor multi-class SVM</td>
<td>Gesture</td>
</tr>
</tbody>
</table>

Personality traits (agreeableness, conscientiousness, extroversion, neuroticism, openness), Sociability.
<table>
<thead>
<tr>
<th>System</th>
<th>Sensing</th>
<th>Social Interaction</th>
<th>Behavioural Cues</th>
<th>Social Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Features</td>
<td></td>
</tr>
<tr>
<td>[66]</td>
<td></td>
<td></td>
<td>(relative) spectral entropy, (maximum) autocorrelation peaks log and value, energy</td>
<td>HMM, EM, MAP, Auditory (conversation, pitch, rate, turn-taking), Social context</td>
</tr>
<tr>
<td>Whistle [133]</td>
<td>Microphone, speaker</td>
<td>auto-correlation, peaks, cross-correlation, maximum peaks</td>
<td>TD2S, TDoA, triangulation</td>
<td>Interpersonal distance, spatial arrangement</td>
</tr>
<tr>
<td></td>
<td>[88]</td>
<td>Accelerometer</td>
<td>mean, variance, magnitude, covariance, energy, entropy, FFT coefficients</td>
<td>J48 adaptive decision tree</td>
</tr>
<tr>
<td>System</td>
<td>Sensing</td>
<td>Social Interaction</td>
<td>Behavioural Cues Features</td>
<td>Method</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>83</td>
<td>Accelerometer</td>
<td>Moving averaged window, mean, std, ZCR, 75% percentile, interquartile range, power spectrum centroid, entropy of the vertical and horizontal components, cross-correlation, amplitude of the vertical components, magnitude of the horizontal components</td>
<td>decision tree, k-means clustering, HMM-based Viterbi algorithm smoothing</td>
<td>Activity</td>
</tr>
<tr>
<td>System</td>
<td>Sensing Interaction</td>
<td>Social Interaction</td>
<td>Behavioural Cues</td>
<td>Social Behaviour</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Visage[113]</td>
<td>Camera, motion (accelerometer, orientation, gyroscope)</td>
<td>moving window on image, gravity direction, motion intensity, mean-variance on direction, eye corners, edges of mouths</td>
<td>Face detection (adaboost), Lucas-Kanade, CAMSHIFT, device posture, Pose from Orthography and Scaling with Iterations (POSIT), Active Appearance Models</td>
<td>Facial expressions, Head pose</td>
</tr>
</tbody>
</table>


[50] W. Hu, G. Cao, S. V. Krishnamurthy, and P. Mohapatra, “Mobility-assisted energy-aware user contact detection in mobile social networks,” 2013 IEEE 33rd


...
**Bibliography**


Bibliography


[161] P. Sundström, A. Ståhl, and K. Höök, “emoto: Affectively involving both body
and mind,” in *CHI ’05 Extended Abstracts on Human Factors in Computing

based mobile services and applications,” in *Proceedings of the 2Nd European Con-
ference on Smart Sensing and Context*, ser. EuroSSC’07. Berlin, Heidelberg:

sensor from smartphone usage patterns,” in *Proceeding of the 11th Annual Inter-

[164] L. R. Goldberg, “The structure of phenotypic personality traits.” *American Psy-

performance: A meta analysis,” *Personnel Psychology*, vol. 44, no. 1, pp. 1–26,

[166] G. Chittaranjan, J. Blom, and D. Gatica-Perez, “Who’s who with big-five: Analy-
zizing and classifying personality traits with smartphones,” in *Proceedings of the

“From nonverbal cues to perception: Personality and social attractiveness,” in *Proceedings of the 2011 International Conference on Cognitive Behavioural Sys-

[168] A. Bogomolov, B. Lepri, and F. Pianesi, “Happiness recognition from mobile
phone data,” in *Social Computing (SocialCom), 2013 International Conference


detection using bluetooth on mobile phones,” in Proceedings of the 2013 ACM
Conference on Pervasive and Ubiquitous Computing Adjunct Publication, ser.
UbiComp ’13 Adjunct, 2013, pp. 1135–1142.

 propagation model derived from fundamental principles: Analytical analysis and
 comparison with measurements,” Antennas and Propagation, IEEE Transactions

327.


emerging domain,” Image and Vision Computing, vol. 27, no. 12, pp. 1743–1759,
Nov. 2009.


