Smart Homes 1

Weihua Luo[1]*, Rabab Alkhalifa[2]*, Jiayun Wang[3]*, Nyasha Masamba[4]*, Boyang Zhao [5]*

wl16595   *ra16121*   *jw16097*   *nm16667*   *bz16563*
33891   *31140*   *32985*   *32928*   *33498*

Abstract
Sensors, in their many forms, have the ability to transform how people live. Introducing sensors into a home could potentially give us the ability to analyse and subsequently predict frequent activities in order to make homes smarter. In this project, we seek to gain an understanding of sensor data obtained from within a home environment. Using data obtained obtained from a wrist-worn accelerometer, video and depth (RGB-D) data and passive environmental sensors in a home context, we aim to gain insights into the actions and activities conducted by the participants. The data is mainly explored from a visualisation standpoint, and the main deliverable of the project is a web application which enables quick, and easy interactive visualisation of the data.

Keywords
data science, sensor, video, data, programming, visualisation, statistics, machine learning

***Department of Computer Science, University of Bristol, Bristol, United Kingdom**

Introduction

This section introduces the problem to be solved, along with associated aims and objectives of the project.

0.1 The Problem

This work will use the SPHERE dataset [4] as the basis for analysis . It has been argued that traditional regimes of healthcare are in need of an overhaul to better the experience of patients, relatives, carers and healthcare professionals. Ambient assisted living (AAL) has been proposed as a potential solution. To this end, the EPSRC-funded "Sensor Platform for HEalthcare in Residential Environment (SPHERE)" Interdisciplinary Research Collaboration (IRC) [1] designed a multimodal system driven by data analytics requirements. The system was tested in one smart home, and will be deployed in a general population of 100 homes in Bristol (UK). The data from the smart home were collected from the following three primary sensor modalities:

- Wrist-worn accelerometer and Received Signal Strength Indication (RSSI) data

- RGB-D cameras (i.e. video with depth information)

- Passive environmental sensors

Accompanying these data are annotations on participant location within the smart home, as well as annotations relating to the activities they were performing at the time. The floor plan of the smart home environment is shown in Figure 1. Nine rooms can be seen in this figure, namely bathroom, bedroom 1, bedroom 2, hallway, kitchen, living room, stairs, study and toilet. In addition, twenty activities of daily living (ADL) are annotated within this dataset. These ADLs are: ascent stairs, descent stairs, jump, walk with load, walk, bending, kneeling, lying, sitting, squatting, standing, stand-to-bend, kneel-to-stand, lie-to-sit, sit-to-lie, sit-to-stand, stand-to-kneel, stand-to-sit, bend-to-stand and turn. The ADLs are further categorised into 1) ambulation activities (i.e., an activity requiring continuous movement, e.g. walking); 2) static postures (i.e., times when the participants are stationary, e.g. standing or sitting); and 3) posture-to-posture transitions (e.g. stand-to-sit, stand-to-bend) [2].

Participants in the smart home experiment were asked to follow a script as they performed a variety of tasks around the house. Each participant created a full sequence of sensor data based on the tasks they performed [4]. The data were received by us as CSV and JSON files. There were testing and training sequences available, but we concentrated on the 10 training sequences as they were complete. Although the data were all collected from the same home, they were largely disjointed and could not be easily understood. Hence, the biggest challenge we faced during this project was how we could extract useful information from the seemingly disjointed sensor data, and what added value this useful information would bring.

0.2 Aims and Objectives

The aim of our project is to visually gain insights about participant activities, using sensor data, in order to subsequently aid machine learning predictions. We realised that with these sensor data, we could infer patterns of behaviour. The information derived from these patterns would be helpful for machine learning practitioners (MLP) when making predictions based on the data. For example, they can cross-check new data with
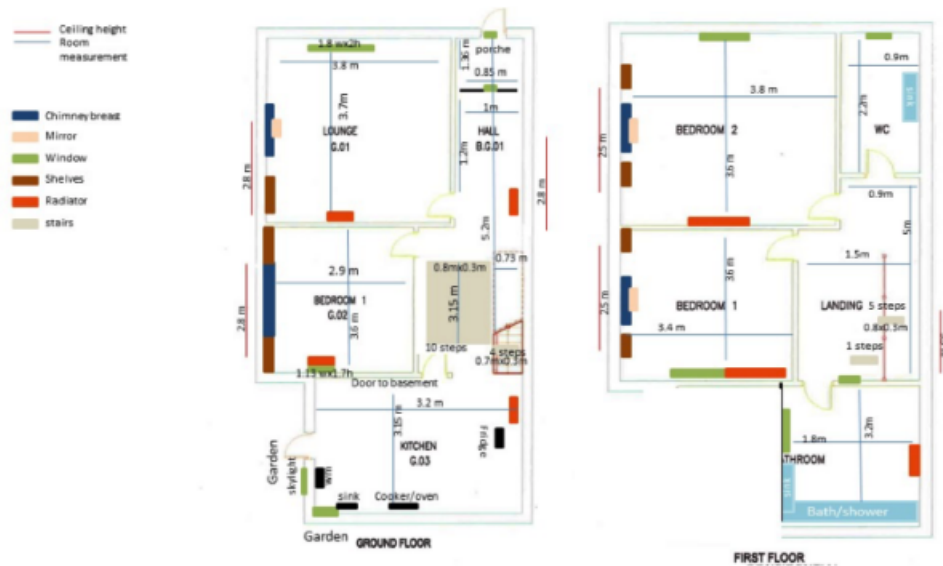
Figure 1. Smart home floor plan [4]

the visualisations to find inconsistencies in the new data; or they can visualise the most probable outcomes before training their models, and hence have an inclination to notice when the their models are wrong/making incorrect predictions. The predictions from machine learning could be used to conduct medical diagnoses remotely, energy could be saved when the home is predicted to be vacant, and the home's layout could be adapted to avoid probable trips and falls. We focused on providing various ways to visualise the data so that a MLP has several tools at their disposal to visualise the data before training a model and making predictions. Our main research questions were:

1. Can the data be visualised in a quick, easy and interactive way?

2. Are there any unsuspected relationships or inconsistencies in the data that a MLP might find interesting?

3. Can visualisations help MLPs gain higher confidence in their models?

These research questions are congruent with the aim of the project. In order to achieve the aim and answer the questions, we started off with three major objectives. First, we would use diverse preprocessing methods to clean the data in order to make it convenient for us to understand the data. Second, we would employ advanced computational methods to store and retrieve the data. Third, we would use interactive visualisation techniques to dynamically visualise the data, which is the main objective towards solving the problem.

0.3 Assumptions and Scope

It was assumed from the onset of the project that we would not be analysing streamed data (from the modalities) in real time. This simplifies the problem and allows us to concentrate on solving the scientific aspect of the problem rather than the infrastructural one. Furthermore, although the solution would be aiming to help MLPs with their work, carrying out the machine learning task is beyond the scope of this project.

1. Data Science

This section explains our understanding of the data science field and how data science experiments should be carried out.

1.1 What is Data Science?

Data Science is the use of statistics, data analysis and related methods to understand or extract information from data. Other overlapping fields include machine learning and data mining. The most common goal of data science is to extract knowledge from data, or process the data into an understandable structure in order to reach informative conclusions. The data science process usually involves stages such as data collection and storage, data pre-processing, data exploration, data modelling and publishing.

1.2 Data Science Tools

As data science has evolved as a discipline, several tools have evolved to serve the needs of data scientists. From proprietary suites such as MATLAB, SPSS and SAS to open source languages such as R and Python, the choices are vast, making it crucial to pick the correct tools for a given project. For our exploratory data analysis, the main tools used were mongoDB (for storage) and Python (for data manipulation). The motivation for selecting these tools is discussed below.

1.2.1 MongoDB

After data has been collected, it needs to be stored in a form that makes it suitable for future querying and analysis. CSV files are not very suitable for this; they do not provide a mechanism for external querying of the data. A database was need for this purpose, and mongoDB was chosen as the most suitable datastore [3]. First of all, the data did not exhibit any complex relationships, so a relational database was not necessary. Furthermore, the 'schemaless' nature of mongoDB meant that we did not need to design a schema, freeing up more time for developing the tools to analyse data. Since there was no need for guaranteed consistency, ACID transactions (as featured in relational databases) were also not required. Trading off guaranteed consistency for eventual consistency means that mongoDB is faster than a traditional Structured Query Language (SQL) based database like mySQL. In addition, mongoDB's indexing capabilities, coupled with its simple query language, made it a great choice when working with several files of time series data. It also handles CSV and JSON file data easily, it is open source, and the data were not being analysed in real time hence there was no need for a real-time datastore. Another option considered for the datastore was CouchDB, but mongoDB seemed more popular and had a larger support community in its favour.

1.2.2 Python

An essential part of data science is the need to 'hack' and manipulate data. For that, a programming language was needed. We chose Python. The statistical language, R, was also considered. However, Python gave us more scope to achieve our exploratory visualisations where R showed some limitations. Python has emerged over the last decade as a first-class tool for scientific computing tasks, including the analysis and visualisation of large datasets. In particular, the usefulness of Python for data science stems primarily from the large and active ecosystem of third-party packages, a number of which were used in this project: Pandas dataframe for manipulation of data ingested from mongoDB, Matplotlib for publication-quality visualisations, Bokeh for interactive web app visualisations (similar to D3.js) and ffmpeg for video data synthesis. Many other tools used during the project, such as Jupyter notebooks, also work with Python and provide the computational environment in which many data scientists work [5].

2. Data Science System Architecture

This section elaborates on how the deliverables were created through each stage in the data science process.

2.1 Storing and loading data

After choosing mongoDB as the database to store the data (as the smart home data is in CSV and JSON format), the advantages of mongoDB started to become apparent as the datastore had a highly flexible storage format. This meant we could handle and store both CSV and JSON data conveniently during the data storage process. The other advantage of mongoDB, i.e. that it performs at a relatively high data I/O operation speed, also came in handy. As we were planning to construct an interactive visualisation tool, the speed of data reading from the database and the strength to ensure the stability of the data's structure were crucial for the performance of the visualisation tool. Figure 2 shows the planned architecture for the web app.
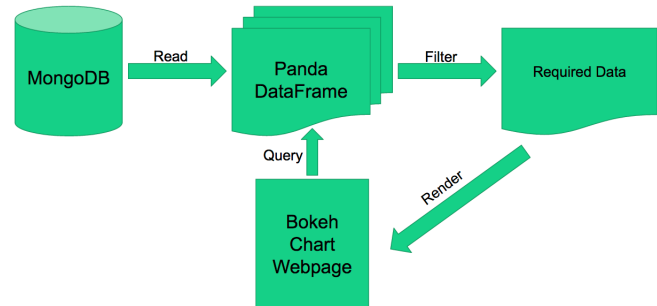

**Figure 2.** System architecture

Data will be ingested from the mongoDB datastore and manipulated using the Pandas dataframe. An incoming query from the web app will produce filtered data from the dataframe, which will subsequently be rendered by the Bokeh library on the web page.

2.2 Data wrangling
2.2.1 Data cleaning

Because the data received from the SPHERE project had already been largely cleaned, the main reason for performing data cleaning at this point during the project was to remove outliers. Our method was to first plot the data, which makes it easy to visually notice the outliers. To clean the acceleration data, of which the x, y, z axes represent the coordinates of the accelerometer, we worked from the assumption that the accelerometer (bound to the wrist) moved within a fixed size space. So, our method of finding the outliers was to plot the coordinates in a 3-dimensional space. Figure 3 shows the resulting plot. The outliers can be recognized distinctly from the figure. Different colours indicate different rooms.

Further work to find the outliers existing in the RSSI signal within the acceleration data file is based on aggregation of the file with location data, which will be discussed in more detail below. The idea is that the sensor reflecting the strongest signal should be the nearest one to the participant. Hence, if we check the signal strength with respect to location, the outlier data will be uncovered by showing the expected and actual sensor presenting strongest signal strength.
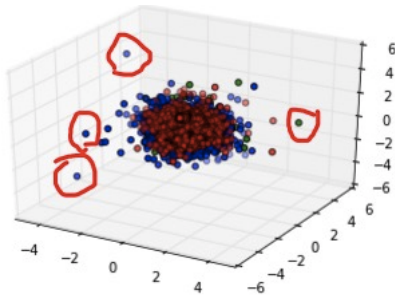
Figure 3. Outlier detection

2.2.2 Missing data

We focused on the acceleration data for missing data management, where the signal reading of sensors in some rooms was missing from the data. After intensive observation of the data, we found that the missing data points were likely due to the fact that there was no signal received from the sensor in question. This means there was somehow a loss of signal data (possibly due to hardware failure, or maybe the participant was completely out of the sensor's range). However, the problem was that these missing data would cause problems in the next steps of our study. By running a statistical count algorithm, we found the missing data of each column to be as shown in Table 1.

Data column	Count of missing data
x	0
y	0
z	0
Kitchen_AP	24,317
Lounge_AP	14,160
Upstairs_AP	20,985
Study_AP	35,710

Table 1. Missing data from acceleration file

With the exception of the x, y, z coordinates in the acceleration data, there is a huge amount of missing data from the other columns. The Study_AP data is completely empty (we were informed to disregard this by SPHERE project members). In order to solve the missing data problem, we calculated the mean, standard deviation, minimum and maximum values of the data. We also found where the first quartile, median, and third quartile thresholds of the original data were. The data in Figure 4 shows our results. We found that the maximum and minimum were not good enough to fill up the gap, because they are extreme values. However, using the mean of the column as the value to fill the gap had the effect of reducing the error and standard deviation, so we used the column mean to address missing values in the data.

2.2.3 Data aggregation

After intently studying these diverse data for a while, we found the key index of all the different data documents is time(t).

Hence we focused on integrating the annotation data and location data together in one collection, based on time. For example, the location file has detailed entries of each record while the annotation data has short entries of each record. The common element between them was time. We realised that if we split each record in the two files by time, the time field would make it convenient for us to combine the two data documents. Such aggregation of data was the foundation for us to research the relationship between the two types of data and to prepare the data for visualisation. An illustration of this method is shown in Figure 5.

We also used time as the key index to aggregate other files such as the acceleration data with location data; and also to aggregate the three video files in each complete sequence. The former was in preparation for finding the relationship between the accelerometer signal and location. The latter was in preparation for finding the flow of video data through each sequence, and the aggregation is depicted in Figure 6. First, we merged all video files (kitchen, living room, hallway), ensuring they were correctly labelled (by location). We noticed from the data that the difference between each data point in the video files was 0.3 seconds. Hence we adapted the time field in the location_0 file to discrete time intervals which were 0.3 seconds apart. Finally, we sorted the entries (by time).

2.3 Data exploration and visualisation

2.3.1 Web app interactive visualisation

We produced an exploratory web interactive visualisation application which is able to dynamically present information about participant activities, as shown in Figure 7. The location of a participant can be seen interactively in the floor plan as time progresses (using a time slider). A Gantt chart is created to show the location of the participant, and a MLP using the app is able to choose which person they want to observe. There are also statistical analyses of the location data and annotation data by 2-dimensional and 3-dimensional histograms. When the page loads, the statistics for all 10 training sequences are loaded. From there, the user can also decide the room or the Person ID whose statistics they would like to see. The 3-dimensional histogram shows the pattern between annotation and location, which can be customised by changing the Person ID or choosing all of the participants for example. Meanwhile, the 2-dimensional histogram shows the statistical data of the annotations/ADLs, also interactively customisable by changing the Person ID and/or location.

To aid clarity on some of the functionality, Figure 8 shows an examplary depiction of the dynamic behaviour of a participant in the living room as time progresses (on the time slider).

2.3.2 Video visualisation

Since video data came in the form of bounding box coordinates (for participants' privacy), the MLP would benefit from

	t	x	y	z	Kitchen_AP	Lounge_AP	Upstairs_AP	Study_AP
count	35710.000000	35710.000000	35710.000000	35710.000000	11393.00000	21550.000000	14745.000000	0.0
mean	910.075092	0.400164	-0.403389	0.238920	-78.26174	-77.575452	-84.933198	NaN
std	526.783375	0.538850	0.475885	0.366541	10.83795	8.776482	10.240221	NaN
min	0.017856	-2.002000	-4.096000	-2.640000	-105.00000	-105.000000	-105.000000	NaN
25%	455.517416	-0.032000	-0.804000	0.048000	-83.00000	-81.000000	-93.000000	NaN
50%	908.063928	0.454000	-0.396000	0.176000	-75.00000	-75.000000	-85.000000	NaN
75%	1367.216440	0.920000	-0.170000	0.508000	-71.00000	-72.000000	-77.000000	NaN
max	1823.866952	4.094000	2.108000	2.384000	-60.00000	-63.000000	-54.000000	NaN

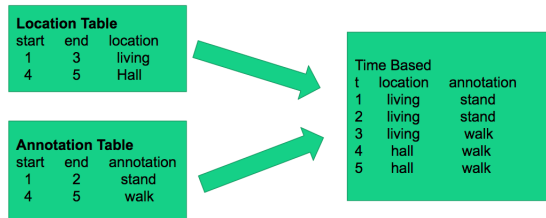**Figure 4.** Descriptive statistics for each data column

Figure 5. Data aggregation by combining tables
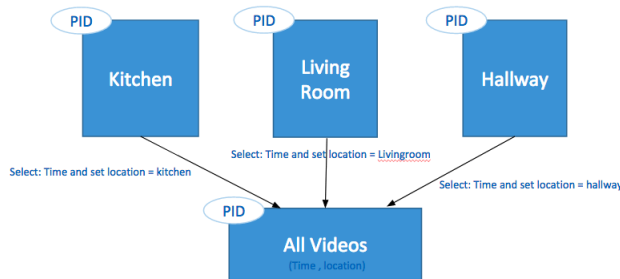

Figure 6. Data aggregation of video files

viewing something more friendly to the human eye to see, in a real-life setting, how some parts of a sequence happened. Hence, the video visualisations were a creative way to solve this problem, particularly considering the location photograph in the background dynamically changes as the model of the participant (a red box based on bounding box coordinates) also moves around the home. Figure 9 shows what the video visualisation tool looks like. Note that in real life, this visualisation shows dynamic backgrounds and motion of the bounding box as time progresses.

2.3.3 Jupyter notebook visualisations

The Jupyter notebooks allowed interactive execution and sharing of Python code, and supplemented the two forms of visualisation above by allowing interactive graphing of the data. Most of the stand alone figures included above were created in Jupyter notebook. Due to the interactivity of Jupyter notebooks, we could write code that could easily be changed for further visualisation by the MLP as and when needed. For example, to similarity between certain sequences, the MLP

could plot two gantt charts next to each other, such as in Figure 10 below.

Plotting RSSI values overlaid with the triggered access point (location) also became possible, as can be observed in the notebook visualisation in Figure 11. This enabled us to see relationships between the RSSI and the triggered access point.

3. Results

This section presents our observations as seen directly from the data science tools we built.

3.0.1 Web app

The web application, which is the MLP's main tool, allows interactive analysis of each person's behaviour, or the behaviour of all the training sequences as a whole. Listed below are a few generalisations made from the interactive visualisation:

1. It is apparent from the dynamic gantt chart that all the participants spend the most time in the living room, followed by the kitchen. They also spend the least amount of time in the toilet, bath room, study room and bedroom 1. Almost all the time, participants walk and stand in these rooms except occasional sitting in the toilet.

2. Participants are sitting in the living room most of the time. What is interesting is that they do not sit continuously for a long time, but rather they frequently stand up and sit down as time progresses, accompanied by a little walking. Conversely, participants spend more time standing in the kitchen accompanied by occasional walking. Combining with the finding in the video visualisation, participants take a lot of time standing before the cooker/doing some kitchen work.

3. A significant finding is the difference in behaviour of participants in bedroom 1 and 2. Participants never lie and sit in bedroom 1 but spend most time sitting and lying on the bed in bedroom2. This finding is able to distinguish the location of the person.
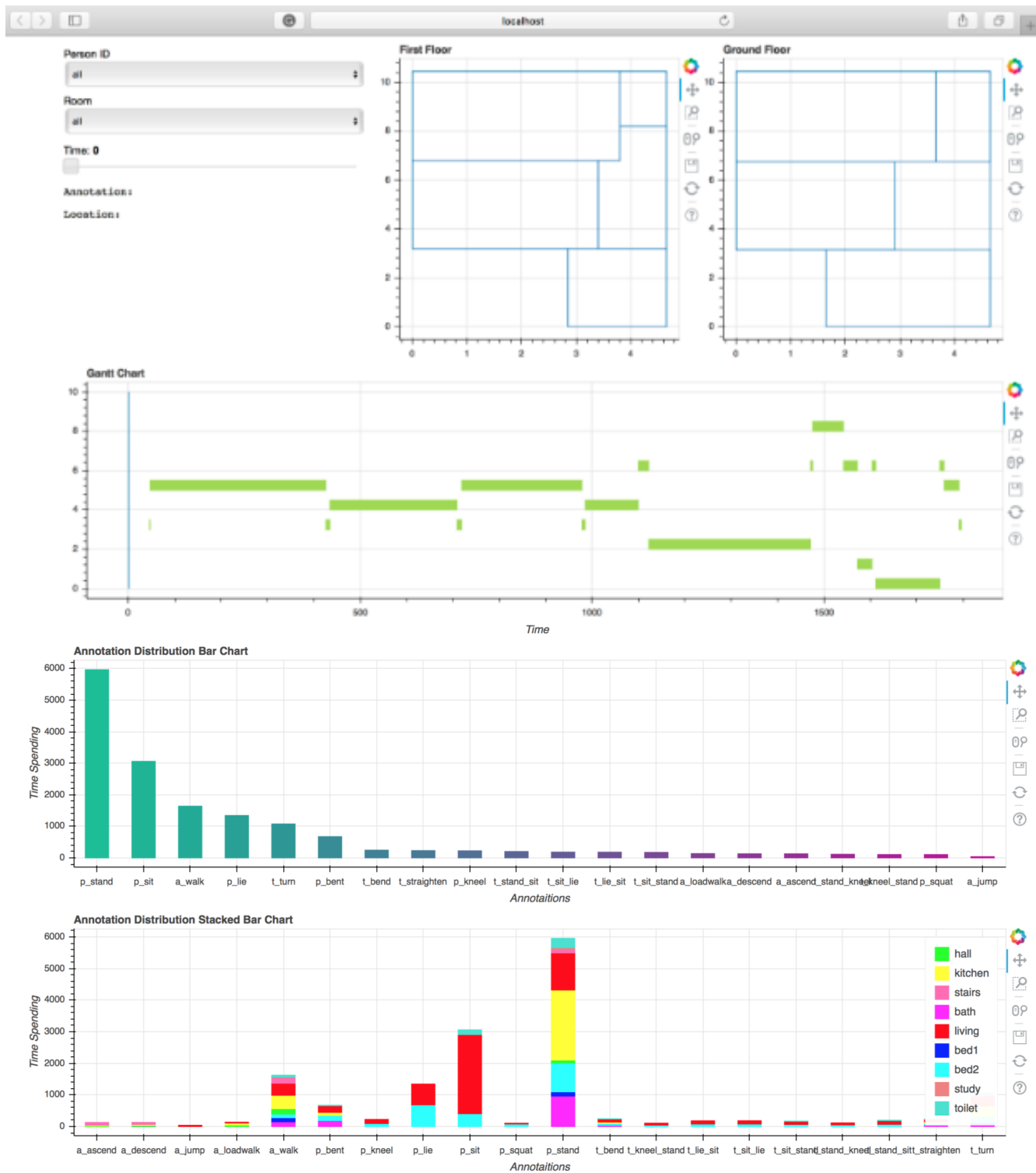
Figure 7. Web application interactive visualisation

4. When the participant shows a behaviour of sitting, he only has the possibility to be in living room, bedroom2 or toilet. But participants sit more than lying in the living room and show opposite behaviour in bedroom 2. Meanwhile, they never lie in the toilet. This finding is able to help distinguishing the location in these three rooms.

5. A unique behaviour of participants in the living room is jumping. From observing the app visualisation, it is determined that participants never jump in other rooms.
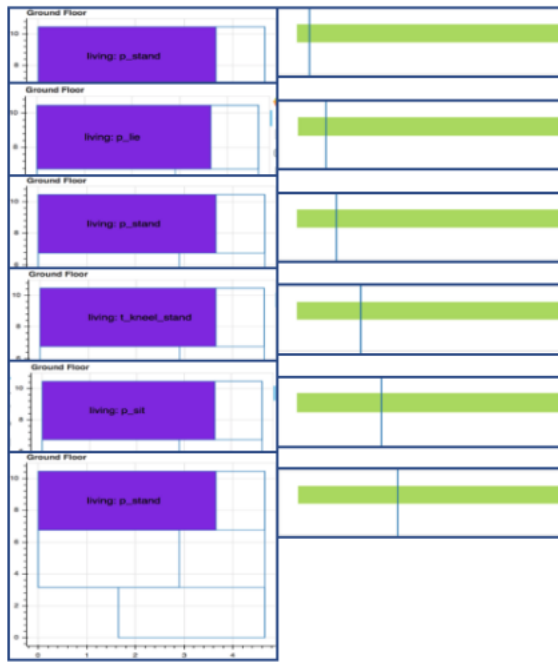
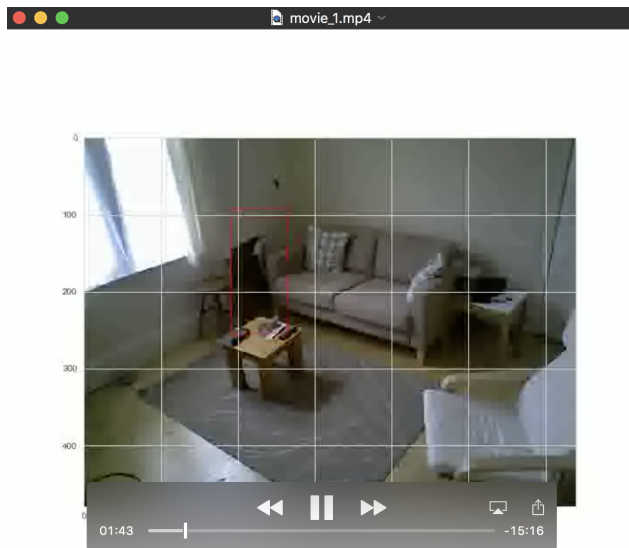**Figure 8.** Illustration of dynamic behaviour with time

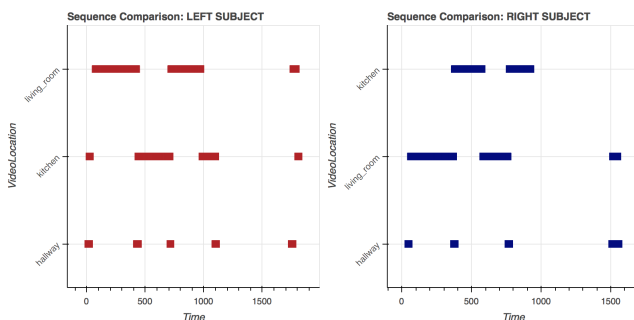**Figure 9.** Still image of video visualisation

Figure 10. Sequence comparison in Jupyter notebook

6. The participants spend little time in bedroom 1. All participants only enter the room, walk around and turn around swiftly then leave the room.

7. Participants generally just walk in the hall without turning around (almost all the time). This is in contrast the behaviour exhibited in bedroom 1.

8. The participants only descend or ascend in the stair area. That is the sole activity in the stair area.

3.0.2 Video

Videos give the MLP a way to see exactly how a person moved around the home in a life-like manner. On observing the videos along with the video data, we observed the following:

1. The camera can only detect the moving part of the body, so the bounding box in the video sometimes can be deceivingly small representing only some partial body detection.

2. Some video data are missing, presumably because the person is in a different room at the time, meaning the bounding box coordinates would have been recorded as zero values and then subsequently removed when the data were cleaned.

3. The camera/bounding box sometimes struggles to capture people's frames accurately, for example when they are standing in front of a window, only the upper part of their body is detected. Maybe that is because only the head is moving (albeit slightly).

4. The video flow was smooth, suggesting the data capture equipment had a high degree of accuracy .

3.0.3 Jupyter notebook

The Jupyter notebooks give the MLP a way to visualise the data interactively. Some of the observations that can be made using this tool are as follows:

1. By plotting two video sequences side by side, it can be observed, that the video data for the participant on the left and the participant on the right are different. Hence, the way in which they performed the script is also different. All the MLP has to do to come up with this visualisation is to change two lines of code, which is very handy.

2. By comparing the merged video data from the three locations against all other location data as measured by the environmental sensors in Jupyter notebook. We observed that because the data are clean, the time sequences are complementary and generally do not intersect. In fact, they show the progression of the time sequence as the person performed the tasks, as seen in Figure 12 below. The coloured shapes represent video data (blue - living room; green - hall; red - kitchen), and the thicker the shape, the more video data available. As
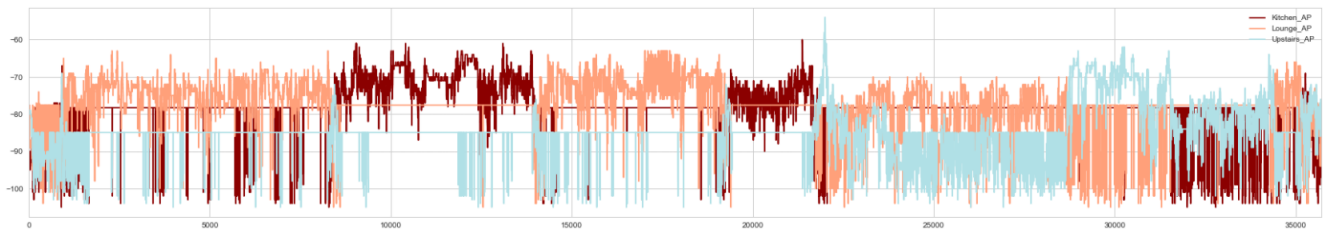
Figure 11. RSSI strength vs location

can be seen from the figure, the person (ID 1) travels from the living room through the hall to the kitchen, then back to the living room and so forth until the sequence completes in the hall.
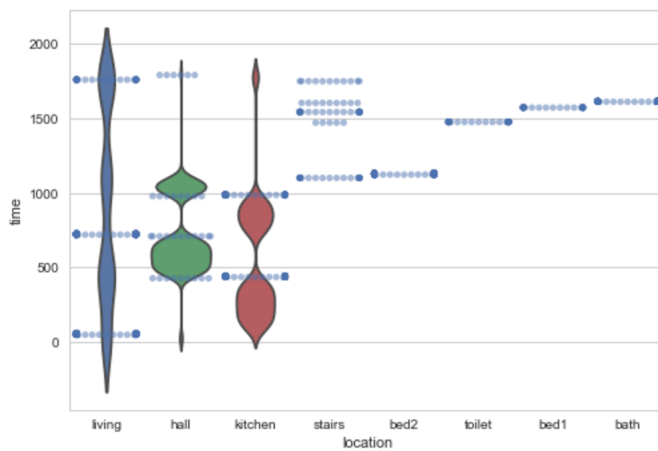

Figure 12. Video data vs location data

4. Insights

This section provides an evaluative analysis of the insights and generalisations we made from our observations.

- By considering the floorplan and the list of ADLs together, it should be clear to see that there are correlations between these: since there is no seating area in the kitchen, it is very unlikely that somebody will sit in the kitchen. However, since there are many seats in the living room, it is much more likely that somebody will sit in the living room than in the kitchen. Additionally, one can only ascent and descent stairs when one is on the staircase. It is also obvious that the signal strength of sensors in different rooms, to a extent, reflects the location of participant, as the nearest sensor should receive the strongest signal.

- In some instances, the location cannot be precisely determined. PIR sensors can elicit false negative and false positive activations - i.e., they do not always trigger when someone is present, and they can trigger when someone is not present due to the infrared radiation

from the sun in warm weather. Video data obtained from across the whole house would have gone a long way to solving this location issue, as the camera does not suffer from the same problems. The MLP can have more confidence about the location of a participant if they are in a location which is fitted with a camera. The association rule might look like the following:

IF (kitchen PIR sensor activated) and (kitchen video activated) THEN Location = Kitchen

Once a location is known, we can consequently have more confidence in the likely activity being done, if we also consider annotations. For example, the association rule in this case would become:

IF (kitchen PIR sensor activated) and (kitchen video activated) THEN ADL = standing

The above could be used by the MLP to infer that the participant could be cooking or washing plates at that particular time, for example. Furthermore, both location and ADL could be targets for machine learning classification.

- After we aggregated the acceleration data and location data by time, we realised there is a relationship between signal strength and location. Combining with the room map and the sensor location, the patterns we extracted are as follows:

IF(the signal strength in lounge sensor is the strongest) and (the signal strength in other two sensors is weak) THEN the location = living room.

IF(the signal strength in kitchen sensor is the strongest) and (the signal strength in other two sensors is weak) THEN the location = kitchen.

IF(the signal strength in lounge sensor is around 100) and (the signal strength is around 70-75) and (the signal strength in kitchen is missing or around 78) THEN the location = bedroom 1

IF(the signal strength in upstair sensors is around 85-100) and (the signal strength in lounge sensors is around 70-90) and (the signal in kitchen is missing) THEN the location = bedroom 2

IF(the signal strength in upstair sensors is around 65-70) and (the signal strength in lounge sensors is around

85-100) and (the signal in kitchen is missing or around 100) THEN the location = toilet

IF(the signal strength in upstair sensors is around 70) and (the signal strength in lounge sensors is around 90-100) and (the signal in kitchen is missing or around 100) THEN the location = stairs

IF(the signal strength in upstair sensors is around 85-100) and (the signal strength in lounge sensors is around 75-100) and (the signal in kitchen is around 70-100) THEN the location = hall

- From behaviours of all participants, several patterns can be extracted from the observation of the web tool above.

IF(the participant is sitting down and standing up alternatively) THEN the location = living room

IF(the participant is standing more than 250 seconds) and (the participant is never sit down) THEN the location = kitchen

IF(the participant spend more time in sitting down more than in lying) THEN the location = living room

IF(the participant spend more time in lying more than in sitting down) THEN the location = bedroom 2

IF(the participant only sit down but never lie) THEN the location = toilet

IF(the participant is jumping) THEN the location = living room

IF(the participant spend most time for walking and loadwalk) THEN the location = hall

IF(the participant has behaviours of descend and ascend) THEN the location = stairs

IF(the participant only stands, walks and turns in a continuous manner) THEN the location = bedroom 1

5. Future Work

Future work will largely be devoted to performing activity and location recognition using machine learning on the data, using the insights gained from this data science process. A number of features make this dataset valuable and interesting to MLPs [2]:

- It features missing data.

- The data are temporal, and correlations in the data must be captured either in the features or in the modelling framework.

- There are many correlations between activities.

- The problem can be modelled in a hierarchical classification framework.

- The targets (ADLs) are probabilistic (since the annotations of multiple annotators were averaged).

- The optimal solution will need to consider sensor fusion techniques.

Given more time, it would have been interesting to tackle the problem further and build a tool for the MLP that analyses the raw sensor data and predicts the activity, based on parameters set by the MLP. For example, the tool could have looked something like the MATLAB suite shown in Figure 13 [6]. It takes in accelerometer data as an input, trains a neural network, predicts/estimates the ADL, calculates statistical correlations between activities, and even produces a pre-trained model in C++ for use outside the MATLAB environment. Such tools would greatly complement the tools we have already provided to the MLP.

6. Conclusion

Like any well-thought out data science experiment, we started off with three central questions in order to guide our aims and objectives. In this section, we will study of these questions to determine if the aim of the project has been achieved.

The first question pondered whether the sensor data we had been given could be visualised in a quick, easy and interactive way. The answer to this question is a resounding yes. In this project, we have delivered a suite of three tools which can visualise different aspects of the data to aid a machine learning researcher make sense of the data before using the machine learning data for their machine learning predictions.

The second question wondered if there were any unsuspected relationships or inconsistencies in the data that a MLP might find interesting. Using the tools we built, we found quite a few of these. Each tool we derived found some interesting patterns. For example, the web app visualisation enabled us to find out that jumping only takes place in the living room, therefore the MLP can say with a high probability that if jumping is detected by the sensors, the participant is likely to be in the living room. The video found that the bounding box is unreliable when the participant is standing next to a window; however the flow is typically smooth so the video equipment generally exhibits a high degree of accuracy. The Jupyter notebooks enabled us to see when sequences were different, enabling the MLP to compare these sequences and get a general sense of how scripts were performed differently by the participants without even touching the data themselves.

The third question finally asked whether the visualisations could help a MLP gain higher confidence in their models. Given that the MLP can get a deeper understanding of the data using the data science tools we are providing, it should be easier for them to spot when their model is behaving badly or if there is an area they need to revise. The results and insights we have provided, which include several association rules derived from the data, could also help clarify some issues
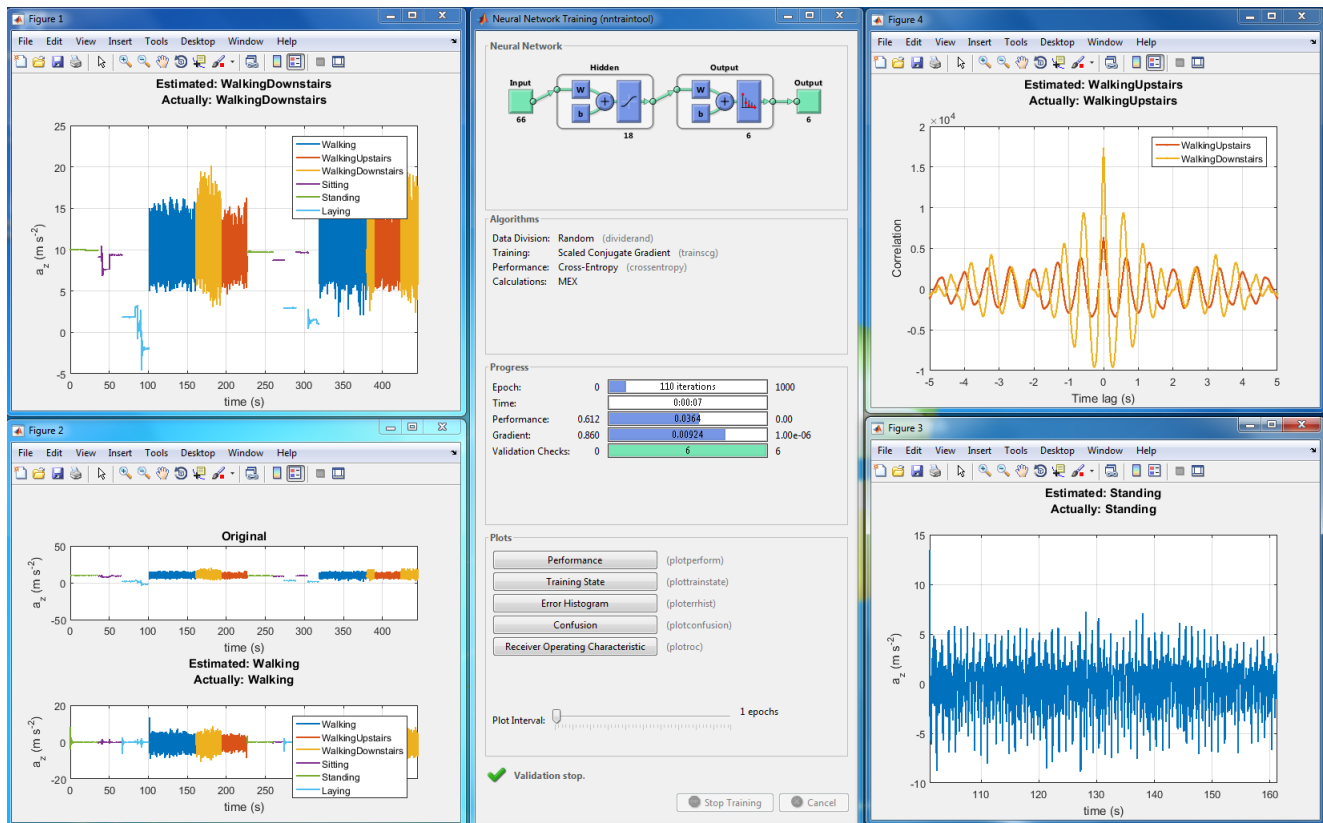
Figure 13. MATLAB neural network activity recognition suite [6]

regarding the targets the MLP might wish to predict using machine learning. Hence, the tools we have delivered should go a long way towards helping the MLP gain higher confidence in their models.

In conclusion, the project has achieved its aim, i.e to visually gain insights about participant activities, using sensor data, in order to subsequently aid machine learning predictions. Given more time, more functionality could have been added to the web app and more quantitative analyses would have been performed to complement the methods implemented in the project. As already alluded to, building tools for the machine learning process would have been our next major challenge, potentially also taking our analysis into the realm of analysing streaming, real-time sensor data.

References

[1] Sensor Platform for HEalthcare in Residential Environment (SPHERE), 2015. *SPHERE Home.* Available from http://www.irc-sphere.ac.uk/ [Accessed on April 2, 2017].

[2] Diethe, T., Twomey, N., Kull, M., Flach, P., and Craddock, I., 2017. *Probabilistic Sensor Fusion for Ambient Assisted Living.* Available from https://arxiv.org/pdf/1702.01209.pdf [Accessed on May 2, 2017].

[3] Moschetti, B., 2014. *MongoDB vs SQL: Day 1 - 2.* Available from https://www.mongodb.com/blog/post/mongodb-vs-sql-day-1-2 [Accessed on March 5, 2017].

[4] Diethe, T., Twomey, N., Kull, M., Flach, P., Song, H., Camplani, M., Hannuna, S., Fafoutis, X., Zhu, N., Craddock, I., and Woznowski, P., 2016. *The SPHERE Challenge: Activity Recognition with Multimodal Sensor Data.* Available from https://arxiv.org/pdf/1603.00797.pdf [Accessed on March 5, 2017].

[5] VanderPlas, J., 2017. *Python Data Science Handbook: Essential Tools for Working with Data.* O'Reilly Media, Sebastopol, CA.

[6] Bunkheila, G., 2015. *Signal Processing for Machine Learning.* Available from https://uk.mathworks.com/videos/signal-processing-for-machine-learning-99887.html [Accessed on April 2, 2017].