Robot-based Evaluation of Bluetooth Fingerprinting

UNIVERSITY OF CAMBRIDGE

Khuong An Nguyen

Queens' College

Computer Laboratory, University of Cambridge

A thesis submitted for the degree of

MPhil in Advanced Computer Science

15-June-2011

Abstract

Indoor tracking is a new research area, which is not yet fully solved. All current indoor tracking systems are either expensive or not accurate enough. In this project, a new affordable indoor tracking system was proposed and implemented for an office room. Through-out the research, a large scale experiment was conducted to verify the stability and the reliability of the system. It was observed that the Bluetooth signal is very stable with typical features of a radio wave signal, and is reliable to be implemented into an indoor positioning system. The fingerprinting method was employed to manipulate the Bluetooth signal at many positions in the office room. In addition, a robot was created to perform the complex and time-consuming data collection process. The system has the benefit of affordable Bluetooth technology and the accuracy of robotics.

Keywords: *indoor tracking, fingerprinting, Bluetooth, robotics.*

I, Khuong An Nguyen of Queens' College, being a candidate for the M.Phil in Advanced Computer Science, hereby declare that this report and the work described in it are my own work, unaided except as may be specified below, and that the report does not contain material that has already been used to any substantial extent for a comparable purpose.

Signed

Date 15-June-2011

This report contains 14,978 words, excluding bibliography, photographs and diagrams but including tables, footnotes, and appendices.

To my parents, for their love and support.

Acknowledgements

I profoundly thank Dr. Robert Harle for his immense day-by-day guidance, for his relentless supportive encouragement, patience, for sharing his insightful wisdom and financial arrangement throughout this research. Without him, this project would not have been a possibility. I also thank Brian Jones for his time in Linux and hardware support.

Finally, I wish to acknowledge the warm and friendly atmosphere within the Digital Technology Group during my six month researching here.

Contents

List of Figures

Chapter 1

Introduction

Indoor tracking is a new research area, where solutions and developments are still forthcoming. The demand for a system as effective as GPS for outdoor tracking has attracted many researchers in recent years. In this thesis, a new indoor tracking system was proposed and implemented for an office room. The system emphasises the affordability, maintainability and efficiency aspects, which are currently lacking in most tracking systems. Through-out the research, a large scale experiment was conducted to verify the stability and the reliability of the system. In addition, a robot was created to carry out the complex and time-consuming data collection process with great accuracy.

1.1 Indoor Positioning Systems

Outdoor positioning has long been a successful technology, thanks to the Global Positioning System (GPS) [6]. The indoor positioning counterpart however, has not been developed to the same scale.

Many attempts to apply the GPS technology into indoor positioning have not been successful [27][28], because the satellite signals cannot penetrate the building's infrastructure. Another difficulty which hinders the GPS deployment into indoor positioning is the 10 m accuracy. Although recent research has improved the accuracy up to 3 m [28], this is still not reasonable for an indoor tracking system, where the accuracy should be as high as possible, and preferably to within centimetres. Thus, there is a need for such an indoor positioning system.

One of the largest obstacles every indoor system faces, is the signal attenuation, caused by the building's small furniture such as tables and chairs. This reason simply opposes any attempt to calculate or to predict the indoor signal strength pattern directly. The Bat system [3][4][5], which was developed at Cambridge University, is the most accurate indoor tracking system. It uses ultrasonic pulses to measure the times-of-flight taken to reach the receivers mounted on the ceiling. However, the system is extremely expensive to install and maintain. Just a single corridor on the second floor of the Computer Lab already needs a minimum of 500 receivers. Another popular indoor tracking system is the RADAR system [7][8]. This is the first system to apply the Wi-Fi signals to estimate an indoor position. However, it is not accurate enough [29].

More recent approaches focus on a method called 'fingerprinting'. This is simply a database of locations, along with some unique traits to distinguish amongst positions in the database. The remaining task is to choose a suitable representative signal to make up the database, and to develop techniques to predict the optimal location from the data within this database. Many attempts have been formulated to apply the Wireless LAN signal for the 'fingerprinting' technique. Amongst those are the Horus system [30], the RADAR system [8] and the Nibble system [9]. These systems have an advantage over the sensor-based system as they utilise the in-built hardware of the building.

1.2 A Better Indoor Positioning System

Those systems above are, however, either very costly to maintain in the case of the Bat system, or inefficient when predicting indoor positions in the cases of RADAR, Nibble. Since indoor positioning is a new technology, there is virtually no complete literature comparison amongst those systems, as well as any standard to rate them. In general, there are three features that a practical indoor positioning system should follow

- **The performance** is the most critical feature to judge any system. Amongst many aspects of the system performance are accuracy, speed and scalability. The system accuracy measures the difference between the predicted position and the actual position. The system speed is the total running time, including the training time and the actual deployment time. The scalability measures how well the system performs with large scale data.

- **The cost and maintenance** determine how practically the system can be deployed in the market. The cost to install and apply should be reasonable, as well as the ease of maintenance over the long term.

Bearing in mind the above features, this project designs a new indoor tracking system with Bluetooth technology and the fingerprinting method. In particular, a robot was designed to maximise the accuracy and the efficiency of the fingerprinting process. This thesis will show that the system has big potential, with good performance, and an affordable cost to install and maintain, thanks to the nature of the Bluetooth devices. In particular, to prove the feasibility of such a system, this thesis addresses the following research questions:

How stable is the Bluetooth signal? Current fingerprinting systems favour the usage of Wireless LAN over Bluetooth technology, because the Bluetooth signal has a shorter range and a long discovery time. This thesis proves that the use of Bluetooth technology is suitable for indoor positioning.

How the Bluetooth signal deviates over long distances? An intuitive expectation would be the further the two devices are away from each other, the weaker the tracking signal between them becomes. However, by considering minor factors such as signal fading, attenuation and human influence, the changes in the Bluetooth signal over short and long distance are studied, as well as how to incorporate them into a practical indoor tracking system.

How the Bluetooth signal changes at different heights? For actual deployment, the Bluetooth tag is not always at the same height as the human body. Thus, the fingerprinting system should be able to cope with such change. This thesis investigates the changes of Bluetooth signals in terms of height.

How the position of Bluetooth device antennas affects the tracking signal? Many current fingerprinting systems only consider a single tracking direction. This is not true for Bluetooth signals, where even a 45 degree rotation would see a difference in signal strength. This thesis tracks down those changes and incorporates them into the system.

Further, the thesis looks at a method called 'histogram' and its benefits to the indoor positioning research, since *very little has been done to fully use the probability distribution in the research of indoor positioning*, according to [29].

1.3 Limitations of Scope

This thesis does not intend to cover an in-depth analysis of all Bluetooth properties, as well as all fingerprinting literatures. The implemented system is an attempt to deliver high quality tracking results, using affordable technology. In particular, the project is only concerned with the following

Bluetooth properties for indoor tracking: Although the Bluetooth technology standard does specify many properties, not all of them are useful in the indoor tracking context. This thesis only discusses the most influential properties and how to tweak them for optimal tracking results.

Fingerprinting algorithms: While there are many algorithms to infer a position with a fingerprinting database, this project only implements the most popular deterministic and probabilistic algorithms in the area. These algorithms have been proved to work efficiently with other fingerprinting systems, thus their results are comparable and verifiable to those in this thesis. Finally, although the histogram methods mentioned in this thesis have been widely applied in statistics, they have not been intensively researched in the indoor positioning area.

1.4 Thesis Outline

This thesis is divided into seven chapters as follows:

Chapter 1 outlines the development of the Indoor Positioning area, including an overview of current indoor positioning systems, which leads to the motivation for a better indoor positioning system, as well as the limitations of the thesis's implemented system.

Chapter 2 covers the background materials of this project, including the deterministic Weighted K-nearest neighbours, the probabilistic Naive Bayesian classifier, and the histogram Kolmogorov-Smirnov, Student's T-Test, as well as a discussion of relevant Bluetooth properties.

Chapter 3 explores the Bluetooth properties for an indoor positioning system in detail, and determines whether Bluetooth technology would be a suitable candidate for such a system.

Chapter 4 describes the data collection process, which includes all the software and hardware, as well as the strategies to collect Bluetooth signals with a robot.

Chapter 5 discusses the fingerprinting idea, and three algorithms to manipulate the database collected in the previous chapter, along with their performances and evaluations.

Chapter 6 verifies how well the implemented system performs in a comparable manner to other current fingerprinting systems, and addresses several issues, which need to be fixed to make the system widely recognised.

Chapter 7 concludes the thesis with a summary of the contributions, a self-evaluation of how well the specified goals are achieved, the empirical problems throughout developing the system, and an outline of the scope for future work.

The logical progression of the project is graphically depicted in figure 1.1.

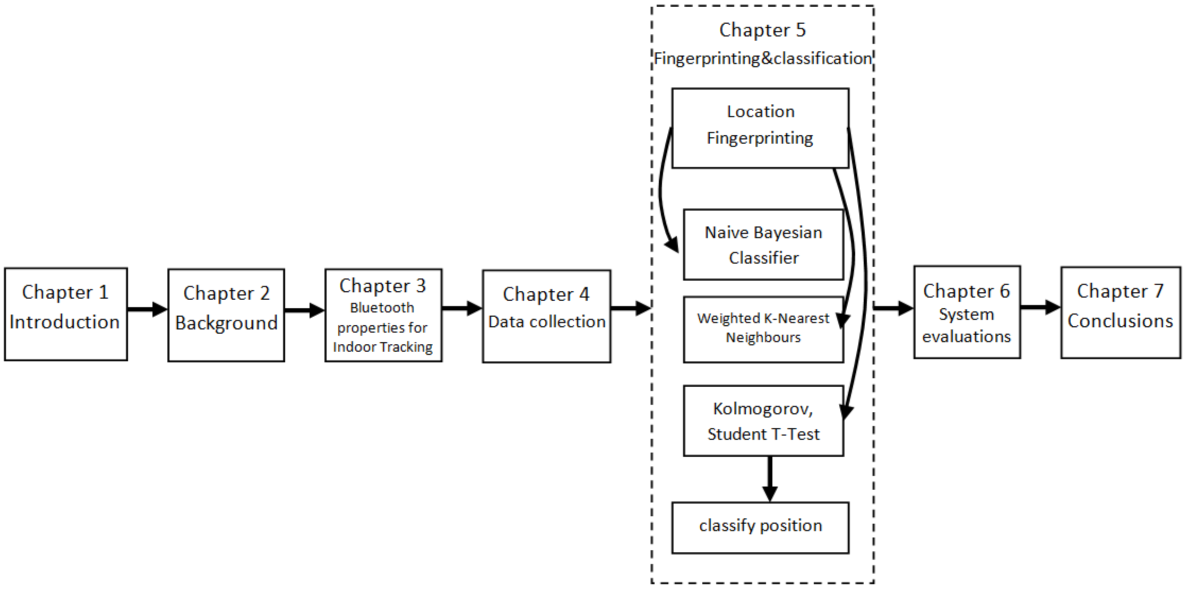

Figure 1.1: Report progression

Chapter 2

Deterministic, Probabilistic Algorithm, Histogram Method and Bluetooth Properties

Location tracking is one of the research areas, which does not yet provide both an affordable solution and an accurate performance. All positioning algorithms, which use a method called fingerprinting, share the same goal of attempting to predict an optimal tracking position, which is closest to the actual position. Although their approaches to solving the problem are different, they all refer to a training database. The algorithms' performance, thus, vastly depends on how good the training data is. Amongst the classic algorithms are the nearest neighbours algorithm, the Bayesian network, the Kolmogorov-Smirnov, and the Student's T-Test.

This chapter covers the background materials used in this project, including the general idea of location fingerprinting, the deterministic algorithm, the probabilistic algorithm and the histogram methods. However, the chapter is not concerned with the applications of these algorithms in the indoor positioning context, which will be discussed later in chapter 5. The chapter concludes with the introduction of the Bluetooth properties related to the location tracking context.

2.1 Location Fingerprinting

The idea of location fingerprinting is to perform a real-time survey of many small positions in a pre-defined space such as an office room, in order that the position's unique characteristics are captured and recorded into a database. Further on, the characteristics of an unknown position are compared against those in the database to estimate a closest co-ordinate for this unknown position. In this project, those characteristics are recorded in the form of the Bluetooth signal strength (RSSI). This section summarises two phases of the fingerprinting procedure: the online stage - where the database is created, and the offline stage - where an unknown position is classified. More details of the fingerprinting process are discussed in the Data Collection chapter.

2.1.1 Online Stage

During the online stage, a fingerprinting database is created, which uniquely maps many physical positions represented by the combination of Bluetooth signal strengths to the Bat position. The Bat position is identified as a (x, y, z) co-ordinate tuple provided by the Bat system. So, given a particular signal strength pattern of an unknown location, the task is simply to estimate this unknown position using the fingerprinting database. To minimise the hassle of the data collection process in this stage, a robot was designed to carry a laptop and record the Bluetooth signal and the Bat data automatically.

2.1.2 Offline Stage

During the offline stage, the system estimates an optimal position using the fingerprinting database created in the online stage. The process involves first recording the current signal strength combination of all nearby Bluetooth base stations, then classifying this combination sample using the fingerprinting database. In general, the accuracy of the estimated position depends on the quality of the fingerprinting database and the effectiveness of the classifier algorithms, as will be discussed later in the thesis.

2.2 Deterministic Algorithms

The deterministic algorithm is usually simple and easy to implement. Given a set of points S, in any dimensional space, and a single point p in the same dimensional space, the algorithm selects the best point t ∈ S, which is closest to the given point p. This section discusses three typical deterministic algorithms in a developing fashion (figure 2.1).

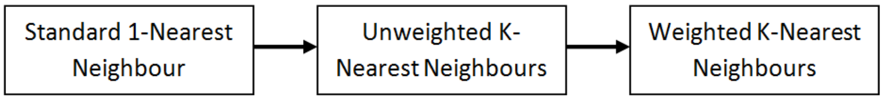

Figure 2.1: Deterministic algorithms development

2.2.1 Nearest Neighbour

The simplest deterministic algorithm to solve this problem is the Nearest Neighbour algorithm. It simply calculates the Euclidean distance between the given point p and every single point t ∈ S, as formulated in 2.1. The point t with the smallest distance is the answer. Figure 2.1 shows an example of the algorithm.

$$Distance(p,t) = \sqrt{(x_p - x_t)^2 + (y_p - y_t)^2} \tag{2.1}$$

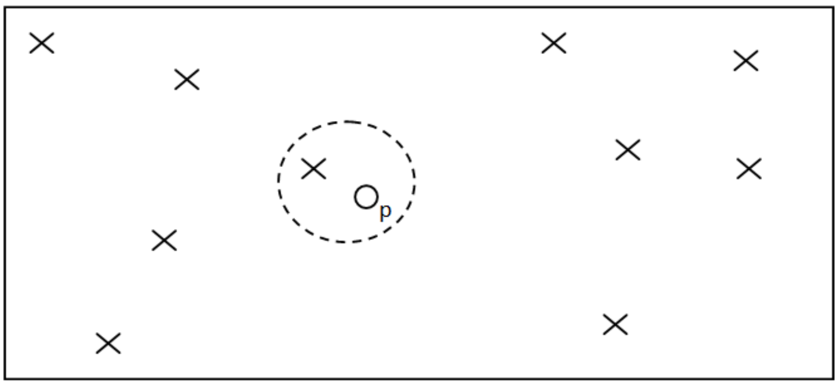

Figure 2.2: 1-Nearest Neighbour

2.2.2 K-Nearest Neighbours

The weakness of the standard Nearest Neighbour approach is the confusion when there are more than one point t $\in$ S with a similar smallest distance to p, since the algorithm could pick any of them. Thus, the K-Nearest Neighbours algorithm is introduced. Instead of picking a single closest point, the algorithm considers K nearest ones. The estimated position is the average of these K points' coordinates. This approach makes more sense as there are usually many points with roughly similar distance to the given point p in a highly dense area. Further, a group of points would represent the estimated position better than just a single one. The only parameter to consider in this algorithm is the number of neighbours K. When K = 1, the task becomes finding a single nearest neighbour as described before. An intuitive thought would be a 'reasonable' number of points around p, because if K is too large, the space in which the algorithm covers also contains irrelevant points far away from p, which results in a decrease in the estimation accuracy. Figure 2.3 demonstrates a 3-Nearest Neighbours case.

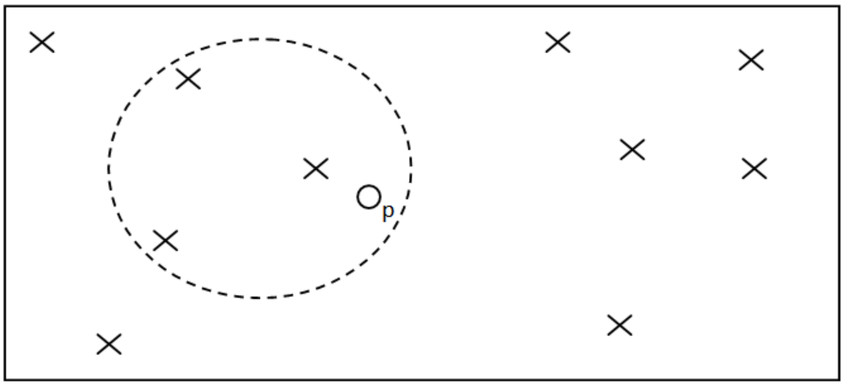

Figure 2.3: 3-Nearest Neighbours

2.2.3 Weighted K-Nearest Neighbours

The above K-Nearest Neighbours algorithm can be further enhanced by giving a weight to each of the K neighbours. This weight is typically the inverse Euclidean distance from that neighbour to the point p. The inverse distance would prioritise those neighbours which are closer to p, over neighbours which are further away. This is a major improvement since an un-weighted K-Nearest Neighbours algorithm simply averages the co-ordinates of all K neighbours. If K is large, the accumulative error caused by points further away would degrade the overall estimated position. Using this algorithm, formula 2.2 calculates the estimated position e, given the Euclidean distance $dist(t_i, p)$ between each of the K neighbours and p - the point which needs to be classified. e_x and p_x are a single co-ordinate in the dimensional space of the estimated position and point p respectively. This formula is repeated N times for each separate co-ordinate, if the space has N dimensions.

$$e_x = \frac{\sum_{i=1}^{K} \frac{1}{dist(t_i, p)} \; p_x}{\sum_{i=1}^{K} \frac{1}{dist(t_i, p)}} \tag{2.2}$$

To demonstrate the efficiency of the weighted points, figure 2.4 shows a simple example with 5 points. The points B, C and D have the same distance to p, while E is further away. Using an un-weighted 4-Nearest Neighbours approach returns an estimated point (1.5, 2.25) which has been dragged over E, while weighted 4-Nearest Neighbours returns (1.99, 1.99) which is concentrated around 3 closest points B, C and D.

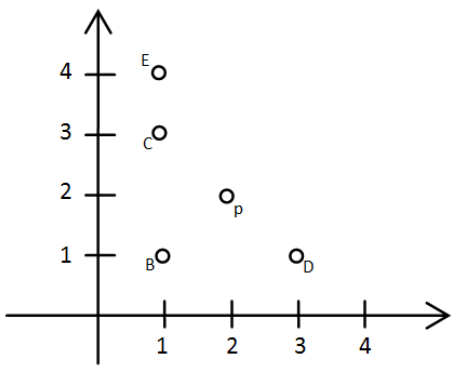

Figure 2.4: Weighted 4-Nearest Neighbours

2.3 Probabilistic Algorithm

A probabilistic algorithm is an algorithm in which the result is obtained from the probability distribution. The probabilistic approach models the data with the conditional probability and uses the Bayesian technique to estimate the correct position. The section first introduces the definition of Bayesian network, then uses this idea to implement the Bayesian classifier. For practical use, a special form of Bayesian classifier known as the Naive Bayesian classifier is introduced, which concludes the section.

2.3.1 Bayesian Network

A Bayesian network is a directed acyclic graph, which represents a set of nodes and their dependencies. A directed edge from node X to node Y (figure 2.5), means Y conditionally depends on X.

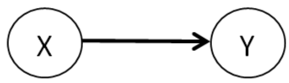

Figure 2.5: Conditional dependence

A node N is said to be conditionally independent of node M if N does not directly connect with M. Each node X is assigned with a probability table, which specifies the distribution over X, given the value of X's parent. Given a classification task, the Bayesian network can be applied to predict the decision outcome. The standard formula of the Bayesian classifier is

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \tag{2.3}$$

P(X | Y) stands for the probability of the event X, given the event Y.

2.3.2 Naive Bayesian Classification

Naive Bayesian classifier is simply the Bayesian classifier with the dependency assumption relaxed. In particular, Naive Bayesian assumes that the presence or absence of any node in the Bayesian network does not affect any other nodes. For example, considering 'wet grass' and 'cloudy' as two nodes of the network, although they both contribute to the probability of the 'raining' event, the existence of the 'wet grass' event does not affect the existence of the 'cloudy' event and vice-versa.

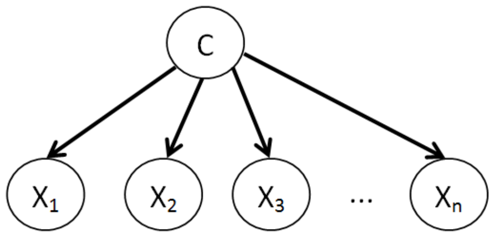

Figure 2.6: Conditional independence

The biggest advantage of Naive Bayesian is the computational overhead reduction, in order to estimate a probability. The quantity P(X | Y) in the formulae 2.3 is often impractical to calculate directly, without any independent assumptions. Since each node $x_1, x_2, \ldots x_n$ is conditionally independent of each other, given a common class C, as depicted in figure 2.6, its probability can be calculated separately, and the combination of separate probabilities of each node can be combined to yield an overall probability of the big event C. The general formula of Naive Bayesian in terms of each separate node can be calculated as:

$$P(X|C) = P(x_1|C)P(x_2|C)\ldots P(x_n|C) \tag{2.4}$$

2.4 Histogram Method

The histogram method looks at the total distribution of two sequences of numbers and derives the difference between them directly. The size of the two sequences does not have to be equal.

2.4.1 Histogram Table

Each sequence is also known as a histogram. Figure 2.7 demonstrates an example of the histogram sequence (-55,-55,-54,-54,-54,-53,-53,-53,-53,-53,-52,-52)

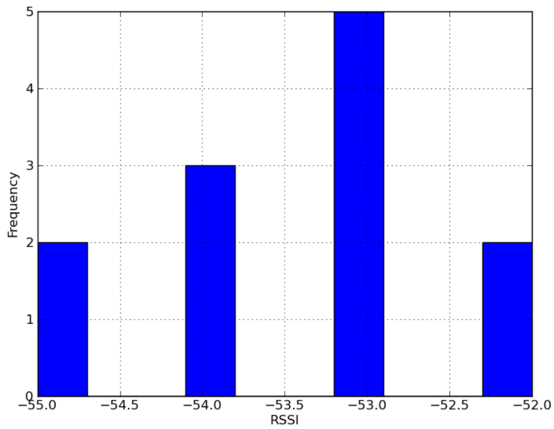

Figure 2.7: Histogram sequence example

Based on this frequency distribution, the difference between a pair of histograms can be measured. This thesis applies two methods, the Kolmogorov-Smirnov (K-S Test) and the Student's T-Test.

2.4.2 Student's T-Test

The T-Test is arguably one of the most popular methods to compare two histograms. It takes any two histograms and returns a t-value representing the difference between them. This method, however, makes two assumptions: that the histogram is normally distributed, and the two histograms must be independent. The formula to calculate the t-value is as follows

$$t = \frac{|\bar{X_1} - \bar{X_2}|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{2.5}$$

where $\bar{X_1}$ and $\bar{X_2}$ are the means of the two histograms, and n_1, n_2 are the size of two histograms. The variance of each histogram s_1^2 and s_2^2 is calculated as followed, with n as the size of histogram and X as the mean of histogram.

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1} \tag{2.6}$$

The t-value calculated above needs to be checked before it can be accepted, because the variance of the two histograms might be completely irrelevant. To verify the correctness of t-value, a p-value and a degree-of-freedom are employed. The p-value is simply the percentage from which the variance is accepted, such as for p = 0.05, the assumption is that only at 5% of the time the two histograms are completely random. In other words, the confidence is 95% that the two histograms are the same. However, there is still 5% this could be wrong. This p-value is also known as the level of significance. The degree-of-freedom is calculated in formula 2.7. It is simply the number of values in both samples that can vary. The calculation leaves out one fixed sample in each histogram, whilst the rest can vary.

$$N = (n_1 + n_2) - 2 \tag{2.7}$$

After selecting the p-value and calculating the degree-of-freedom N, a t-distribution table (appendix) is consulted. This table lists all reference t-values under a certain degree-of-freedom N and a p-value. Using the reference t-value from the table, the calculated t-value above is rejected if it is greater than this reference value.

2.4.3 Kolmogorov-Smirnov Test

Similar to the Student's T-Test, the Kolmogorov-Smirnov Test (K-S Test) also measures the difference between two histograms by calculating a d-value. It also assumes that the two histograms are independent. However, the major difference is that it makes no assumption about the normal distribution feature. This comes at one cost that the Student's T-Test is more accurate when the histogram is normally distributed [1][2]. Besides, the K-S Test reports if the histogram seems to have a normal distribution or lognormal distribution, which is an advantage to decide if the Student's T-Test can be applied instead.

For each value s in the whole combination of the two histograms, K-S Test first calculates the number of values $F_1(s)$ and $F_2(s)$ in the two histograms, which are less than or equal to s, n_1 and n_2 are the sizes of the two histograms

$$F_1(s) = \frac{Number\ of\ values\ x_i \leq s}{n_1} \tag{2.8}$$

$$F_2(s) = \frac{Number\ of\ values\ x_i \leq s}{n_2} \tag{2.9}$$

The d-value is calculated as followed, with sup(S) function returns the least element, which is greater than or equal to every value s.

$$d = sup|F_1(s) - F_2(s)| \tag{2.10}$$

This d-value is compared to a reference d-value from the d-distribution table (appendix), as similar to the Student's T-Test.

2.5 Bluetooth Features

Bluetooth is a popular piece of technology and can be found in countless devices these days such as mobile phones, laptops, headsets. This part covers the most influential Bluetooth features, which are relevant to the indoor tracking context.

2.5.1 Bluetooth Technology

Bluetooth is a means to wirelessly transfer information over short distances, using a short wavelength radio signal (figure 2.8). Every Bluetooth device contains a unique 48-bit address to identify itself, such as 00-19-0E-06-CF-24.

Figure 2.8: Official Bluetooth Trademark

The Bluetooth devices are divided into three classes, as described in table 2.1. Class 1 devices have the longest range, but consume at the highest power level. Class 2 devices are the most widely used dongles, and can be found in all mobile phones and laptops. Class 3 devices, however, are obsolete and are not manufactured any more.

Table 2.1: Bluetooth classes comparsion

	Working range	Power consumption
Class 1	100 metres	100 mW
Class 2	10 metres	2.5 mW
Class 3	1 metre	1 mW

2.5.2 Bluetooth Connection

To discover a target dongle, the inquiry dongle broadcasts a 'discovery message', then waits for a reply. Upon receiving the 'discovery message', the listening dongle responds with the necessary information to set up a connection. The listening dongle has two options to decide whether to respond to a 'discovery message'. The 'Inquiry Scan' option controls the dongle's discoverability. The 'Page Scan' option controls the dongle's connectivity. If the 'Inquiry Scan' is switched on, the dongle can be discovered. If the 'Page Scan' option is on, the dongle will accept any connection request from other dongles.

2.5.3 Frequency Hopping

All Bluetooth devices operate in the 2.4 Ghz frequency band, the same band as the Wi-Fi 802.11. However, Bluetooth technology divides the 2.4 Ghz band into 79 channels and employs a technique called channel hopping. This approach minimises the interference amongst the Bluetooth dongles, by relentlessly switching to different channels 1,600 times every second. It is worth noting that Wi-Fi 802.11 also divides the 2.4 Ghz band into 14 channels (fewer than Bluetooth), but a Wi-Fi network will pick a particular channel and stick with it through-out the session. Each Wi-Fi channel is 5 Mhz wide, while each Bluetooth channel is only 1 Mhz wide.

When two Bluetooth devices are communicating, they will both switch to the same channel at the same time. If two Bluetooth devices which are not communicating directly, happen to be on the same channel at the same time, the packet will be lost in the collision. This packet needs to be re-transmitted later. Since Bluetooth 1.2 version, the Bluetooth device tries to avoid a high-interference channel when performing hopping.

Chapter 3

Bluetooth Properties For Indoor Tracking

Bluetooth technology is widely integrated into many ubiquitous devices such as mobile phones, laptops,... which strongly benefit many Bluetooth-applied applications. In this project, Bluetooth signals were used to implement an indoor tracking system. To achieve this purpose, this chapter delves into relevant Bluetooth properties, which are important to the indoor tracking context, to verify that Bluetooth technology is a suitable back-bone infrastructure. The chapter opts for an experimental-approach, in which the Bluetooth properties were confirmed through many practical experiments. They were either natural experiments, where the environment was left as usual, or artificial experiments, where the conditions of the environment were tweaked to produce different noises and effects.

The chapter has four sections corresponding to four research questions. First, the Bluetooth signal's robustness is experimented with. Second, the change of signal over long distances is investigated. Third, the effect of the device's height is considered. Finally, the influence of the device's antenna to the signal strength is inspected. The last section concludes on the suitability of Bluetooth signals for deployment with an indoor tracking system.

3.1 How Stable is the Bluetooth Signal?

The first experiment researches the robustness of the Bluetooth signal between two fixed locations. The idea of this experiment was to set up two computers in two steady positions, preferably on the floor or on the table, opposite and parallel to each other. Then the Bluetooth signal strength between them was observed over a long time. Even when two Bluetooth devices were within range, there was no guarantee that one device can discover the other in less than 10.24 seconds from one full scan. Thus, the Bluetooth readings between the two computers were averaged every 10.24 seconds. When the two computers were 30 cm away, scaled by a ruler with +/- 1 mm error, an observation of the Bluetooth signals over 24 hours showed that the signal variation was less than 3 dBm over the complete duration with the Belkin Class 2 Bluetooth dongle (figure 3.1). It was surprising to see that the signal had a strange shape over the first five hours, after that the signal became relatively stable with just 1 dBm variation. Since there was no logical explanation in the Bluetooth document about this change, a best guess would be this is the time needed for the device to settle down into a stable state. However, this settling interval was within the 3 dBm range stated above.

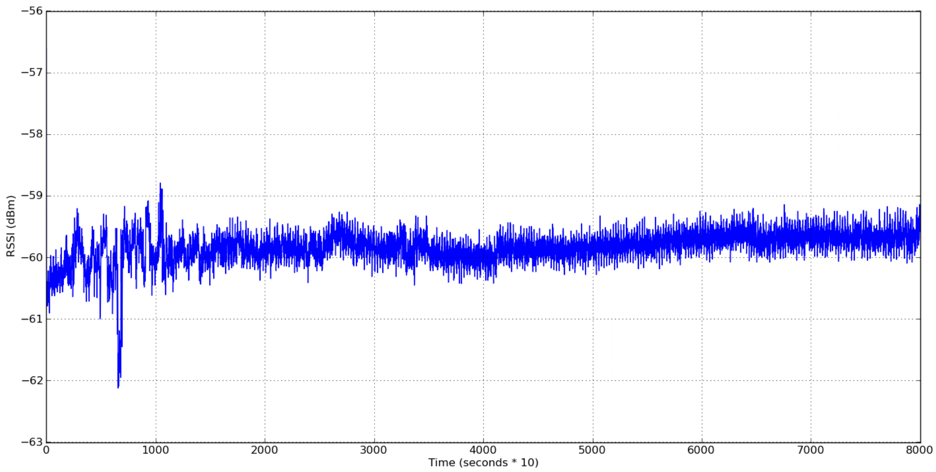

Figure 3.1: Bluetooth RSSI at 30 cm from 11am-8th-March to 10am-9th-March with Belkin dongle

Further, to have a closer look at each separate RSSI, figure 3.2 graphs 510,210 individual Bluetooth RSSI collected in the same period of time above with the same Belkin dongles. It was very interesting to observe that there was no isolated RSSI through-out the whole duration. The whole Bluetooth histogram could be as high as 30 dBm and were not totally normalised. This result was also considerably higher than the 10 dBm range reported with the Wi-Fi signal [30].

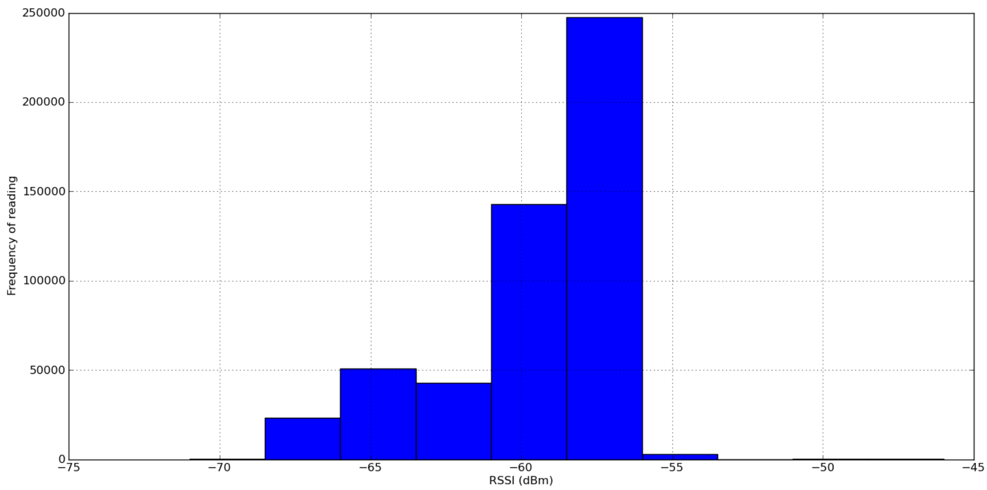

Figure 3.2: Individual RSSI in 30 cm from 7am-2nd-April to 8am-3rd-April

However, one session alone would not be sufficient to confirm the stability of the Bluetooth signal. Thus, the experiment's parameters were varied, including the environment's conditions which were changed as follows:

- **The distance** between the two computers was shifted to 60 cm and 1 m.
- **Different computers** were swapped around to make sure that the Bluetooth signals were not jeopardised by a broken computer.
- **Different Bluetooth dongles from the same company** were replaced to make sure the signals were not affected by a broken dongle.
- **Different dongles from different companies** were experimented with.

- **Different locations** were selected to set up the experiment.

- **Different days** were chosen to perform the experiments to confirm that the Bluetooth signals were not affected by the weather, sun, and humidity.

Each above condition was experimented separately as described below. First, when the distance between the two computers was increased to 60 cm and 1 m (figure 3.3), a similar period of up to a five hour interval was needed to stabilise, as well as the signal variation being well within the 3 dBm range, with the Belkin Bluetooth dongle. This result showed that the signal robustness can still be maintained over different distances.

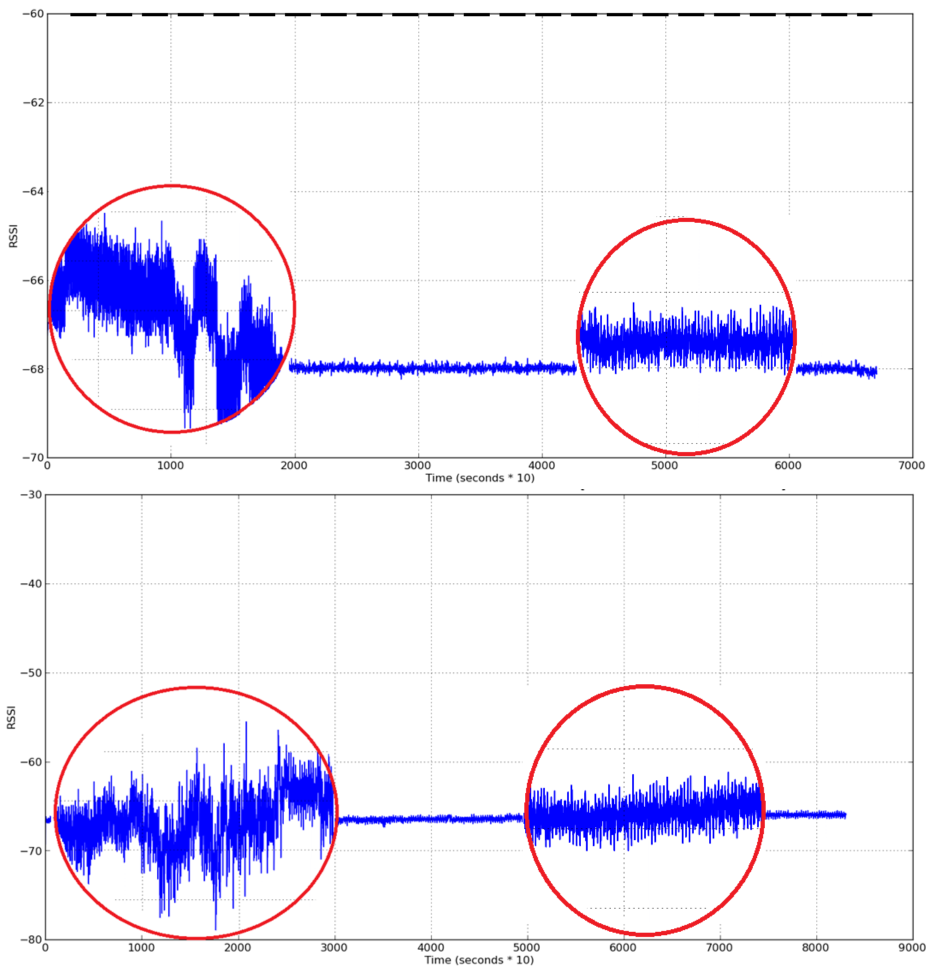

Figure 3.3: Bluetooth signals at 60 cm and 1 m

To make sure that the above results were not jeopardised by hardware problems, two new computers were swapped, and two new Belkin Bluetooth dongles purchased in the same batch from the same company were replaced to confirm the result's reliability. Although it cannot be guaranteed that the two dongles would be totally identical, and the two computers' configuration being exactly the same, figure 3.4 shows an un-surprisingly similar 3 dBm variation and up to a five hour settling time.

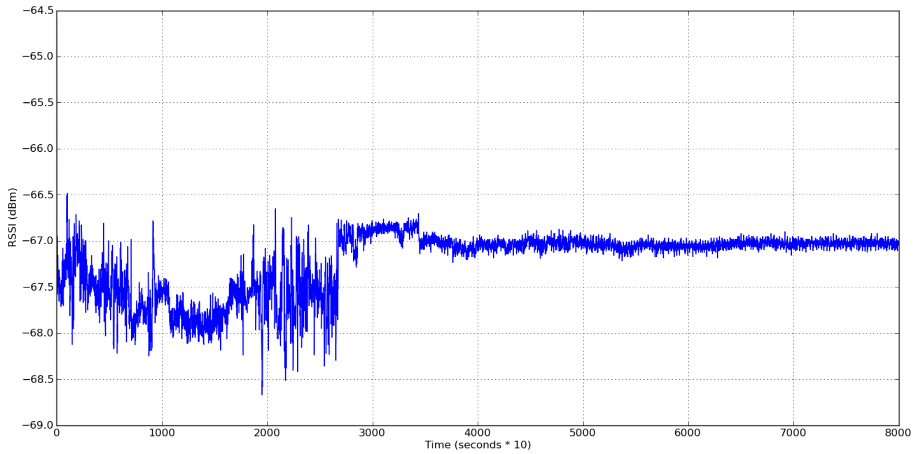

Figure 3.4: Bluetooth signals at 1 m with new dongles and new computers

Further, to guarantee that the above reports were not just correct for a particular Belkin dongle, three additional Bluetooth dongles from three different companies with the highest reviews on Amazon UK have also been tested. The Nexxus and Bluenext are Class 2 dongles, while the Bluewalker is a Class 1 dongle. Figure 3.5 shows the Bluetooth signal strength of the three dongles at 1 m over 24 hours. The experiment was carried out over 3 continuous days during the weekend, when there were no people around. In particular, even in the same position with the same computers, the three dongles' average RSSI in the stable state were slightly varied at -68 dBm, -68.3 dBm and -67.5 dBm for the Nexxus, Bluenext and Bluewalker dongles respectively. The Bluewalker dongle required nearly 8 hours to stabilise, while the Bluenext dongle took 5.5 hours and the Nexxus dongle only needed 4 hours. Overall, because of these differences, it is very important to use the same dongle for every environment and every experiment. Despite the difference between the three dongles' hardware, it was distinguishable that the signal variation was still within the 3 dBm range reported with the Belkin dongle previously, as well as 4 to 5 hours were still required to stabilise as expected, which cancels out the manufacturer's issue.

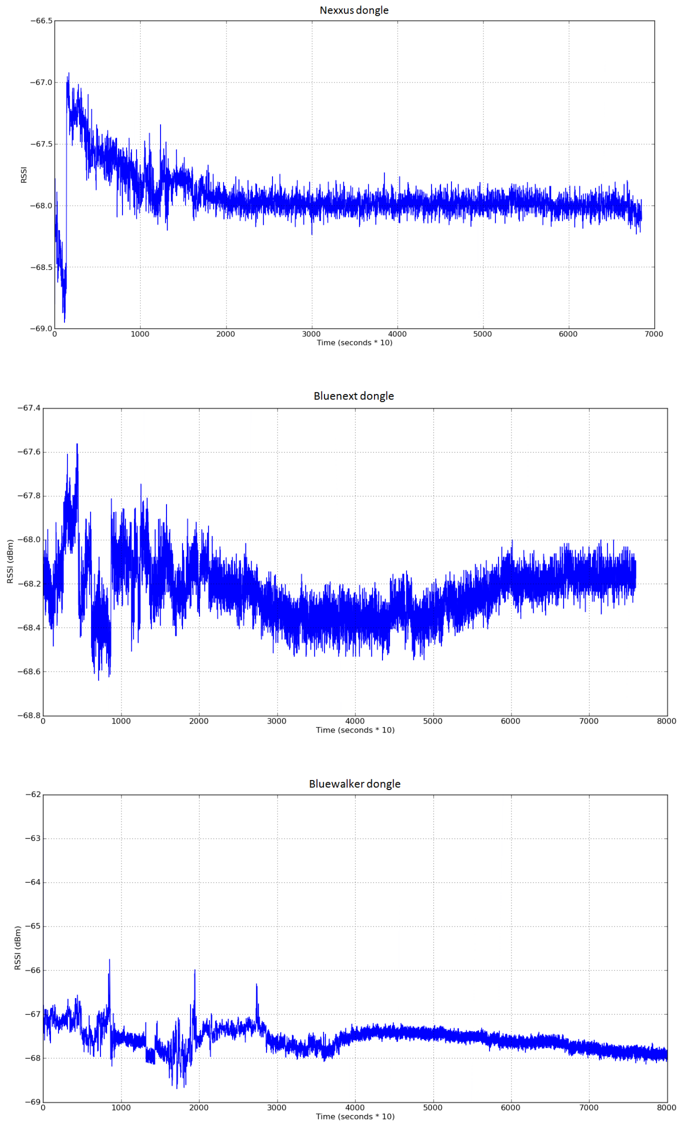

Figure 3.5: Bluetooth RSSI at 1 m with Nexxus, Bluenext & Bluewalker dongles

To test the reliability of the Bluetooth signal in different environments, the two computers in the above experiment were moved to several different locations in the Computer Lab. Figure 3.6 shows a similar expected result with 3 dBm variation and up to 5 hours to stabilise, when the experiment was moved to another room.

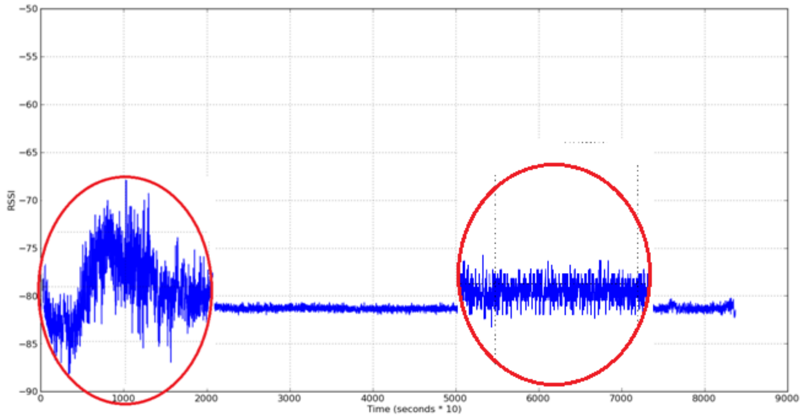

Figure 3.6: Bluetooth signals in a different room

Up till now, the experiment was performed over a wide area, with very little interference from the surroundings. The next two tests experience the Bluetooth signal in two extreme conditions with the Belkin dongle. First, an artificially dense Wi-Fi environment was created with four computers placed at four corners in the same room continuously accessing the Wi-Fi signals by pinging a website. Figure 3.7 identifies the result with the Wi-Fi signal floating around, and in the same position when the four computers no longer accessed the Wi-Fi signals during the weekend, when there were no people around. Since the Bluetooth technology applies the frequency hopping technique to communicate between devices as discussed in the background chapter, the two results looked very similar. It was noticeable that the signal variation was just 0.5 dBm when there were no Wi-Fi signals and was slightly higher at 0.75 dBm when the Wi-Fi signals were around. Both cases required around 5 hours to stabilise. After that, the signal seemed to be more stable when the Wi-Fi signals were not around at +/-

0.15 dBm variation, compared to +/- 0.25 dBm variation when the Wi-Fi signals were around. Overall, thanks to the frequency hopping technique, the Bluetooth signal was virtually immune to other radio signals in the air.

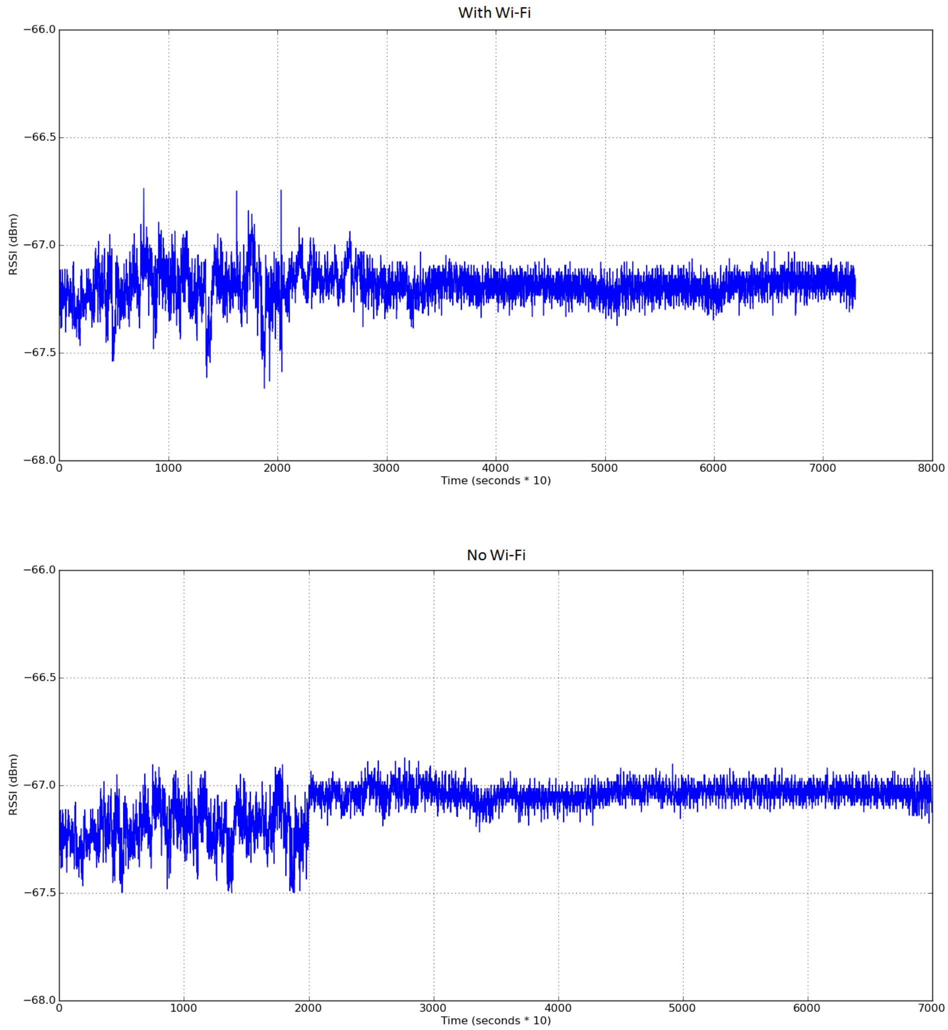

Figure 3.7: Bluetooth signals at 30 cm with Wi-Fi and without Wi-Fi

In addition, to simulate a busy environment during office hours, the experiment was set up near the front door. During the daytime, there were many people walking in and out at different times, and people returned home around 5 pm to 6 pm. An interesting finding was the fact that the Bluetooth signal struggles to stabilise with so much change happening around in the environment. Figure 3.8 demonstrates the findings with two Belkin dongles. It was interesting to observe that the signal cannot stabilise within 9 hours from 8 am in the morning until 5 pm in the afternoon. When people had already gone home around 5 pm to 6 pm, the signal took another 4 hours to get into the stable state around 9 pm, which results in a flat stream from 9 pm to 9 am the following day, when people returned back to work. The signal variation seemed to be much larger too. However, a closer look confirms that most signals were well within the 3 dBm range, with occasional signals popping out of the range.

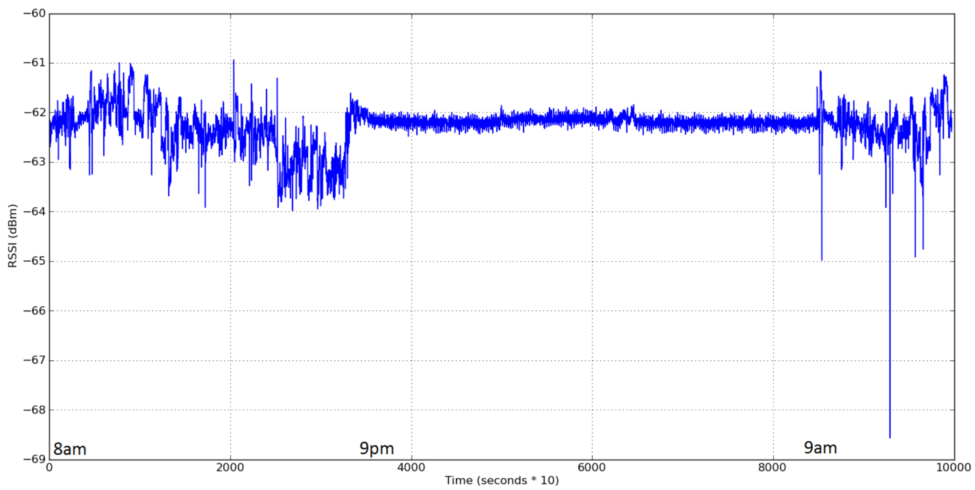

Figure 3.8: Bluetooth signals at 1 m during office hours with Belkin dongles

Finally, all the experiments mentioned above were repeated over different days within a six month period from January-2011 until June-2011, to confirm the robustness of the Bluetooth signal despite the weather effects such as sun, humidity, ... The performances were well-expected as the above results imply. In general, the experiments in this part showed that the Bluetooth signal is stable enough to be used as location identification. The frequency hopping technique is a unique feature of Bluetooth technology, which minimises the interference with other radio signal sources and helps to maintain the robustness of the Bluetooth signal.

3.2 How the Bluetooth Signal Changes Over Long Distances?

In an ideal world, an intuitive expectation would be the further the two devices are away from each other, the weaker the Bluetooth signal between them becomes. To verify this theory, two different environments were chosen to set up the experiment. The large ground in front of the Computer Lab can be considered as an ideal space with no objects or signals interferring in the early morning (figure 3.9). The DTG corridor on the second floor of the lab was the second experimental environment, with a small width and other potential signal interference (figure 3.10).

Figure 3.9: The large ground in front of the Computer Lab

In both environments, a computer was put at a fixed position, and a robot was located near this computer. This robot carries a laptop to measure the Bluetooth signal strength while moving perpendicularly further away from the computer. The robot stops every 10 cm to measure 50 Bluetooth signal strength readings (RSSI). More details about the robot constructions and the experiment scales are discussed in the next chapters. Figure 3.11 demonstrates the Bluetooth signal observation over 12 m in the DTG corridor. The first impression was although

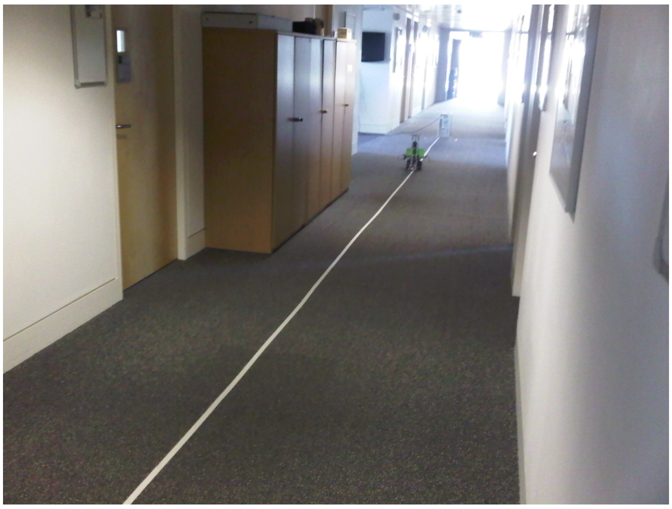

Figure 3.10: The DTG corridor

a 10 m Belkin Class 2 dongle was used, the robot can still catch the Bluetooth signal at a 12 m range. In the same corridor, the Bluetooth dongle's discoverable distance was experimentally found to be up to 18 m, which is almost twice the working range specified in the documentation. Overall, the Bluetooth signal strength (RSSI) decreases quadratically with the distance between the robot and the computer, according to Friis' free space transmission equation 3.1. This is called the free-space loss because the Bluetooth signal loses power as it moves through the air in the form of radio waves.

$$P_{RX} \quad = \quad P_{TX} \; G_{TX} \; G_{RX} \; (\frac{\lambda}{4\Pi d})^2 \tag{3.1}$$

with P_{TX} = Transmission power of sender
P_{RX} = Remaining power of wave at receiver
G_{TX} = Gain of transmitter
G_{RX} = Gain of receiver
λ = Wave length
d = Distance between sender and receiver

In addition to the free-space loss, the Bluetooth signal also suffered from multipath fading over long distances. This fading occurred as the signal was easily

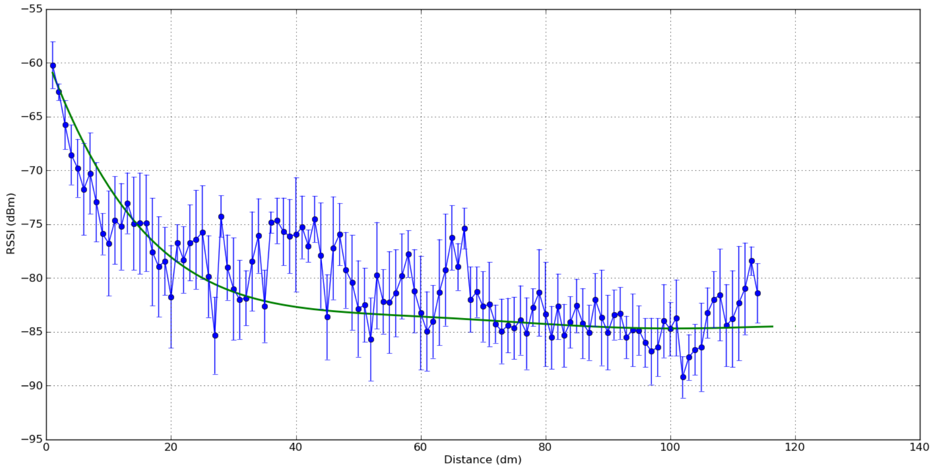

Figure 3.11: Bluetooth signals vs. Distance in the DTG corridor

interfered with by many influential factors such as the reflection from metal objects, polarisation of the electro-magnetic fields, diffraction around sharp corners, as well as scattering off the wall, floor or ceiling, which result in the Bluetooth signal reaching the destination from more than one path. Further, there were always minor noises in the surroundings such as humidity, temperature, etc... which affected the Bluetooth signal.

The green line in the above graph demonstrates the overall trend of the signal in a theoretical situation. The Bluetooth signal, however, becomes relatively flat after 3 m, which can be explained by the fact that the original Bluetooth signal was lost within all the noises after this point. At a distance of 3 m and more, for any two positions with less than 20 cm in between, the difference of the RSSI becomes as small as 1.5 dBm. However, the difference can reach 5 dBm for two positions at a distance of less than 3 m.

Next, considering the second environment at the large ground in front of the Computer Lab, the Bluetooth signal also resembles a quadratic trend (figure 3.12). However, the signal shapes up more nicely and is very close to the ideal green line, as there was much less interference, in comparison with the previous experiment.

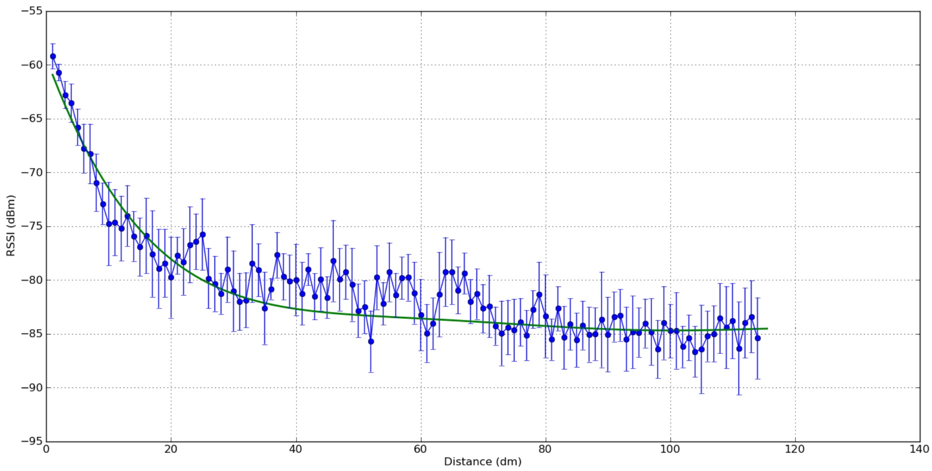

Figure 3.12: Bluetooth signals vs. Distance in the large ground

The signal variation was noticeably smaller than in the DTG corridor too, because this is a wide area with no obstacles, so that the Bluetooth signal can easily reach the destination in a quick unobstructed straight line. The discovery distance was experimentally found to be up to 21 m, which was also longer than the 18 m distance in the DTG corridor. This result can be explained similarly, since the signal did not become weaker and weaker by bouncing off the obstacles, thus it could reach a longer distance.

3.3 How the Bluetooth Signal Changes at Different Heights?

The weakness of the robot design was that the highest height the robot was able to reach was limited to 50 cm off the ground - to be discussed in chapter 4. Although the Bat tag and the Bluetooth dongle can freely move within 40 cm - the typical distance from a human's chest to the trousers' pocket, it would be ideal to experiment from 90 cm to 130 cm off the ground, which is the actual height in the human's scale. Due to the lack of materials and the balancing issue, the robot's height was temporally restricted at 50 cm. Thus, up till now, the base stations were put on the floor so that the Bluetooth dongle on the laptop and the Bluetooth dongle on the base station were parallel and at the same height off the ground.

This section experiments with the changes in the Bluetooth signal in terms of height, also to prove that there is not much difference when the whole system is shifted to the actual human's height. First, the base station and the robot were put on the floor. The robot manually shifted the Bluetooth dongle upward every 10 cm, and measured the Bluetooth RSSI at each point. This 10 cm distance was scaled again by a ruler with +/- 2 mm error, to manually fix any error caused by the robot's motor. The idea of this experiment was to experience multipath fading, which results from signals reaching the destination from more than one path. Usually, the strongest signal will arrive at the destination through the shortest unobstructed straight path. The other signals bounce off other objects such as the floor and change their directions before reaching the destination (figure 3.13).

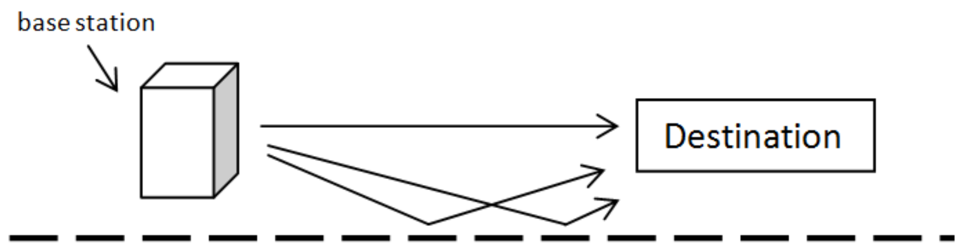

Figure 3.13: Signal mutipath between two positions

In some cases, a Bluetooth wave hits the floor and changes its direction. This signal meets the strongest straight-line signal and causes either constructive interference or destructive interference. The constructive interference happens when two waves of signals are inphase, and both use the same frequency. When these two waves meet in the air, they will complement each other to form a new stronger wave. The destructive interference is the opposite situation. When the two waves are outphase, and they are both in the same channel, they will cancel each other out.

The purpose of this experiment was to find out if there was any single RSSI with a significant high or a significant low dBm in the case of signal interference, by observing the changes of the Bluetooth RSSI every 10 cm over one hour, from 9 cm to 50 cm off the ground. Figure 3.14 graphs 2,000 individual Bluetooth RSSI collected over one hour when both dongles were 9 cm off the ground, opposite and parallel each other. Whilst figure 3.15 graphs 2,000 individual Bluetooth RSSI in the same position, however one dongle was at 9.10 cm off the ground higher than the other dongle. It was surprising that neither constructive interference nor destructive interference could be fully observed. The signal varies from -64 dBm to -54 dBm when the height was 9.00 cm, and varies from -68 dBm to -54 dBm when the heigh was 9.10 cm. There was no isolated RSSI observed in both graphs. An explanation would be because of the frequency hopping technique, which makes a Bluetooth wave change its frequency 1,600 times per second. Although the two Bluetooth waves come from the same source, they were not presented on the same frequency channel over a long period of time to cause either constructive interference or destructive interference. The above experiment was repeated from 90 cm to 130 cm off the ground, to confirm that the constructive and destructive interference could not be re-produced at this height either.

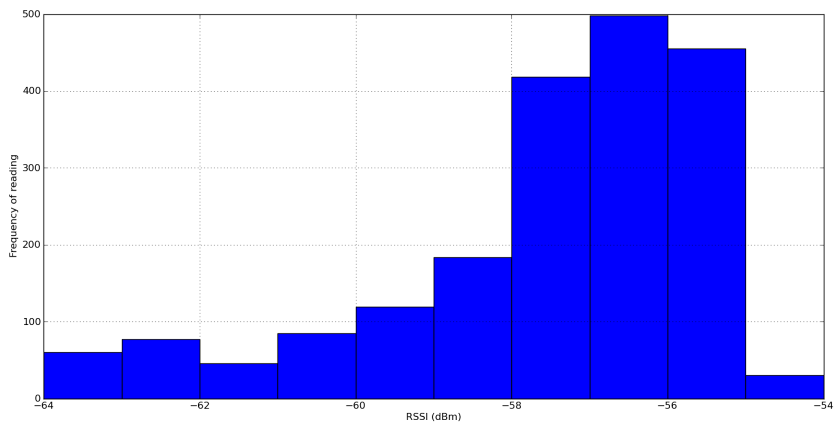

Figure 3.14: Bluetooth RSSI at 9.00 cm off the ground

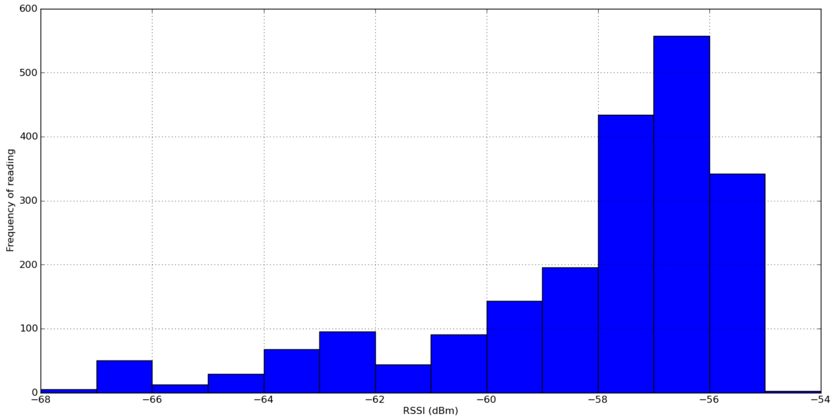

Figure 3.15: Bluetooth RSSI at 9.10 cm off the ground

Once it was clear that the constructive and destructive interference would not have a huge influence on the experiment in terms of height, the total ranges of the RSSI from 9 cm to 50 cm and from 90 cm to 130 cm off the ground were compared to prove that the Bluetooth signal would be the same when the whole system was shifted from 9 cm to 90 cm off the ground. It was observed that the RSSI ranged from -72 dBm to -55 dBm when the dongle was moving from 9 cm to 50 cm off the ground. The range was similar with -71 dBm to -53 dBm when the dongle was moving from 90 cm to 130 cm off the ground. In fact, the small difference between them could be explained by the natural changes in the environment noises such as the humidity.

3.4 How the Position of the Device Antenna Affects the Bluetooth Signal?

Up till now, most experiments in this project assumed that the two Bluetooth devices are positioned opposite and parallel to each other. However, a research question arises if a slight change in the direction the Bluetooth device faces would affect the signal reading. In this experiment, the testing space was divided into eight different directions parallel to the floor (figure 3.16).

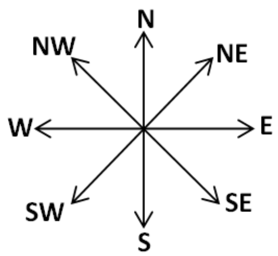

Figure 3.16: 8 directions in space

A computer and a robot were positioned in a straight line 30 cm away, at the same height. There was a clear view without any obstacles between the station and the robot. The robot rotated clockwise parallel to the ground in one spot and stopped at every 45 degrees to measure the signal strength at each direction. It was observed that the signal was stronger when the two Bluetooth devices were facing each other, and became weaker when they were turning their backs to each other (figure 3.17).

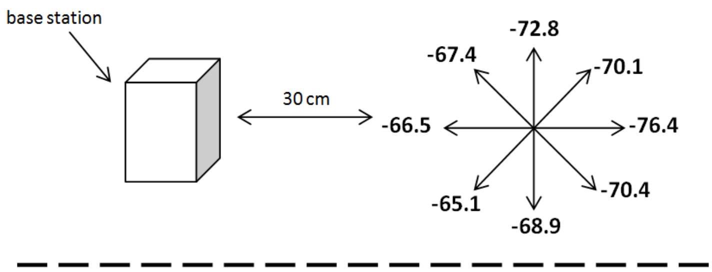

Figure 3.17: Bluetooth signal orientations

Similar results were expected at different locations in the same experimental space above (figure 3.18).

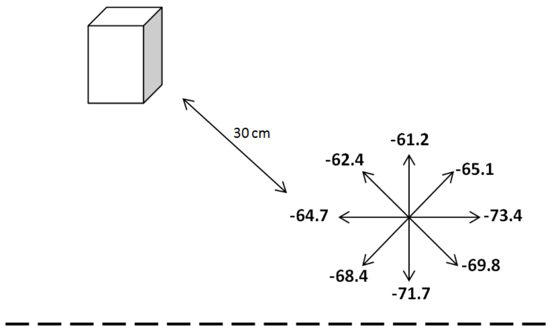

Figure 3.18: Bluetooth signal orientations in a different location

The reason for such changes were because of the design and the position of the internal antenna of the Bluetooth device. Observing the internal design of the Belkin Bluetooth dongle revealed the position of the antenna (figure 3.19), which shows that the signal was broadcasted at the front of the dongle. Hence, it was expected that the signal was stronger when the two devices were facing each other.

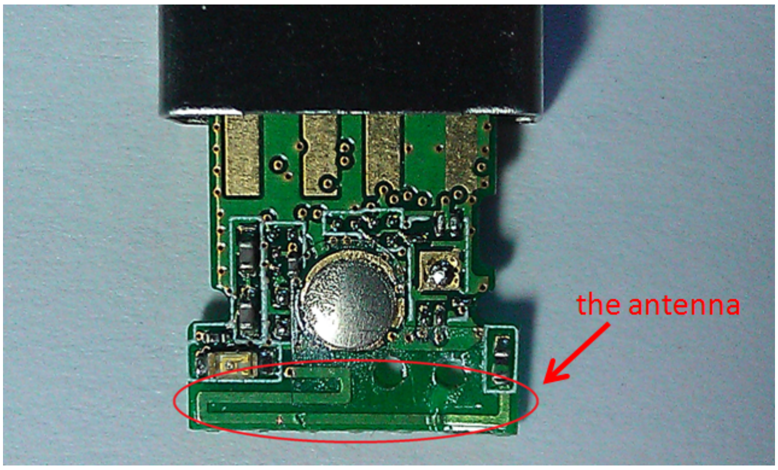

Figure 3.19: Internal antenna design of Belkin Bluetooth dongle

3.5 Bluetooth Properties Summary

From the above experiments, it was observed that the average Bluetooth signal strength only varied around 3 dBm in fixed positions. However, the signal required up to five hours to stabilise, which is an interesting factor requiring further analysis. Further, due to the nature of a radio signal, the Bluetooth signal also suffered from free-space loss, multipath fading and other environmental noises. When moving further away, the signal strength had a quadratic shape. At a distance of 3 m and more, for any two positions with less than 20 cm in between, the difference of the RSSI became as small as 1.5 dBm. However, the difference can reach 5 dBm for two positions at a distance of less than 3 m. When the two Bluetooth devices were lifted together from the floor to 90 cm off the ground - the typical distance in real deployment, there was not much difference in signal readings compared to when they were on the floor, thanks to the frequency hopping technique. Finally, the orientation of the Bluetooth device itself did have an impact on the signal reading, which was stronger when the two devices are opposite and parallel to each other.

In this chapter, the Bluetooth signal has been shown to be very promising with regards to its robustness, reliability and other properties as expected in a standard radio wave signal. These results provided a core foundation for the indoor tracking system to be implemented in this project. Based on the properties investigated within this chapter, special strategies to collect the Bluetooth signal will be described in the next chapter.

Chapter 4

Data Collection

Having already learned the characteristics of relevant Bluetooth properties for indoor tracking, this chapter discusses the strategies to collect those Bluetooth signals into a database in a particular office room. This database will be a valuable tool to be employed in the next chapter.

The chapter opens with a description of the environment, where the Bluetooth signals are collected. Following this, a section is dedicated to the hardware used for collecting the signals, including the Bat system and a robot created specifically for this purpose. Finally, a step-by-step guide of the data collecting process is covered in the last part.

4.1 The Environment

The 'DTG Meeting Room' on the second floor in the Computer Lab, University of Cambridge was chosen as the experimental area (figure 4.1). This room has the best Bat signal coverage in the Computer Lab. A clear area of 12 square metres (5 m x 2.5 m) within the room was covered by Bluetooth signals by placing five Bluetooth base stations as shown in figure 4.2. Since most indoor tracking systems were expected to work well within 2 m, the above area was sufficient enough to implement the system in this project.

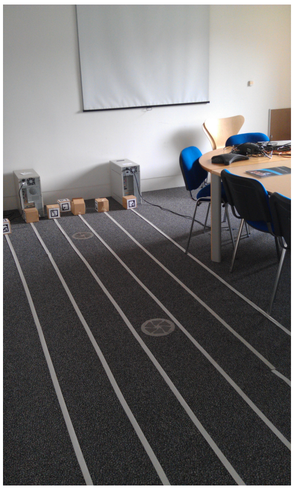

Figure 4.1: DTG Meeting Room

For this particular area, five bluetooth base stations would be sufficient. Although in real deployment, the more stations that are presented, the better the tracking performance will be, since any point in the area would be covered by the Bluetooth signals from many stations.

The stations' positions were strategically chosen to satisfy three conditions. First, any point within this space must properly see at least three stations. In other words, any random place in this area has a distance of less than 2.5 m to at least three base stations nearby. Since, for any two locations further than 3 m from a Bluetooth station, their differences in signal readings are too small to compare easily, as discussed in the previous chapter. Second, to guarantee that every position had a unique set of station signals, at least three stations must combine

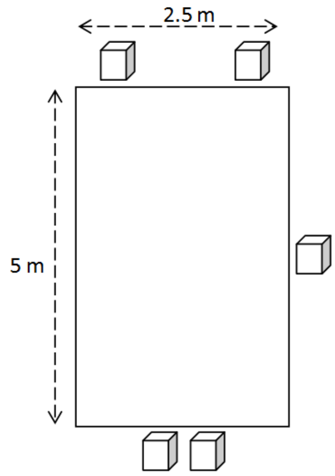

Figure 4.2: DTG Meeting Room - a top down sketch

together to represent a single position. Figure 4.3 shows a common situation with location A and location B both having the same combination of signal from the two stations, which causes a problem in identifying the two locations.

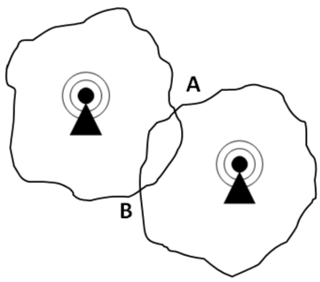

Figure 4.3: Combination of only two base stations

Third, the base stations were arranged in a way to guarantee a certain degree of signal randomness amongst all positions within the area. Figure 4.4 shows a bad arrangement of five base stations for the same above area. Considering any two points in the area such as X and Y, when they move closer or further away to the stations, the signal strengths reflecting their postions are also increasing or decreasing at the same time.

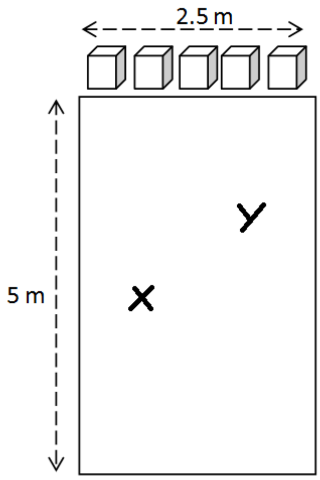

Figure 4.4: Bad arrangement of five base stations

4.2 Hardware Implementations

The data collection process involved four main pieces of hardware. The Bat system contains the tags, the central server and the Firestoker computer. The Bluetooth stations were strategically positioned to broadcast the Bluetooth signals. The robot, which carries the laptop on its back, was created especially for this project.

4.2.1 The Bat System

The Bat system, as discussed in the background chapter, provides an accurate reading up to 3 cm of the current position [3]. There are three important components of the system. First, the tags (figure 4.5) emit strong ultrasonic pulses, which are caught by the receivers mounted on the ceiling to determine the current position of the tags. Second, a central server which controls all the receivers, processes the signals to determine the correct position of a particular tag, and broadcasts the result on a multicast stream. A local computer named Firestoker, which connects directly to the central server, responds to any wireless request asking for the position of a particular tag. This computer acts as an intermediary for easy access, as the central server cannot be queried wirelessly directly.

Figure 4.5: The Bat tags

4.2.2 The Bluetooth Stations

The Bluetooth station were simply a CPU with an attached USB Bluetooth dongle (figure 4.6). The computer ran Ubuntu Server 10.1.

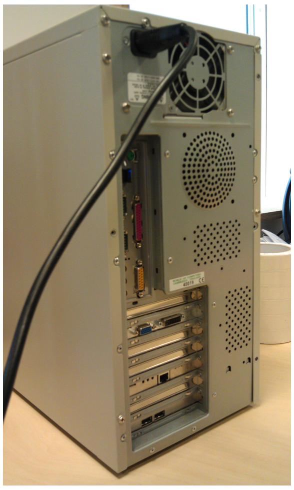

Figure 4.6: The Bluetooth station

4.2.3 The Laptop

A laptop was used to communicate with the Bluetooth stations and saved the Bluetooth signal strengths onto a database on the hard drive. An external Bluetooth dongle was connected to the laptop via a USB extension cable. This allowed freedom of movement, since the USB dongle is relatively small. This laptop also wirelessly connected to the Firestoker computer mentioned above to query the location of the Bat tag, and saved this information into the database. Finally, a direct USB connection was established between the laptop and the robot - to be discussed in the next part, giving driving instructions. The connectivity of the laptop to different components in the system are shown in figure 4.7.

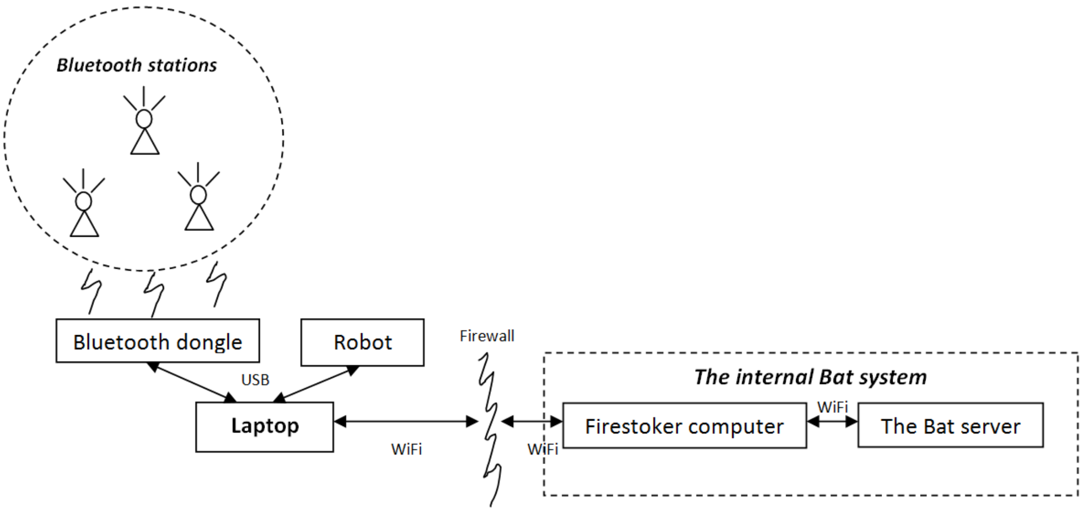

Figure 4.7: The laptop connectivity

4.2.4 The Robot

It takes much effort for a human single-handedly to collect the Bluetooth signals, since a large area of many separated positions must be covered. Not only does the accuracy decrease over long period of times due to natural human mental fatigue, but the water inside the human body also greatly attenuates the Bluetooth signals [26]. The application of robotics solves all these problems. To collect the Bluetooth data efficiently, a robot was designed with Lego Mindstorms NXT (figure 4.8).

This robot can carry a laptop on its back, as well as changing the height of the Bat tag and the Bluetooth dongle up to 40 cm - the typical distance from the human's chest to trouser pockets, which are the probable positions to carry a tag. The design of the robot guarantees two factors:

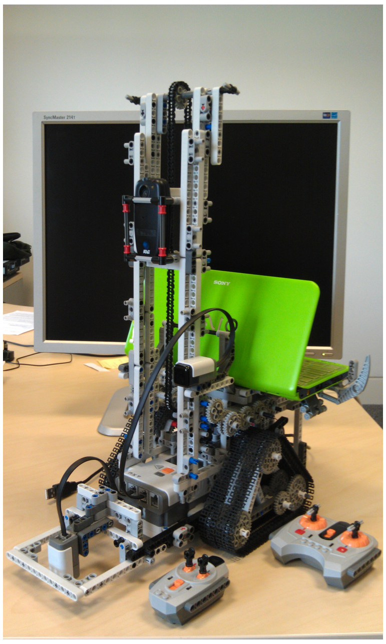

Figure 4.8: The robot

- **The robustness** is the most important condition of any robot. This robot has been able to operate through a long three hour data collecting process without any problem. To achieve this stability, the whole design ensures that the main gravity point is precisely in the centre of the robot. Another challenge is the total weight of all robot materials plus the laptop and the Bat tag exceeding 3 kg, which immensely bend down all plastic axles supplied in the kit. Thus, all the default axles were replaced with custom aluminium ones.

- **The compactness** is very important for this project, as the robot must not occupy too much space, and if the robot is to be used in real deployment, it must be able to travel through the small areas amongst the furniture in any typical office room. After many optimisations such as changing from the normal wheels to the tank treads, exchanging the larger gears for smaller ones, the robot's width is only 15 cm, and the length is 24 cm. Taken into consideration the above factors, this size is acceptable as the enviromental space is divided into blocks of 10 cm - to be discussed in the next part.

This robot also integrates a light sensor, which helps tracking a white line on the floor - to be discussed in the next part, and an infrared receiver which allows the robot to be manually controlled. The rechargeable battery allows the robot to operate continuously for 3.5 hours from a single charge.

4.3 Data Collecting Process

To collect the Bluetooth signals efficiently, the room was divided into blocks of 10 cm squared. This is the most practical scale, since it would take around 8 hours - as discussed below, to fully scan this environment. Also, the 10 cm accuracy was already very desirable for an indoor tracking system. At the beginning, the robot was placed at one corner of the room, it then exhaustedly traversed the room in a zig-zag pattern (figure 4.9).

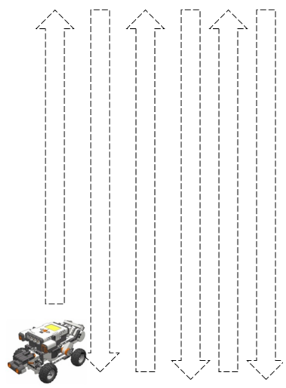

Figure 4.9: The robot goes zig-zag

A line made of removable white masking tape was drawn on the floor to guarantee that the robot moved as straight as possible. The robot tracks this line using a cheap light sensor (£5). This sensor has a strong infrared LED, which dominates most ambient lights. The sensor returns a positive number presenting the current light intensity. By putting this sensor 2 cm off the ground, the reflection of the infrared light is strong enough to distinguish between different surfaces.

By combining this light sensor with the PID controller algorithm, the robot

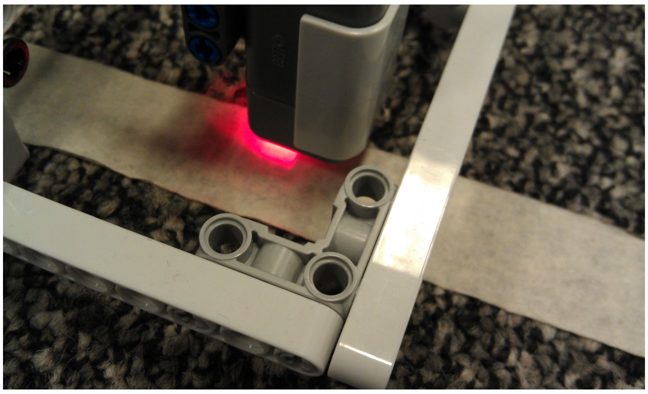

Figure 4.10: The light sensor in action

was able to follow a line precisely. The idea of the PID controller is to convert the light sensor reading into the motor power, and to maintain a given initial state. At the beginning, the carpet's surface and the white line were calibrated, the sensor readings are 34 and 49 (light intensity measurement) respectively. By taking the average of these two numbers, 41.5 would be the number representing the initial state position of the light sensor exactly at the rear of the line and the carpet (figure 4.10). While moving, due to friction, sliding and other minor factors, the robot might not be able to maintain a straight movement. So, the PID controller constantly checks the sensor reading for two possibilities, assuming that the sensor is positioned at the left rear of the line.

- If the number falls within the [34, 41.5] range, it means the robot is going off the line toward the carpet on the left, hence, PID decreases the power of the right motor and increases the power of the left motor to shift the movement to the right.

- If the number falls within the [41.5, 49] range, it means the robot is moving into the white line on the right, hence, PID decreases the power of the left motor and increases the power of the right motor to shift the movement to the left.

The amount of increasing and decreasing power is proportional to the sensor reading. When the reading is exactly at 41.5, the PID controller is off.

While traversing every position in the room, the robot stops at every 10 cm, then notifies the laptop to initiate the scanning process. The laptop first sends a query wirelessly to the Firestoker machine to acknowledge the current position of the Bat tag, which is also the position of the laptop and the robot. This step usually takes 2 seconds and can take up to 5 seconds if the Bat tag is near the wall. Then, it takes another 8 seconds for the laptop to collect the Bluetooth signals at any single place. Further, the robot rotates 360 degrees in one spot, so that the laptop can pick up the Bluetooth signals in eight different directions. The laptop notifies the robot to continue moving, when the Bluetooth information and the Bat position have been saved onto the hard drive. Overall, it took around 8 hours to collect all the necessary information in the 'DTG Meeting Room'. Unfortunately, because of the limited capacity of the rechargeable battery, the robot and the laptop must be recharged every three and a half hours.

Chapter 5

Fingerprinting and Optimal Position Classifications

One of the problems of indoor tracking is signal multipath fading, caused by signals bouncing off the walls and indoor objects. The signal gets weaker and weaker over time, until it becomes too unreliable to identify. The fingerprinting technique was designed to tackle this problem. It uses the location database collected in the previous chapter to cover the signal characteristics at each small position in an office room. The classification algorithms were then implemented to estimate the most probable position from this database, given the tracking signal of an unknown location.

This chapter has two parts. First, the fingerprinting technique is discussed, which covers the online stage, the offline stage in addition to a special testing stage for this project. Then, three classifiers which manipulate the fingerprinting database are discussed, together with their performances and their disadvantages. They are the Weighted K-Nearest Neighbour algorithm, the Naive Bayesian classifier, and the Histogram methods. The chapter concludes with a summary of the three classification algorithms, and argues which classifier is suitable to use.

5.1 Location Fingerprinting

The idea of location fingerprinting is to perform a real-time survey of many small positions in a pre-defined space such as an office room, in order that the position's unique characteristics are captured and recorded into a database. Further on, the characteristics of an unknown position are compared against those in the database to estimate a closest co-ordinate for this unknown position. In this project, those characteristics are recorded in the form of the Bluetooth signal strength (RSSI). As discussed in chapter 3, the Bluetooth signal strength differs from position to position. Thus, any position can be uniquely identified by combining multiple Bluetooth signal strengths together from many Bluetooth base stations, as discussed in the Data Collection chapter. This technique easily solves the multipath fading problem, caused by the reflection and the scattering of the wireless signal in a dense indoor environment with a large spread of small furniture obstacles. Because, there is no importance attached to how a signal reaches a destination, its strength is fully captured in the database. However, an obvious challenge is how to generate a quality fingerprinting database to cover 'enough' small positions, as well as how to capture those 'unique characteristics' at each position efficiently.

This section describes two standard phases of the fingerprinting procedure: the online stage - where the database is created, and the offline stage - where an unknown position is classified. Further, a testing stage is purposely included into the project to generate a test set to evaluate the system's performance, with the aid of a robot.

5.1.1 Online Stage

During the online stage, a fingerprinting database is created, which uniquely maps many physical positions represented by the combination of Bluetooth signal strengths to the Bat position, as discussed in the previous Data Collection chapter. The Bat position is identified as a (x, y, z) co-ordinate tuple provided by the Bat system. So, given a particular signal strength pattern of an unknown location, the task is simply to estimate an unknown position using this fingerprinting database.

5.1.2 Offline Stage

During the offline stage, the system estimates an optimal position using the fingerprinting database created in the online stage. The process involves first recording the current signal strength combination of all nearby Bluetooth base stations, then classifying this combination sample with three algorithms - to be discussed later in the chapter. In general, the accuracy of the estimated position depends on the quality of the fingerprinting database and effectiveness of the classifier algorithms.

5.1.3 Testing Stage

The testing stage is an extra step of this project. To evaluate the performance of the system, a test set was created with the robot. The set contains 263 random Bat positions, along with their Bluetooth RSSI, in the same 'DTG Meeting Room' environment used to generate the fingerprinting database in the online stage. The data was collected over 3 hours on the 14th-April-2011 (figure 5.1).

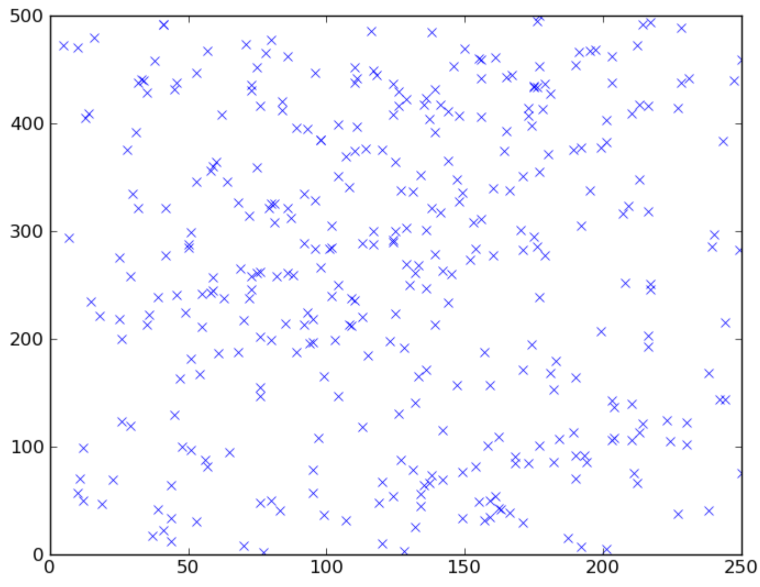

Figure 5.1: Random test points in 'DTG Meeting Room'

The robot was slightly modified to attach an ultrasonic sensor, which can detect the presence of any nearby object. First, the robot was put in the middle of the room, so that it has a large space around itself. It was programmed to drive in a random direction. When a wall was detected within 10 cm, the robot randomly changed its direction to avoid hitting the wall. The robot stopped at random times to record the Bat position and the Bluetooth RSSI of this location onto the hard drive. The robot also changed its direction randomly to increase the location coverage. This robot did the same job of a human holding a laptop and walking around the room to record the Bluetooth signal and the Bat data. Figure 5.2 shows the traversing path of the robot in the above environment.

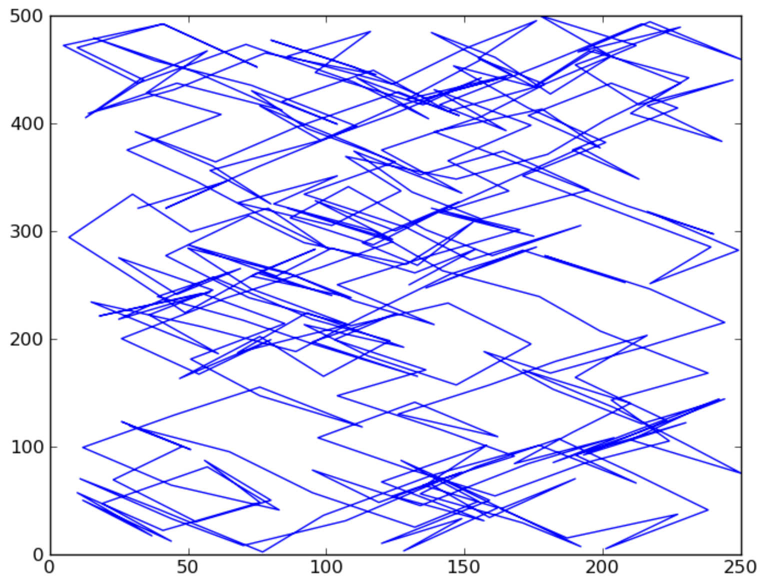

Figure 5.2: The traversing path of the robot

Those 263 positions in the test set are considered as the unknown positions to test the three classifier algorithms in the next part. The output positions of these classifiers are compared to the Bat position references, which are also recorded in the test set. Thus, the accuracy of these algorithms can be evaluated.

5.2 Classifier Algorithms

Having already obtained a fingerprinting database, the next important step is to work out the optimal Bat co-ordinate for an unknown position using this database. Three separate typical algorithms were applied for this purpose. The deterministic weighted K-nearest neighbours algorithm computes the average co-ordinates of the locations. The probabilistic Bayesian algorithm uses the probability of the signal strength, whereas the histogram methods use the mean and the variance of the signal strength distribution to estimate an unknown position. Their performances and evaluations are explained in this section.

5.2.1 Weighted K-Nearest Neighbours

The general Weighted K-Nearest Neighbours algorithm as discussed in detail in the background chapter, selects the K nearest neighbours of a given point in the database. Then, based on the distance between this point and each point of the K neighbours, a new estimated position is calculated, which is hopefully close to the actual position.

This algorithm can be applied directly into location fingerprinting. To classify an unknown position, the signal strengths to all visible Bluetooth base stations at this unknown location are measured as an n-tuple X = $(b_{x1}, b_{x2}, \ldots, b_{xn})$ where n is the number of installed base stations, b_{xi} is the RSSI measurement from the Bluetooth station i, if the station i is not discoverable, then $b_{xi} = 0$. Since an ideal fingerprinting database guarantees to cover every small position in the environment, in theory, this tuple X should appear somewhere in the database. However, since the database in this project is a collection of many 10 cm squared blocks, an unknown location may slip in somewhere between these blocks (figure 5.3). So, it is impossible to find an exact tuple in the database, which completely matches this tuple X of the unknown location.

A solution is to find K tuples in the database, which are 'nearest' to the tuple X. These K tuples, along with their Bat positions, can combine to produce a new estimated position e, which is hopefully close enough to the actual unknown position. To find K nearest tuples in the database, the distance between each tuple Y = $(b_{y1}, b_{y2}, \ldots, b_{yn})$ in the database and tuple X = $(b_{x1}, b_{x2}, \ldots, b_{xn})$ is

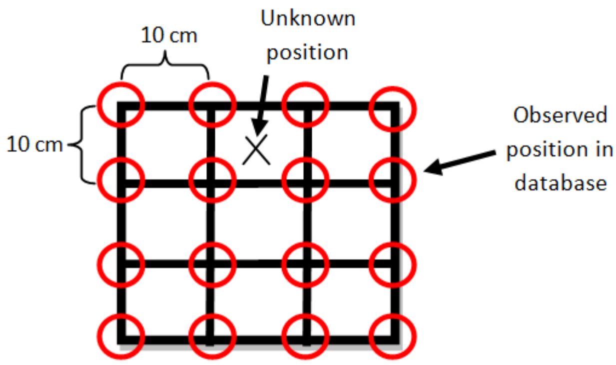

Figure 5.3: Some unknown positions between 10 cm blocks

calculated directly from their individual RSSI, as follows:

$$dist(X, Y) = \sqrt{(b_{x1} - b_{y1})^2 + (b_{x2} - b_{y2})^2 + \ldots + (b_{xn} - b_{yn})^2} \quad (5.1)$$

It is worth noting that although the above formulae adapts the same Euclidean distance approach, the 'signal distance' for the X, Y tuples is totally different from the usual 'Bat distance' in the 3-dimensional space (x, y, z) format.

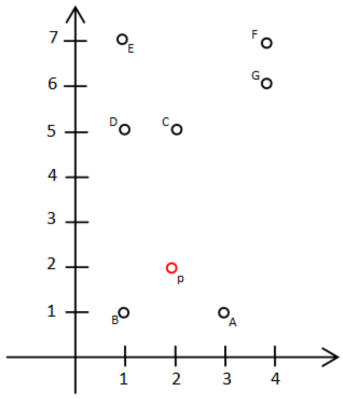

Figure 5.4: 7-Nearest neighbours

One of the problems of the indoor signals is that many locations far away might have similar combinations of RSSI, because of signal multipath. Figure 5.4 suggests such a situation, with 7 nearest neighbours, 5 of them (C, D, E, F,

G) have similar signal strength combinations, although they are not close to the actual correct position p. It is clear that those positions are too far away, and may affect the final estimation if included in the K nearest neighbours. Fortunately, by considering the weight, corresponding to the inverse 'signal distance' between each neighbour and the unknown position, the final estimated position would not be significantly affected by those positions further away. The reason to invert these 'signal distances' is to prioritise closer neighbours over those which are further away. Formula 5.2 calculates the estimated position e, given K nearest neighbours and their Bat positions, as discussed in the background chapter.

$$e_x = \frac{\sum_{i=1}^{K} \frac{1}{dist(t_i, p)} \; p_x}{\sum_{i=1}^{K} \frac{1}{dist(t_i, p)}} \tag{5.2}$$

The final problem is to determine the value for the constant K. When K = 1, the task is to ultimately find a single nearest neighbour, which does not take any advantage of the weighted distance. When K is very large, the K neighbours might contain many irrelevant positions far away. The most 'reasonable' K value was found experimentally at 430 for the above environment. However, for a larger K value, the performance only changes by 2%-3% (figure 5.5).

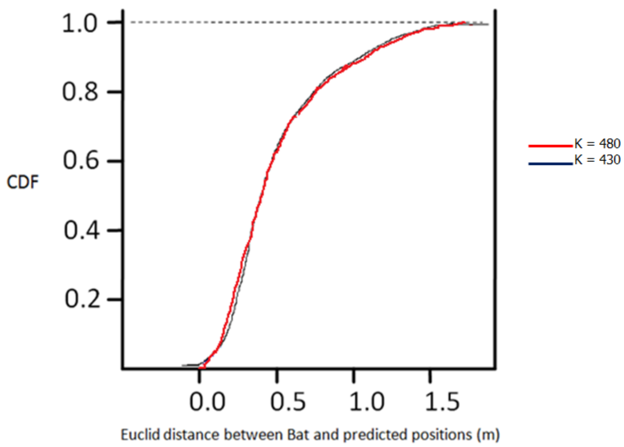

Figure 5.5: Performance comparison for different K values

Comparing this algorithm to the un-weighted version of the original algorithm implemented in the RADAR system [7][8] shows a significant performance boost of 21% (figure 5.6).

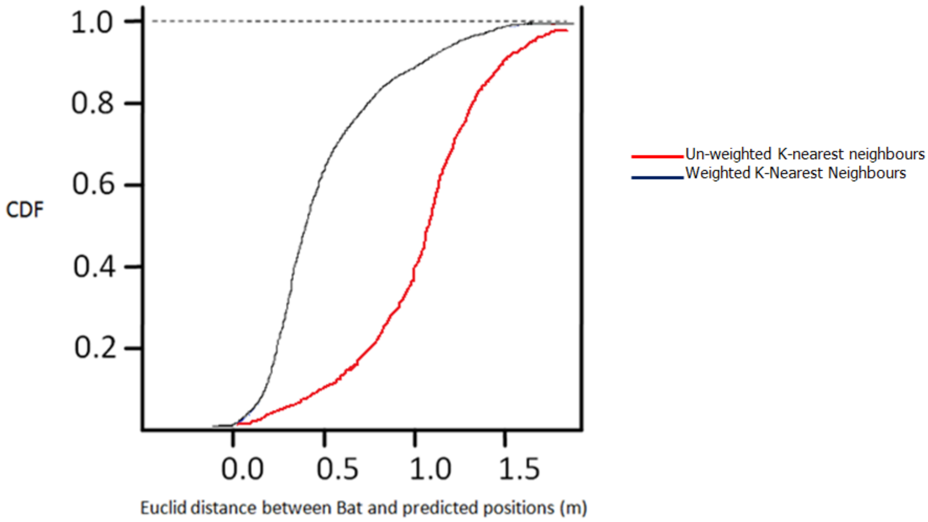

Figure 5.6: Performance of weighted algorithm and un-weighted algorithm

For this system, the value of K is expected to be much larger than other systems due to the nature of the Bluetooth signal. When the distance to a base station is more than 3 m, the difference amongst signal strengths is too negligible as discussed in chapter 3. It is also worth noting that there is no optimal K value for every position. The 430 value above for K gave a relatively high performance across all 263 test positions. However, some positions in the test set can achieve better results with a different K. Take for example a position in the actual test set with the Bluetooth combination [-60, -51, -54, -43, -62], which had a bad performance error at 1.1 m for K = 430. However, when K is increased to 620, the error drops to 60 cm.

In conclusion, the 'most reasonable' value for K depends on the signal properties and the deployed environment. For the Bluetooth signal, the constant K is expected to be larger than 100 for a good prediction of most positions, due to signal multipath. This is also the weakness of the algorithm, since the K parameter must be manually experimented in each environment to find a suitable value. However, the algorithm is very simple and easy to implement.

5.2.2 Naive Bayesian Classifier

While the weighted K-nearest neighbours algorithm computes an average of all nearest locations to estimate an unknown position, the Bayesian classifier algorithm picks up exactly one position in the fingerprinting database, which has the highest probability, to represent an unknown position.

The idea of probability was introduced through a chart called a histogram table. Since the Bluetooth signal fluctuates even in fixed positions as discussed in chapter 3, many readings were taken for a particular position during the offline stage to cover these variations. All these RSSI readings were then stored in the histogram table. Figure 5.7 shows a snapshot of the histogram table for the Bat position (98.12, 23.45, 7.33)

00:19:0E:06:C6:BF	00:36:CF:12:0A:34	00:19:0E:18:C9:38	00:19:12:D2:C1:15	00:0F:03:16:C7:0B
-73	-75	-60	-65	-47
-72	-74	-60	-66	-47
-73	-74	-59	-65	-48
-73	-75	-59	-65	-48
-73	-76	-60	-65	-47

Figure 5.7: Histogram table snapshot

So, given a single RSSI reading and a particular position in the database, it is possible to calculate the probability of this RSSI reading occurring at that location, based on the fact of how often that RSSI reading appears at that location. The formula 5.3 calculates the probability of a signal strength s happening at the location L.

$$P(s|L) = \frac{number\ of\ times\ s\ appears}{total\ number\ of\ readings\ taken\ at\ location\ L} \tag{5.3}$$

However, a location is actually represented by a tuple of the combination of RSSI from many base stations such as $(s_1, s_2, \ldots, s_n)$ where s_i is the RSSI of the station i measured from this location. To calculate the probability of an arbitrary RSSI tuple $(s_1, s_2, \ldots, s_n)$ at a particular location L, it is not very useful to adapt a similar direct approach as formula 5.3 above suggests, as the whole exact tuple might not even appear at that location L, because one of the stations was not discovered, or the signal variations were not fully collected in the histogram table. For this reason, applying the formula 5.3 would return zero for the probability $P((s_1, s_2, \ldots, s_n)|L)$. In fact, it should return a non-zero probability even when some RSSIs were missing or not perfectly matched. One way to achieve this is to look at each individual RSSI of each base station. Fortunately, by assuming that every base station is independent, it is possible to calculate the probability that a tuple $(s_1, s_2, \ldots, s_n)$ would occur at location L, based on the probability of each individual signal strength s_i at the same location L, as follows:

$$P((s_1, s_2, \ldots, s_n)|L) \quad = \quad P(s_1|L)\, P(s_2|L) \ldots P(s_n|L) \tag{5.4}$$

Each separate term $P(s_i|L)$ can be calculated independently using formula 5.3. This assumption is also known as the Naive Bayesian approach and is strongly supported by the fact that Bluetooth technology uses the frequency hopping technique, which minimises the interference between wireless signals in the air. Thus, the Bluetooth signal from one base station does not influence the Bluetooth signal from other base stations in the same tracking space.

However, the actual problem to be solved is to find out the probability that the estimated position is in fact position L, given a particular RSSI tuple. This is exactly the reverse of the probability $P((s_1, s_2, \ldots, s_n)|L)$ calculated above. This reverse probability $P(L|(s_1, s_2, \ldots, s_n))$ can be easily computed using the Bayesian formula.

$$P(X|Y) \quad = \quad \frac{P(Y|X)\, P(X)}{P(Y)} \tag{5.5}$$

By substituting L and $(s_1, s_2, \ldots, s_n)$ into the above formula

$$P(L|(s_1, s_2, \ldots, s_n)) \quad = \quad \frac{P((s_1, s_2, \ldots, s_n)|L)\, P(L)}{P(s_1, s_2, \ldots, s_n)} \tag{5.6}$$

Now considering the above formula, for any position L, the probability $P((s_1, s_2, \ldots, s_n)|L)$ can be computed with the formula 5.4. The probability P(L) of a location L itself is always $\frac{1}{N}$ where N is the total number of positions in the database. The probability $P(s_1, s_2, \ldots, s_n)$ is the number of times the tuple $(s_1, s_2, \ldots, s_n)$ appears in the whole database, divided by the total number of tuples in the database. In other words, P(L) and $P(s_1, s_2, \ldots, s_n)$ are two constants, and can be ignored in the computations.

Finally, given an RSSI tuple of an unknown position, using the formula 5.6, it is possible to calculate the probability of a location L $P(L|(s_1, s_2, \ldots, s_n))$ for every position recorded in the fingerprinting database. The position with the highest probability is obviously the estimated position.

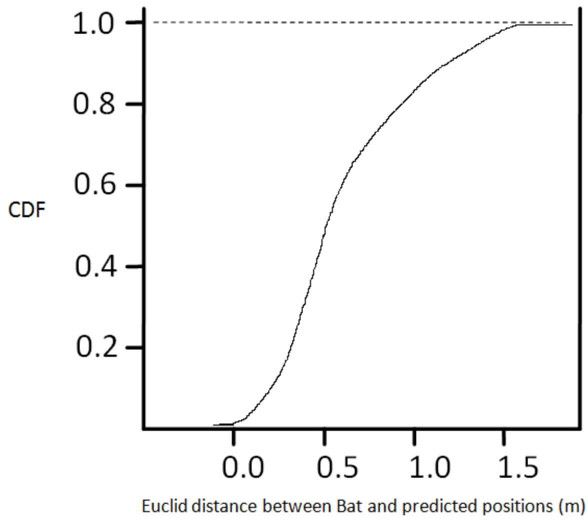

Figure 5.8: Bayesian classifier accuracy

Figure 5.8 identifies the accuracy of the Bayesian classifier for the above test set. The obvious disadvantage of this algorithm was that the database density must be high to cover enough positions in the environment, because the algorithm chooses only one position in the database with the highest probability to represent the estimated position. Ideally, any position in the whole environment should have been covered in the database, or the distance from the unknown position to any nearest position in the database should be as small as possible. On the other hand, the algorithm takes into account the RSSI variation of each individual position, rather than just averages of every co-ordinate as in the nearest neighbour algorithm.

5.2.3 The Histogram Methods

The histogram method adapts a similar approach as the Bayesian classifier by using the histogram table, which covers the signal variation at every position in the database. Besides, it takes a further step by considering the whole signal variation of the unknown position too, rather than just using a single RSSI reading in the online stage as in the Bayesian algorithm.

The method simply took many signal readings at an unknown position, to create a histogram of this unknown position. Then, this histogram was directly compared to the histogram of every position in the database to establish a closest match. For this purpose, a method called 'Student's T-Test' was applied to compare the two histograms, as discussed in the background chapter. This method returned a positive t-value representing the different level between the two histograms as follows: where $\bar{X}_1$ and $\bar{X}_2$ are the means of the two histograms, and n_1, n_2 are the size of two histograms:

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{5.7}$$

The only issue is the fact that this T-Test makes two assumptions at the beginning. The first assumption is the two histograms must be independent, which is correct for this system, since the signal strength of the unknown position does not affect the signal strength of any position in the database. The second assumption asserts that the two histograms are normal distributions, which does not totally satisfy all the Bluetooth histograms. However, the results from chapter 3 shows that although the Bluetooth histogram does not possess a perfect Gaussian bell-curve shape, it does not have any complete isolated RSSI for 95% of the time. The experimental result for the test set showed that the performance does not degrade too much by this violation of the second assumption (figure 5.9).

However, in different test sets and different environments, one may find a complete non-normal distribution histogram. In this case, the Komolgorov-Smirnov test (K-S Test) can be a substitution. This method calculates a d-value, which also represents the difference level between the two histograms. However, the major difference is that it makes no assumption about the normal distribution

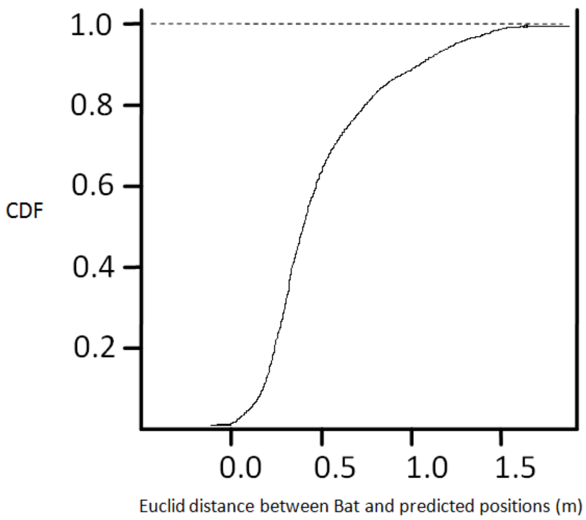

Figure 5.9: Student's T-Test accuracy

property. This comes at one cost where the T-Test is more accurate when the histogram is normally distributed or nearly normally distributed [2]. Besides, the K-S Test reports whether the histogram seems to be normally distributed or log-normally distributed, which is an added advantage to assess if the T-Test can be applied instead.

After choosing either the T-Test or the K-S Test to use, the next step is to perform a significant test to determine if the t-value or the d-value can be accepted or not. More details of the significant test can be found in the background chapter. The main idea is to choose a p-value, for example p = 0.05, which means the confidence is 95% that the two histograms are the same. This 0.05 value is a standard measurement for most statistical tests, and was experimentally confirmed to be accurate with the test set in this project (figure 5.9). After that, a t-distribution table is looked up and the calculated t-value or d-value is rejected if it is greater than the reference value in the table.

The obvious disadvantage of the histogram methods is the fact that it needs a signal strength histogram of the unknown position to provide an estimation. For real-time tracking with fast moving objects, the user cannot afford to wait until sufficient signal strength variations have been collected.

5.2.4 Summary of Classifier Algorithms

In general, all three algorithms perform equally well as shown in figure 5.10.

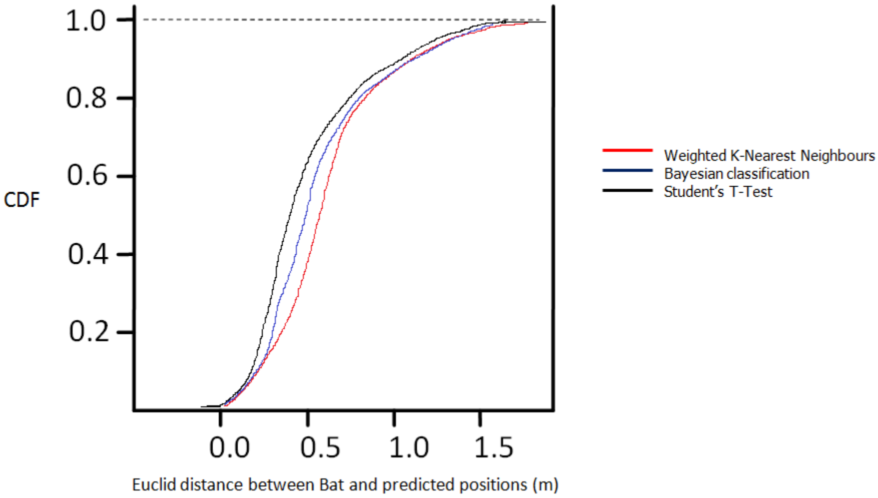

Figure 5.10: Three algorithms accuracy

The weighted K-nearest neighbours algorithm is very simple and easy to implement, but the biggest challenge is to identify the optimal K parameter, which can only be achieved experimentally for each environment. The histogram method seems to edge out of the other two algorithms, since it takes into account the whole range of signal strength variation. However, to obtain many readings for a single location in a short time during real-time tracking is challenging, especially for the Bluetooth signal. The last algorithm, the Bayesian classifier uses just one reading of the unknown position, and chooses one position from the database with the highest probability as the estimated position. This algorithm does not have any real weakness, and has relatively good performance. Thus, this is the most popular choice for location tracking. The Bayesian classifier can be further enhanced by combining with the weighted K-nearest neighbours. Instead of just picking one location in the database, the algorithm selects K locations with the highest probabilities, then applied the weighted K-nearest neighbours algorithm for those locations. In other words, the RSSI measurement for all locations in the database is replaced with a single number representing the probability of the location, then the weighted K-nearest neighbours is applied as usual.

Chapter 6

System Evaluations

Having already understood the infrastructures and the algorithms underpinning the system, this chapter evaluates the overall performance of the system, in a comparable manner to other indoor tracking systems, as well as addressing several issues to make the system widely recognised.

The chapter opens with a comparison of the overall system's accuracy to the indoor tracking systems described in other research papers. Several aspects of the system such as the running time, the accuracy and the scalability are considered. Next, several weaknesses of the system which arose during the development phase are addressed, together with some possible solutions.

6.1 System Performance

As discussed in the previous chapter, all three algorithms to estimate the unknown position perform equally well. Of all the weaknesses and the advantages, the Bayesian classifier is one of the best options, since it can predict an unknown position with just one tuple of RSSI reading, rather than requiring a whole range of tuples in the case of the histogram methods, and it does not have to tweak the K parameter for each environment in the case of the weighted K-nearest neighbours algorithm.

Overall, the system's accuracy applying the Bayesian classifier is very promising, with less than 1.5 m error, 88% of the time (figure 6.1). The system can achieve less than 50 cm error, 43% of the time. Compared to other popular indoor tracking systems using fingerprinting technique such as the RADAR system [7] with 2 m error, the Horus system [30] with 1 m error, 85% of the time, this performance is very promising, considering the affordability feature of the system.

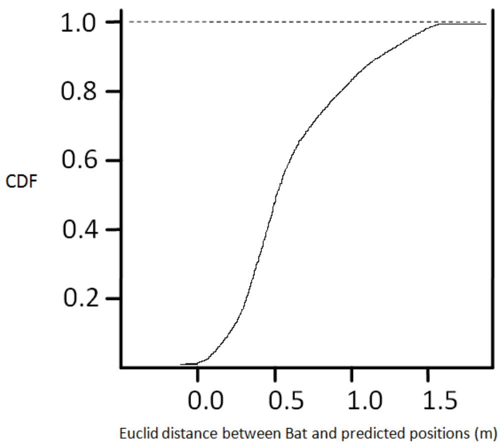

Figure 6.1: Overall system performance accuracy

The overall speed of the system is also very acceptable. As mentioned in the Data Collection chapter, it took around 8 hours to collect offline fingerprinting data in the 'DTG Meeting Room'. The database size is about 24 MB. For this size of data, the classification speed for the whole test set was relatively fast with 11 seconds for all 263 test positions with the Bayesian classifier on a 1.6 Ghz computer. Thus, it only took an average of 0.04 seconds to estimate a single unknown position. Figure 6.2 compares the classifying time of the three implemented fingerprinting algorithms for the test set. In general, the running time of the Bayesian classifier and the Histogram method were very close, while the deterministic algorithm was much slower, because of the big number of neighbours to be considered as discussed in the previous chapter.

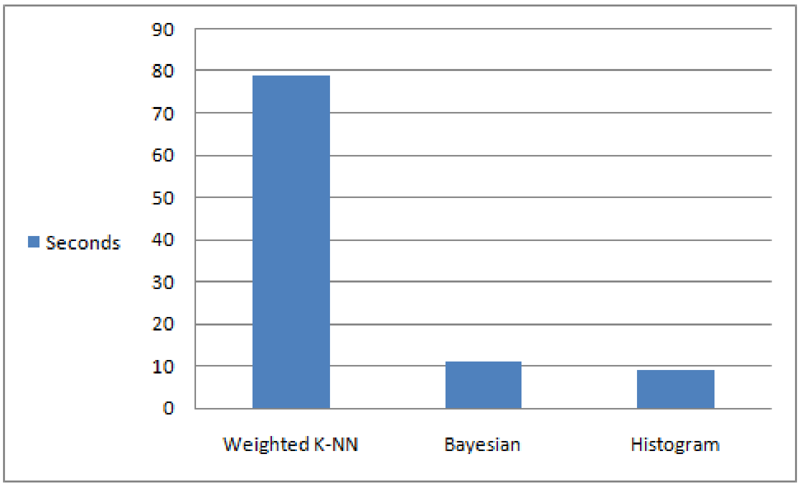

Figure 6.2: Comparison of the classifying time

Further, the reference Bat positions recorded during the offline stage were very important too, as they decide the accuracy of an estimated location. Since the Bat system has a 3 cm error, 95% of the time [3][4][6], there is still a 5% chance that the position error is large. Especially when the robot was near the wall, the receivers struggled to identify the position of the Bat tag. By inspecting the Bat position data of the 'DTG Meeting Room' after one round of scanning, there were 4 points (1.6%) with position errors greater than 20 cm. These noises were removed from the database before the offline stage was performed.

6.2 System Issues

Although the implemented system works well at a certain degree, there are still some issues to be addressed so that the system can be widely recognised in real life. Particularly, these problems are concerned with the scalability, system maintenance, and the ability of the system to adapt to unusual circumstances.

6.2.1 System Scalability

During the online stage, the system must search through the whole fingerprinting database to predict the current unknown position. Since the database is extremely large as to cover a huge range of positions, it takes a substantial time and computational effort to go through every record. Even a 12 m squared area in the 'DTG Meeting Room' already contains over 350 small positions. For real-time tracking in a big indoor building, the querying time should be as small as possible if not instantaneous. A solution is using data clustering, which groups the positions with similar base stations into clusters [30].

6.2.2 System Maintenance

The robot tends to gather dust amongst the gears forming the tank tread. This dust comes from the carpet, and greatly affects the robot's straight movement, which constantly turns the PID controller on. Hence, the battery can be conserved longer, if the gears are cleaned after the online stage is performed.

The Bluetooth dongles can be easily replaced if damaged. The price of a new Belkin dongle is relatively cheap at £7. However, to maintain the database correctness, the offline stage should be re-performed whenever a new piece of hardware replaces another to reflect the new changes in the system.

6.2.3 System Adaptation in Unusual Situations

Since the whole system relies on the fingerprinting database to estimate a position, the position of the base stations must be fixed through-out both offline and online stages to maintain the integrity for the database. However, the system should be able to self-heal itself under different circumstances. Occasionally, a base station presented during the online stage may completely disappear during the offline stage, because of a power issue or being moved out of range. The effect of this situation may be very severe. For the Bayesian classifier, the whole probability calculation for any position involving this base station will return to zero, as the equation is a product of the RSSI at each individual station. A solution for this problem is to give every base station a very small probability such as 0.00001 to prevent the zero result.

6.2.4 Tracking Fast Moving Objects

One of the biggest disadvantages of the system is the slight delay in Bluetooth discovery. In some situations, it might require a full scan of up to 10.24 seconds to discover a base station. Even after customising the Bluetooth dongle's inquiry parameters to reduce the scanning time, there is no guarantee that a signal strength reading can be made every one second. More details of this problem will be discussed in the next chapter.

Chapter 7

Conclusions

In this project, an affordable indoor tracking system using Bluetooth technology and the fingerprinting technique was implemented. The performance of the system is very promising when implemented for the 'DTG Meeting Room' in the Computer Lab. Through-out the development of the system, relevant Bluetooth properties for the indoor tracking context were fully investigated in the form of four research questions, which sought to verify the robustness of the implemented system.

In this chapter, the results obtained within the project are summarised and re-assessed. A section for future research within the same domain concludes this chapter.

7.1 Research Contributions

This section summarises the project contributions in the indoor tracking domain, focusing on the affordability of the system, and the research properties of Bluetooth technology.

7.1.1 The Need for an Affordable Indoor Tracking System

Since most current indoor tracking systems are either expensive such as the Bat system [2][3], or inaccurate such as the RADAR system [5], the deployment of those systems is limited to academic research only. This project specifically addresses the affordability issue by applying the Bluetooth technology, as the widespread use of Bluetooth-enabled devices such as mobile phones, laptops, headsets is a major advantage. The Belkin dongle used in this project can be purchased for as little as £7 on Amazon UK.

7.1.2 Bluetooth Properties for Indoor Tracking

To confirm that the Bluetooth technology is a suitable back-bone infrastructure and to prove that the system is reliable over the long-term, four Bluetooth properties were inspected through many practical experiments.

- The stability of the Bluetooth signal in fixed positions is confirmed to be as small as 3 dBm for the Belkin Bluetooth dongle. Thanks to the frequency hopping technique, the Bluetooth signal is virtually immune to other RF signals such as WiFi.

- The greater the distance is, the weaker the Bluetooth signal becomes. However, when the distance to a base station is further than 3 m, the difference of the Bluetooth signal between positions is as small as 1.5 dBm, which is unrealistic for the fingerprinting technique, since the positions are not clearly distinguishable.

- The height of the base stations does not seem to play an important role in this project. The signal's multipath fading problem, which causes constructive interference and destructive interference cannot be reproduced, because the Bluetooth device automatically changes the frequency channel 1,600 times per second.

- The orientations of the device's antenna hugely influence the Bluetooth signal strength. At the same height off the ground, the signal is stronger when the two devices are totally opposite and parallel each other.

7.1.3 Fingerprinting Algorithms

To tackle the multipath fading problem of the indoor signals, the fingerprinting technique was applied to create a location database covering many small positions in an office room. A robot was designed to tackle the hassle of the data collection process. This robot was also used to survey the system performance by randomly walking around the room to collect Bluetooth data and Bat data. Three typical algorithms were evaluated in this project. The deterministic weighted K-nearest neighbours algorithm selects K closest positions, and computes an average estimated position of these K positions. This algorithm favours the closer positions to the unknown position when computing the estimated position. The algorithm is very simple to implement, but the cost to compute the weight of the K neighbours might be high, as well as the K parameter needs tweaking for each specific environment. The probabilistic Bayesian classifier uses the signal strength frequency to predict the probability of a signal strength tuple. The position in the database with the highest probability is chosen as the estimated position. Finally, the histogram method compares two signal strength distributions directly, as based upon the significant testing. The mean and the variance are manipulated to conclude whether the two distributions are different or not. In general, all three algorithms performed equally well for the test set used in this project.

7.2 Future Work

Although the system implemented in this project is performing well to a certain degree, additional research goals must be achieved before the system can be widely recognised in real life. This section addresses four research issues to improve the system's practicality.

7.2.1 Tracking Fast Moving Objects

One of the biggest disadvantages of the system is the slight delay in the Bluetooth discovery. In some situations, the Bluetooth device might require up to 10.24 seconds for a full scan. Even after customising the Bluetooth dongle's inquiry parameters to reduce the scanning time, there is no guarantee that a signal strength reading can be made every one or two seconds, which still falls sort of the almost instaneous response time with a Wi-Fi system. The problem with this delay is the RSSI reading at a particular moment might not reflect the actual current position. Figure 7.1 shows an example where the Bluetooth reading lags by 2 seconds. While the user is moving from left to right, station A is not discovered until two seconds later.

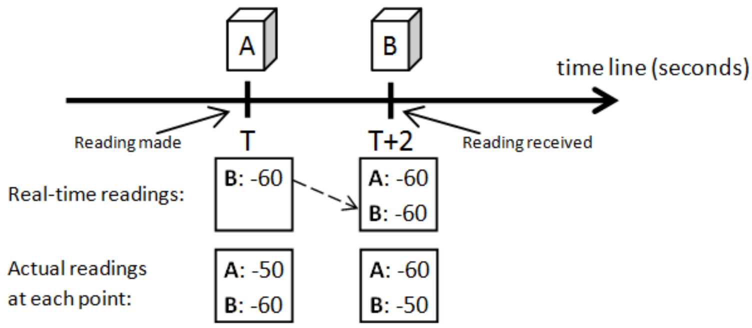

Figure 7.1: Bluetooth signal strength lags while moving

The above issue is not seen with Wi-Fi tracking since the Wi-Fi signal strength can be queried almost instantly. However, when the moving speed is relatively fast and the distance between the tag and the base stations are far, the tracking accuracy decreases rapidly to as much as 10 m [19]. For the scope of the system in this project, the delay in the discovery of the Bluetooth signal introduces a level of uncertainty into the Bluetooth reading. Thus, a solution is to introduce an error measurement into every RSSI reading. This error value is proportional to the base stations' discovery time. Sometimes, a rare Bluetooth reading can be made in less than a second, thus this reading has a very low error level, and can be trusted to increase the confidence in the position estimation for this location.

7.2.2 Exploring Different Domains

The environment in this project was restricted to the 'DTG Meeting Room' inside the Computer Lab, University of Cambridge, where Bat coverage was required. However, to prove that the system is suitable for real deployment, different environments should also be tested such as company offices, warehouses, hospitals, domestic houses, etc ... More research is needed to understand if any modifications are required to adapt the system to the new environments.

7.2.3 Implementing Customised Antenna for Multi-directional Tracking

In the same spot, different orientations of the Bluetooth device's antenna will give different signal readings. Most current indoor tracking systems only implement one-directional tracking for the tag during the online stage [15][16][18]. An ultimate idea would be to have eight Bluetooth dongles implemented in one tag, with each dongle points to a different direction. Thus, a much more precise reading of signal strength combinations can be made at any single place during the online stage. Comparing this reading combination to each location in the database will certainly provide a position with a higher level of accuracy. The problem with this design is the eight Bluetooth dongles will interfere with each other while operating simultaneously. And it would be impractical to turn on each dongle one at a time, since the scanning time would be too long. A better approach is to connect all eight dongles' antennas together to form one big antenna.

7.2.4 Improving the Robot Design to Human-scale

One of the problems already identified during the project development is the height of the robot. Due to the weight, balance, and the lack of materials, the robot can only change the Bat tag and the Bluetooth dongle's height up to 40 cm. Although this range can accurately reflect the distance between a human's chest to the trouser pockets, it would be more beneficial to build a robot with a similar height to that of a human. Thus, the base stations could then be set up in their actual positions in real deployment, rather than currently on the floor in the project.

7.2.5 Understanding Bluetooth Settling Time

The Bluetooth's settling time reported in this project still cannot be explained by just inspecting the Bluetooth specification. The shape of the Bluetooth signal during this interval cannot be predicted either. Although this settling interval does not affect the overall system accuracy, it would still be more beneficial to skip this process to proceed to the stable state as soon as possible. Further research would be to analyse this state to predict the possible signal trend.

7.3 Summary

A promising affordable indoor tracking system with Bluetooth technology and the fingerprinting technique was implemented in this project. The application of a robot to collect the Bluetooth data more accurately and more efficiently, as well as the use of the Bat system have improved the system performance. The study of relevant Bluetooth properties for the indoor tracking context confirms the reliability and the robustness of the system over the long-term. Although the system does have some weaknesses, which hinder the real-life deployment, future research plans show that after overcoming those short to mid-term technological challenges, the existence of such a system on the market is a near certain reality.

Appendix

Table 1: Histogram Distribution Table Reference

Degrees of freedom	0.1	0.05	0.01	0.001
1	6.31	12.71	63.66	636.62
2	2.92	4.30	9.93	31.60
3	2.35	3.18	5.84	12.92
4	2.13	2.78	4.60	8.61
5	2.02	2.57	4.03	6.87
6	1.94	2.45	3.71	5.96
7	1.89	2.37	3.50	5.41
8	1.86	2.31	3.36	5.04
9	1.83	2.26	3.25	4.78
10	1.81	2.23	3.17	4.59
11	1.80	2.20	3.11	4.44
12	1.78	2.18	3.06	4.32
13	1.77	2.16	3.01	4.22
14	1.76	2.14	2.98	4.14
15	1.75	2.13	2.95	4.07
16	1.75	2.12	2.92	4.02
17	1.74	2.11	2.90	3.97
18	1.73	2.10	2.88	3.92
19	1.73	2.09	2.86	3.88
20	1.72	2.09	2.85	3.85
21	1.72	2.08	2.83	3.82
22	1.72	2.07	2.82	3.79
23	1.71	2.07	2.82	3.77
24	1.71	2.06	2.80	3.75

Table 2: Histogram Distribution Table Reference

Degrees of freedom	0.1	0.05	0.01	0.001
25	1.71	2.06	2.79	3.73
26	1.71	2.06	2.78	3.71
27	1.70	2.05	2.77	3.69
28	1.70	2.05	2.76	3.67
29	1.70	2.05	2.76	3.66
30	1.70	2.04	2.75	3.65
40	1.68	2.02	2.70	3.55
60	1.67	2.00	2.66	3.46
120	1.66	1.98	2.62	3.37
∞	1.65	1.96	2.58	3.29

References

[1] F. J. Massey: *"The Kolmogorov-Smirnov test for goodness of fit"* (1951).

[2] H. W. Lilliefors: *"On the Kolmogorov-Smirnov test for normality with mean and variance unknown"* (1967).

[3] R. Want, A. Hopper, V. Falcao, J. Gibbons: *"The active badge location system"* (1992).

[4] A. Ward, A. Jones, A. Hopper: *"A new location technique for the active office"* (1997).

[5] A. Harter, A. Hopper, P. Steggles, A. Ward, P. Webster: *"The anatomy of a context-aware application"* (1999).

[6] P. Enge, P. Misra: *"Special issue on GPS: The global positioning system"* (1999).

[7] P. Bahl, V. N. Padmanabhan, A. Balachandran: *"Enhancements to the RADAR user location and tracking system"* (2000).

[8] P. Bahl, V. N. Padmanabhan: *"RADAR: An in-building RF-based user location and tracking system"* (2000).

[9] P. Castro, R. Muntz: *"Managing context for smart space"* (2000).

[10] N. B. Priyantha, A. Chakraborty, H. Balakrishnan: *"The Cricket location-support system"* (2000).

[11] N. J. Muller: *"Bluetooth Demystified (McGraw-Hill Telecom)"* (2000).

[12] G. Chen, D. Kotz: *"A Survey of Context-Aware Mobile Computing Research"* (2000).

[13] M. Addlesee, R. Curwen, S. Hodges, J. Newman, P. Steggles, A. Ward, A. Hopper: *"Implementing a sentient computing system"* (2001).

[14] P. Castro, P. Chiu, T. Kremenek, R. Muntz: *"A Probabilistic Room Location Service for Wireless Networked Environments"* (2001).

[15] J. Hightower, G. Borriello: *"Location systems for ubiquitous computing"* (2001).

[16] A. Smailagic, D. P. Siewiorek, J. Anhalt, D. Kogan, Y. Wang: *"Location Sensing and Privacy in a Context Aware Computing Environment"* (2001).

[17] R. Battiti, M. Brunato, A. Villani: *"Statistical learning theory for location fingerprinting in wireless lans"* (2002).

[18] Filiz Bektas: *"Investigation of Antenna Diversity Techniques for Bluetooth Communication"* (2003).

[19] M. A. Youssef, A. Agrawala, A. U. Shankar, S. H. Noh: *"A probabilistic clustering-based indoor location determination system"* (2002).

[20] A. M. Ladd, K. Bekris, A. Rudys, G. Marceau, L. E. Kavraki, D. S. Wallach: *"Robotics-based location sensing using wireless ethernet"* (2002).

[21] W. Stallings: *"Wireless Communications and Networks"* (2002).

[22] D. Pandya, R. Jain, E. Lupu: *"Indoor location estimation using multiple wireless technologies"* (2003).

[23] H. Balakrishnan, R. Baliga, D. Curtis, M. Goraczko, A. Miu, B. Priyantha, A. Smith, K. Steele, S. Teller, K. Wang: *"Lessons from Developing and Deploying the Cricket Indoor Location System"* (2003).

[24] R. K. Harle: *"Maintaining World Models in Context-Aware Environments"* (2004).

[25] A. Smith, H. Balakrishnan, M. Goraczko, N. Priyantha: *"Tracking moving devices with the cricket location system"* (2004).

[26] K. Kaemarungsi: *"Properties of indoor received signal strength for WLAN location fingerprinting"* (2004).

[27] A. Ali, L. A. Latiff, N. Fisal: *"GPS-free indoor location tracking in mobile ad hoc network (MANET) using RSSI"* (2004).

[28] G. Dedes, A. G. Dempster: *"Indoor GPS positioning - challenges and opportunities"* (2005).

[29] K. Kaemarungsi: *"Design of indoor positioning systems based on location fingerprinting"* (2005).

[30] M. Youssef, A. Agrawala: *"The Horus WLAN Location Determination System"* (2005).

[31] K. W. Kolodziej, J. Hjelm: *"Local Positioning Systems: LBS Applications and Services"* (2006).

[32] R. Grossmann, J. Blumenthal, F. Golatowski, D. Timmermann: *"Localization in Zigbee-based Sensor Networks"* (2007).

[33] A. S. Huang, L. Rudolph: *"Bluetooth Essentials for Programmers"* (2007).

[34] V. Honkavirta, T. Perala, S. Ali-Loytty, R. Piche: *"A comparative survey of WLAN location fingerprinting methods"* (2009).

[35] G. Yanying, A. Lo, I. Niemegeers: *"A survey of indoor positioning systems for wireless personal networks"* (2009).

[36] S. Hay, R. Harle: *"Bluetooth tracking without Discoverability"* (2009).

[37] K. Curran, E. Furey, T. Lunney, J. Santos, D. Woods, A. Mc Caughey: *"An Evaluation of Indoor Location Determination Technologies"* (2011).