
# Reducing Domestic Energy Consumption Through Behaviour Modification

Thesis submitted for the degree of Doctor of Philosophy
Engineering Department, University of Oxford

Rebecca Ford
Christ Church, University of Oxford
Supervised by Dr Malcolm McCulloch

Trinity Term 2009

Rebecca Ford
Christ Church
Submission for D.Phil
Trinity Term 2009

# Abstract

This thesis presents the development of techniques which enable appliance recognition in an Advanced Electricity Meter (AEM) to aid individuals reduce their domestic electricity consumption. The key aspect is to provide immediate and disaggregated information, down to appliance level, from a single point of measurement.

Three sets of features including the short term time domain, time dependent finite state machine behaviour and time of day are identified by monitoring step changes in the power consumption of the home. Associated with each feature set is a membership which depicts the amount to which that feature set is representative of a particular appliance. These memberships are combined in a novel framework to effectively identify individual appliance state changes and hence appliance energy consumption.

An innovative mechanism is developed for generating short term time domain memberships. Hierarchical and nearest neighbour clustering is used to train the AEM by generating appliance prototypes which contain an indication of typical parameters. From these prototypes probabilistic fuzzy memberships and possibilistic fuzzy typicalities are calculated for new data points which correspond to appliance state changes. These values are combined in a weighted geometric mean to produce novel memberships which are determined to be appropriate for the domestic model.

A voltage independent feature space in the short term time domain is developed based on a model of the appliance's electrical interface. The components within that interface are calculated and these, along with an indication of the appropriate model, form a novel feature set which is used to represent appliances.

The techniques developed are verified with real data and are 99.8% accurate in a laboratory based classification in the short term time domain. The work presented in this thesis demonstrates the ability of the AEM to accurately track the energy consumption of individual appliances.

# Acknowledgments

First and foremost I would like to express my sincere gratitude to my supervisor, Dr Malcolm McCulloch, for his continued support and guidance through the duration of my research. During the last three years he has been the source of invaluable advice, useful suggestions and constant encouragement.

I would also like to express my appreciation to my examiners, Dr Chris Stevens and Dr Gareth Taylor, for both their time and their constructive comments and insights.

I am very much indebted to Dr Marcus Leong, Mr Jim Donaldson and Dr Tim Woolmer for their persistent recommendations, helpfulness and friendship during the time I spent researching and writing my thesis.

Finally, I would like to thank my family for their unending support, and in particular I would like to thank my brother, Benjamin Ford, for the time he spent and the effort he took in proof reading this thesis.

# List of mnemonics, abbreviations and acronyms

| | |
|---|---|
| ADC | Analogue to digital converter. |
| AEM | Advanced Electricity Meter - the name assigned to the device intended to implement the set of techniques developed in this thesis |
| BBN | Bayesian belief network. |
| $C$ | Capacitance. |
| $CL$ | Confidence level. The confidence to which a particular classification is made. |
| $CO_2$ | Carbon dioxide. |
| $dz/dt$ | The time based differential of signal $z$. |
| $\mathrm{d}(\mathbf{x}_i, \mathbf{c}_j)$ or $\mathrm{d}_{ij}$ | The distance between the $i^{th}$ feature vector $\mathbf{x}$ and the $j^{th}$ cluster center $\mathbf{c}$. |
| ECI | Energy consumption indicator. |
| FCM | Fuzzy C-means clustering. |
| FEB | Feature extraction block of the AEM system design. Developed in Chapter 6. |
| FSM | Finite state machine. |
| HMM | Hidden Markov model. |
| Hz | Frequency measure used in the UK. |
| $i$ or $i(t)$ | Time dependent current signal. |
| IM | Interface model. Interface models are developed in this thesis and are indicative of a particular type of electrical interface. |
| $L$ | Inductance. |
| $\mathbf{m}$ | Denotes the membership vector for a particular observed feature vector. |
| $m_i(j)$ or $m_{ij}$ | Denotes the membership value which indicates the level to which feature vector $i$ belongs to known appliance $j$. |

| | |
|---|---|
| MGB | Membership generation block of the AEM system design. Developed in Chapter 5. |
| NALM | Non intrusive appliance load monitoring. Denotes a series of techniques which are used to monitor appliance load profiles in a non-intrusive manner. |
| NN clustering | Nearest neighbour clustering. Clustering method which classifies by assigning new data points to the nearest known cluster. |
| PC | Personal computer. |
| PHC, NHC, TC | Positive half cycle, negative half cycle, total cycle respectively. |
| $P(z)$ | The probability of obtaining particular outcome $z$. |
| $P(z\|y)$ | The probability of obtaining particular outcome $z$ given a set of conditions $y$. |
| $R$ | Resistance. |
| RF | Radio frequency. Describes signals with frequency between about 3Hz and 300GHz. |
| RMS | Root mean square. The RMS value of a signal is the square root of the mean of the squares of the values that make up the signal. |
| $t$ | Time. |
| THD | Total harmonic distortion. |
| v-section | A segment of a time series which exhibits significant variation. |
| $v$ or $v(t)$ | Time dependent voltage signal. |
| $V_n$ | The magnitude of the $n^{th}$ voltage harmonic. |
| WPT | Wavelet packet transform. |
| $\mathbf{x}$ | This denotes the feature vector which is extracted from an individual appliance. |

| | |
|---|---|
| $\Delta i\,(t)$ | Change in current signal observed when an individual appliance changes state, i.e. indicative of the current drawn by that individual appliance. |
| $\Delta p\,(t)$ | Change in power signal observed when an individual appliance changes state, i.e. indicative of the power drawn by that individual appliance. |
| $\omega$ | The frequency, in radians, of the time dependent signals. For the mains voltage in the UK this value is approximately $100\pi$ radians (i.e. 50Hz). |
| $\sum_{j=1}^{n} a_j$ | The sum of $a_1 + a_2 + \ldots + a_{n-1} + a_n$. |
| $\sum_{j} a_j$ | The sum of $a_1 + a_2 + \ldots$ for variable $a$ over all values of $j$. |
| $\prod_{j=1}^{n} a_j$ | The product of $a_1 \times a_2 \times \ldots \times a_{n-1} \times a_n$. |

# Contents

# Chapter 1

# Introduction

The aim of this thesis is to develop a set of techniques to be implemented in an Advanced Electricity Meter (AEM) which gives homeowners useful information about the way in which they use electricity in the home so that they may be provided with the means and motivation to reduce their domestic energy consumption.

This chapter examines the broad context of the thesis aims, including an examination of the links between domestic energy consumption and global warming, and develops a roadmap which enable these aims to be satisfied.

Section 1.1 identifies the reasons behind the drive to reduce global warming which is shown to be linked to carbon dioxide emissions. Mechanisms by which carbon dioxide emissions may be reduced are explored in Section 1.2 and it is shown that a large reduction is possible by taking energy efficiency measures.

The importance of the domestic sector in potential savings from energy efficiency is discussed in Section 1.3. This sector accounts for approximately one third of all the energy consumed in the UK and the home is a place where individuals have a large amount of control over the way in which energy is consumed. Therefore, any significant savings made through energy efficiency measures taken in the domestic sector may result in significant reductions of carbon dioxide emissions in the UK.

Having explored the context within which this research is carried out, Section 1.4 identifies the set of objectives which must be met in order to establish an effective set of techniques which enable individuals to reduce their domestic energy consumption.

## 1.1 Why save energy? Global warming and the challenge to reduce carbon dioxide emissions.

This section is concerned with understanding why there is a current need to reduce energy consumption. This is done by illustrating the links between global warming (which has potentially catastrophic consequences), carbon dioxide emissions and energy consumption.

Global warming is the increase in the air temperature and ocean temperature of the earth. Figure 1.1 shows the smoothed annual temperature anomaly from 1850 to 2000 which clearly indicates a significant rise over the period in question. The catastrophic effects due to a global temperature increase are beyond the scope of this report, but more information can be found elsewhere [1].

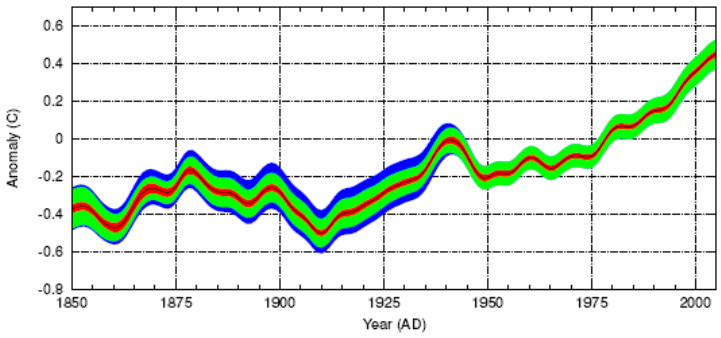


Figure 1.1: Temperature anomaly from 1850 to 2000. The zero level on the graph is the mean temperature from 1961 to 1990. This image is taken from a report by P. Brohan et al. [2].

These increasing levels of temperature that the world is experiencing due to global warming are closely linked to the increasing levels of carbon dioxide in the atmosphere as can be seen in Figure 1.2. This graph illustrates a comparison between solar variation, amounts

of the isotope $^{18}O$ of Oxygen gas, levels of methane, relative temperature and levels of carbon dioxide. The increase in concentration of these greenhouse gases is thus inevitably linked to the greenhouse effect, global warming and the effects of climate change the world is now experiencing. When the levels of carbon dioxide are high, the temperature variation is also high. Thus to reduce the rate of global warming, or even to stop it, the levels of carbon dioxide in the atmosphere must be reduced.

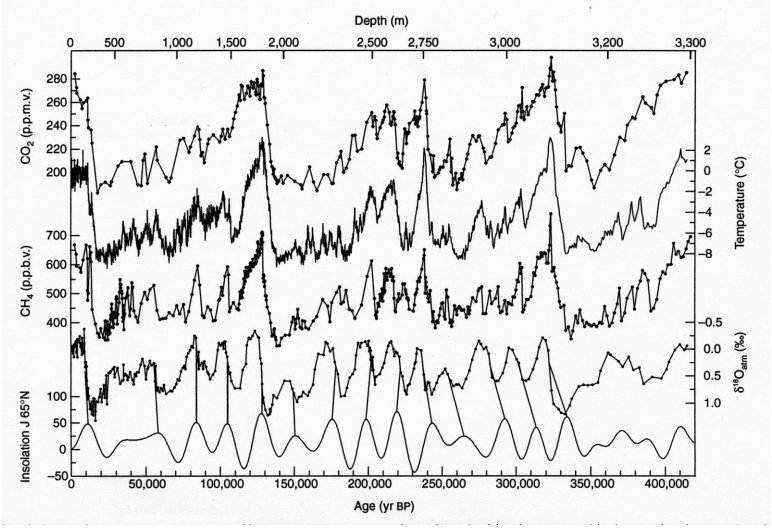


Figure 1.2: Vostok Ice Core Data. This image is taken from [3].

Carbon dioxide is present in the atmosphere due to a number of natural processes [4], however, since 1860 the effect of human behaviour on global warming has been so great that the world entered a new era; the Anthropogenic Era [5]. When fossil fuels are burnt carbon dioxide is released into the atmosphere and this disturbs the near balance of natural transfers of $CO_2$ between the oceans, sediments, terrestrial biosphere and atmosphere, thereby causing levels in the atmosphere to rise [6]. Currently over 70% of the energy used in the UK comes from fossil fuels including coal, oil and natural gas as illustrated in Figure 1.3. With current trends relating to the increase of energy use year on year, it is expected that the world demand will increase by more than 50% by 2030 [7]. If the energy generation mix remains as it is, this increased demand will lead to increased levels of carbon dioxide in the atmosphere, higher temperatures and raised levels and frequency of catastrophic incidents [1]. Actions must be taken to cut carbon

dioxide emissions and curb the potential effects of global warming. The need for action is so strong that it is backed by government policy, with a call to cut carbon emissions by 60% by the year 2050 [8].

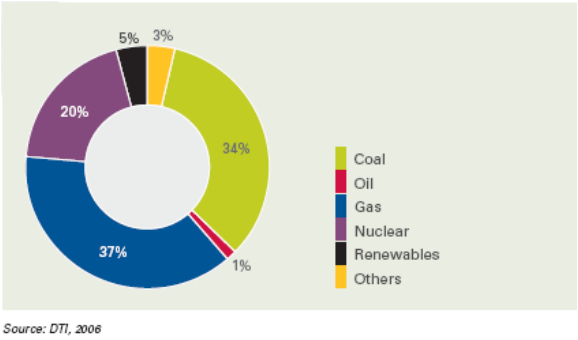


Figure 1.3: UK Energy Generation Mix in 2005. This image is taken from the DTI Energy Review [9].

## 1.2 Mechanisms to reduce carbon emissions

This call to cut carbon emissions can be answered by a number of measures, one of which involves decreasing the use of fossil fuels and consequently increasing the use of renewables. The government has set out a framework in the DTI Energy White Paper 2007 [8] in order to ensure an increase in the proportion of renewables in the energy generation mix to 20% by the year 2020. This will provide some reduction in the amount of carbon emissions, but this is not the entire solution. Other reductions will come from improved technologies and efficiencies in the transport sector, a carbon trading and offsetting scheme and energy efficiency measures. Table 1.1 shows how the government expects these measures to impact the reduction of carbon dioxide. Energy efficiency measures account for about 30% of the total reduction of carbon emissions by the year 2020, higher than the effects expected to be seen by improving the energy supply mix and transport infrastructure put together.

Energy efficiency measures include both the reduced emissions of products with newer, more energy-efficient technologies, but also an improved efficiency in the way in which

| Measures | MtC abated in 2020 |
|---|---|
| Energy Efficiency | 6.8–11.5 |
| Energy Supply | 0.7–2.1 |
| Transport | 1.8–5.1 |
| Trading and Offsetting | 14.1–14.3 |
| Total | 23.4–33.0 |

Table 1.1: Estimated carbon impact from various measures. Data from the DTI White Paper [8].

energy is used. Both of these measures are important, although the improved technologies will only affect people who buy these new products to replace older, less efficient ones. Improving the way in which energy is used is a measure which can be adopted by everyone in all aspects of their life and this thesis explores the mechanism by which individuals can reduce their energy consumption through efficiency measures in the domestic environment.

## 1.3 Energy reductions in the domestic sector

Energy efficiency measures play an important role in the reduction of $CO_2$ released into the atmosphere and the domestic sector represents an arena with the potential for large savings in energy for two reasons. The first reason is simply due to the magnitude of this sector as the domestic sector represents a significant proportion of the carbon dioxide emissions by end use in the UK. This is illustrated by Figure 1.4 which shows the UK emissions of $CO_2$ by end user for 2004.

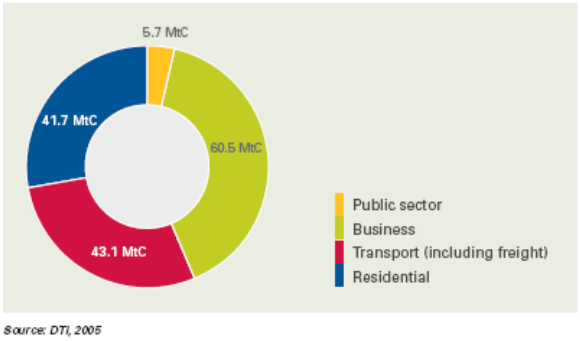


Figure 1.4: UK Energy Use By Sector in 2004. This image is taken from [9].

The second reason that the domestic sector can allow for a large energy saving due to improved behavioural efficiency is because the home is a place where the homeowner has control over a large number of factors relating to how they use their energy. Aside from refrigeration devices, the homeowner has control over when and how often appliances are used. Even refrigeration devices can be controlled to some extent by setting the desired temperature. By providing homeowners with the means to become more energy efficient there is a large potential for savings to be made. For example, in one particular study [10] up to a 39% reduction in energy use of a cooking appliance has been observed simply by making users aware of the implications of their actions.

## 1.4 Thesis roadmap

It should be recalled that the objective of the AEM is twofold. The first is to provide individuals with the motivation to reduce their energy consumption and the second is to provide individuals with the means to do so.

Therefore, before the techniques implemented in the AEM are developed, a set of key questions are posed in order to understand the mechanism by which such a device may satisfy its objectives. The answers to these questions are contained in the work in the following chapters and this section outlines the path taken by this research.

The first questions is:

**How can individuals be motivated to change their behaviour patterns?**
This question is answered in Chapter 2, which considers the factors surrounding motivation. It is shown that individuals who are intrinsically motivated, i.e. those who are motivated to perform a specific task for reasons which relate to the actual task alone (rather than motivation coming from some external reward or punishment scheme), are more likely to perform that task with greater endeavor and persistence than individuals motivated through some other reason. It is therefore desirable to encourage individuals

to develop an intrinsic motivation towards a task and this may be cultivated by making the task interesting, novel and challenging. Furthermore, individuals must perceive themselves to be in control of their actions (rather than taking actions to please some external body) and the must have useful and timely feedback about their actions so that they may observe the consequences of any actions undertaken.

Having established the mechanism by which individuals may be motivated to reduce their energy consumption, the second question is:

**How can individuals be enabled to reduce their domestic energy consumption?**
In order for an individual to reduce their domestic energy consumption they must possess both the means and the motivation to do so. Consequently, the factors which affect the ability of an individual to reduce their domestic energy reduction are explored in 7 different studies in Chapter 2. The factors investigated include:

- The links between knowledge base about the way different appliances in the home use energy and feedback provided to the user about their own energy consumption patters.
- The levels of user interaction with feedback mechanisms and the ability of the user to set themselves goals in the task of reducing energy consumption.
- The time intervals at which feedback is presented to individuals.
- The quality of the information provided to the individual about their energy consumption patterns.

It is shown that individuals do better at reducing their energy consumption when they are provided with automated feedback about their energy consumption on a daily basis in a manner which allows them to investigate the consumption patterns of individual appliances in the home. This, along with commercialisation considerations of such an appliance, enables the proposition of a set of key design criteria for the AEM.

**What are the implications of the design criteria?**

The implications of the design criteria are explored in Chapter 3. It is shown that the AEM should track the daily energy usage of individual appliances so that the user is provided with information which both encourages them and enables them to modify their behaviour. It is also identified that the data should be collected in a non-intrusive manner so that the AEM remains a usable device. Furthermore, to satisfy cost constraints, it is established that the AEM must identify the energy use of individual appliances using simple hardware to monitor the the total energy consumption of the home and complex software to identify individual appliances within the total load.

It is also identified in Chapter 3 that the research in this thesis is only concerned with the implications of the design criteria which relate to the mechanism by which individual appliances may be tracked within the total load. Any implications which relate to the presentation of feedback are not dealt with in this thesis.

**Is there any existing state of the art?**

The existing state of the art is discussed in Chapter 3. Two types of analysis are explored from which other existing systems stem. These types of analysis include steady state non-intrusive appliance load monitoring (NALM) and transient non-intrusive appliance load monitoring, although this is considered in this work to be too sophisticated for the domestic sector. NALM analysis monitors the total power consumption on the home and, when a change in level occurs, a signature relating to that change is extracted from the signals monitored. In the existing state of the art this signature contains 2 features – the real and reactive power of the observed change, which should correspond to the real and reactive power of some appliance in the home. There are, however, limitations to the NALM analysis and these are mainly due to errors in identifying appliances caused by a lack of information extracted when a step change in power is observed. Therefore, a scheme is proposed for the AEM which attempts to overcome the limitations of the NALM analysis by extracting a more informative signature when a step change in power

is observed.

**What information is used to form the appliance signature?**

The information which contributes to the appliance signature is deduced in Chapter 4 by considering the current drawn by an appliance when it is in use, as well as accounting for the ways in which appliances are used. It is identified that one possible set of features may be extracted from the current drawn by an appliance. When a voltage is applied to a set of electrical components a current will be drawn which is indicative of that particular combination of components. Therefore, features which are extracted from the current signal are representative of the particular electrical construction of an appliance. Further features are identified by considering the ON–OFF state diagrams of appliances. These features relate to the Finite State Machine (FSM) model of appliances. For some appliances, e.g. a kettle, there may just be ON and OFF states, however, other appliances, e.g. a washing machine, may have a number of states which are cycled through when the appliance is in use. Finally, a feature which depicts the influence of the time of day on the use of an appliance is identified. Therefore, it is shown in Chapter 4 that the appliance signature is comprised of three different sets of features, or feature domains: the short term time domain features, the FSM domain features and the time of day domain features.

**How does the AEM use the appliance signature in a classification framework to identify appliances in the total load?**

The mechanism by which the appliance signature is used to identify the appliance from which it is drawn is developed in Chapter 4. The appliance signature is comprised of three different types of features which must therefore be considered separately. By considering the type of information present in the features and identifying the desired framework capabilities, a novel classification framework is proposed. This framework analyses each feature domain separately and the analysis results in the generation of a set of membership values. These values represent the degree to which the features

belong to each of the appliances which the AEM has been taught to identify. Once these membership values have been generated, it is identified that they should be combined in a weighted geometric combination to produce a final membership value to enable classification.

The analysis developed in this chapter poses 2 further questions. The first relates to the mechanism by which the memberships are generated and the second to the mechanism by which the features are extracted from the data. It is also established at this point that the remainder of the work in this thesis focuses only on the development of the short term time domain features and corresponding memberships.

**How are memberships generated for the short term time domain features?**

The mechanism by which memberships are generated for the short term time domain features is explored in Chapter 5. It is identified that the work in this thesis focuses only on development of a membership generation algorithm in the scenario that the AEM may be trained to know the set of appliances which it must later identify from the features generated. Therefore, the membership generation is broken down into two regimes, the training phase and the working phase.

By considering the data available and the desired properties of the membership function, a set of criteria is developed. These criteria are used to guide and assess mechanisms which enable the AEM to evaluate appliance prototypes[1] and subsequently to generate membership values for feature vectors extracted during the working phase.

By considering the possible multi-modal operation of appliances, a novel algorithm is proposed for use in the training phase. This algorithm determines the number of modes of operation of the appliance before determining prototypes for each identified mode. To satisfy the criteria it is proposed that the prototypes consist of the appliance label, a typical feature vector, and a measure of the spread of data.

[1]Appliance prototypes are a set of parameters which fully represent the set of possible features generated by an appliance when in use, using as little information as possible.

A novel algorithm is also proposed for the generation of membership values for feature vectors not used in the training phase. It is shown that these memberships are based on both the proximity and relative proximity of the feature vector to each of the existing cluster prototypes which enables memberships to represent the typicality of a data point in the clusters whilst also taking into account the mutual exclusivity of the occurrence (i.e. only one appliance is assumed to change state between subsequent observations of the system).

**How are the short term time domain features extracted from the data?**

Chapter 6 explores the mechanism by which features are extracted from the data. It is determined that the features, $\mathbf{x}$, must be compatible with the method of membership generation developed in Chapter 5. It is also identified that the selected features should separate different appliances within the identified feature space[2] and to allow vectors from the same appliance to occupy a compact and disjointed region within the space. This enables the membership generation algorithm to generate meaningful results which may later be used to identify appliances from their feature vectors. Consideration of the signals which are monitored leads to the proposition of a set of criteria to which the identified feature set must adhere. Guided by these criteria, a novel method of feature extraction is proposed whereby the features are based on a model of an appliance's electrical interface. The electrical interface is defined as the set of electrical components which constitute the front end of an electrical appliance, i.e. those which are 'seen' by the supply. For example, fan heaters or hairdryers may be modelled as a single resistive component connected in series or parallel with an inductor between the live and neutral of the supply[3]. Features extracted in this way are shown to satisfy all the feature set criteria.

[2]Feature space is a $d$-dimensional space, where $d$ corresponds to the number of features, so that feature vectors are represented as data points within that space.

[3]For diagrams of the electrical interfaces modelled in this thesis refer to Figure 6.21 in Chapter 6.

**How are the individual objectives met to enable the development of techniques which will aid users to reduce their domestic energy consumption?**
Chapter 7 uses the proposed methods of feature extraction and membership generation for the short term time domain features in a real world example in order to demonstrate the classification ability of the developed analysis. A kettle, a fan heater, a soldering iron and a hairdryer are selected for use. The voltage and current of each appliance are monitored and used to generate 2 data sets. The first is used to generate prototypes for each appliance considered and the second data set is used along with the appliance prototypes to generate membership values for each feature vector $\mathbf{x}$ which represent the degree to which is belongs to each of the known prototypes. These membership values allow appliance identification based on the short term data alone and, although it is beyond the scope of this report, it is proposed that they are then used in the classification framework identified in Chapter 4. This enables a more robust classification which includes FSM and time of day features. The identification of appliance state transitions allows the energy use of individual appliances to be tracked and this information can then be provided to homeowners in order to aid them to reduce their domestic energy consumption.

## 1.5 Novelty value of the work

The opportunity is taken here to highlight the novelty introduced by the work presented in this thesis. There are four main aspects which are discussed briefly in this section.

The first novel aspect of the work produced in this thesis is in the development of the classification framework in Chapter 4. The framework enables each feature domain to be analysed separately to produce a set of membership values whereby each membership represents the degree to which the corresponding feature set belongs to a particular appliance. These memberships are then combined in such a way that allows each feature

set to strengthen the reliability of the classification whilst letting more rigorous features play a stronger role in the final result.

The second novelty is in the mechanism by which the AEM is trained to know a set of appliances in the home. Standard techniques are combined in an innovative way in order to determine appliance prototypes from a set of labelled training data in which the number of modes of operation is unknown.

The third novelty is in the mechanism by which memberships are generated to represent the level to which a feature vector belongs to each of the appliances to which the AEM has been trained. This membership enables the degree of belongingness to depend not only on typicality of that feature vector to the known appliance groups, but also to be influenced by the assumption that the feature vector must have been generated by a single appliance.

The final novel aspect developed in this research is in the method by which the short term time domain features are extracted from the data. These features are dependent on the electrical interface of the appliance, which is deduced by observing particular aspects of the current drawn by that appliance when in use. This enables the features for each mode of operation of each appliance to be represented by a set of parameters which, when plotted for a large amount of training data within the selected feature space, form compact and disjointed clusters.

An initial IP search has been performed to identify any similar work in this area. The search did not discover any work which is similar with regards to the specific techniques developed in this thesis for identifying individual appliance energy consumption. A patent application was made early on in the development of the work and was later dropped due to its similarity to existing state of the art. However, this apparent similarity was caused by a lack of detail in the original application and the option to re-apply with a more detailed description of the work is currently under consideration.

# Chapter 2

# Behaviour Modification to Reduce Consumption

This chapter aims to identify the mechanisms by which individuals may be provided with both the motivation to change their behaviour patterns and the means by which they may reduce their domestic energy consumption.

The amount and type of motivation an individual possesses for a particular task will influence the effort to which they undertake that task and the persistence with which they carry it out. Section 2.1 considers the factors which cultivate an intrinsic motivation (whereby the individual is motivated by factors inherent to the task in itself), since this type of motivation results in increased effort and greater persistence.

Section 2.2 explores 7 studies that investigate the ability of individuals to reduce their domestic energy consumption in order to examine the influence of various factors. These factors include the links between feedback about energy consumption and knowledge base about how appliances in the home use energy, the links between levels of user interaction with feedback mechanisms including the ability of users to set themselves goals, the time intervals at which feedback is provided to the user and the quality of the feedback provided. This allows the theories relating to motivation, effort, persistence and knowledge to be considered in the context of the domestic environment.

By considering the implications of the different studies with reference to motivational

theory, a set of criteria are developed in Section 2.3. These criteria relate to the type, frequency and presentation of information that should be given to individuals in order to most effectively encourage and enable them to reduce their domestic energy consumption. This forms the basis for the design of the AEM.

## 2.1 Influences on motivation

In order to enable an understanding of the factors which influence an individual's motivation to reduce energy in the domestic environment, this section introduces the concept of motivation, including the different types of motivation a person might possess for a task and the consequences in task performance. Having established the type of motivation which results in increased effort and persistence towards a task, the factors which cultivate this type of motivation are examined. This allows a set of conditions to be proposed which attempt to facilitate an environment to enhance an individuals motivation and subsequent task performance.

Motivation can be split into two broad categories, intrinsic motivation and extrinsic motivation [11]. Intrinsic motivation is a type of motivation which is due to the task in itself, whereas extrinsic motivation is caused by a separate outcome of the task. Intrinsically motivated people tend to have more interest and excitement for the task compared to those who exhibit extrinsic motivation which is externally controlled and this intrinsic motivation results in better task performance, persistence and creativity [12].

Extrinsic motivation, although dependent on some external outcome, varies along a scale of how internalised that behaviour has become. Figure 2.1 presents the different types of motivation and the associated perceived locus of causality[1] and regulatory processes.

External regulation of extrinsic motivation is a motivation which is brought on by some

[1]The locus of causality [13] is an individuals perceptions as to the cause of their performance in a task. This may be internal, whereby is it due to factors inherent to the individual performing the task, or external, whereby it is due to some influence over which the individual has no control.

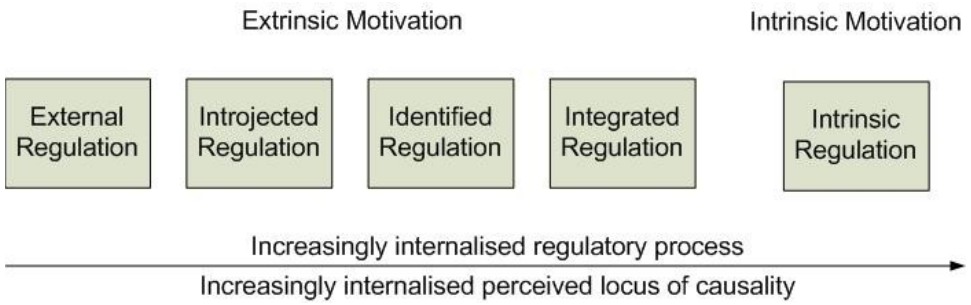


Figure 2.1: Types of motivations, regulatory styles and loci of causality. Information taken from [12]

external system of punishment or reward. The motivation for the particular task comes either from the reward to be achieved by carrying out the task, or in order to avoid some punishment for non-participation. Slightly further along the scale towards internalisation is introjected regulation of extrinsic motivation. This type of motivation induces behaviour which is undertaken in order to achieve some internal reward or avoid an internal punishment (for example, guilt or anxiety). Further along the scale still is identified regulation of extrinsic motivation. In this case the individual identifies with the particular values of the task and undertakes the behaviour due to its personal importance. The most internalised form of extrinsic motivation is integrated regulation whereby the individual evaluates the regulations and brings them on-board with their other values and needs. This type of motivation is very similar to that of intrinsic motivation, however, the difference is due to the reasons for which the behaviour is undertaken. In extrinsic motivation the behavior is undertaken in order to attain the outcome of the task (even though the individual has taken on board the values and meanings of the task as their own) rather than for the essential enjoyment of the task in itself.

Increasing internalisation of motivation leads to increased persistence of the task, an increased engagement with the task and more positive self-perceptions, with intrinsic motivation at the top of this scale [11]. Therefore, in order to achieve better results in the task of reducing energy consumption in the home, an intrinsic motivation and an interest in the actual undertaking of the task should be cultivated.

Whilst intrinsic motivation will only occur for a task which holds intrinsic interest for that individual, including an appeal of novelty, challenge or aesthetic value, certain environments can be cultivated in order to facilitate this type of motivation rather than undermine it. Feelings of competence can enhance intrinsic motivation but only when accompanied by an internally perceived locus of causality, whilst external rewards will diminish this motivation if viewed as an external controller of behaviour [11].

Internalisation of extrinsic motivation can also be facilitated in certain environments. Extrinsically motivated behaviour is normally undertaken when the behaviour is prompted or valued by a significant other. The perceived levels of competence are an important factor as an individual is more likely to adopt activities that are valued by others if they view themselves as being competent at them. In order for this extrinsically motivated behaviour to become internalised the individual must perceive a level of autonomy in carrying out the task and its meaning must be understood and taken on board as their own [12].

Perceived competence and self-efficacy[2] in the activity has an important influence on the subsequent motivation levels and performance [15]. Hence, in order to improve task motivation and subsequent performance, an individual's perception of their competence and self-efficacy in undertaking that particular task must be raised.

An individual's perception of their own performance based on previous experiences will affect the beliefs they hold about the current task. This is also affected by certain social influences specific to the task. These task specific beliefs are related to the individuals perceived ability to perform the task, the difficulty level of the task and the specific goals that the task holds. These task specific beliefs in turn affect the individual's belief about the expected task outcome and the amount to which they value that task. The perceived outcomes will influence the choices the individual takes in terms of performance, effort

[2]Feelings of competence and perceived self-efficacy are judgments of how well a person views them-self of being able to execute courses of action to deal with prospective situations [14].

and persistence applied to the task. This is illustrated in Figure 2.2.

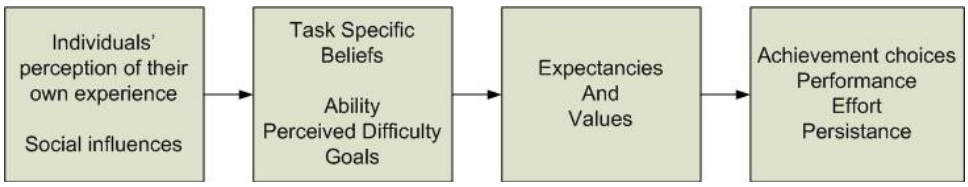


Figure 2.2: Specific beliefs affecting the way in which a task is performed [15]

Therefore, in order to improve an individual's performance, effort and persistence in a specific task, their perceived self-efficacy levels must be heightened, since an increase in perceived self-efficacy leads to increased coping efforts. These expectations of personal efficacy can be affected by 4 factors. Previously attained successes will heighten perceived self-efficacy whilst repeated failures will reduce it, especially if these failures appear early on [14]. These performance accomplishments based on personal experiences are the most influential of the 4 factors which can affect the expectations of personal efficacy. The other 3 include observation of the success of others, verbal persuasion and states of psychological arousal including anxiety and stress [16].

External incentives may also increase personal efficacy. Whilst they have been seen to diminish motivation for those people intrinsically motivated for a task [11], positive incentives can verify existing competencies when they are received for attaining successes in desired outcomes through challenging performances. These authentications of self-efficacy may subsequently promote further interest in the task [14].

Performance accomplishments and external incentives may well increase motivation for a subsequent task, but they cannot induce motivations for the current task in a backward propagation. Instead, visualisations and representations of these future events allow likely outcomes to be anticipated and goals to be set in order to affect the desired outcome of the future event [17].

Goal setting and intermediate goal setting is a way in which self incentives can be set and task performance can be improved [14]. Feedback allows the individual to observe

whether their levels of actions match their expected outcomes, which in turn enables them to take appropriate action to achieve the desired standard. Once this standard has been reached the goals need to be re-adjusted in order that the effort does not diminish [17].

Individuals tend to be influenced more by perceived successes than actual successes [14], so it is important to ensure that the efforts are noticeable when presented in feedback. This is particularly important in this application as the individual may feel that they are unable, as a single person, to have an effect on climate change, although within a certain community the effects may be noticeable. In order for the collective efficacy of a group to increase, the individuals must believe that they have an ability to influence the outcome and to sustain this collective efficacy the efforts of the group must be noticed without long delays [14].

Therefore, in order to improve task performance, effort and persistence, the task should be made challenging, novel and interesting so that an intrinsic motivation is cultivated for that task. Perceived task ability will also affect the levels of motivation for the task and this can be improved through well presented feedback relating to task performance. The feedback given should be done so in a timely manner so that individuals can see the results of their efforts and the way in which the feedback is communicated should be positive and show the overall impact that they are making through their actions. There should be a method of allowing individuals to set themselves goals and targets for reaching their objectives and opportunities to revise them once initial standards have been met. This self regulated feedback is an important factor in task performance and Section 2.2 will continue to explore how this may be implemented in the home environment to improve performance in the task of reducing domestic energy consumption.

## 2.2 Feedback in the domestic environment

As outlined in the previous section, feedback can be a useful tool in aiding homeowners to reduce their energy consumption, but only if presented to them in a useful and timely manner. This section will explore seven studies which investigate various mechanisms of feedback, including the tools used and information provided, in order to determine their effectiveness.

Section 2.2.1 considers the dual influence of feedback and information. The feedback is the data presented to the user about their own energy consumption and the information is data relating to tips and hints on how to conserve energy. Section 2.2.2 considers interaction between the user and the feedback system in order to explore influences on task enjoyment and consequently effort and persistence. Section 2.2.3 considers the time lapse between actions and feedback. It shows that timing of information plays an important role not only in allowing the user to interpret the feedback, but also allows them to see the consequences of their actions. Finally, Section 2.2.4 considers the quality of the feedback presented and the effect of this on the ability of the user to reduce their energy consumption.

### 2.2.1 Mutual supportiveness of feedback and information

In order for a behaviour change to be adopted, both information and feedback need to be present, since regardless of the way in which feedback is provided, if a person has no understanding of how to make a change, their willingness to do so will not become realised. A London study set out to identify links between the amount of understanding a person possesses about energy use in the home, their energy literacy and the levels of energy conservation behaviour undertaken in the home [18]. This study reports that the differentiating factor between those individuals who do make attempts to conserve energy and those who do nothing is related to their knowledge about energy saving

techniques and their beliefs about the effect of these savings and how much they are actually contributing. Hence, to encourage non-savers to reduce their energy usage and put in place more conservative actions, these incorrect beliefs must be challenged and the individuals must also be presented with some knowledge on how to conserve in combination with incentives encouraging them to adopt conservative actions.

These ideas linking energy literacy to conservation behaviour were put to test in a trial consisting of 160 London homes all of which had gas fired heating. The 160 homes were split into four groups consisting of a control group and three experimental groups. One of the experimental groups were provided with daily feedback, another was provided with a leaflet containing energy saving information and the third was provided with both the daily feedback and the information. The daily feedback consisted of a chart on which the participants had to record their daily energy use by looking at their energy meters. The trial lasted for 6 weeks though the energy use statistics were recorded over the following 12 weeks in order to study the longer term effects of the trial. Interviews were conducted following the trial to establish what conservation measures were undertaken and the effect of the trial on energy conservation beliefs of individuals.

The results of the trial compared the increase of weekly energy use in weeks 1-3 (i.e. during the first half of the trial) to that in weeks 7-18 (the 12 weeks following the experimental period). These are illustrated in Table 2.1 for gas and electricity (data taken from [18]). The most prominent observation is that for all the groups, both the electricity consumption and the gas consumption increased over the time period of the experiment. This is because the experimental period began at the start of November and the decrease in temperature over the winter months (and other conditions which affected all the households) is reflected by the increase in energy use. The data relating to the increase in energy consumption of the no feedback or information group is therefore used as the baseline for comparison.

| | Gas | | Electricity | |
|---|---|---|---|---|
| | Feedback | No Feedback | Feedback | No Feedback |
| Information | +24% | +32% | +3% | +5% |
| No Information | +41% | +46% | +15% | +12% |

Table 2.1: Increase in gas and electricity use from weeks 1-3 to weeks 7-18.

It can be seen from the results that those in the feedback plus information group saved the most in both gas and electricity consumption suggesting a mutual supportiveness of the two strategies. The feedback generated an initial interest in the subject, the information provided the means for the individual to change their behaviour and further feedback allowed the results of behavioural changes to be observed and pro-conservation beliefs to be strengthened. There are also suggestions from the trial that those in the feedback and information group and those in the information only group may have undergone longer term changes in behaviour. This is because the increase in energy used by these groups when compared to the control indicates savings made in the post-experimental period, suggesting some longer term effects after the short 6 week experimental period.

From the interviews conducted after the experimental period, it was established that savings were made by homeowners turning down central heating, reducing the time it was on for and trying to use less hot water. It was concluded that the savings made were because of more pro-conservation beliefs and increased knowledge about conservation in parallel with the ability to see results of conservation measures undertaken. This ability to see the results of energy saving endeavours is a very important factor when encouraging individuals to modify their behaviour and attempt to reduce their consumption.

It can also be seen from the results that those in the feedback only group made no significant savings, suggesting that the results from the feedback were difficult to interpret. From the interviews it was also established that the members of this group had little ability to understand where the energy was being used in their home and they had further problems in observing how changes in usage affected the overall consumption. It was also suggested that the method in which the feedback was collected became a burden. The

disengaging nature of the feedback method may have prevented the users from spending time trying to understand the feedback they obtained. From information presented in Section 2.1 it is postulated that for an effective behaviour change to occur, users should become interested in the particular task they are undertaking by making it challenging, enjoyable and novel. Perhaps the disengaging method of collecting feedback, combined with the lack of understanding between actions undertaken and energy saved, prevented individuals in the feedback only group from reducing their energy consumption.

Therefore, any mechanism designed to enable individuals to reduce their energy consumption should be made enjoyable by removing the burden associated with collecting the data. It should also ensure that the links between energy saving endeavors and the effects of endeavors undertaken are clear to the user.

### 2.2.2 The importance of self regulation and user interaction

This section explores the ideas presented in Section 2.1 which link increased levels of perceived self efficacy, competence and autonomy to a more effective task performance.

Goal setting within a task allows the homeowner to feel that their actions are internally regulated and allows them to clearly see the effects of any actions they take (recalling that performance accomplishments are the most influential factor on an individual's perceived self efficacy). A study in the Netherlands [19] set out to understand the effect of goal setting and electronic daily feedback on the natural gas energy use of the involved participants.

In total 325 homes participated in the experiment, split into 3 experimental groups and 3 control groups. The first experimental group was provided with an Indicator device. This device displayed the daily temperature-corrected gas use in comparison to some standard set by the user and was placed in a visible location. They also received conservation information. The second experimental group received monthly external feedback about

their gas use, along with conservation information. The third experimental group received a self monitoring chart as well as conservation information. The first control group received only conservation information. The second control group were unaware of their participation and the third control group was comprised of those homes unwilling to participate.

The results are presented in Table 2.2 which shows the percentage change in gas use in the experimental and post-experimental periods from the baseline. The baseline gas use is that observed during 1983, the experimental period is defined from February 1984 to February 1985 and the post-experimental period is February 1985 to February 1986.

| Condition | Change from baseline during | |
|---|---|---|
| | experimental period | post-experimental period |
| IND | -12.3 | -2.1 |
| MEF | -7.7 | -3.2 |
| SMO | -5.1 | -1.5 |
| INF | -4.3 | +1.4 |
| C2 | -0.3 | -2.2 |
| C3 | -0.2 | -2.9 |

Table 2.2: Change in gas use from the baseline in the experimental (1984) and post-experimental (1985) periods. Baseline is taken as the gas use in the pre-experimental period 1983, in which there is no statistical difference between the groups. NOTE: IND = indicator group, MEF = monthly external feedback, SMO = self monitoring, INF = information only i.e. control group 1, C2 = control group 2, C3 = control group 3. Results are taken from study conducted by J H Van Houwelingen et al. [19]

From the results it can be seen that for the 3 experimental groups and the information only control group, a significant reduction in gas use is observed during the experimental period, with no significant savings made by control groups 2 and 3. This indicates the effectiveness of conservation strategies which include goal setting, feedback about energy consumption and conservation information. Furthermore, the energy conserved for the indicator group is significantly larger than that conserved by either the MEF group or the SMO group [19].

The ability of the goal setting and daily electronic feedback in enabling users to take control of their actions and to clearly see how their actions affected their behaviour in

comparison to a desired behaviour, resulted in effective conservation measures. The electronic feedback was more effective than the self monitoring charts, echoing the negativity associated with acts which are considered a burden presented in the study [18] discussed in Section 2.2.1.

It is also observed from Table 2.2 that there are no significant savings made by any of the experimental groups in the post experimental period. The ability of the feedback in enabling individuals to reduce their consumption by observing the effects of their actions is not an effective conservation strategy once the feedback is removed. This may be in part due to a lack of intrinsic interest in the task resulting in reduced actions when there are no attainable results and may be in part due to the inability of individuals to assess their energy consumption without any feedback information. This implies that individuals need a constant reminder in order to conserve and the links between timely feedback and conservation are explored further in Section 2.2.3.

The effectiveness of cultivating an intrinsic interest for an activity and thus increasing an individual's effort and persistence in a task is further explored in a study in Bath [20] which investigated various methods of providing consumption feedback to homeowners. This study consisted of 120 households split into 6 experimental groups plus a control group. The study was conducted over a 9 month period during which the energy consumption of each home was compared to weather corrected energy consumption data from the previous year.

The feedback was provided in the following manner to the different experimental groups. The first group received a comparison of their own energy use to their own previous usage, the second had a comparison between their consumption and the consumption of other homes, the third received feedback about energy consumption in financial terms, the fourth in environmental terms and the fifth received feedback about their energy use as well as a leaflet presentation about actions they could take to reduce their consumption.

The sixth group had a computer installed in their home which included graphical displays about their current energy consumption and their consumption in the previous year, a questionnaire about general energy saving and a directory of energy saving advice.

Table 2.3 presents a comparison between the number of households reducing and increasing their energy consumption. Values for the mean percentage increase/decrease along with the standard deviation are also presented. A binomial test carried out in the study suggested that only the group with the computers installed in their homes showed a significant difference in the numbers of homes reducing their energy consumption.

A focus group carried out after the study suggested that despite concerns for environmental issues, there was little desire to change behaviour with regard to lighting, heating, cooking and washing. One of the main issues was money – conservation became a trade off between cost and comfort. A further point discussed concerned the visibility of energy and the need for a constant reminder to conserve energy. Those in the computer group had this constant reminder as the computer was set up in a visible location, whilst those in the other groups only received monthly information. Those in the computer group also had the ability to receive feedback on demand and to interact with the data through the programs installed on the computer. This interaction made the task novel and interesting and the immediacy of the feedback allowed the users to see clearly the effect of any energy saving endeavors they made.

### 2.2.3 The importance of timely feedback

Because motivation is affected by perceived successes it is important to ensure that the feedback provided enables the user to see the effects of their energy saving endeavors in order that they do not become disheartened with their efforts. Therefore this section explores the links between energy conservation measures undertaken and the ability the see the results of these measures.

| | Groups | | | | | | |
|---|---|---|---|---|---|---|---|
| | Computer | S v O[a] | S v S[b] | Leaflet | Control | Money | Environment |
| Mean increase | -4.31% | -4.60% | 1.50% | -0.39% | 7.78% | -4.84% | 4.46% |
| Standard deviation | 20.80 | 18.60 | 19.30 | 12.10 | 33.60 | 13.50 | 18.00 |
| No who increased | 3 | 7 | 10 | 8 | 12 | 6 | 8 |
| No who decreased | 12 | 9 | 8 | 11 | 10 | 7 | 9 |

[a]Self vs Others
[b]Self vs Self

Table 2.3: Comparison of energy increase and decrease in the homes participating in the study. Data taken from [20].

Currently the main method of feedback for energy consumers in the UK is their bill, which is often received quarterly and generally provides just an averaged estimate. It therefore provides little information about how the energy is being used and the timing is such that the information provided is so long after the act that users are unable to understand the link between their actions and their energy use. A study in Norway, [21], set about looking at the potential of energy savings through a more informative energy bill.

The study considers the impact on savings from frequency of billing, presentation of information and comparative contexts. The experiment lasted 3 years and consisted of 1286 participating homes, split into 4 groups (including 1 control group and 3 experimental groups). All 3 experimental groups were presented with more frequent bills. Experimental group 1 received feedback about their actual consumption, group 2 received graphical feedback about their current consumption as compared to previous consumption and group 3 received feedback about their current consumption compared to some standard, as well as energy saving tips. It was found that at the end of the 3 years the experimental groups averaged[3] a saving of 10% over the control group, up from the saving of 7.6% at the end of the second year, showing a persistence of the effects induced.

These results suggest that the crucial factor in this particular study is the frequency with which the bill is presented. The experimental groups reported that the actions undertaken in order to conserve energy were mostly related to heating - the experimental groups were more likely than the control group to turn down heaters when airing homes and waited for lower outside temperatures to prompt heating use.

Figure 2.3 illustrates a typical bi-monthly energy bill compared to the standard. From this new billing system (presented to all three experimental groups) the relationship between increased heating (during the winter months) and the increased energy use is

[3]The average values are discussed as there was no statistical difference between the saving of the different experimental groups.

very obvious. However, the effects of other actions undertaken by the user on their total energy consumption is not obvious from this bill. The time lapse between the feedback and actions other than those relating to heating is too long to observe any effects of the actions on energy consumption. Therefore more frequent billing is unlikely to help users reduce consumption of energy other than that used for heating. As there are very limited actions which can be undertaken to reduce the heating energy consumption, this may explain why the three experimental groups all conserved the same amount of energy.

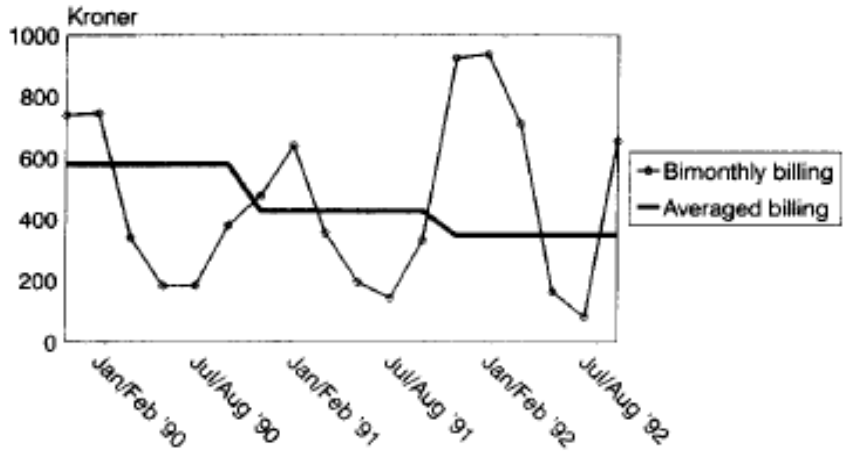


Figure 2.3: Ability to distinguish seasonal variation in energy use with bimonthly billing as opposed to average billing. Picture taken from [21].

The study also allowed the following relationship to be identified. The increased feedback led to an increased awareness and/or knowledge about energy use which in turn prompted changes in energy use behaviour and decreases in consumption.

The studies considered so far have all looked at total energy consumption. Two of these studies ([18], [21]) have identified through focus groups that the main actions undertaken in order to conserve energy are those relating to heating and hot water. This is due to the lack of information available in the feedback which can be linked to other appliances in the home. It is quite obvious from observing the total energy use what effect varying the heating and hot water has on the total energy use as these either tend to be on a separate system (many UK homes use gas for heating and hot water only) or the uses can be seen clearly through the seasonal variations. The ability to identify the energy consumption

of other appliances from the total energy observations is harder. The following section therefore considers how the quality of information affects the ability of homeowners to reduce their consumption.

### 2.2.4 The importance of feedback quality

A utility bill which presents the homeowner with only their overall energy use can be likened to a shopping bill where the final amount is known, but where none of the individual items are priced [22]. Consumers therefore find it hard to identify what actions they can undertake in order to save energy and many are under the impression that turning off lights would result in the same amount of energy savings as using less hot water. When questioned as to their ideas about which appliances were the largest energy consumers many homeowners identified the washing machine to be in the top three. However, lighting, freezing and dishwashers are in fact the top energy consumers [10]. This lack of understanding of energy use results in the consumer having little idea as to the effect of various energy savings actions they may take.

Furthermore, if only the total energy consumption is displayed to users, any energy savings made with regard to a specific appliance may seem very small when considered in the context of the entire energy bill and consumers may doubt their conservation actions [22]. Therefore, so that individuals are provided with greater understanding as to the energy consumption patterns of individual appliances and so that they may observe the effects of any conservation actions undertaken with regards to a particular appliance, it is desirable to provide them with feedback about the energy consumed by each individual appliances.

Cooking is one domestic activity which has a large possibility for behaviour modification. Up to 50% variation in electricity consumption has been observed between 6 chefs all cooking the same meal [10] and therefore the following study investigates the behaviour

modification possible when providing users with detailed information about how their cooker consumes energy.

In a field study carried out to investigate the effect of providing consumers consumption information about the electricity use of their cookers, a total of 44 households were split into 4 groups. The first group of 12 (although 3 had to later be excluded from the study as they were single households with non-typical cooking patterns) represented the control. Group 2 was provided with a 17 page, A4 pack containing information about the consumption patterns of electric cookers and tips on how to reduce the energy used when cooking. Group 3 was provided with an electronic energy consumption indicator (ECI). This indicator allowed consumers to investigate their energy use of specific events as well as the consumption of the cooker during the day, the previous day, the current week and the previous week. Group 4 was presented with both an energy consumption indicator and an information pack.

The groups were monitored for 12 months. Summer use was taken as the baseline and the electricity used by the control groups was was 11%, 9% and 22% higher in spring, autumn and winter respectively. This allows for seasonal adjustment of the results. Table 2.4 shows the overall energy savings, compared to the control group, obtained from the three experimental groups.

| Group | No of Homes | Average Change | % that saved | Savings made |
|---|---|---|---|---|
| Info Pack | 12 | -3% | 66% | 6.4% |
| ECI | 10 | -15.2% | 80% | 20% |
| ECI and Info | 8 [a] | -8.9% | 75% | 14% |

[a] Two of the homes has to be dropped from this group. One was due to problems when transferring data and the other due to a distinct lifestyle change during the experimental period which drastically affected the results.

Table 2.4: Comparison of energy savings made in the experimental groups. The total average consumption changes are listed for each group, along with the percentage of homes in each group that saved and the savings made by this percentage. Data taken from [10].

The ECI seems to be far more effective than the information pack in aiding consumers

to reduce the energy consumption of a single activity. However, when questioned, the members of the ECI group had less idea about energy saving practices that could be undertaken than those in the information pack group. Those in group 4, information plus ECI, also had very little knowledge about energy saving techniques when cooking, suggesting that they paid far more attention to the ECI than the information booklet. These findings suggest that substantial savings can be made simply by providing an interaction between the user and the energy consumption of an individual appliance.

The ideas relating to energy savings through feedback about the energy consumption of individual appliances is investigated on a larger scale in 2 studies carried out in Japan [23] [24]. These studies investigated the impacts on energy consumption of putting energy display systems into a number of homes consisting of a married couple with between one and three children. The first study conducted in Kyoto in early 2000 consisted of installing an information terminal in nine residential homes (though one was removed from the results due to the presence of a photovoltaic system). The information terminal displayed to the users both the total house energy consumption and the energy consumption of individual appliances (given in terms of Japanese Yen in order for the participants to understand the values) averaged over 30 minute periods. Users were able to see the daily usage curve of the appliances, the daily changes over 10 days and a comparison to past data. They were also provided with energy saving hints for the various appliances and were asked (through buttons on the display) to select whether they would take the advice given or not.

The second study was conducted in late 2002 and differed from the first only slightly. The new energy measuring devices were installed in 10 homes and users were provided with additional information relating to room temperature and city-gas consumption of the home. There was also a control group of 9 homes involved in this study in order to compare the energy savings made through changes to space heating (as this varies with external temperature).

The results of the trials were two fold. The first was to do with the interaction between the user and the system. This was observed by monitoring the buttons pressed on the display, which allowed the energy consumed by appliances and the users' response to tips to be investigated. The trials showed a very strong initial interest with large number of user interactions. This interest decreased gradually as the trial continued. From the second trial it was seen that those houses which exhibited a large initial interest in the display continued during the experiment to keep their interest high. Those who were not that interested initially decreased their interaction more considerably over the period of the experiment. The most investigated usage was that concerned with the daily load curves of individual appliances.

The second result was concerned with considering the actual decrease in power and energy consumption of various appliances. Two time periods are defined as Period 1 and Period 2. In trial 1 these consist of 40 weekdays before and after installation of the display respectively. In trial 2 it is reduced to 28 weekdays before and after the installation of the display system.

The first trial showed that power consumption of the whole house reduced on average by 9% following installation of the monitoring device. Power consumption for heating the living room dropped to 77% of the initial value and dropped to 84% for the other rooms. A 12% reduction in power consumption of displayed appliances was observed, along with a 5% reduction in the power consumed by non-displayed appliances.

The second trial showed that those homes with energy monitors reduced their average power consumption by 5150Wh/day (17.8%) after installation whilst those without the monitors only reduced by 891Wh/day (4.7%). Installation of the energy monitoring device resulted in an average reduction rate of total energy consumption of 12%, with a 20% reduction in the energy consumption for space heating, one of the largest contributors to the total energy consumption.

From these trials it can be seen that the provision of detailed energy feedback including the energy use of individual appliances allows the user to significantly reduce the energy consumption of those appliances. It also increases their energy awareness and as such they reduce the energy consumed by appliances which are not monitored (though by a much smaller amount). The trials also showed that the users were interested in the daily load curves of the appliances, as this may be where they can most clearly see how their actions affect the energy consumed by the appliances. The conclusions from these trials and from the others considered in Sections 2.2.1, 2.2.3 and 2.2.2 allow a set of criteria to be established for the design of the AEM to ensure that the user is presented with useful, timely and interactive feedback.

## 2.3 Design criteria

The AEM is a tool which aims to empower homeowners to reduce their domestic energy consumption by providing them with feedback relating to their energy use. The content of the feedback should enable users to understand their energy demands and consequently see how they might reduce their consumption. It should also allow them to observe the effects of any actions they undertake. The presentation of the feedback should be in such a manner as to attempt to cultivate an intrinsic interest. Therefore Section 2.3.1 explores the implications of the results from the trials discussed in the previous sections alongside a reference to motivational theory. Consideration of these implications alongside a commercial constraint allows a set of design criteria for the AEM to be established. These design criteria are presented in Section 2.3.2.

### 2.3.1 Implications of the studies

From these studies and the theory relating to behaviour change presented in section 2.1, it can be seen that for individuals to effectively reduce their domestic energy consumption, they must possess a motivation to modify their behaviour. They must also be presented

with information relating to pro-conservation behaviour. Whilst feedback alone is not enough to allow the user to understand what changes they need to make in order to reduce their energy consumption, care must be taken not to overwhelm them with too much information as in the ECI study presented in Section 2.2.4.

In order to cultivate an intrinsic interest for the task of reducing energy consumption the feedback should be gathered and presented to the user in a such a way that it is not considered a burden. This implies an automated system which captures the information and a feedback system that is interesting, challenging and novel. It should also allow the user to set themselves goals in order that they perceive the task to be internally regulated. By allowing interaction between the user and the monitoring system further interest in the task is cultivated and this results in more effective energy conservation.

The feedback should be on a timescale to allow the users to see the effects of any conservation actions they take for 2 reasons. The first is in order that they gain further knowledge as to the energy saving consequences of various actions. The second is so that they do not become disillusioned with their efforts.

Feedback provided on a bi-monthly basis is effective in aiding users to reduce their energy consumption for space heating. However, in this time frame users cannot see the effect of other energy saving endeavors and therefore feedback should be presented on a daily basis as in the Japanese studies presented in Section 2.2.4.

The results of modifying behaviour with regards to heating is obvious when considering the total energy bill as heating represents a large amount of the total home energy usage. However, users have little knowledge about which appliances are the largest electrical energy consumers and about the potential savings from various energy saving actions. To observe these savings and to prevent users from becoming disheartened with their endeavors they should be able to see the detailed energy consumption patterns of individual appliances. This implies providing the user with daily disaggregated feedback.

These implications are useful in the design of the AEM in order that individuals are provided with feedback about their energy use in such a way that encourages as large a reduction in consumption as possible.

### 2.3.2 Design of the AEM

For the purposes of this research the design of the AEM will focus only on the reduction of electrical energy consumption. The reason for this is due to the relative simplicity of the gas framework and complexity of the electricity framework within most homes. The reduction of gas consumption does not required detailed feedback, as large savings are possible from feedback about the total gas use as illustrated in the studies above, [18] [21]. However, the reduction of electricity consumption is not so easy when only the total electrical energy use statistics are provided[4] . This analysis will require a more complex feedback system which is able to provide individuals with information about the energy use of each appliance and therefore this thesis is concerned with electrical energy consumption reduction.

On top of the ideas presented above in Section 2.3.1, the AEM should be usable. The method of monitoring should be such that it is easy to install. The AEM should also be relatively cheap so that it is an attractive option to the user. A one year payback period is desirable, so assuming the system will save users on average between 10 and 20% of their electricity bill, the AEM should cost a total of no more than 20% of the average UK electricity bill. In 2008 this was £405 for standard credit payments, £376 for direct debit payments and £424 for users with prepayment meters [25]. The average electricity price for the same year was £0.1263 per kWh [25], so, if it is assumed that users will save up to 20% of their electricity bill, this corresponds to an annual saving of up to approximately 51kWh or 50kg of $CO_2$ [26].

[4]This is because the user is unable to determine the energy consumption of individual appliances and is unable to determine or observe the energy saving endeavors of individual actions.

The feedback presented to the user should allow the user to act on the information and therefore the user must be able to understand how to subsequently reduce their energy use. The feedback should also allow user interaction, perhaps via goal setting. This allows them to perceive themselves as being in control of the task and helps cultivate an intrinsic interest.

Finally, in order that the feedback allows the user to see the consequences of their energy saving endeavors, the feedback should be instantaneous. It should also allow the user to investigate individual appliance energy use.

To summarise:

1. The system should be easy to install as difficult measures, that are considered a burden, disengage people from the task.
2. The system should be cost effective in order that it is an attractive option to the user. The average UK energy bill is £383 and the target price is between 10% and 20% of this to allow a one year payback, therefore the AEM should be commercially available for less than £70.
3. The information must be useful and allow the user to act on the information.
4. The user must perceive themselves to be in control of the task, their actions and consequent results.
5. The user must be able to compare their current performance with a goal they have set themselves based on their own previous consumption and average community consumption.
6. The system must provide instantaneous feedback.
7. The system must allow the user to get accurate in-depth information about their energy consumption (i.e. by appliance information).

## 2.4 Chapter summary

In this chapter it is established that in order to reduce energy consumption in the home changes need to be made towards the way in which energy is consumed. For homeowners to take more pro-conservation energy behaviour on board they need to change the way in which they perceive and consequently use energy. This behaviour modification requires both motivation and knowledge.

In order that the changes are persistent and effective, the individual should possess as near an intrinsic motivation for the task as possible, which can be accomplished though useful, timely and interactive feedback about how energy is being consumed. The feedback should allow the user to understand where they are using energy in their home, along with tips on how to conserve in specific areas. It should also allow the user to see the effects of all conservation measures they take. In an electricity framework, this implies providing disaggregated feedback so that the user can see which appliances are consuming electricity. In order that the system is usable, the AEM should be easy to install and cost no more than £70.

The dual implications of a monitoring system which is cheap and easy to install, and which can provide interactive and disaggregated feedback to individuals about their energy consumption forms the basis for the rest of this thesis.

# Chapter 3

# Critical Analysis of Existing State of the Art

The aim of this chapter is to identify the implications of the design criteria (proposed in Section 2.3.2 at the end of Chapter 2) and to determine whether there is any existing relevant state of the art.

Before the implications of the design criteria are discussed, Section 3.1 begins by providing a short discussion about smart metering in the UK. Whilst the objective of this thesis to develop a tool to aid consumers reduce their electrical energy consumption, it is noted that there are further considerations in the broader context of smart metering. Section 3.1 outlines the general requirements of smart meters, the current drive to replace existing meters with smart meters and the implications of this on the design of the AEM.

Having established the role of the AEM within the context of smart metering, a discussion into the implications of the design criteria of the AEM is presented in Section 3.2. This results in a proposed system designed to track individual appliance energy consumption which monitors the total energy consumption of the home using simple hardware. Complex software is then employed to identify individual appliances within the total load and to subsequently track their energy consumption.

This thesis considers three possible mechanisms of extracting characteristic signatures which may be used to identify which appliances are drawing power from the supply at

any one time. These mechanisms, which are explored in detail in Section 3.3, include probing the supply and analyzing the returned signature, tagging appliances so that they generate a particular characteristic signature when they are turned on and monitoring the current drawn by appliances when in use and identifying a signature from this signal. By considering the consequences of each mechanism, it is decided to identify appliances in a non-intrusive manner by identifying characteristic signatures observed in the electrical current drawn by an appliance.

Having identified the mechanism by which the energy consumption of individual appliances is tracked within the total load, Section 3.4 considers the current state of the art in the context of the proposed system design. The discussion is constrained to two papers which present work in the area of non-intrusive appliance load monitoring.

The limitations of the existing research is explored in Section 3.5. This enables the proposition of a possible mechanism designed to overcome some of these limitations. This mechanism forms the basis for the work in the following chapters.

## 3.1 Smart Metering in the UK

### 3.1.1 What are smart meters?

Smart meters are not universally defined but at their most basic they must contain a measurement system with some amount of data storage (to monitor the electrical energy consumed and store the data) and a communications link (to communicate the information to another device). This communication link may be one way or two way. A one way system only allows the energy information to be transferred from the meter to some external data collection point allowing for remote accurate meter readings, but a two way system allows for automated meter management with greater possibilities including remote switching between credit and prepayment [8] and remote tariff-change and time-of-day tariffs [27] which may aid in the reduction of the peak load demand.

The key capabilities of a smart meter are therefore defined as [27]:

- The ability to measure the electrical energy consumption at defined time intervals.
- The ability the record the electrical energy consumption at a 'billing-level' accuracy.
- A two way communications link.
- The ability to store the time-interval electrical energy consumption data and transfer information remotely to a data collection point.
- The ability to store and display both the electrical energy consumption information and the price tariff information.

### 3.1.2 Government expectations

In the 2007 White Paper on Energy [8] the government states that it intends to 'roll forward a package of measures in Great Britain which will change the way in which energy use is communicated to customers'. The expectation is that, by 2017, all domestic energy consumers will have smart meters with visual displays installed in their homes. The government also states its belief that all interested customers should have access to a real-time electricity display which enables them to observe their immediate energy consumption. It is proposed that from May 2008 all homes in which electricity meters are being replaced and all newly built homes should be provided, free of charge, with a real time electricity display. Furthermore, until March 2010, any households requesting a real time electricity display should be provided with one free of charge by their energy supplier. These displays provide real time feedback about the energy consumption in the home and the associated costs at a minimum performance requirement of 95% accuracy.

These statements made by the government tackle two key issues. The first is in the installation of smart meters in UK homes which will provide a large range of benefits to the supplier (as outlined in the following section). However, because the requirements of

a smart meter do not have implications on the feedback of information to consumers, the roll out of smart meters may not provide much potential for energy (and money) savings to consumers. This is tackled by the requirement for a real time display, which has been shown in the previous chapter to help individuals reduce their energy consumption.

### 3.1.3 What are the benefits of smart metering?

This section briefly outlines the potential benefits of smart metering for both the supplier and the consumer. Supplier benefits were identified by considering smart metering projects undertaken by other countries and following discussions with suppliers [27], [28]. These include:

- Reductions in meter reading costs due to communications links enabling automated remote meter reading.
- Reductions in fraud, theft and debt - remote meter readings eliminate need for estimated bills, better use of prepayment options, automated detections of meter tampering.
- Improved cash flow - more frequent accurate meter readings result in reduced time periods between electricity usage and accurate payments.
- Reductions in customer service costs - more accurate billing and hence reduced billing query calls, reduced problems with pre-payment meters hence fewer call outs.
- Increased ease of supplier switching with more accurate data transfer between companies.
- Possibilities of time of day price tariff to help reduce peak demands which results in lower costs to suppliers and better security of supply.
- Possibilities for users to switch remotely between credit and prepayment methods

to help reduce costs.

- Possibilities for companies to better direct retail packages at individual users - e.g. time of use tariffs; energy services, micro-generation, increased dual fuel offers.

Consumer benefits are more focused towards improved cost management schemes and include [27]:

- More accurate bills. This has been shown to result in some amount of savings (in both money and energy), particularly with regards to seasonal changes in use of appliances (i.e. the increased use of heating and lighting in winter than summer).
- Better consumer information about energy consumption from a display in the home. This has been shown to result in significant energy and cost reduction from more efficient use of appliances.
- Improved ease of supplier switching and benefits of new offers from suppliers.

However, despite the potential benefits to both the supplier and consumer, and the potential reduction in energy use and carbon dioxide emissions, there is not a clear business case for widespread deployment of smart meters [27] and policy or regulatory intervention is needed if the targets set out in the Energy White Paper are to be met. The next steps must therefore be a full analysis into the possible smart meter deployment options, with the government, Ofgem and the industry working hand in hand [28].

### 3.1.4 Possible technical options for smart meters

Recall that smart meters must electrically monitor and store electricity consumption data (at a billing level accuracy) for defined time period durations. It must also have a two way communications link to enable this information to be fed to both the supplier and a feedback display within the consumers home and to allow automated meter management. Figure 3.1, taken from a paper by Marvin, Chappells and Guy [29], il-

lustrates the communications links which may be considered for smart meters, and the environmental opportunity and smart applications systems which may subsequently be harnessed. These environmental opportunities and smart applications are key in realizing the potential benefits to both consumer and supplier which may be delivered from smart metering.

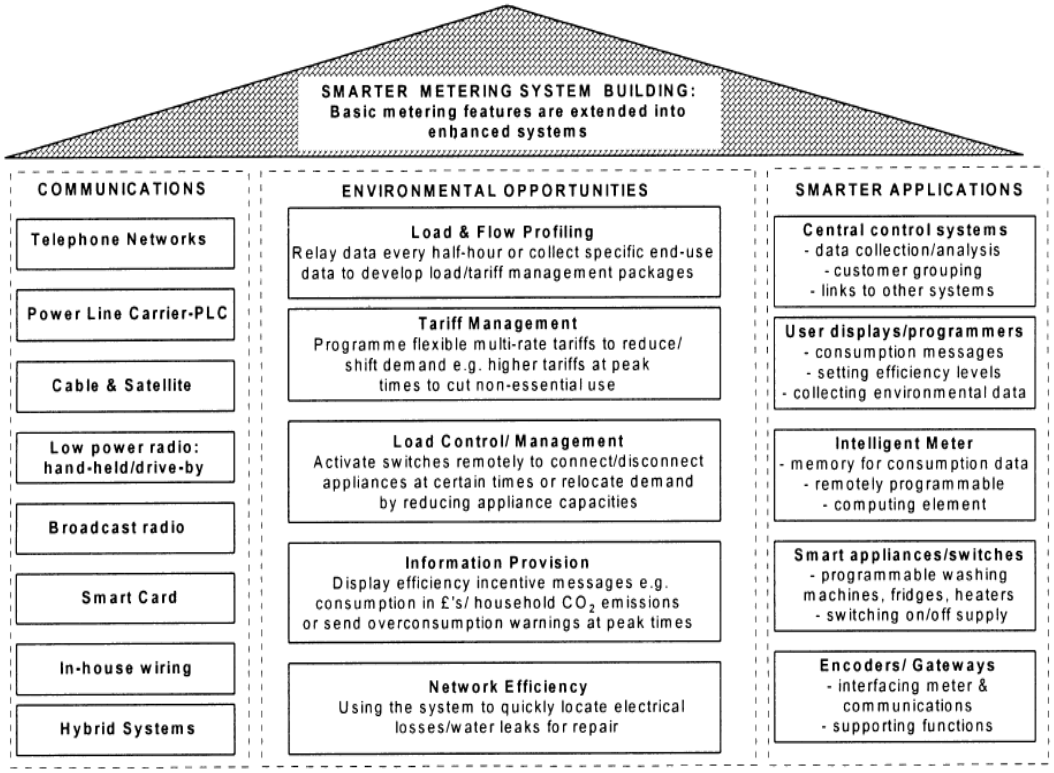


Figure 3.1: Illustration of the possible technology options for smart metering systems, image taken from [29].

### 3.1.5 The role of the AEM in smart metering

Meeting the objectives set out in this thesis lies in the development of a device which will enable individuals to reduce their domestic electricity consumption. It does not focus on the other potential benefits which smart meters may deliver, though these should be kept in mind as possible sideline opportunities of the AEM. Therefore, the AEM should facilitate effective feedback of information to individuals, whilst retaining the potential to provide this information to energy suppliers.

The AEM shall be viewed as a smart application opportunity of smart metering in lieu of Figure 3.1. The AEM shall monitor the electricity consumption (though it may be possible that this data is read in from the actual smart meter), manipulate this information and communicate some form of intelligent feedback to the consumer. This feedback of information will add to the Information Provision environmental opportunity of Figure 3.1. A more detailed outline of the design of the AEM is established by considering the implications of the design criteria of Section 2.3.2.

## 3.2 Implications of the design criteria

This section considers the implications of the design criteria, established in Chapter 2, on the mechanism by which the AEM monitors the domestic energy consumption in order to provide useful feedback to individuals.

Design criteria 1 and 2 relate to the physical construction and installation of the AEM. The AEM must be easy to install and cheap to manufacture (in order that it is commercially available for less the £70) if it is to have an impact on the UK domestic market. This places constraints on the construction of the AEM and on the data processing, memory and communications specifications of the device.

Design criteria 3, 4 and 5 have implications on the method by which the information (once collected and analysed) is presented to the user. For the information to be useful it must allow the user to understand where they are using the energy in their homes. From the studies carried out in Japan [23] [24] it is observed that the information used the most frequently by participants relates to the energy consumption patterns over the course of a day. This implies that the AEM must provide the capabilities of storing the necessary information. This may require links to a PC if the standalone AEM device is to cost less than £70 to the consumer. To allow the user to compare their energy use to a community average, a link via the PC to some connected website is possible.

The implications of these design criteria are important in enabling the structure of the AEM to be considered, however, further consideration into these feedback mechanisms are beyond the scope of this thesis.

Design criteria 6 states that the user must be able to get instant feedback about the energy consumption in their home. This implies links between the monitoring system and a display in a visible location which allows the user to instantaneously observe any changes in their consumption patterns. Again, the method by which this information is provided to the user is beyond the scope of this thesis.

Finally, design criteria 7 states that the user must be able to obtain in-depth information about which appliances are drawing power from the supply at any one time. This implies some method of tracking the state of every appliance in the home so that at any one time it is known which combination of appliances are consuming energy.

There are two potential methods which enable the tracking of individual appliance energy consumption.

1. Monitor each appliance individually using complex hardware. Use simple software to collate these results and present to the user.

2. Monitor the total energy consumption using simple hardware. Use complex software to identify individual appliances within the total load and having identified the states of the appliances use this information to track their energy use.

If Method 1 is used to track the energy consumption of individual appliances the problem of identifying which appliances are drawing energy from the supply at any one time becomes a very simple problem. However, this method implies the use of complex hardware which is likely to violate design criteria 1 and 2. If each appliance is to be monitored individually, a tracking device must be placed between the socket from which each appliance draws energy and the actual appliance. Whilst this may be a simple problem for some appliances, it is not so for built in appliances (such as ovens or electric showers), which

tend to be the most energy consuming. Furthermore, if multiple monitoring devices are required, which all have communication capabilities, it is unlikely that the total product will be in the price bracket determined by design criteria 2. This method is therefore not currently suitable for use in the domestic environment.

Method 2 proposes the use of a single monitoring device which measures the energy consumption of the whole home. By ensuring that the proposed monitor is compatible with standard UK electricity meters it is possible to design a system with hardware which satisfies design criteria 1 and 2. The software implemented in the system must enable individual appliances to be identified within the total load and therefore a complex system of software is required.

Figure 3.2 illustrates the system design of the AEM under Method 2. Data relating to the total energy use of the home is monitored at a single point. This may be data collection part of a smart meter, if it is found that the required information for the AEM is identical to that captured by smart metering. Otherwise, the data shall be collected at the user side of the smart meter, before the branching of the supply into smaller circuits. Once the data is captured is it stored. Again, if the two storage requirements are identical, it may be that the storage part of the smart meter will form the storage solution, otherwise a separate data storage unit will be needed. Once this information has been captured and stored, the main body of the AEM processes this data. This is in order to establish the energy consumed by individual appliances. Communications links between the various parts of the system are indicated by arrows on the diagram. These communications links may be any of those described by Figure 3.1.

Once the AEM has manipulated the information, it is sent to a real-time display, a PC and to the energy supplier. The real-time display is one of the requirements set out by the government in the Energy White Paper [8] and this will enable individuals to see how much energy in total they are currently using, with comparison to other total

energy use stats. The links to the PC allow the user to delve more deeply into the energy consumption statistics - it is here that they will be able to interact with the data and observe the energy consumed by individual appliances. There are further possible links to an on-line community. This will not only allow the user to compare their energy consumption with others in the area, but will also enable software upgrades for the programs on the PC. Because the AEM main unit also has communications capabilities it is possible that automated upgrades will be available for this part of the system. Communications links to the energy supplier allow the supplier to see how particular users are consuming energy, and it may enable them to offer various energy packages to encourage peak demand reductions. However, before online or supplier communication links are solidified, it is important to evaluate data security and privacy issues.

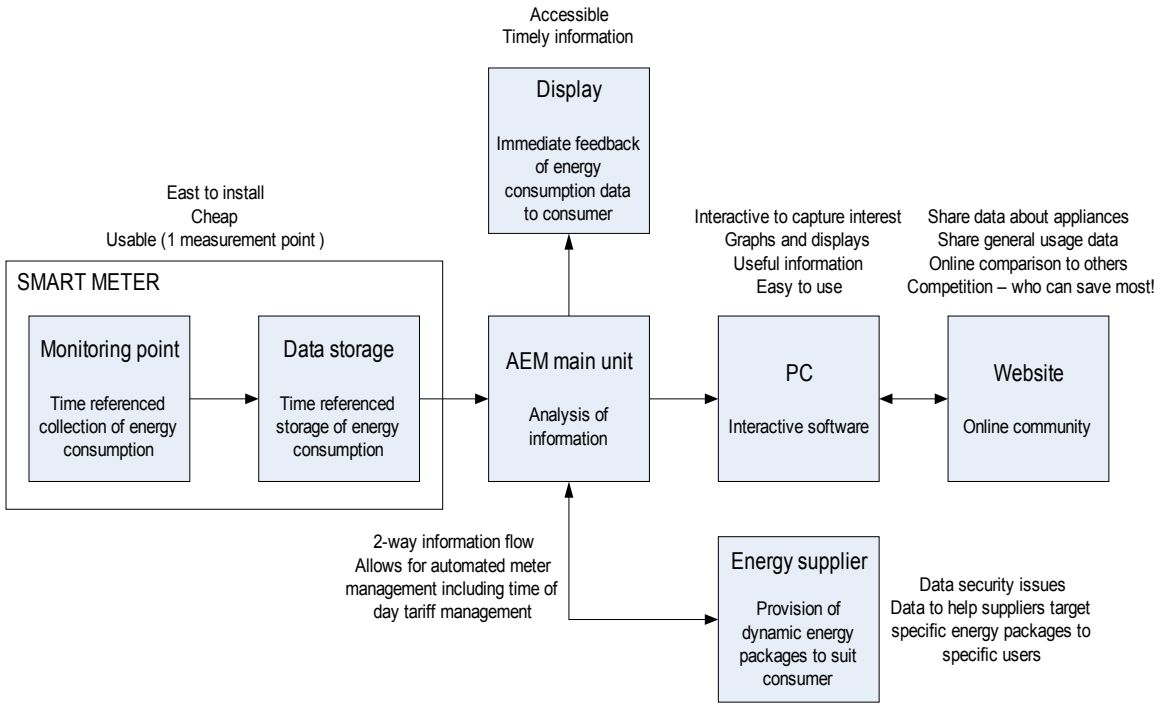


Figure 3.2: System design overview for the AEM.

The research presented in this thesis is concerned with the monitoring point, data storage, AEM main unit and the mechanisms by which individual appliance usage can be tracked using the proposed hardware. The aspects of the system beyond this point are not considered further in this research. In order to identify methods for tracking individual

appliance usage, the domestic environment is first modelled. This provides an insight into the plethora of available information.

## 3.3 The domestic model

The monitoring point tracks the total load, which, for the domestic environment, can be likened to a set of appliances connected to the supply (between the live and neutral lines) in parallel, as illustrated in Figure 3.3.

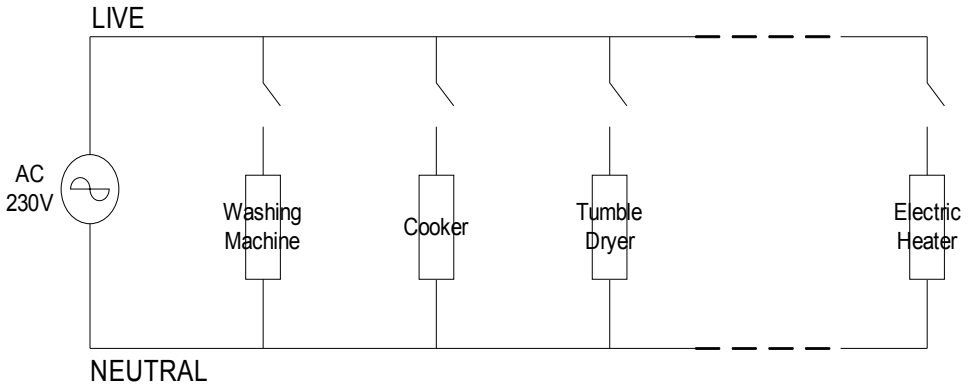


Figure 3.3: Illustration of the connection of various loads within the domestic system.

When an appliance is switched off, the switch between the supply and the appliance is opened and it appears disconnected. However, when an appliance is on, the switch is closed and characteristics inherent to that appliance may be investigated.

This thesis considers three possible mechanisms for investigation:

1. Electrically probing the appliance with some input signal and analysing the returned output.

2. Tagging appliances with some external device which is connected in such a way that it emits a particular signal when the corresponding appliance is switched on.

3. Observing the current and power drawn by the appliance when it is in use. These signals are individual to the particular electrical construction of the appliance.

The following sections consider each of these mechanisms in greater detail and explore their implications with reference to the key design criteria.

### 3.3.1 Electrical probing

When a signal is applied to an electrical appliance it reacts in a way determined by the components which constitute its electrical interface.

Three common linear electronic components include resistors, inductors and capacitors, denoted $R$, $L$ and $C$ respectively. These are included in the construction of a large number of electrical items and therefore the mechanism of electrical probing may be explored by considering the different outputs which may be observed by electrically probing these simple components. Under a 50Hz voltage source, $v = V_0 \sin(\omega t)$, the current drawn by the three components will be $i_R = \frac{V_0}{R} \sin(\omega t)$, $i_L = \frac{V_0}{\omega L} \cos(\omega t)$ and $i_C = -V_0 \omega C \cos(\omega t)$ for the resistor, inductor and capacitor respectively, where $\omega$ is the fundamental frequency of the voltage. However, if the voltage source injected contains some component of the 3rd harmonic, so that $v = V_0 \sin(\omega t) + V_3 \sin(3\omega t)$, the current now drawn by the resistor, inductor and capacitor will be $i_R = \frac{V_0}{R} \sin(\omega t) + \frac{V_3}{R} \sin(3\omega t)$, $i_L = \frac{V_0}{\omega L} \cos(\omega t) + \frac{V_3}{3\omega L} \cos(3\omega t)$ and $i_C = -V_0 \omega C \cos(\omega t) - 3V_3 \omega C \cos(3\omega t)$ respectively.

Therefore, features which are indicative of the components contained in an appliance's electrical interface may be deduced by probing the appliance with some constructed signal (for example amplified harmonics of the voltage signal or transient pulses) and observing the changes in the current drawn by the appliance [30].

This method of analysis allows the system to be interrogated at times determined by the user and makes no constraints as to the number of appliances which may have changed state between two inspections of the system load. It simply considers the appliances which are in use at any one time.

However, a system which attempts to monitor the actual state of all the appliances

in the home suffers from a complexity problem. If the system contains N appliances, which exhibit only ON/OFF behaviour, there is $2^N$ possible combinations of appliances to consider. For a home with 20 appliances[1] this results in 1,048,576 combinations of appliances. This requires large computational power which is not available in the proposed system and therefore a system which monitors the changes occurring in the system would do better.

There are further problems which relate to interference and power quality [30] which discourage utility companies from backing this form of appliance interrogation. For these reasons this mechanism is not explored further in this thesis.

### 3.3.2 Tagging appliances to generate features

It is possible to tag appliances with radio frequency (RF) devices which emit a particular RF signal when the appliance is in use [31]. This signal is carried over the power lines to a read-write device (RWD) which identifies the appliance from the features of the signal received.

The tag is attached to each individual appliance. This requires an intrusive approach, although this only needs happen one time unless the appliance or tag needs to be replaced or repaired. In the case of portable appliances the tag is attached to the power cord to enable identification even if the appliance is moved. Once the tag is installed it emits a signal over the power lines whenever the appliance is in use.

This approach allows easy identification of appliances if each tag is individual and even allows two identical appliances to be identified as separate appliances. However, this method of appliance identification currently invalidates design criteria 1 and 2.

The proposed system is no longer easy to install as tags need to be added to each appliance in the home. Furthermore, to prevent interference on neighboring electricity

[1] 20 is selected for this example as it represent the typical scale of magnitude for the number of appliances in the home.

supplies, filters must be added to the system. The system now comprises a number of pieces of hardware (the tags, filters and monitoring devices) which are unlikely to satisfy the cost constraints. For these reasons this method of generating features is not explored further in this thesis, however, it is acknowledged that at some point in the future smart appliances may well be produced with integrated tags, removing all problems discussed here.

### 3.3.3 Extracting features from the current

Each appliance draws current from the power supply in a way which is governed by its electrical interface. The observed current is therefore a signature which is indicative of the appliance's electrical interface and consequently of the underlying appliance. A set of three current signals are show in Figure 3.4. The method by which this data, and in-fact all the data used in this research, was captured is detailed in Appendix A. From Figure 3.4 it can been seen that a large variation in current may be observed, even for two appliances which are similar by end use. It can be seen that the current drawn by TV1 is largely resistive whilst that drawn by TV2 is non-linear and perhaps caused by some kind of switching behaviour. The current drawn by the CD player is different again, exhibiting positive peaks at 90° and 180°, and negative peaks at 270° and 360°. This illustrates the plethora of information available in the current drawn by individual appliances when in use.

By extracting features from the current observed when an appliance is in use, appliances with different currents are represented by different parameters. From Figure 3.4 it can be seen that there is a plethora of information which can be extracted by considering the shape of the current and all this information is available for use in aiding the differentiation between various appliances. The features should be chosen so that different appliances have parameters which clearly distinguish them from other appliances (Chapter 6 explores the type of available information in much greater depth).

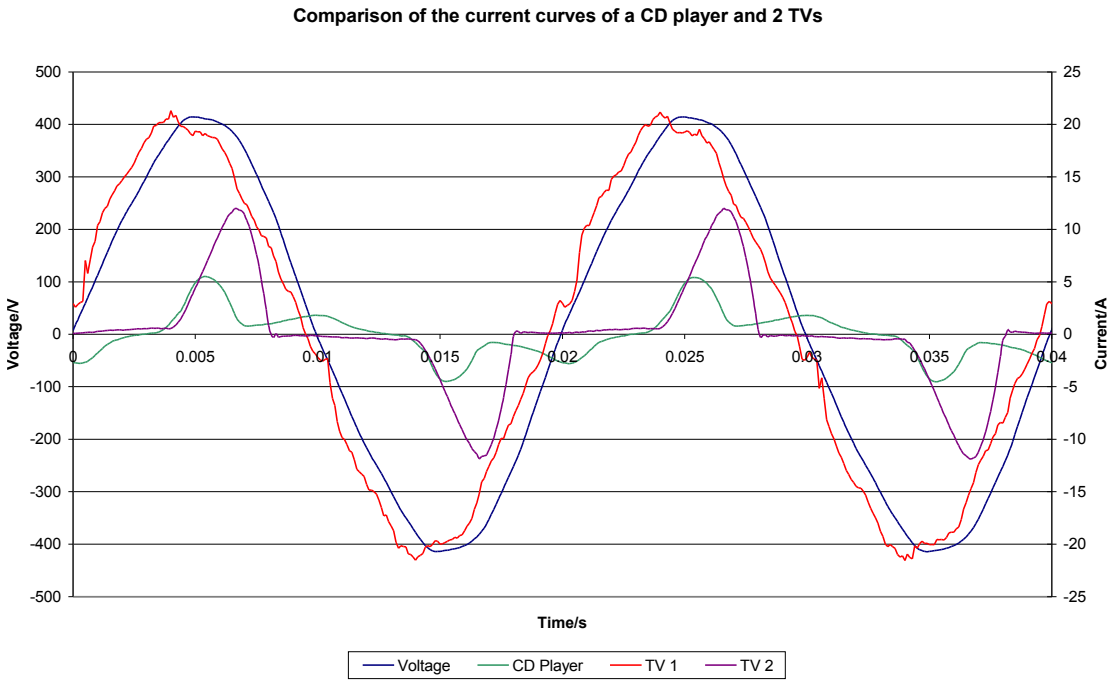


Figure 3.4: Comparison of the current drawn from different appliances. Illustrated here is a CD player and 2 TVs illustrating that even appliances which can be classed with the same name (i.e. TV) may have very different features depending on their electronic construction.

Unlike the previous two mechanisms of identifying feature sets, this method is non-intrusive and does not require the use, or complex installation, of any expensive hardware. Therefore, it is possible to implement this type of analysis without invalidating the design criteria.

The difficulty of this method lies in identifying individual appliances within the total load. This can be demonstrated using the following simple example. Consider the scenario that the features identified from the signals monitored correspond to the power consumption of appliances and that the home contains 5 appliances with the following power consumptions: 100W, 200W, 300W and 401W and 500W. If the observed total power consumption of the home reads 500W, then it is possible that either the 200W and 300W appliance are switched on, or the 500W appliance is on. Furthermore, if the reading suddenly changes to 501W, it would now appear that the 100W and 401W appliance are on, despite the unlikely ability of all 3 or 4 appliances to switch on/off

simultaneously. It is more likely that there is just some variation within the reading due to noise, measurement errors or voltage variation (the voltage in the UK is set at 230 V +10% -6% [32]).

This example demonstrates two problems with the proposed analysis. The first is its inability to consider multiple loads if the feature set does not provide enough information. This is also worsened by the fact the number of appliances in the home is more likely to be in the region of 20 , resulting in a possible 1,048,576 combinations of appliances which may be on or off. The second problem is due to the variation within the features emitted by a single appliance. The system must be able to ignore small changes which occur and are likely to be caused by noise, the variation in voltage or errors in measurements.

Both problems may be overcome by considering changes in the feature sets observed, which correspond to the parameters of an individual appliance, rather than the feature sets in themselves.

### 3.3.4 Identifying appliance state changes

This section considers the mechanism by which the current drawn by an individual appliance is isolated from the total current drawn by the home in order that features may then be extracted from this isolated current.

When an appliance switches on or off (or changes between states if it has more than one mode of operation) a change in the power signal will be observed. In most cases this will correspond to a step change in the power as seen in Figure 3.5. By comparing the current waveform before and after the occurrence of the step change, the difference is isolated and it is this difference which represents the current consumption of the individual appliance.

However, the observation of a step change must be considered with reference to other surrounding transitions in order to avoid misrepresentation of the change in current observed. This is because the step change in power may occur in two, rather than one

transition, which can be seen by observing the data points in Figure 3.5. The first step change in power is from 0W to 1120W and the second from 1120W to 2250W. As seen in the graph, 2250W is the working power of the appliance. The reason that the step change in power occurs in two transition is because the appliance is switched on mid voltage cycle, as illustrated in Figure 3.6.

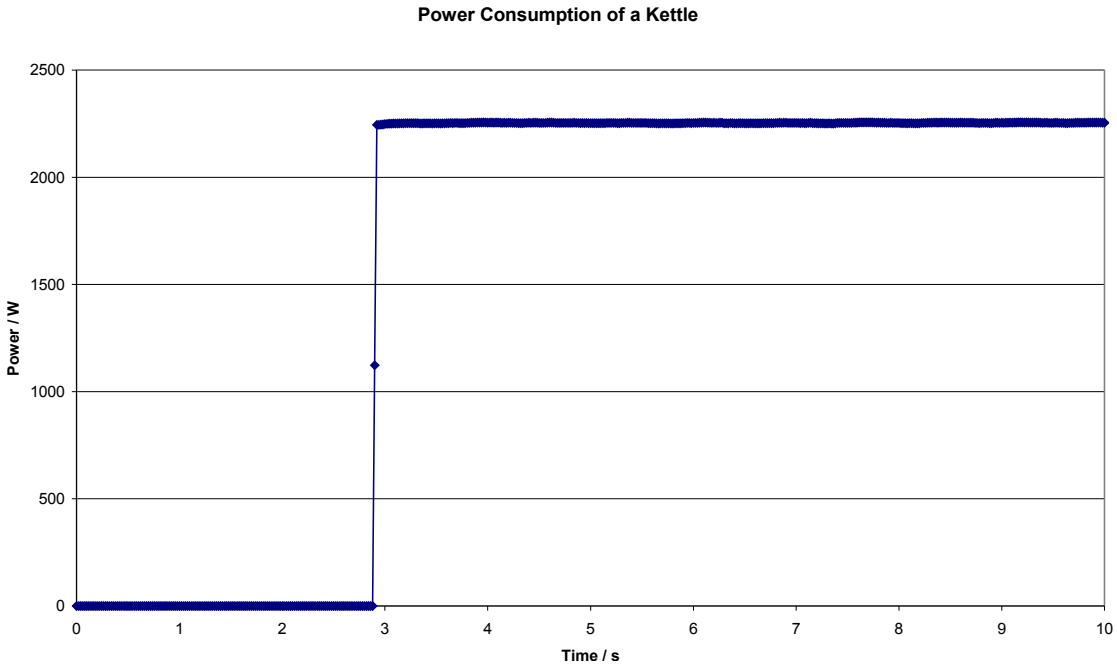


Figure 3.5: A plot of the observed power being drawn by the kettle before and after the kettle is switched on. Each data point corresponds to the power of an individual current and voltage cycle.

It is also possible that the power may rise above the steady state value before settling back down due to some transient behaviour. For instance, this will occur during switch on if there is a large capacitor in an AC-DC power supply. A capacitor draws current according to the equation $i = C\frac{dv}{dt}$. Therefore, if the voltage seen across the capacitor at switch on is larger than that already stored on the capacitor, the value for $\frac{dv}{dt}$ will be large and consequently there will be a current surge whilst the capacitor charges up. This will cause a transient peak in the power for typically one cycle.

The current of the individual appliance is calculated by taking the difference in current waveforms before and after the state change is observed. Because the features are based

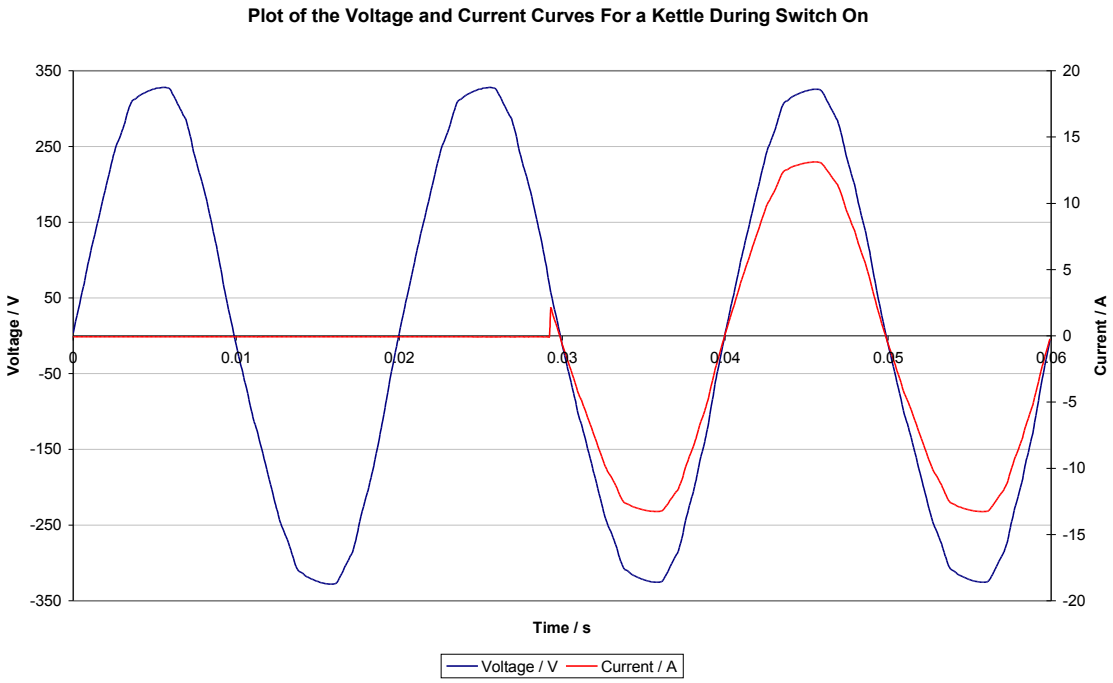


Figure 3.6: A plot of the current being drawn by the kettle and the voltage across it before and after the kettle is switched on.

on the steady state behaviour of appliances, any transient behaviour should be ignored when calculating the current waveform for the individual appliance. For this reason, when a step change in the power is observed, the first cycle after the step change is ignored (as the chance of switch on being at the exact start of a cycle is 1 in 256) and the difference in current and power consumption is only measured once the power is observed to settle into steady state.

Steady state is defined as a constant current (above the noise floor) for a defined period of time, $\tau$. $\tau$ is defined by considering appliance state changes – $\tau$ must be small enough to ensure that other appliances do not switch on during this time frame, but must be large enough to avoid problems caused by transients being misclassified as steady state. Therefore $\tau$ is defined as 0.1 seconds, or 5 cycles. Assuming that the current and voltage monitors contain 8 bit ADCs (analogue to digital converters) which must account for positive and negative signals, there are effectively 7 bits with which to measure the signals. The smallest bit is assumed to contain noise error and therefore the conversion to digital

may contain an error of approximately 1%. Assume that there may be approximately this much error also present in the analogue side of the monitoring system for both the current and voltage signals. Therefore, define a constant current to be one for which the power varies by less than 5% of the average power over 5 cycles.

It should be noted that only short timescale transients will be removed by considering a steady state as 5 cycles and longer timescale transients will not be accounted for. One very common transient which occurs over a longer time scale than that considered here is due to the heating effect in resistors. This causes the resistance of a device to increase during use and the power consumption to decrease, as observed in Figure 3.7. Initially the power consumption is at 1250W, but this decreases to 1185W over time. The switch on characteristics will correspond to those features identified at the 1250W operation, but the switch off characteristics (assuming that the hairdryer is in use for longer than about a second which is the time taken for the change in resistance to occur) will correspond to the usage at 1185W. It is not possible to remove these longer duration transients due to the increasing probability with time that further appliances will switch on, and therefore these transients must be accounted for by the system (this is explored further in Chapter 5).

This method of detecting changes in the observed power above a certain threshold value removes the problem of identifying features of individual appliances within the agglomerated current and also removes the problems of apparent changes in features due to noise.

One fundamental problem with this method of analysis is the inability to detect appliances which do not have a step change when switched on or off (i.e. appliances with a ramping function) and the inability to detect more than one instance of the same appliance. However, these issues are unlikely to be overly problematic in the domestic environment. This is because there are not many appliances in the home which exhibit

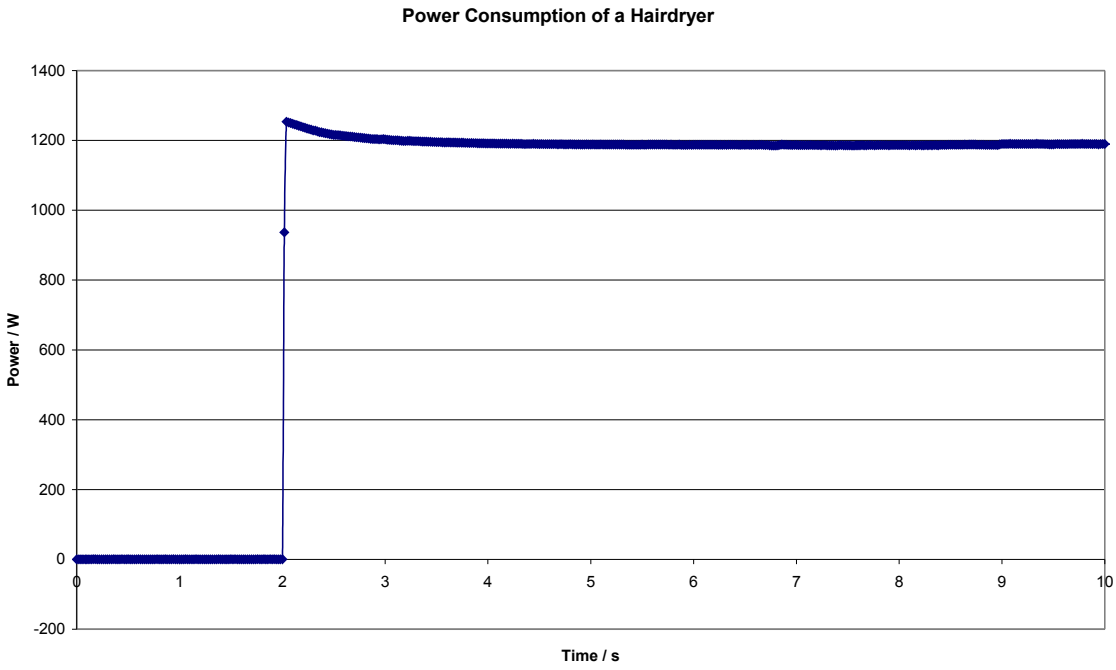


Figure 3.7: A plot of the power consumption of the hairdryer before and after switch on.

ramping power curves – the most common is dimmer switches on lights, but these tend to have step changes followed by ramps in the power. It is also unlikely that the inability to detect between separate instances of the same appliance will be problematic when considering the context in which this research is carried out – if the user does not know which one of two identical appliances are consuming energy it is unlikely to affect their energy conservation actions a great deal.

Having identified the mechanism by which the energy consumption of individual appliances is tracked within the total load, the following section considers the current state of the art.

## 3.4 Existing mechanisms which monitor the energy consumed by individual appliances

This section considers existing systems which have been designed to identify the energy consumed by individual appliances whilst non-intrusively monitoring the total load and extracting relevant features from the signatures observed.

Whilst it has been acknowledged in the previous section that the mechanism for detecting the energy consumption of individual appliances in the home should consider step changes in the power to highlight the state changes of those appliances, Section 3.4.1 begins by exploring a method which tracks the states of appliances by considering the harmonic content in the steady state signals monitored. This work is included for completeness, and the problems with this type of method are presented in Section 3.4.2. This reiterates the need to consider only the step changes observed in the system.

When a step change is observed in the power consumption in the home it is assumed to be caused by an appliance switching states. Once the power has settled into steady state, the real and reactive power changes may be calculated. Section 3.4.3 presents a paper published by George Hart in 1992 [30] which explores the method by which these real and reactive powers may be used to identify appliance energy usage. Two options are presented. The first is in the scenario that the system has undergone a manual training procedure so that the observed features may simply be matched to a set of known features. The second is when this manual training is not present, and state diagrams are deduced for appliances under pseudo names. The method by which these state diagrams may be deduced is presented in work by Baranski and Voss in 2004 [33], however, this is determined to be too complex for the work presented in this thesis and for the time being it is assumed that there shall be a manual training procedure.

Section 3.4.4 considers work which uses the presence of transient features to identify appliances. Sections of the power signal where the total load is varying rapidly, denoted v-sections, are used to help identify the underlying appliance causing the transient. Two methods are presented. The first is in a paper published by Leslie Norford and Steven Leeb in 1996 [34]. The second method is formalised in a patent application by General Electric Company, filed in 2007 and published in 2009 [35].

The limitations of the existing state of the art are explored in Section 3.5. By considering

the reasons behind these problems a set of solutions are proposed which may be used to overcome the existing limitations, and this forms the basis for the computational design of the AEM.

### 3.4.1 Harmonic content of steady state signals

This section considers work carried out by Katsuhisa Yoshimoto in the Komae Research Laboratory in Japan [36]. This work is based on the ability of a non-intrusive load monitoring system, attached to an external revenue meter, to infer the states of individual appliances within the home. It is particularly targeted towards appliances which have load consumption fluctuations, such as inverter driven systems present in many homes in Japan.

When appliances are in use they drawn current from the supply in a way unique to that appliance. Because the current signal is repetitive in the 50Hz frequency domain, it may be decomposed into an infinite harmonic series. Therefore, when an electrical appliance is switched on, a set of harmonics are observable which are unique to that appliance. These harmonics may therefore be used to infer the appliance which generated them.

The methodology applied by the researchers in the Komae laboratory was to train a neural network to learn the correlation between the harmonics observed and the load consumption of appliances. More information on neural networks is available in books [37], [38] and is not presented in this thesis due to spacial constraints. A set of training data was generated for the neural network by variously combining 5 typical household appliances including an inverter-driven air conditioner, an inverter-driven refrigerator, an incandescent lamp, a fluorescent lamp and a television. An experimental verification of the algorithm was then performed with a correct inference rate 99% of the time. Therefore, the algorithm developed may correctly identify the energy consumption of these 5 appliances, combined variously, with 99% accuracy.

### 3.4.2 Limitations of the harmonic approach

The main limitation of the approach described above is due to the time and data concerns with neural networks in general. Because the system is not considering state changes of appliances and instead monitors the signal generated by any combination of appliances, every time a new appliance is added to the system the network will need to be entirely retrained to incorporate the new appliance as the harmonics observed in the total signal will be different. The time taken to retrain the network will be proportional to either $N * d$ or $N^2 * d$ depending on the type of algorithm used in the network, where $N$ is the number of appliances and $d$ is the number of harmonics used to characterise each appliance [37]. Therefore as the number of appliance which are connected to the system increases, and as the number of features used to represent appliances increases, so does the time to train the network and the amount of data needed to train the network .

By considering only step changes in the system, characteristics indicative of individual appliance may be isolated from the background noise and consequently, when the system needs to be trained to a new appliance, the effect of all the other appliances connected to the system may be ignored. This strengthens the case for identifying the step changes in the power signal and subsequently extracting a set of features which are indicative only of the appliance which has just changed state.

### 3.4.3 Non-intrusive appliance load monitoring

The proposed system uses a non-intrusive appliance load monitoring (NALM) mechanism in order to determine the energy consumption of individual appliances. The appliances considered in the analysis are limited to those which are represented by ON/OFF state diagrams and those which are represented as finite state machines (FSM).

The NALM system monitors the whole house current and voltage waveforms once every second. The power is calculated and normalised to account for changes in magnitude

of the voltage waveform (which may be up to 10%). When an appliance in the home changes state there is a corresponding change in the observed power consumption of the total load. Edge detection is used in order to track these changes and the features which correspond to this change are isolated. In this analysis the features, extracted from the current and voltage, correspond the the real and reactive normalised power of the change. The output from this step of the analysis is a list of the features identified at every step change in power consumption. The task is to use these features, which form the appliance signature, to identify the appliance from which they were drawn.

If the NALM system has undergone a training procedure during installation, there will be an on board database corresponding to the specific signatures of the appliances contained within the home. This allows the set of observed features to be labelled with the appliance name of the corresponding known signatures. The ability to track the energy use of individual appliances is now reduced to a simple problem of tracking the on and off times of individual appliances and tabulating statistics.

However, in the event that the list of known appliances is non-exhaustive, or in the case of a NALM system which has not undergone a manual training procedure, the observed signatures are classified in the following way. The set of observed step changes are plotted in feature space (a 2-dimensional space including the real and reactive power) and clusters are formed from this data. Each cluster corresponds to a single transition of a single appliance. The analysis then uses these clusters to determine the ON/OFF and FSM diagrams based on the magnitudes of the observed changes. Having identified which change belongs to which model, these changes are tracked under pseudo names and the final requirement is to match these pseudo names to real names for the individual appliances. In the work presented by Hart the methodology by which the FSM diagrams are determined is not so clear and it has been further developed by Baranski and Voss [33]. However, for simplicity, it is decided to constrain the work presented in this thesis to those times when manual training is possible (though it is noted that extension to

an automated training regime is possible) and therefore a more detailed discussions into Baranski and Voss's work is not presented here.

### Limitations of the NALM system

The NALM system has limitations in both the information that is collected and the mechanism by which that information is analysed. Whilst the system proposes sensible methods for physical data collection (so that the hardware is easy to install and cheap to manufacture) and training procedures (in both the manual and automated cases), it is not so clear as to how an appliance signature, once identified, is matched to the appliance from which it is drawn. Consider the case in which the system is trained manually during installation. In this scenario the system will already have a collection of typical signatures for the appliances within the home. Once the NALM system is working in appliance tracking mode (as opposed to training) it will identify signatures which correspond to every observed step in power consumption. However, due to noise on the system or variation in the voltage signal (discussed later in Section 6.1), these signatures may be slightly different to those measured during training. There is no clear way in which these signatures are subsequently matched to the set of known signatures. This classification task of identifying the underlying appliances from which a set of observed signatures is drawn must therefore be further developed if the non-intrusive mechanism of monitoring appliances is to become successful.

A further limitation of the NALM becomes evident when considering the information present in the identified appliance signature. The analysis used considers only two features, the real and reactive power of the appliances. Therefore, if two different appliances have similar power levels they may be classified as the same appliance. This problem is particularly evident when there is no training regime and the analysis must use the observed transitions to build FSM models of appliances. A cluster analysis procedure is used and there will therefore be problems of misclassification if two or more clusters lie

close together in feature space (cluster analysis is discussed at much greater length in Chapter 5).

Further problems are introduced when considering appliances which have transient or ramping behaviour. If an appliance has a power which ramps up or down during use, the on and off characteristics may not mirror one another. If the appliance ramps from zero power, the system may miss the appliance all together.

If an appliance exhibits transient behaviour during start up, the NALM system may wait some time for steady state behaviour before determining the changes occurring. If another appliances changes state during this time, the system will observe an appliance with switch on characteristics which are a sum of the two individual appliances.

Different appliances have different turn on transients due to inherent differences in start up. Therefore, the observed transients may be useful in adding a further dimension to the appliance signatures, which will aid more accurate tracking.

### 3.4.4 Using transients to identify appliances

This section considers the way in which transient behaviour may be added to the NALM method of analysis in order to aid classification when tracking individual appliance energy use from the total load.

This method of analysis considers the start up signal of an appliance to be a segmented time series. In each segment the power may be changing rapidly, or it may be fairly constant. The transient detector identifies the segments which have significant variation, denoted as v-sections. Each v-section has a characteristic shape associated with it.

There is a training regime which is necessary for this analysis, during which the transient detector segments the start up time series to determine a particular sequence of v-sections which is representative of a class of loads, for example, an induction motor.

When the power signal of the home changes, rather than wait for the steady state conditions, the transient detector searches for a precise time pattern of v-sections and these are used to identify particular characteristics of the underlying appliance. This section considers two different sets of research which employ this analysis. The first was developed by Norford and Leeb and published in 1996, the second by Durling et al. and published in the US Patent Application Publication in 2009.

In the method employed by Norford and Leeb [34], two transversal filters are used to positively identify particular sequences of v-sections.The first filter identifies particular shapes in the data stream $x_{a.c.}$. The inner product is determined for $x_{a.c.}$ and the template shape $t_{a.c.}$, with an output of 1 indicating a perfect match. Due to noise on the system and small variations in the repeatability of v-sections, some amount of error is allowed for and therefore any output value within a designated tolerance limit of 1 is considered to be a match. The second filter considers the magnitude of the v-sections to ensure that any noise which appears to resemble an individual v-section in shape (but not magnitude) is not mistaken for a v-section.

The work published by Durling et al. [35] does not go into detail into the methods used to match the v-sections identified in the working phase to templates built during training, but rather builds on the work done by Norford and Leeb. The transient detector monitors the voltage and current signals and when transients occur particular parameters (including amplitude, before and after differentials, harmonics and phase) which are used to characterise that transient are recorded. The pre-programmed look up table is used to determine the transient index for the occurrence, denoted $\mathrm{TI}_m$, from a set of known transient patterns $TI_1, \ldots, TI_M$. Whilst the method by which this is done forms the bulk of the paper published by Norford and Leeb, it is not detailed in Durling's patent application, perhaps because the techniques employed are similar to those previously developed.

Once the transient occurrence has been identified, the task of the transient event detector is to match this to a particular appliance. Durling uses Bayesian theory to perform this task. A set of conditional probabilities $\Pr(TI_m|A_n)$ for appliances $A_n$ $(n = 1, \ldots, N)$ are determined in offline lab experiments. These denote the probability of observing a particular transient in the scenario that the event is known the be caused by a particular appliance. A Bayesian classifier is then employed to choose the appliance which maximizes the joint probability $\Pr(TI_m, A_n)$, where $\Pr(TI_m, A_n) = \Pr(A_n|TI_m) \,.\, \Pr(A_n)$. $\Pr(A_n)$ is approximated using a time series analysis by considering the historical time series of on and off events for particular appliances and thus determining the probability of the time for the next event for that appliance.

**Limitations of transient monitoring**

Transient monitoring will run into problems if a second appliance switches on during the v-section segment of the start up period of the first appliance. However, if the first appliance has a short transient this is unlikely due to the rate of occurrences being low in comparison to the time constant of the transient. If the first appliance has a long transient, it is unlikely that the whole transient will contain v-sections and therefore it is also unlikely that a second appliance will switch on during a v-section time segment. However, the event detector may have to unravel a set of v-sequences which belong to more than one particular load.

The analysis also required significant training, although most of this may be done offline as the transient represents a type of load and not a particular appliance. Furthermore, the same appliance may exhibit a different transient dependent on where in the voltage cycle the switch on occurs.

Transient monitoring also requires a more sophisticated monitoring system than steady state analysis as it needs to cope with the higher frequencies observed during transient behaviour, as well as greater memory capacity to cope with the additional information

requiring storage. This type of system may well be useful in small and medium enterprises (SMEs) where variable loads cause problems for the steady state detection system, however, it is likely to be more sophisticated and costly than necessary for the home environment. Therefore, the transient option is not considered further in this research, although it is acknowledged that it may be useful in more sophisticated systems.

## 3.5 Limitations of the existing state of the art

This section considers the limitations of Non-intrusive Appliance Load Monitoring systems in more detail so that a possible mechanism which overcomes these limitations may be uncovered.

The main limitations of the NALM system include:

1. Inability to detect appliances which have power consumptions which ramp up and down.

2. Errors if two appliances are too close to one another in feature space. This will occur because there is some amount of variance in the power consumption of the appliances and this may cause the features of two appliances to overlap and create problems of classification.

3. Errors if more than one appliance switches on between subsequent observations of the system load. This is particularly prevalent when the power of the two appliances matches the power of another single appliance.

4. Lack of classification framework - this problem has been tackled in Durling's research for the particular information present in that system, however, this needs to be further explored for the type of information present in steady state appliance load monitoring.

The problem in detecting appliances with power consumptions which ramp up and down

is inherent to the method of analysis used in the NALM system. This could potentially be overcome by considering the entire appliance behaviour as a slow transient and using transient analysis to detect these events. However, for the purposes of this thesis, it will be assumed that appliances which exhibit this type of behaviour shall not be detected.

The next two limitations can be overcome by considering the source of the errors. Limitation 2 is caused by a lack of features chosen to represent appliances. Figure 3.4 in Section 3.3.3 illustrated the large amount of available information in the monitored current. Therefore, by extracting more useful features, this problem of misclassification can be overcome.

Limitation 3 is due to more than one appliance changing state between subsequent inspections of the system load. This may be due in part to the NALM system monitoring the load at a slow rate of 1 Hz, but may also be due to slow transients of appliances turning on - the appliance load is only considered once it is in steady state. A simple solution is to increase the rate at which the system is monitored, although this will not overcome the problem of slow transients. If such a problem occurs, the NALM system attempts to identify the two separate instances by considering the possible features when two events are combined. In some cases this may be successful, but the analysis falls over when the summation of the two feature sets appears to belong to a third appliance. By choosing a more useful and larger set of features (recall that NALM uses just the real and reactive power) it is possible to overcome this problem by reducing the likelihood of the summation of the two feature sets matching a third appliance.

The final limitation with the NALM system is the lack of classification framework which enables identification of the underlying appliances from the observed features. This problem is overcome by identifying the desired capabilities of the framework with reference to the available information.

Therefore, a system is proposed in the following chapter which enables appliance identi-

fication using non-intrusive methods, satisfying the identified system layout and design criteria. Furthermore, this system will overcome some of the limitations of the NALM analysis by extracting a larger and more useful set of features and utilising a classification framework to enable the features to be used in a sensible manner so that the underlying appliance may be identified from the features observed.

## 3.6 Chapter summary

This chapter set out to provide a discussion into the existing state of the art. This was accomplished by first considering the implications of the design criteria in order that the discussion was limited to those areas of research relevant to the work presented in this thesis. This resulted in a proposed system designed to track individual appliance energy consumption using simple hardware which monitors the energy use of the home from a single point. Complex software is needed to subsequently identify the individual appliances within this total load.

Having determined the system design, the domestic model is then considered so that possible mechanisms of tracking individual appliance energy use may be identified. From the three methods explored, it was proposed to identify appliances based on features observed in the current and voltage of an appliance when it is in use.

Having identified the area of the research in this thesis, the relevant state of the art was then explored. This discussion was constrained to three proposed systems of analysis. The first considered the steady state harmonic content of the system and was discarded due to the high demands of training required for this system. The other two considered step changes in the power signal which are assumed to correspond to a single appliance changing state. Two types of analysis were then explored - the first considered the steady state load profile of that individual appliance, the second monitored transient behaviour due present at switch on/off.

The limitations of the existing state of the art were then explored and used in order to propose a mechanism of analysis for the AEM which allows most of the limitations to be overcome in a manner which does not invalidate the design criteria. This proposition forms the basis for the work presented in the remainder of this thesis.

# Chapter 4

# The Domestic Environment - A complex system and the need for a suitable framework

This chapter aims to identify a set of features, based on non-intrusive appliance load monitoring, which form an appliance signature to represent appliances so that the problems of the existing state of the art are overcome. Furthermore, an innovative framework for classification is proposed which allows the appliance signature to be used in a sensible manner in order to identify appliances changing state within the total load.

Since the proposed system monitors the total home energy use and identified features from the non-intrusive signals observed, it satisfies the design criteria established in Chapter 2. The discussion in Chapter 3 identified the limitations of the existing systems which employ this type of analysis, which are mainly due to the lack of information extracted from the data. Therefore, Section 4.1 considers the set of potential data which may form the appliance signature.

Further limitations in the existing state of the art relate to the identification of the underlying appliance once the features have been extracted. There seems to be no clear framework within which this classification takes place, and therefore no protocol developed for the mechanism by which an identified feature set is assigned to a particular appliance. In the work presented in this thesis there is a strong need for this framework

due to the extended appliance signature developed. Therefore Section 4.2 considers the features contained in the signature and the desired capability of the framework.

Initially a Bayesian belief network is proposed to enable appliance identification to use of all the feature sets contained in the appliance signature, however, this network does not satisfy all of the desired framework capabilities. Consequently, a novel framework which generates membership values which denote the amount that a particular set of features belongs to particular appliances (i.e. the level of belongingness) is proposed in Section 4.2.3. The memberships for each set of features are considered independently, and these are combined in a novel manner in order to allow all the identified feature sets to play a role in enabling appliance identification within the domestic environment.

## 4.1 Appliance signatures

It was identified in Chapter 3 that the limitations of the NALM system were due in part to the lack of information extracted from the monitored current and voltage. This analysis considered only the real and reactive power of the appliances, however, as observed in Figure 3.4 in the previous chapter, the shape of the current and power drawn is able to provide a much greater amount of information.

Therefore, it is possible to extract features from the current which contains information not only about the power drawn by the appliance, but also about the shape of the waveform, which may provide information about the electrical interface of the appliance. These features may include information as to the linearity of the waveform, or the presence of harmonics or peaks in the signal. These features are drawn from the monitored 50Hz current and voltage signals and form the short term time domain feature set. These features, which are only a subset of the appliance signature, are explored in greater detail in Chapter 6.

A further aspect of the appliance signature is due to the finite state machine (FSM)

behaviour exhibited by appliances. Any appliance which has step changes in power consumptions and therefore a discretely defined number of states can be modelled as an FSM. For a large number of appliances this may be a 2 state model, with just on and off modes, however, other appliances may cycle through a specified number of modes, for example a washing machine. These state diagrams are illustrated in Figure 4.1.

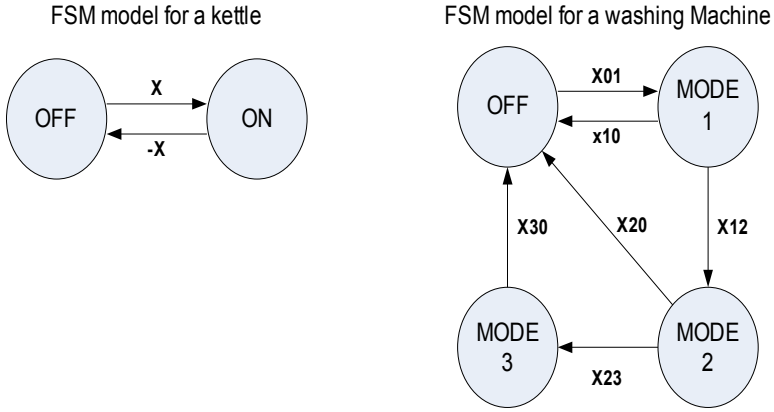


Figure 4.1: FSM models for a kettle and a washing machine.

When the appliance exists in a particular state, there will be a set of possible state transitions, and each of these will have an associated probability of occuring. This probability is dependent solely on the current state that the appliance is in. Furthermore, each state in itself is unobservable, rather there are a set of features which are observed when the appliance is in a particular state. These features relate to the short term time domain features noted when the appliance changes state. This type of system is similar to a Hidden Markov Model, and therefore a short discussion into this type of model is provided in Appendix B.

One final point to note about the FSM models is their dependency on time. Whilst this is not true of HMM behaviour in general, the probabilities of a particular state transition occurring will be dependent on the time passed since the observation of the current state. This is more prevalent for some appliances than for others as the time between state transitions determines the time of use of the appliance within a particular state. Some appliances have a large range of time over which they may be operational,

for example a TV. In this case the user has total control over the time of use of the appliance, which may range from less than a minute up to a number of hours. Other appliances, such as a kettle, are more restricted in their usage times. In this case there is only a small amount of user influence (changing the amount or initial temperate of water to heat) and the usage time is governed by the appliance. By considering the average usage times and the typical variance for a set of appliances, the FSM models can be developed to incorporate this feature as a function which determines the probability of observing a particular state transition given the previous state. The average usage times are depicted for some typical household appliances in Table 4.1.

| Appliance | Average Duration of Use in Minutes |
|---|---|
| Refrigerator | 10 |
| Hob (electric) | 21 |
| Oven (electric) | 43 |
| Kettle | 2 |
| Microwave | 6 |
| Dishwasher | 75 |
| Washing Machine | 60 |
| TV | 30 |

Table 4.1: Average duration of use in minutes of typical domestic appliances. Data taken from [39]

Furthermore, the probabilities of observing an appliance turn on when it is previously seen to switch off may be affected for some appliances by the length of time for which the appliance is off. Whilst this will add little extra information in the case of a kettle, consider the operating conditions of a refrigeration device, or an electric heater. These appliances will be on for a long period of time as far as the user is concerned, but if they are working correctly they will be switching on and off rather frequently at appliance defined times. This adds to the appliance signature and may aid the classification procedure.

The final aspect of the appliance signature takes into account the time of day in which the appliance is used. Refrigeration devices are equally likely to be on at any time of the day, but all other appliances have typical load curves which vary with time. In a study considering typical loads of various types of appliances [39] it is shown that cooking

devices are generally used between midday and 8pm, with the largest use between about 4pm and 7pm, whilst wet appliances have their largest usage between 6am till 11am, with decreasing load from 11am until midnight. Lighting is mainly used between 6pm and midnight, and between 7am and 10am, with smaller consumption during the rest of the day. Therefore, by noting the time of day during which the appliance is in use adds a further feature to the appliance signature.

This section has demonstrated the large amount of information which forms the novel appliance signature and which is available for use in matching an observed power change to the appliance responsible for the change. This information is present in 3 different feature domains, where each domain contains a different type of information with a different analysis procedure. The first domain relates to the short term features observed in the current when an appliance changes state. The method by which this is analysed is discussed in Chapter 5. The second domain is that information which relates to FSM behaviour. It is analysed in a HMM type network, which allows the observations of particular states to determine the time dependent probabilities of future observations. The final domain relates to the time of day during which the appliance is observed to be in use. This domain of information is user dependent and will vary between households. The details which govern the way in which this domain is analysed is not considered further in this thesis.

Since the short term domain contains the most rigorous information, this thesis concentrates on identifying useful features within this domain. This work is presented in Chapter 6. However, when developing the classification mechanism for the AEM it is important to identify the effects of the multiple feature domains in the appliance signature on the methods used in classification and the subsequent results. This enables the development of a system which has the ability to cope with the full set of information available.

## 4.2 A framework for classification

In Section 3.4.3 it was identified that one of the limitations of the NALM system was the lack of development of a classification system which could be used to aid identification from appliance signatures. This aspect of a non-intrusive monitoring mechanism is particularly important for the AEM, in which information is present in multiple domains, all of which are useful in appliance identification. Therefore, this section explores the classification mechanisms which are used in the AEM in order to identify an appropriate framework to handle the information present.

The mechanism by which energy tracking of individual appliances takes place is as follows. The current and power consumption of the home are monitored. The power is normalised by the voltage as in the NALM analysis [30] in order to allow for changes in the voltage waveform. When the power is steady, there are no state changes occurring within the system. Any appliances which are known to be on remain on, and those which are off remain off.

When a significant change is observed in the normalised power waveform, it is assumed to be caused by an appliance changing states. By tracking just the changes that occur on the system a single appliance is isolated from the total load. In order to simplify the complex domestic system, the following assumptions are made:

- Only a single appliance changes state between two inspections of the load.
- When an item changes state a step change in power consumption occurs. This ignores the problem of identifying appliances with ramping power functions which is not dealt with in this thesis.
- If two identical appliances exist within the home, they shall be classified as the same appliance.

Once a step change in the normalized power consumed by the home is observed and the

power returns to steady state behaviour, the difference in the current and power signals before and after the change is calculated. This allows the steady state behaviour of the appliance which has just switched on to be captured. These differential signals are denoted $\Delta i\,(t)$ for the current and $\Delta p\,(t)$ for the power, and a set of short term domain features denoted $\mathbf{x}$ are drawn from these signals. Analysis of $\mathbf{x}$ allows the appliance electrical interface to be deduced, which goes some way to identifying the underlying appliance.

This feature vector will also allow the FSM model to deduce the state transition which has just occurred (for a particular appliance). In some cases the feature vector may be indicative of a number of possible state transitions. This allows the time dependent probability distribution of observing the subsequent state transition to be determined. Once this subsequent transition occurs, the probability assumes a particular value rather than a distribution.

Having deduced the transitions which are assumed to be occurring, the probability of observing that particular appliance during the observed time of day can be identified.

All this information is available for use in order to identify the appliance corresponding to the observed step changes in power. This overcomes the problems in the NALM analysis of misclassification due to overlapping clusters in power space, whilst still satisfying the constraints imposed by the design criteria. However, the existence of multiple feature domains requires a formal framework in which classification is performed in order to allow the entire feature set to have an input into the appliance identification.

**The available information**

- Set of features identified from the 50Hz current and voltage waveforms.
- FSM behaviour consideration including duration of use information.
- Time of use (during the day).

**The desired capability of the framework**

- Each feature domain must have some input into the final classification result.
- More rigorous domains should have a greater influence on the final result.
- The addition of a domain should strengthen the final result rather than weaken it.
- The observation of a particular transition within an FSM model should trigger the analysis to expect to observe the following mode at some defined time period after the first observation. For example, once the kettle has been seen to turn on, there should be an increased time dependent probability of observing a step change in power corresponding to it turning off.
- The output should indicate the identified appliance along with a term which depicts the strength of the classification.
- Any new items should be output as unidentified as the AEM has not yet been trained to learn their feature sets.

It should be recalled that for the purposes of this thesis it is assumed that the AEM will operate in a mode which enables a manual training capability. However, as described in the NALM analysis, it is possible to extend the capabilities of the AEM to enable an automated training regime on observations of step changes in the power waveform.

### 4.2.1 Bayesian belief network for classification

A Bayesian belief network (BBN) is a graphical model which allows the probability of obtaining a particular output (i.e. the probability of the observation belonging to a particular appliance) to be calculated by considering the probabilities of obtaining the observable inputs. By defining probabilities for each observation, probability theory is used in order to allow all the individual domain information to contribute to the resultant probability of the output. This allows the output (i.e. the individual appliance) to be

dependent on all 3 feature domains. There is much literature on this topic both in books [38] [40], in academic journals [41] and on the world wide web.

Each feature domain is considered in a separate node in the BBN. The observable inputs are considered for each domain at it's corresponding node. Each node takes any one of a discrete number of inputs, each with an associated probability. For the domestic system there will be three nodes. The first will relate to the short term time domain. The features identified in this domain relate to the underlying electrical interface of the appliance and this feature set therefore gives rise to the probability of obtaining a particular class of appliance (determined by similarity of electrical construction) rather than a particular appliance type (determined by similarity of operation). This distinction is important, because, as observed in Figure 3.4 in the previous chapter, appliances which are similar by operation may have very different operating conditions.

The second node considers the FSM behaviour. This node will have inputs relating to the deduced state of the appliance (determined by considering the feature set $\mathbf{x}$), the previous state that the appliance was known to be in, and the time lapse between the two observations. The probabilities generated therefore link together the two consecutive states of the appliance.

The third node accounts for the time of day during which the appliance was known to be in use, and therefore generates probabilities of the observation belonging to a particular appliance given the time of day of the observation.

The probability of observing a particular appliance is determined by the inputs into all 3 nodes of the BBN, and these dependencies are illustrated in Figure 4.2.

The probability of the observed data set belonging to a particular appliance is calculated by looking at $P(Appliance|A, B, C)$, where $A$ represents the time of day variable, $B$ represents the FSM behaviour and $C$ represents the class of device based on the electrical interface.

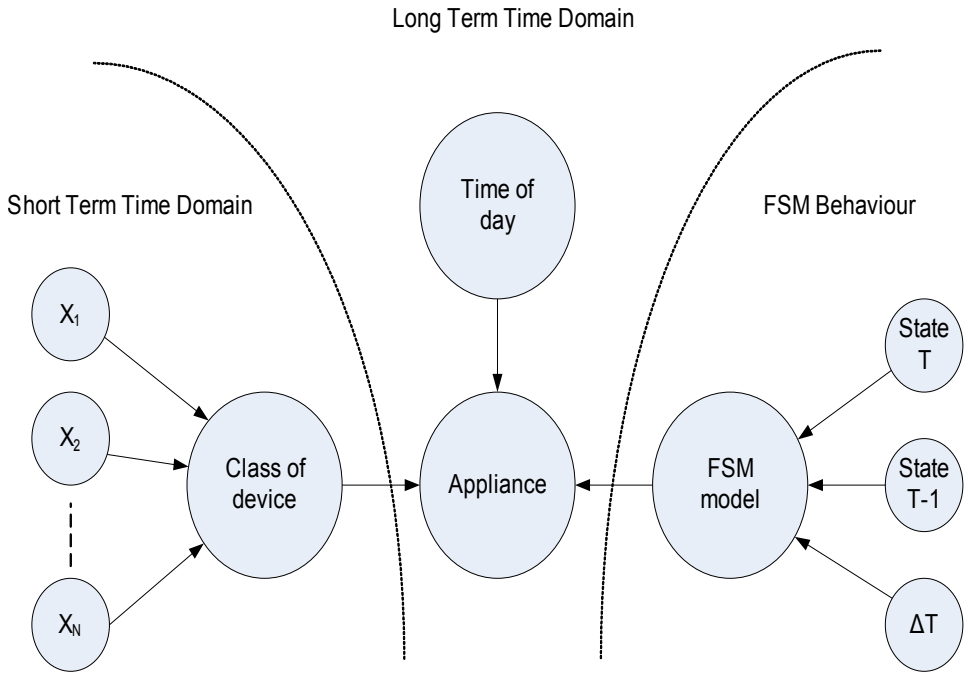


Figure 4.2: Bayesian Belief Network for Classification Algorithm.

To calculate the probability $P(Appliance|A, B, C)$, some amount of training data must be captured. Bayes theory is used to expand and manipulate this probability to decipher the necessary inputs and ensure that these values are noted in training. To simplify matters begin by considering just a single node. Let the appliance be represented by $W_i$.

$$P(W_i|A) = \frac{P(A|W_i)P(W_i)}{P(A)} \tag{4.1}$$

The probabilities on the right hand side of the equation can be calculated by looking at training data. Adding a level of complexity to the model by considering 2 nodes:

$$P(W_i|A, B) = \frac{P(W_i)P(A|W_i)P(B|W_i, A)}{P(A)P(B)} \tag{4.2}$$

As node $B$ is not dependent on node $A$ this can be simplified:

$$P(W_i|A, B) = \frac{P(W_i)P(A|W_i)P(B|W_i)}{P(A)P(B)} \tag{4.3}$$

and extended to more variables:

$$P(W_i|A, B, C) = \frac{P(W_i)P(A|W_i)P(B|W_i)P(C|W_i)}{P(A)P(B)P(C)} \tag{4.4}$$

Having established the information that must be collected, the nodes themselves are now considered in more detail to understand how this information is used. Recall that each node must exist in any one of a pre-defined discrete number of states.

**Node A** This node represents the time of day of use. Recall from Section 3.3.3 that different types of appliance (classed by similarity of use rather than construction) have different typical usage times, with some appliances more likely to be used at different times of the day than others [39].

By considering appliance usage patterns, along with a total daily usage curve for various types of households (working/non working, winter/summer etc.) the time of day is split in order to represent the typical usage patterns most accurately. For this reason it was decided to split the time of day into 5 time periods based on existing research in this area [39]. These are:

- Time Period 1: Midnight until 6am
- Time Period 2: 6am until 10am
- Time Period 3: 10am until 4pm
- Time Period 4: 4pm until 7pm
- Time Period 5: 7pm until midnight

Therefore, the probabilities $P(A)$ and $P(A|W_i)$ can be estimated by considering the proportion of the entire day that each time period represents, along with the typical usage curve of the various appliances with reference to the time periods.

**Node B** Node $B$ corresponds to the FSM behavior. The inputs into this node are features relating to the duration of use of the appliance, the current assumed state of the appliance, and the known or assumed previous state. Whilst the data relating to the state of the appliance already exists in a discrete number of possibilities, the data

relating to the duration of use needs a little more consideration.

As mentioned earlier, the length of time for which an appliance is observed will have some effect on the probability of occurrence of the observed transitions. For some appliances the time period is clearly defined by the appliance, whilst the time period for other appliances is more ambiguous due to user interaction.

The time periods are therefore defined by considering a set of training data. Training data is data that is gathered which relates to the particular observation in question, where all the information is known about the data. By making observations about the various lengths of times that different appliances are used for it is possible to determine natural time splits in the data. These might correspond to:

- Less than 10 minutes
- Greater than 10 minutes but less than 1 hour
- Greater than 1 hour but less than 5 hours
- Greater than 5 hours

However, actual data would be needed to justify these times. Once the data has been divided, the training data can be used to generate certain probabilities for the appliances. For example, the probability of observing a kettle switch off if the time of use is greater than 10 minutes should be low, though it should be much larger if the usage time is less than 10 minutes and more in the typical time frame of use. These dependencies can be accounted for by this node.

**Node C** Node $C$ is a little more complex. This represents the class of the appliance, where the grouping is based on the short term feature set data rather than the appliance type data. Appliances which have similar feature sets, such as a kettle and an electric heater, will belong to the same class $c_1$ in $C$, whilst belonging to different appliance type

groups. The features $X_1$ to $X_N$ (where $N$ is the total number of features) are generated by each appliance during use. By monitoring individual appliances, feature sets are evaluated (this is discussed in more detail in Chapter 6) and, by making various assumptions about the distribution of the features, the probability of observing a particular feature set can be calculated. Therefore, the probability generated is that of observing a particular feature set, $\mathbf{x}$, given the appliance class $c_i$. Mathematically this is written $P(\mathbf{x}|c_i)$. However, the data needed for the classification network is $P(c_i|\mathbf{x})$. This is the probability of actually having appliance class $c_i$ given that the observed feature set is $\mathbf{x}$. Bayes theory tells us that:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{\Sigma_j P(X|C_j)P(C_j)} \tag{4.5}$$

which allows training data to be used in order to determine the class of appliance to which the short term data belongs. Once this class has been identified it can be used in order to represent the dependency of the appliance on the electrical interface observed.

### 4.2.2 Limitations of the Bayesian network

A Bayesian belief network satisfies the following desired framework capabilities:

- Each feature domain must have some input into the final classification result.
- The observation of a particular transition within an FSM model (e.g. a washing machine) should trigger the analysis to expect to observe the following mode at some defined time period after the first observation.
- The output should indicate the identified appliance along with a term which depicts the strength of the classification.

It does not directly allow some domains to have a greater influence on the result (however this can be done by broadening the expected distributions of the less rigorous features) and the addition of more information may not strengthen the result. Also, because of the constraints imposed by probability theory, the sum of the probabilities of the feature

set belonging to one of the known appliances must be 1, i.e. $\sum_{\text{all } i} P(W_i|\text{featureset}) = 1$. If the feature set observed actually belongs to an unknown appliance, or is due to some transient or noise on the system then there will be a resultant misclassification because the network does not allow the option of an unidentified or noise group.

A further limitation of the Bayesian network (BBN) is due to the way in which the information present in the appliance signature is processed. In the Bayesian network the nodes must exist in one of a discrete number of possibilities. For example, the duration of use of an appliance exists for each appliance as a distribution function with mean and covariance as established during a training procedure. However, in the BBN this is reduced to 4 categories which may not entirely match the distribution.

Therefore, it is necessary to develop a system which matches the useful capabilities of the BBN, whilst allowing different domains to contribute to the classification result at differing strengths, with each subsequent domain adding strength to the classification, as well as allowing for the scenario in which the output appliance is unknown. Furthermore, the developed system should allow the information to remain in its original format rather than constricting it to discretised groupings.

### 4.2.3 A model for memberships

This section proposes a novel method by which the 3 feature domains in the appliance signature are combined in a framework which satisfies all the desired framework capabilities. Because the construction of the Bayesian network is appropriate to the domestic model, the Membership Model is based on the network proposed in Figure 4.2. The ensures that all 3 feature domains have an input into the final appliance classification, and allows FSM type behavior to be accounted for. However, if the framework is to overcome the limitations of the BBN, the data presented at each node and the combination of the data in each node must be dealt with in a different manner to that described in Section

4.2.1.

One of the major problems of the BBN is the inability of the system to allow for an unknown appliance. This is due to the constraints imposed by probability theory which states that the sum of the probability of the observed feature set in all appliance groups must be 1. By relaxing this constraint, and dealing with memberships instead of probabilities, this limitation can be overcome. Rather than each node generating a probability $P(W_i|A)$, a membership is generated which represents the degree to which the observed feature set belongs in each known appliance group.

The second major problem with the BBN is the representation of the information at each node. In the BBN the features are divided into a discrete number of groups at each node, for example, the duration of use node is split into 4 groups. However, for each appliance the duration of use is a continuous distribution function and inaccuracies are introduced when this data is sectioned into discrete groupings. Therefore it is important that during classification the data must be present at each node in it's original format and this raw data should be used to generate memberships.

The proposed classification method is illustrated in Figure 4.3. A single monitoring point collects data about the voltage and current. A step change in power indicates the presence of an occurrence (i.e. some appliance connected to the system changing states) and the current drawn by this individual appliance, $\Delta i\,(t)$, is isolated from the total load, $i\,(t)$. Features are extracted from these short term time domain signals and the mechanism by which this is done is discussed in Chapter 6. The memberships generated from these short term time domain features are discussed in detail in Chapter 5. Memberships are also generated for the FSM and Time of Day features, though the details of this are not dealt with in this thesis. Therefore the proposed classification system uses the identified feature set values to generate memberships at each node which represent the degree to which the observed feature belongs to each of the appliance classes.

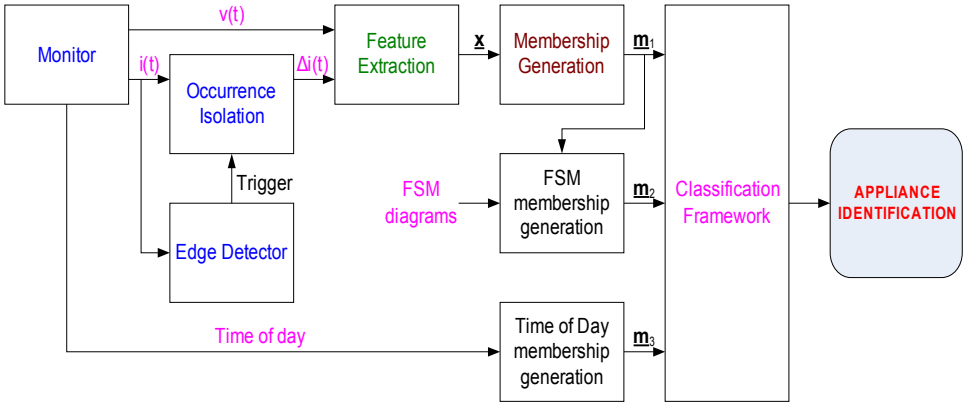


Figure 4.3: Illustration of the system design of the AEM. Parts in blue are explored in Chapter 3, pink corresponds to Chapter 4, dark red to Chapter 5 and green to Chapter 6. The FSM membership generation and Time of Day membership generation blocks are not developed in this thesis.

The remainder of this chapter is concerned with identifying a mechanism by which the Classification Framework block in Figure 4.3 combines the memberships in such a way that allows the different domains to have a differing level of input into the final classification and so that the addition of information does not weaken the result. Furthermore, the classification network must be able to cope with signatures from appliances which are not yet known.

The following sections explore two well known methods of combining classification results, before an adapted mechanism is proposed for use in the AEM. These two known methods include membership combination via the arithmetic mean and the geometric mean [42].

### Arithmetic mean

Let the individual membership functions be denoted by $m_i(j)$ for the $i^{th}$ feature set, where $i = 1, 2, \ldots, N$, in the $j^{th}$ appliance group. The resultant membership function is then given by:

$$m_{total}(j) = \frac{1}{N} \sum_{i=1}^{N} m_i(j) \tag{4.6}$$

The highest and lowest values obtained place boundary limits on the final value, so that the resultant membership cannot be greater than the highest individual membership,

or smaller than the lowest one. It is also possible to give different membership values different weightings in order to allow some of the values to have a greater effect on the overall membership than others.

However, it is desirable that the time of day and duration of use values only have a small effect on the resultant membership unless their values are close to zero. This is not possible using an arithmetic combination. Even if one of the membership functions is set to zero indicating an impossible scenario, the final membership value will not be zero unless all the other membership values are also.

### Geometric mean

Using the geometric mean the resultant membership function is generated by:

$$m_{total}(j) = \left( \prod_{i=1}^{N} m_i(j) \right)^{1/N} \tag{4.7}$$

As with the arithmetic mean, the highest and lowest values obtained for the individual membership functions place boundaries on the resultant membership function. Whilst a membership value close to zero in this scenario will bring the final membership value toward zero, indicating a better model than the arithmetic mean for this particular application, all membership functions have equal weighting. The following paragraph proposes a mechanism of combining the different membership functions so that some are allowed to play a stronger role in the final membership value than others.

### Varying the weightings in the geometric mean

It is possible to vary the weightings of the various elements in the geometric mean, so that the resultant membership is actually calculated according to:

$$m_{total}(j) = \left( \prod_{i=1}^{N} m_i^{a_i}(j) \right)^{1/\sum_{i=1}^{N} a_i} \tag{4.8}$$

where the value $a_i$ determines the weighting of the individual memberships.

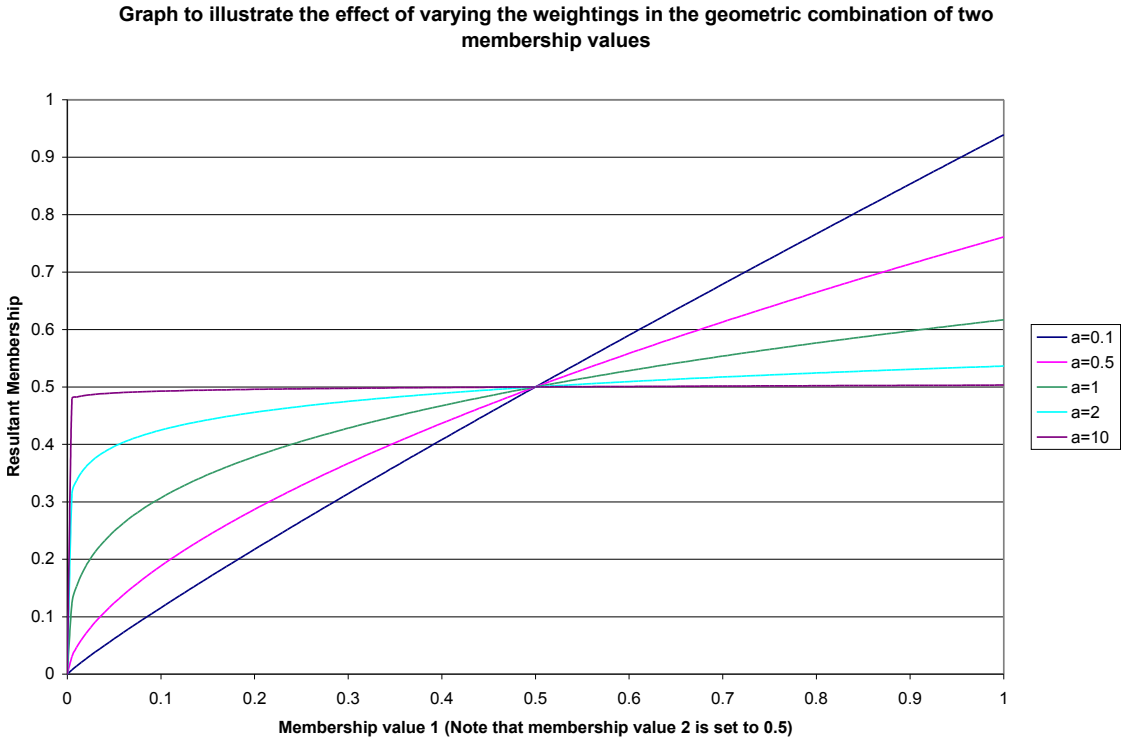


Figure 4.4: Graph showing the effect of changing the weighting of the individual membership functions.

Figure 4.4 illustrates the effect of varying the relative weightings of two membership functions. The first membership function is given a weighting of 1. This membership value is plotted along the x-axis. The second membership value is set at 0.5, and its weighting $a$ is varied according to the values in the legend. It can be seen that when the value $a$ is small in comparison to 1, the second membership value plays a much smaller role in determining the final membership then when its value is large in comparison. By changing the value of $a$ the different feature domains are given different weightings in the final classification.

The actual values chosen for the weightings are subject to further research and are not considered in this report. This is due to the limitation of the discussion presented in this this to those features in the short term domain only.

Once the memberships have been combined there will be an overall resultant membership of the total feature data in each appliance group $M(W_i|\mathbf{data})$. This can then be used

to classify the appliance. In an ideal situation there will be only one appliance with a significant membership value, however, this may not be the case. Consequently it may also be desirable to develop a confidence level for the classification based on the absolute and relative membership values. This confidence level indicates the strength of the classification.

### 4.2.4 Developing a confidence level

The objective of a confidence level should be to determine the strength of the classification. The first assumption is that a final classification is made. Assume that so long as the final membership value is above 0.2 (the reason that 0.2 is chosen as the critical value is explained in Section 5.3.7), then a classification can be made so that the appliance with the highest membership value is the classified appliance. The strength of this classification will depend not only on the final classification result (a higher membership value will result in a stronger classification), but also on the gap between the highest classification and the second highest [43]. The larger the gap between the highest membership value and the second highest, the stronger the classification. This section proposes a mechanism for generating a confidence level based on the final top two membership values generated in the classification process.

Following the classification process the information present is a set of membership values corresponding to the likelihood of the data set belonging to each of the 'known' appliances that the system has been trained to recognise. The classification is a good one if the membership value of the data in the 'chosen' appliance group is high, and if the membership values of the data set in all the other appliance groups is low. If either the gap between the 2 strongest classifications (i.e. the two largest membership values) becomes small, or the actual final membership value for the 'chosen' appliance becomes small, then the resultant confidence level should be small. This scenario implies a geometric combination. Let the final membership value be denoted $M_1$ for the highest membership,

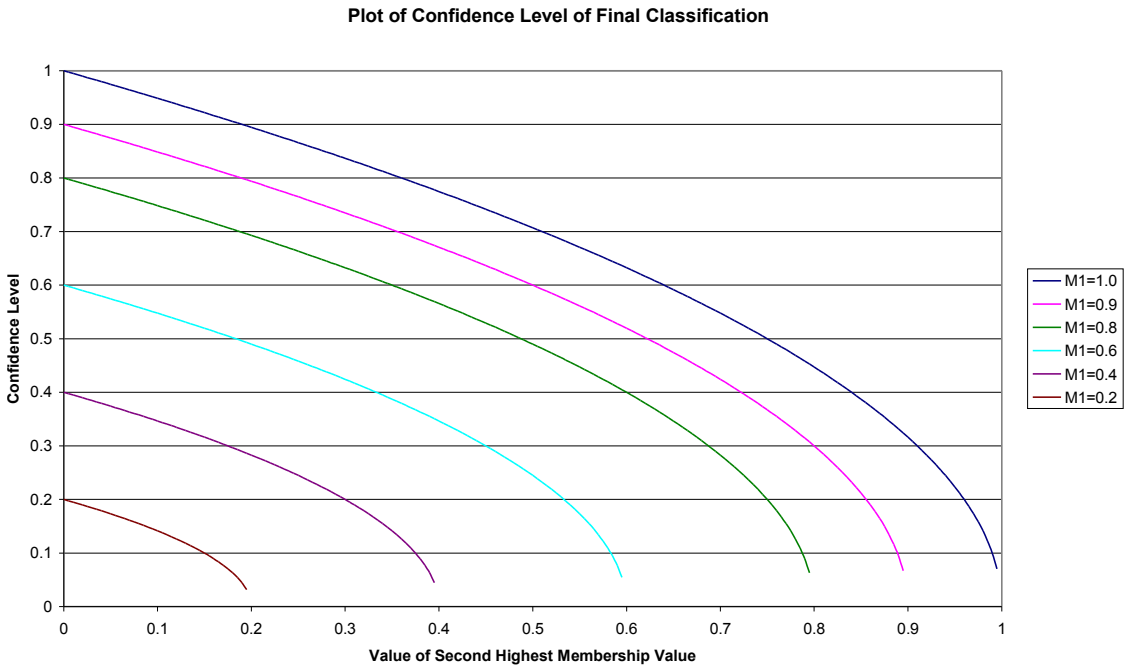


Figure 4.5: Confidence level based on membership values of the two highest final membership values in the classification network. The plot shows the final confidence level for a number of values of $M_1$ and all values of $M_2$ up to the value of $M_1$.

and $M_2$ for the second highest. The confidence level, $CL$ is defined such that:

$$CL = \sqrt{M_1 \left( M_1 - M_2 \right)} \tag{4.9}$$

The variation in confidence level with membership values $M_1$ and $M_2$ is shown in Figure 4.5. From this figure it can be seen that the value $M_1$ and $M_1 - M_2$ provide the boundary for the confidence level. The confidence in classification will be no higher than the highest resultant membership value, and it will be no lower than the difference between the two. Therefore, a strong classification results as the highest membership value approaches 1, and when the second highest approaches 0.

The classification system proposed in this section allows for all the various features to be taken into account, whether their memberships are based on probability theory or otherwise. They can be combined in such a way to give stronger weighting to certain features which may be deemed more reliable, so that less rigorous features only have the ability to rule out an appliance group, rather than significantly strengthen the result. It

also allows for the inclusion of the unknown if new appliances are added to the system, and an ability to determine the confidence strength of the final classification. All these facts make this system a very versatile one which satisfies the criteria imposed and which is therefore a good model for the classification framework of the AEM

## 4.3 Chapter summary

At the end of this chapter both a novel appliance signature and an innovative framework have been proposed in order to allow individual appliances to be accurately identified within the total load in a manner which satisfies the design criteria developed in Chapter 2.

The appliance signature was developed by considering the possible information available for use in characterising appliances. Three domains of interest were identified which correspond to the short term time domain information present in the 50 Hz current and voltage signals, the time dependent FSM behaviour of appliances, and the time of day during which the appliance is used.

Having established the type of information available, and the method by which it is extracted from the observed signals, a classification network was proposed in order to allow the different domains of information to all play a part in the identification of appliances from their signatures. The available information and the desired capabilities of the framework were determined and subsequently used to analyse the performance of the proposed solution.

Initially a Bayesian belief network was proposed to account for the dependency of the appliance on the multiple domains of information, however, this network did not quite match the available data or the desired capabilities. Consequently a novel framework was proposed in order to overcome these issues. The proposed framework allows each domain to generate a membership value which represents the degree of belonging of the observed

feature set within the appliance group. These membership values are then combined geometrically to enable the different domains to each contribute to the classification.

Finally, a confidence level was generated for the classification. This was determined by the strength of the classification (i.e. the resultant membership value) and the strength of the second highest membership value.

The work explored in this chapter has therefore identified the way in which an appliance signature is extracted from the observed data when a state change occurs, and subsequently used in a structured framework to track individual appliance energy consumption.

# Chapter 5

# Mechanisms For Classifying Short Term Data

This chapter is concerned with the Membership Generation block (MGB) in the AEM system design, as illustrated in Figure 5.1. The aim of this chapter is to establish a mechanism by which membership values are generated for the short term time domain feature vector $\mathbf{x}$, which represent the level to which that feature vector belongs to each of the known appliances. The membership value is therefore a vector of length equal to the number of known appliances.

This chapter begins by making assumptions about the spread of data generated by each appliance and assessing the objectives of the MGB in order to develop a set of criteria to which the developed membership generations methods must adhere. This is presented in Section 5.1.

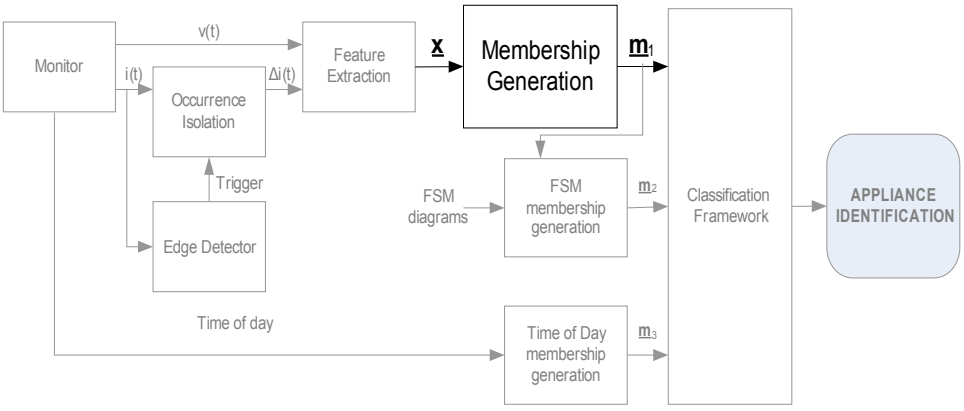


Figure 5.1: Membership Generation Block.

For the purposes of this thesis the AEM is based on the NALM system (see Chapter 3) that chooses to incorporate a manual training procedure and Section 5.2 explores the method by which the AEM is trained to know a set of appliances. Each mode of operation of each appliance is represented by a set of parameters which characterise that appliance using as little information as possible.

The number of modes of operation for a particular appliance may be unknown and the first objective of the training procedure is to determine the number of modes (or number of clusters), M. This is presented in Section 5.2.1. Once the number of modes of operation has been determined, Section 5.2.2 explores the method by which the number of samples, N, which are needed to characterise each mode are determined. Each mode of each appliance is then represented by the mean, covariance and average intracluster distance of the N samples.

Section 5.2.4 applies the developed training mechanism to a set of data representing a fan heater operating in 3 different modes. The accuracy of the predicted cluster centres is calculated and the training algorithm is found to represent the modes of operation to an average accuracy of greater than 98%.

Once the AEM has been trained to know a set of appliances, the method of generating memberships for a new set of feature vectors is considered in Section 5.3. The memberships are based on the proximity and relative proximity of the new vectors to the set of existing clusters.

The proximity measure used in this thesis is defined in Section 5.3.1. The Mahalanobis proximity measure is used which allows the distance between the unlabelled vector and existing clusters to be influenced by the spread of data within each cluster and therefore accounts for the varying size of the clusters.

Because the membership value determines the degree to which a data point belongs in a set of clusters, the membership values must exist along a graded scale from 0 to 1, where

0 represents no level of belonging to a cluster and 1 represents complete belonging to a cluster. This scale of memberships implies the use of fuzzy clustering, which is introduced in Section 5.3.2.

Probabilistic fuzzy clustering and possibilistic fuzzy clustering membership generation schemes are considered in Sections 5.3.3 and 5.3.4 respectively and the benefits and problems of each method are discussed. Sections 5.3.5 and 5.3.6 explore existing methods which attempt to alleviate the problems of probabilistic and possibilistic fuzzy clustering by generating memberships and cluster centres which tie together the two individual schemes. These existing methods are, however, not appropriate for the type of information considered in this chapter and a consideration of their limitations leads to the proposition of a novel scheme for membership generation, presented in Section 5.3.7.

The novel method of membership generation is applied to a real set of data in Section 5.3.8 and analysis of the results shows promising performance, with a total of 99.5% of the new set feature vectors having significant memberships (i.e. greater than 0.2) in the clusters to which they belong and negligible memberships in the clusters to which they do not belong.

## 5.1 MGB citeria

This section aims to develop a set of criteria to which the MGB must adhere and which can be used as a benchmark for identifying and assessing any mechanisms proposed. These criteria are developed with consideration of the type of information presented to the block and the objective of the membership function.

The vector $\mathbf{x}$ which forms an input to the Membership Generation block is representative of the short term time domain signals, $v\left(t\right)$ and $\Delta i\left(t\right)$. The way in which the features are generated from these signals is considered in Chapter 6, however, it is important to acknowledge the dependency of this feature vector on the electrical interface of the

appliance from which it is drawn. Therefore, the following assumptions are made:

- Each appliance will generate feature vectors which form compact and disjointed clusters in feature space.
- Each individual mode of operation of appliances with more than one ON state will generate feature vectors which form compact and disjointed clusters in feature space.

It should be recalled from Chapter 3 that the NALM system, upon which the AEM is based, has two modes of operation. The first incorporates a manual training procedure into the initialization of the device, whilst the second uses a clustering procedure to enable the identification of likely FMS models of unnamed appliances before deducing the corresponding appliance labels. This thesis is concerned with identifying a mechanism for membership generation in the case that the system undergoes an offline manual training period, with the extension to automated training being beyond the scope of this research.

Therefore, the MGB has two objectives. The first is to enable a training phase during which accurate representations are generated for known appliances. These representations must allow each mode of operation of each appliance to be represented separately and must also allow all possible operating conditions to be accounted for. (For example, during use, the resistance of an appliance may increase due to the heating effect, causing the current drawn by the appliance to vary. The representation should ensure that the entire spectrum of possible resistance values are accounted for.) Furthermore, due to the hardware constraints of the AEM, the representations should aim to use as little memory as possible.

The second objective is to enable the generation of membership values for unlabelled data points. These membership values indicate the degree to which the unlabelled feature vector belongs to each of the known appliances and therefore the membership value should be based on the proximity of the unlabelled data point to the appliance representation.

It should also take into account the spread of the feature vectors belonging to each appliance as some modes of use may have larger ranges of operating points than others. This spread may be measured during training.

The membership value should also take into account the assumption that only a single appliance changes state between subsequent observations of the system. This can be done by considering the relative proximity of the unlabelled feature vector to all the existing appliance representations. Finally, the membership value must allow for the option of the unlabelled data point belonging to a new appliance to which the system has not been trained.

Therefore the following criteria can be identified:

- For the training procedure
    1. Allow every mode of every appliance to be represented as a separate group.
    2. Represent each group with a few parameters as possible.
    3. Ensure that each representation encompasses all possible operating points.
    4. Ensure that the representation accounts for the spread of the data.
- For the membership generation procedure
    1. Membership values must exist as a value between 0 and 1, with 0 representing no degree of belonging to a cluster and 1 representing complete belonging.
    2. Membership values should account for the proximity of the unlabelled data point to the appliance representations.
    3. The proximity measure should take into account the spread of the data in the appliance representation.
    4. The membership value should take into consideration the relative proximity of the data point to the appliance representations.

5. It must be possible to allow all membership values to be small in the case that the unlabelled data point belongs to none of the known appliance representations.

These criteria are used to guide and assess the mechanisms proposed to first enable appliance representations to be generated during the training phase and subsequently to generate membership values for unlabelled feature vectors.

## 5.2 The training phase

This section explores the mechanism by which appliance representations may be generated. This process is done for each appliance individually and the first step is to determine the number of modes of operation, $M$, exhibited by each appliance. It is assumed that the appliance data within each mode forms a compact and disjointed cluster.

Once the number of modes has been established, the training procedure assigns feature vectors, generated during determination of $M$, to the clusters in order to generate cluster representations. The feature vectors are assigned under a nearest-neighbour scheme (a nearest-neighbour mechanism is fast and has small memory requirements) until a stopping criterion is met. This stopping criterion is used in order to prevent a cluster from becoming over-represented at a particular operating point, due to the presence of slow transient behaviour in the data.

Once there are sufficient feature vectors representing each cluster, the final step of the training is to deduce a cluster prototype so that the amount of memory required to represent each cluster is minimal.

### 5.2.1 Determining the value of M

The value $M$ represents the number of modes of operation of a particular appliance. It is likely that each mode of operation will draw a different current and because the short

term features are based on the current, it is fair to assume that each mode of operation will generate different feature vectors and as such each mode of operation should be represented separately. Furthermore, the modes of operation may not be controlled by the user and therefore the data generated by an appliance and the corresponding feature vectors extracted, may not be separated into their individual modes.

The number of clusters within a given data set may be determined using a a hierarchical clustering algorithm. This algorithm produces a nested set of clusters and by considering the error present when the data is split into a given number of clusters, the natural number of clusters is determined. This procedure uses a standard set of techniques and is presented in Appendix C.

Whilst the techniques for determining the number of clusters may be simple, the difficult part is in ensuring that all the modes of operation and represented in the data set which is input into the hierarchical clustering algorithm. For appliances with defined 'on' times, such as a washing machine, the training should take place over the entire time frame of operation in order to ensure that all modes are captured. Other appliances for which the duration of use is user controlled must be operational for at least their average usage time (as discussed previously in Chapter 4) to try and ensure that each mode is operational for a sufficient time duration.

For an appliance which is in use for 60 minutes, this results in a total of 180,000 feature vectors being generated (assuming that a feature vector is generated to represent every 50Hz cycle). Because of the memory and processing constraints in the AEM, this amount of data cannot be either stored or fed into the hierarchical clustering algorithm.

To overcome these processing limitations, whilst still ensuring all modes of operation are fully represented, a two-fold solution is proposed. The first part of the solution reduces the amount of data by a factor of 10 by only storing every 10th feature vector generated. This is valid because it is unlikely that any substantial changes will occur

over a period of 10 cycles. The second part of the solution is to section the feature vectors into smaller groups as they are generated and determine the number of clusters present and corresponding cluster centre of each group. The set of initial data may then be discarded, freeing up space in memory. Once the entire training data set has been processed in this way, the cluster centres generated may then be fed into the hierarchical clustering algorithm one last time to determine the actual number of modes of operation of the appliance. This is summarised as follows for groups of size $n_h$.

- Switch the appliance on and begin generating feature vectors. Throughout this procedure store every 10th to memory space 1.
- Once $n_h$ feature vectors have been generated, copy them into the program memory and begin to collect another $n_h$ vectors.
- Use the data set in the program memory in a hierarchical clustering algorithm to calculate the number of clusters present in the data and the corresponding cluster centers. Store this information in memory space 2.
- Repeat steps 2) and 3) until data has been captured over the entire time frame.
- Form a further data set from the cluster centres which have been determined during training so far. Use this new data set to determine the true number of clusters in the data (and corresponding cluster centers) using a single linkage hierarchical clustering algorithm (see Appendix C) to account for any underlying trend including the heating effect of resistors.

It is possible to establish a value for $n_h$ by considering the limitations of the processing capacity and the number of feature vectors generated. It is found in practice that a maximum of 500 feature vectors should be input to the hierarchical clustering algorithm and therefore a maximum value of 500 is placed on $n_h$. However, if an appliance is only on for 5 minutes then only 1,500 feature vectors will be generated and using $n_h = 500$ will result in only 3 groups. If each group contains only 1 cluster, it is not then possible

to use the hierarchical clustering algorithm as some of the assumptions in the analysis become invalid, so when the number of samples is small it is more appropriate to have a value of $n_h = 100$. Therefore, the value for $n_h$ is assigned in the following manner:

$$\begin{aligned} n > 4000 \qquad & n_h = 500 \\ 2000 < n < 4000 \quad & n_h = 200 \\ n < 2000 \qquad & n_h = 100 \end{aligned} \tag{5.1}$$

This method ensures that all modes of operation are accounted for, even if they only represent a small proportion of the total usage time. It also reduces the program time complexity and memory requirements.

### 5.2.2 Splitting the data set into M clusters

This section considers how a data set of feature vectors may be split into M clusters in order to use the data set to generate cluster representations. It is important that these representations account not only for the mean of the data, but also for the variability[1] of the feature vectors within each cluster.

Accounting for the variability within each cluster does ensure that the entire operating range of each appliance is considered, however, this goes hand in hand with a loss of information about specific trends exhibited by appliances when in use. For example, the same variance representation may be generated by an appliance which draws power in any of the 3 ways illustrated in Figure 5.2.

For the purposes of this thesis any underlying steady state trends shall be ignored and it shall be assumed that the loss of information is unimportant in the ability of the AEM to classify appliances. Therefore, a set of feature vectors are required to represent each appliance in order to ensure that the training regime incorporates all operating points.

[1]This variability is due to some changes in the parameters of the electrical interface of the appliance, which may be caused by automatic effects such as the heating effect, causing resistance values to change, or manual effects, such as load variation of a motor.

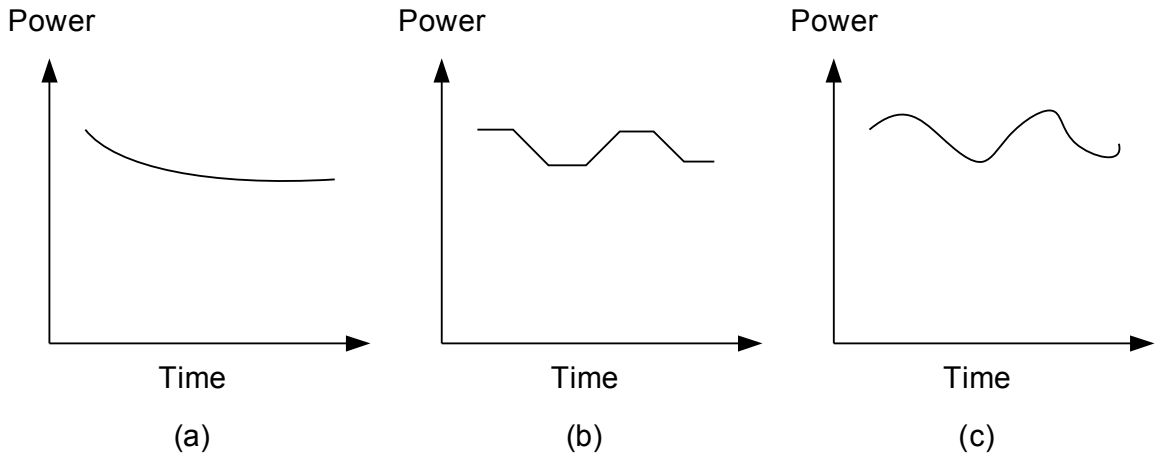


Figure 5.2: Plot of possible power variations with time.

Having determined the number of clusters present in the data and the corresponding cluster centers, the training procedure continues by assigning feature vectors generated and stored in memory during training to each of the clusters.

The feature vectors are assigned to each cluster based on a nearest neighbour (NN) clustering procedure [44]. The reason for using a NN clustering rule is due to the simplicity and minimal time complexity of the process. It can also be done one vector at a time, rather than using a pre-determined amount of data and feature vectors are assigned with complete certainty to a single cluster. This makes the NN scheme suitable for the training procedure of the AEM.

The NN clustering procedure assigns a feature vector to the cluster for which $\|\mathbf{x} - \mu_i\|$ is smallest for $1 < i < M$. This assumes that all the clusters are accounted for (which is known to be true according to the way that $M$ is calculated) and that they have similar and hyper-spherical spreads of data. Clustering is continued until a stopping criterion is met and the $j^{th}$ cluster contains $n_j$ data points.

The stopping criterion must ensure that all operating conditions (and therefore the entire spread of data) are represented within each cluster. It is also important that the clusters are not over-represented, which may occur if the appliance exhibits an underlying trend as illustrated in Figure 5.2 (a). As the number of data points used to represent the

cluster are increased beyond a certain limit, the variance of the cluster reduces and data points that lie furthest from the mean will not appear to be representative of that cluster. This may be problematic if the AEM is presented with a feature vector to identify which comes from the start of use of that appliance, at which point the feature vector obtained will lie the furthest from the mean. Therefore, the remainder of this section explores a mechanism by which the critical number of data points required to represent a cluster may be established.

The only information which is available about the entire population of individual clusters is the mean, as this value is returned from the hierarchical clustering algorithm. Therefore, one method in determining the point at which the number of data points in the cluster representation is truly representative of that cluster (and consequently no more data point should be used as this will result in over-representation) is to observe changes in the cluster mean as more data points are added. This is done using Student's t-test [45] to determine whether any observed changes are deemed significant.

The t-test is used to determine whether the mean of the first $n$ samples of the data is statistically different from the mean determined during hierarchical clustering which is assumed to be accurate given the large amount of data used to determine this value. If these two values are statistically the same then $n$ samples is adequate to represent that mode of operation, whilst a difference means that the $n$ samples are not adequate and further samples are necessary.

The t-test is performed by comparing a calculated value, $t$, to some critical value. The critical value is dependent on the number of samples tested, $n$, and the percentage level, $P$, at which the test is performed and may be found in look-up tables [46]. The $t$ value is determined according to the following equation:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \tag{5.2}$$

where $\bar{x}$ is the mean of the data, $\mu_0$ is the cluster centre as determined by hierarchical

clustering, $n$ is the number of samples and $s^2$ is the sample variance.

The null hypothesis is that the 2 values are equal, which can be rejected at the $P$ percent level if $|t| > t_\nu(P)$ where $t_\nu(P)$ is the critical value for $\nu$ degrees of freedom ($\nu = n - 1$). Therefore, the stopping criterion for the NN clustering scheme is when the null hypothesis is no longer rejected.

The process by which clusters are formed during training is illustrated in Figure 5.3. To reduce the process time, feature vectors are considered in sets of 10 rather than individually and it is assumed that the feature vectors have already been separated by cluster so that all 10 belong to the same cluster. Once the number of data points in a cluster increases by 10, hypothesis testing, as outlined above, is used to determine whether or not the addition of the new feature vectors results in significant difference between the cluster mean, produced from hierarchical clustering and the mean of the data points in that cluster. Under the null hypothesis the two values are not statistically different. If this is found to be true, the stopping criterion is met and no further data point need be added to the cluster. If the values are different then the process is repeated with further sets of feature vectors until the stopping criterion is met.

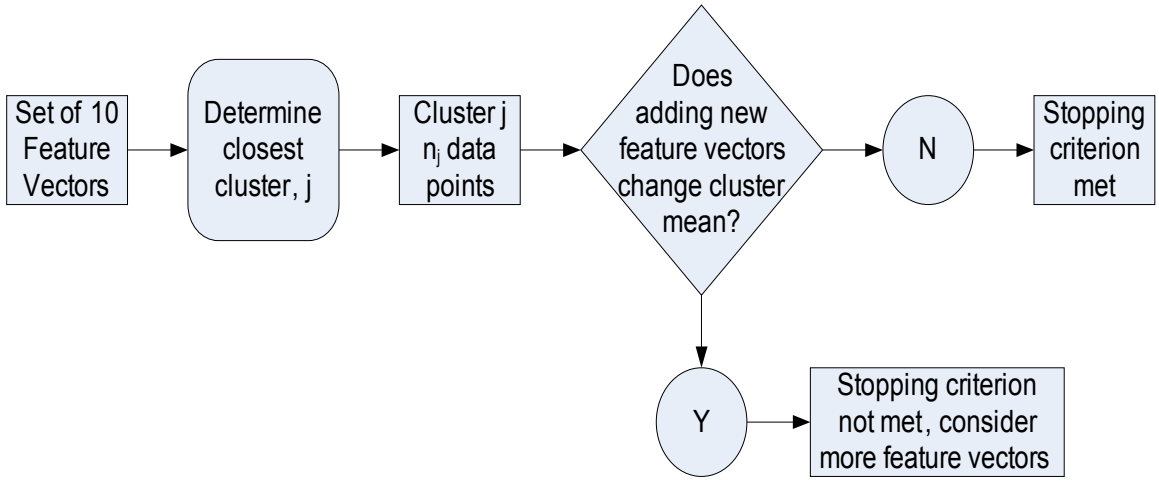


Figure 5.3: Method for generating $M$ clusters which may be used to represent all modes of operation of an appliance.

### 5.2.3 Generating cluster prototypes

The objective of a cluster prototype is to describe the properties of individual clusters using as few parameters as possible. The prototype must also allow membership values to be generated for future unlabelled feature vectors based on the proximity of the feature vector to the cluster, using a proximity measure which accounts for the spread of data within the cluster.

For these reasons the cluster prototypes shall consist of the mean feature vector, $\mu$, the covariance matrix ,$\mathbf{Q}$ [47], and the average intracluster distance $\eta$ [48]. For a cluster containing $n$ feature vectors, these values are calculated according to the following equations:

$$\mu = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k \tag{5.3}$$

$$q_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_{ik}-\bar{x}_i\right)\left(x_{jk}-\bar{x}_j\right) \tag{5.4}$$

where $q_{ij}$ is the element in $\mathbf{Q}$ which represents the covariance between features $i$ and $j$ and

$$\eta = \frac{1}{n}\sum_{k=1}^{n} d\left(x_k,\mu\right) \tag{5.5}$$

Where $d\left(x_k,\mu\right)$ is the Mahalanobis distance measure (see Section 5.3.1).

The proposed method of training allows every operational mode of each appliance to be represented individually, with no reference to other modes of operation. Each representation encompasses all possible operating points, due to the consideration of the changing mean of the data. The prototypes generated from the representations account for the spread of the data using the covariance and average intracluster distance and each prototype uses only 3 parameters (though the actual size of the cluster centre and covariance matrix depends on the number of features used, discussed further in Chapter 6). Therefore the proposed training regime for the AEM satisfies all 4 criteria established in Section 5.1.

### 5.2.4 Application of the training algorithm

The proposed training algorithm is applied to a set of training data which contains feature vectors corresponding to 3 different modes of operation of a fan heater. Due to the limitations of the data set size (caused by memory constraints in the laboratory equipment used) there are 1370 feature vectors generated in total. These are split into 13 groups each containing 100 feature vectors.

The first step is to determine $M$, the natural number of clusters in the data. This is done for each group individually and the cluster centres produced are input into a final hierarchical clustering algorithm. This allows the true number of clusters in the data and corresponding cluster centres to be determined.

Once the number of modes of operation and the mean of each mode is known, the feature vectors are assigned to the clusters according to a NN clustering rule. A t-test is performed to determine the critical number of feature vectors and this enables the cluster representations to be produced and from these the cluster prototypes are calculated.

**Determining $M$** The results for this part of the training regime are shown in Figure 5.4. The entire data set of 1370 is plotted in cyan. From this is can be clearly seen that 3 modes of operation exist in the data. It can also be seen that there are some outliers, most probably caused by transient behaviour resulting in misrepresentation. Therefore, the objective of this part of the training algorithm is to correctly identify the existence of 3 clusters and to correctly determine the cluster centers, without including the outliers.

The blue data points in Figure 5.4 represent the cluster centres established for each group of 100 data points. From this it can be seen that the outliers have been removed as the algorithm is designed to remove clusters which contain an insignificant number of data points (i.e. indicating the presence of a noise or transient data point).

The red data points in Figure 5.4 represent the final cluster centres determined when the

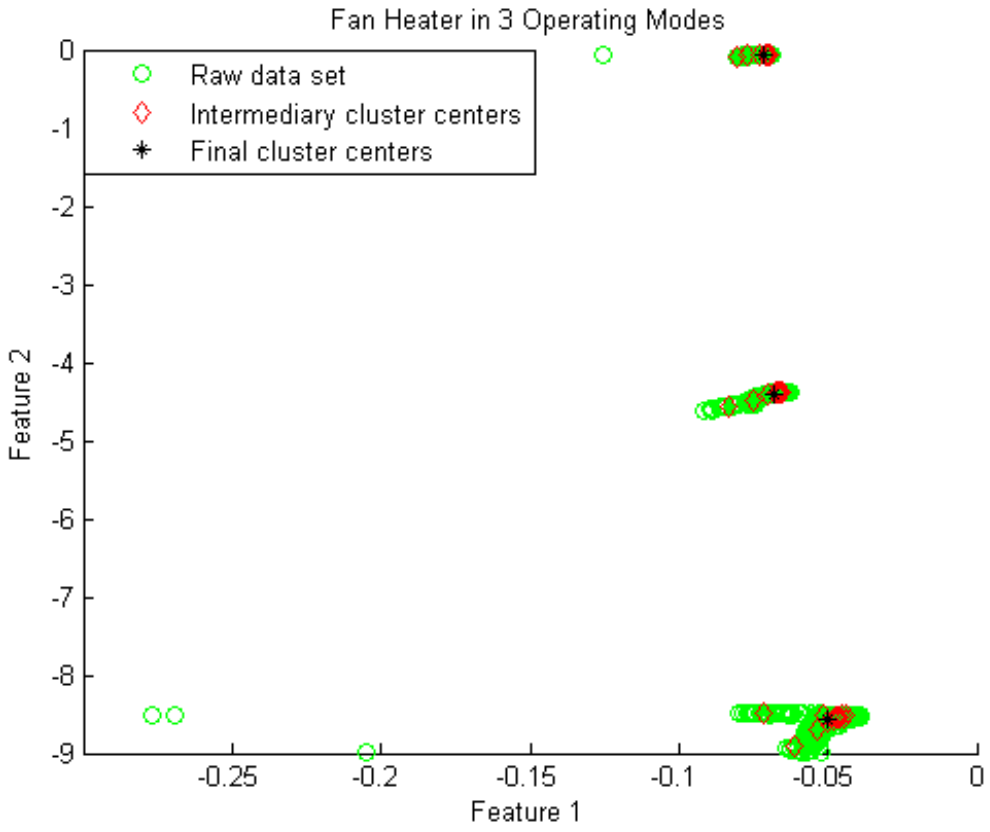


Figure 5.4: 2 dimensional representation the cluster data and cluster centres resulting from the training algorithm.

blue data points are fed into the hierarchical clustering algorithm. It can be seen that the data set has been correctly partitioned into 3 clusters. The error between the cluster centres calculated (as illustrated by the red data set in Figure 5.4) and the true cluster centres has a mean value across all dimensions of 1.17% and a maximum of 7.56%[2]. The cluster centres as calculated by the proposed method are therefore accurate to within 10% for each feature individually and on average are accurate to within 1.5%.

**Determining** $N$ This part of the training algorithm is concerned with determining the critical value of data points required to represent each cluster. For each cluster the cluster mean is known. 10 feature vectors are added to each cluster representation at a time and the mean of the representation is calculated. The significance of the difference

[2]The reason that the error is significantly larger for one feature than the others is due to the magnitude of the feature. Whilst the absolute size of the error does not vary significantly, the small magnitude of one particular feature results in a larger percentage error.

between the two values is analysed using Student's t-test [45] at the 1, 5 and 10 % level.

The number of data points required to represent each cluster at these different percent levels is given in Table 5.1. The cluster representations are calculated using the number of data points needed at the 10% level. The 10% level is chosen as this is the worst case critical percentage level for the null hypothesis being rejected in the comparison between the true cluster centers and those returned from the hierarchical clustering algorithm.

| Cluster | Percent level | | |
|---|---|---|---|
| | 1% | 5% | 10% |
| 1 | 450 | 450 | 450 |
| 2 | 360 | 390 | 400 |
| 3 | 310 | 340 | 360 |

Table 5.1: Number of data points required to represent each cluster as determined by the t-test at the 1, 5 and 10% levels.

**Generating cluster prototypes** The cluster prototypes are calculated from the cluster representations and consist of the cluster centres (which are calculated during hierarchical clustering), covariance matrices and average intracluster distances. These prototypes are stored in memory and are available for use when generating memberships for unlabelled feature vectors.

## 5.3 Membership generation for new feature vectors

The aim of this section is to develop a mechanism by which memberships are generated for a new set of data points in each of the clusters to which the AEM has been trained. The data available is the location of the new feature vectors and the cluster prototypes, generated during training, which include the cluster centers, covariance matrices and average intracluster distance for all of the known appliances.

It should be recalled that the membership value should be based on the proximity and relative proximity of the new feature vector to the existing clusters and should take into

account the spread of the data within each cluster. Section 5.3.1 therefore considers how the proximity measure is calculated to ensure that the spread of the data is accounted for.

Having determined the set of proximity values, Sections 5.3.3 and 5.3.4 explore two well developed methods of membership generation. The benefits and limitations of each method are discussed and this gives rise to two schemes which attempt to overcome the problems of each individual method of membership generation. These schemes are explored in Section 5.3.5 and 5.3.6 and whilst they provide an interesting insight into membership generation, the mechanisms proposed are not entirely appropriate for the work presented in this research.

By considering the information available and the desired capability of the membership function, a novel mechanism of membership generation is proposed and assessed against the key criteria in Section 5.3.7. It is then applied to a real data set, the results of which are analysed in Section 5.3.8.

### 5.3.1 Proximity measures

The proximity measure should define the distance between feature vectors which are considered to be data points in d-dimensional space. This proximity measure is denoted as $\mathrm{d}(\mathbf{x}_i,\mathbf{c}_j)$, or $\mathrm{d}_{ij}$ and represents the distance between data point $i$ and cluster centre $j$ in feature space. This section examines three well known methods of calculating distances between two data points.

**Euclidean distance** The Euclidean distance [49] is defined by:

$$\mathrm{d}_2(\mathbf{x}_i, \mathbf{c}_j) = \sqrt{\sum_{k=1}^{d}(x_{i,k} - c_{j,k})^2} \tag{5.6}$$

Where the number of dimensions in feature space is equal to $d$. This distance measure is quite commonly used although a drawback is the assumption that all the features have

the same unit. Recall that the feature vector is extracted from the current and voltage and the individual features may take different units, for example, time, watts, amps, ohms etc. If the Euclidean distance is to be used as a similarity measure then these features would need to be standardised to the same unit.

This distance measure is a particular instance of the Minkowski metric.

**Minkowski metric** This is defined as [50]:

$$\mathrm{d}_p(\mathbf{x}_i, \mathbf{c}_j) = \sqrt[p]{\sum_{k=1}^{d} (x_{i,k} - c_{j,k})^p} \tag{5.7}$$

The drawback with the use of the Minkowski metric as a distance measure is the over domination of the larger features, although it is possible to normalise the features so that this does not occur. However, this type of distance measure predicts hyper-spherical clusters in feature space and may become distorted due to linear correlation amongst the features. Furthermore, the Minkowski metric does not account for the differences in spread of data within individual clusters.

**Mahalanobis distance** The Mahalanobis distance is defined by [51]:

$$\mathrm{d}_m(\mathbf{x}_i, \mathbf{c}_j) = (\mathbf{x}_i - \mathbf{x}_j)\mathbf{Q}_j^{-1}(\mathbf{x}_i - \mathbf{x}_j)^T \tag{5.8}$$

Where $\mathbf{Q}_j$ is the covariance matrix of cluster $j$. By using the Mahalanobis distance as a distance measure, any correlation between features and differences in values and spread amongst features can be taken into account. It does not assume hyper-spherical clusters as the Minkowski Metric does and therefore it satisfies the criteria that the proximity measure take into account the different spreads of data within each cluster.

The one drawback with this proximity measure is the assumption that the data is already split into clusters so that the covariance matrices of each cluster may be determined. However, for the particular application of calculating distances between new feature

vectors and existing cluster, this drawback is not encountered. Therefore the Mahalanobis distance is used as the proximity measure in the following sections.

### 5.3.2 Fuzzy clustering

Fuzzy clustering is a method of clustering the data points so that each point is associated with every cluster by a membership function varying between 0 and 1 rather than an absolute membership (as in hard clustering where the data point belongs to a single cluster with complete certainty) [52]. This allows the data points to belong with a limited certainty to more than one cluster, which may be appropriate for some of the appliances considered.

Fuzzy clustering is an unsupervised clustering technique, which is normally applied when there is no other information about the data available apart from the location of the data points and the number of clusters which it should form. The mechanism by which it is performed is as follows [53]:

1. Choose $K$ individual data points to coincide with $K$ cluster centers.
2. Generate memberships for each data point in each of the $K$ cluster centers. The method by which these membership values are generated is discussed in Sections 5.3.3 and 5.3.4.
3. Recalculate the cluster centres based on the assigned memberships.
4. Repeat steps 2 and 3 until some stopping criterion is met.

Because the AEM has been through a training routine the cluster centres are already known and therefore standard fuzzy clustering techniques are not entirely appropriate. However, the mechanisms by which the memberships are generated in Step 2) of the fuzzy clustering algorithms are worth considering and the following sections explore two well developed methods.

### 5.3.3 Probabilistic fuzzy clustering

This section explores the Fuzzy-C Means (FCM) clustering algorithm whos membership generation is based on probability theory. Probabilistic fuzzy clustering uses the membership function to determine the relative degree of belonging of each data point to each cluster. It places the following constraint on the membership matrix **U** [54]:

$$u_{ij} \in [0,1] \quad \text{for all } i \text{ and } j \tag{5.9}$$

$$0 < \sum_{j=1}^{N} u_{ij} < N \quad \text{for all } i \tag{5.10}$$

$$\sum_{i=1}^{M} u_{ij} = 1 \quad \text{for all } j \tag{5.11}$$

where $M$ is the total number of classes. In the FCM algorithm [55] memberships are assigned so that the objective function $J_m$ is minimised whereby:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{M} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad 1 \leq m < \infty \tag{5.12}$$

which, along with the criterion imposed by equation 5.11, implies membership values generated according to:

$$u_{ij} = \frac{1}{\sum_{k=1}^{K} \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \tag{5.13}$$

In the above equations, $m$ is called the fuzzifier and it determines how rapidly the membership function decays as a result of distance from the cluster center. This is dealt with in greater detail in Section 5.3.4. For the FCM algorithm, a value of $m = 2$ is known to produce good results [56].

This method of membership generation satisfies only the following 3 of the 5 criteria:

- Memberships may exist between 0 and 1 according to their level of belonging in a cluster as is the case in all fuzzy clustering schemes.
- Memberships account for the varying spread of data in a cluster as the Mahalanobis distance is the proximity measure used.

- Memberships take into account the relative proximity of data points to clusters, as the distance between the data point and every cluster is used to determine every membership value.

Two of the criteria are not met. The first is that the membership values are based only on the relative distance to the clusters rather than the absolute distance and therefore do not take into account the actual proximity measure. This causes a problem which is illustrated in Figure 5.5. Points A and B are equidistant from the cluster centre of Cluster 1 and should therefore have the same membership in the cluster, however B will have a higher membership than A in cluster 1 due to A's proximity to cluster 2. For the same reason point C will have the same membership as point B in cluster 1, even though it should have a higher value as it is more typical.

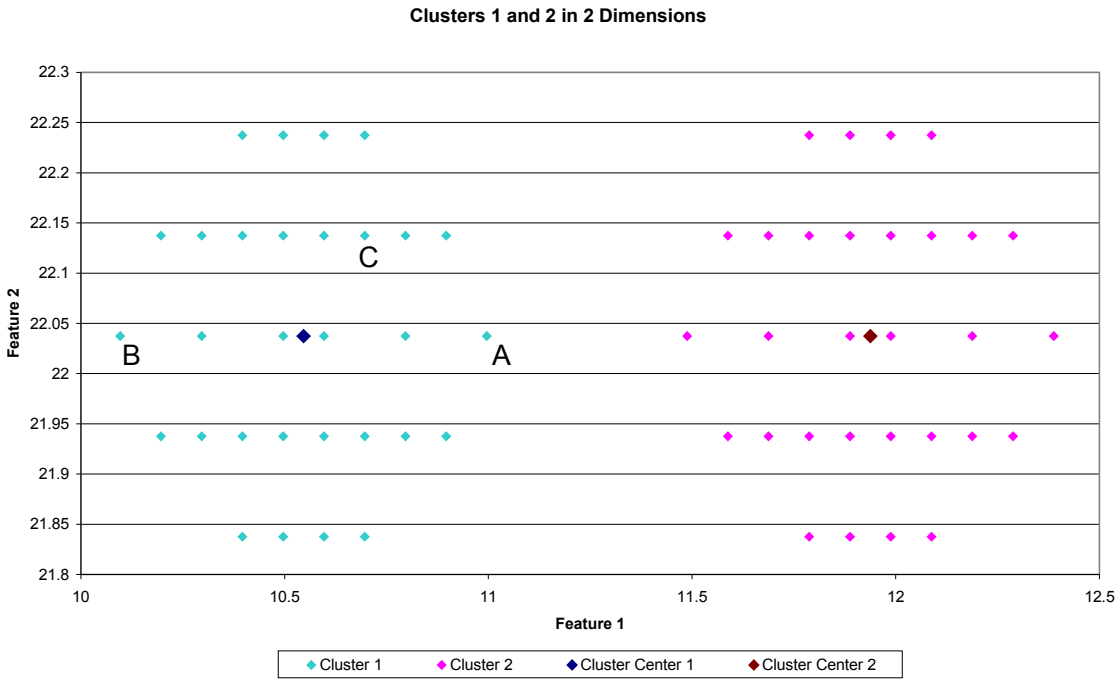


Figure 5.5: 2 dimensional representation of Cluster 1 and Cluster 2

The second criteria which is not met is the ability of the memberships to account for the data point belonging to none of the known clusters. This is due to the constraint that the sum of the memberships for a data point across all clusters must be equal to 1. This means that a feature vector which lies a great distance from any cluster may

be given membership values which do not accurately represent the level to which that vector belongs in the clusters. This may occur due to noise or transients on the system, or due to the addition of new appliances to which the system has not yet been trained.

Both these problems are overcome by relaxing the constraint that the cluster memberships must sum to 1 for each data point (as stated mathematically in Equation 5.11), as is the case in possibilistic fuzzy clustering.

### 5.3.4 Possibilistic fuzzy clustering

This section explores the mechanism of membership generation based on possibilistic fuzzy clustering. In this type of clustering the constraint imposed by equation 5.11 is relaxed so that it becomes [54]:

$$\sum_{i=1}^{K} u_{ij} \leq 1 \quad \text{for all } j \tag{5.14}$$

In order to avoid the trivial solution of all memberships being equal to zero when minimising the objective function $J_m$, the objective function must include a term to increase the membership values of all representative data points, whilst allowing the membership values for data points which are not representative of the data to tend to zero [54]. This can be achieved with an objective function [48]:

$$J_\gamma = \sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{\gamma} d_{ij}^2 + \sum_{i=1}^{K} \eta_i \sum_{j=1}^{N} (1 - u_{ij})^{\gamma} \tag{5.15}$$

$d_{ij}$ is the distance from the $j^{th}$ data point to the $i^{th}$ cluster center, $\gamma$ is the fuzzifier (discussed below) and $\eta_i$ is the average intracluster distance. The minimisation of this function implies a membership generated by [54]:

$$u_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_i}\right)^{\frac{1}{\gamma - 1}}} \tag{5.16}$$

The membership value $u_{ij}$ is therefore solely dependent on the normalised absolute distance, $d_{ij}^2 \big/ \eta_i$, between the $i^{th}$ feature vector and the $j^{th}$ cluster centre. Figure 5.6 illus-

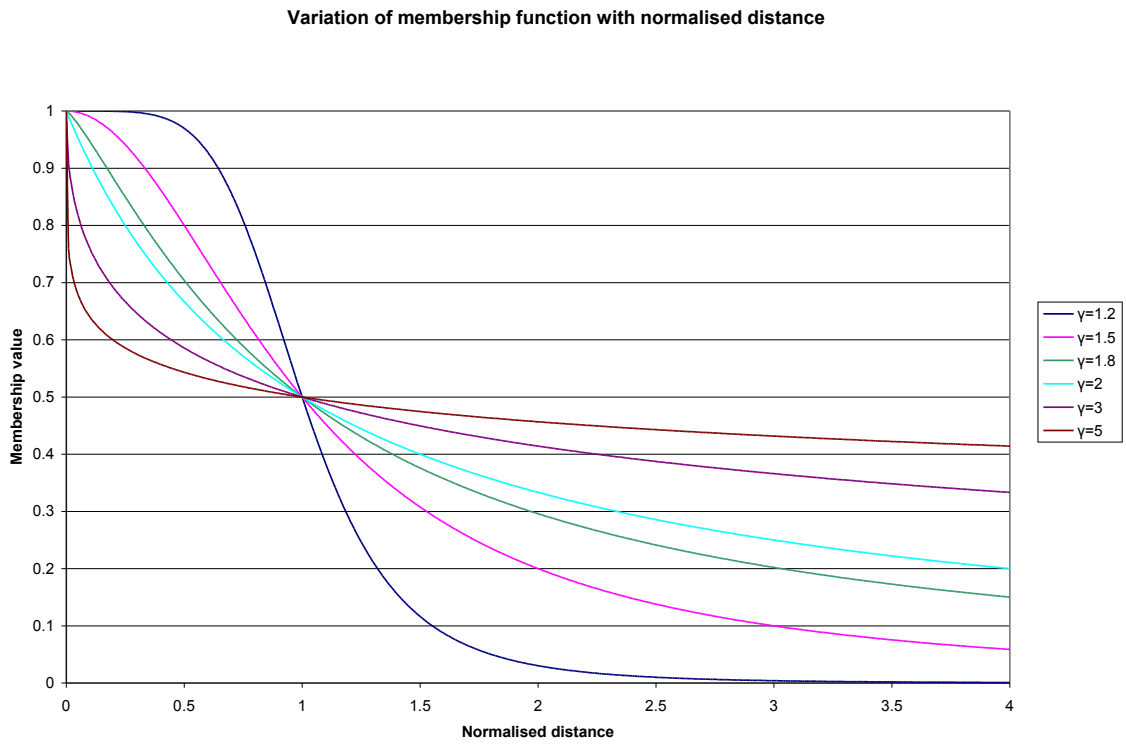


Figure 5.6: Plot of the variation of membership value with increase in normalised distance from the cluster centre for a range of $m$ values

trates the variation of the membership value with normalised distance from the cluster center.

From this figure it can be seen that the membership value is reduced to 0.5 at a normalised distance of 1, regardless of the value of $\gamma$. Therefore, $\eta$ is a measure of the 'elasticity' of the cluster - a compact cluster will have a lower value of $\eta$ than an elongated cluster.

It can also be seen from Figure 5.6 that the value of $\gamma$ has a large influence on the way in which the cluster membership value varies with normalised distance. As the value for $\gamma$ increases, the rate of decay of the membership value with distance decreases. It was suggested for the FCM algorithm that a value of $m = 2$ produced sensible results [56], however, it can be seen from figure 5.6 that a value of $\gamma = 2$ decays at a fairly slow rate with normalised distance.

A suitable value of $\gamma$ can be estimated by considering the probability distribution within a hyper-shperical normal distribution. Under these conditions, 95% of data points are

contained within a distance of $2\sigma$ from the mean. Therefore, a value of $\gamma$ which results in $u_{ij} > 0.05$ for data points within $2\sigma$ of the mean is appropriate.

Approximating $\eta_i$ by $\sigma^2$, a distance of $2\sigma$ from the mean corresponds to a normalised distance, $d_{ij}^2 \big/ \eta_i$, approximately equal to 4. From Figure 5.6 it ca be seen that a membership value of 0.05 at a normalised distance of 4 occurs at $\gamma = 1.5$ and therefore this value is deemed a more suitable fuzzifier in possibilistic fuzzy clustering [56].

The memberships generated through a possibilistic fuzzy clustering scheme do better than the probabilistic memberships as they satisfy 4 of the 5 criteria. However, the criteria that the membership must be based on the relative proximity of the data point to a cluster is not satisfied because the memberships are generated by considering only the absolute distance between a feature vector and a cluster center.

This is analogous to considering memberships based on mutually independent events [57], which causes problems with the types of memberships considered here. A significant membership value in a particular appliance group (which implies the observation of that particular appliance changing states) should rule out the possibility of a significant membership value being generated in another appliance due to the assumption made in Chapter 4 that only a single appliance changes state between two inspections of the system. However, because possibilistic memberships for one cluster are calculated without any consideration to the data points proximity to other clusters, it may be possible that significant membership values are generated in multiple clusters.

Therefore, a method of generating memberships needs to be developed whereby the memberships are based on mutually exclusive events (as in probabilistic fuzzy clustering [57]) but which overcomes the problems of misclassification based on the constraint that the sum of the memberships must be 1.

### 5.3.5 Graded membership values

This section explores a method of generating memberships which is representative of a system with events which lie somewhere between being mutually exclusive and mutually independent, which is more representative of the domestic system than either option individually.

The probabilistic membership values are under the constraint that the total membership of a feature vector across all clusters must be equal to 1, whereas the possibilistic membership values are under no such constraints. Therefore, a graded mechanism is proposed whereby the total membership value is constrained to some degree which is dependent on the level to which the system resembles the probabilistic and possibilistic systems.

When considering a possibilistic membership function, the only constraint on the values that the membership must take is such that each membership value lies between the limits of 0 and 1 and the sum of this membership across all clusters must be less than or equal to 1. Considering a case where there are $K$ clusters, the sum of the membership values for a single data point across all clusters must therefore lie somewhere in the region between 0 and $K$. The membership function can be thought of as a unit hypercube in $K$ dimensions, where the membership function can occupy any area of the hypercube [57]. Reducing the argument to two dimensions for illustration purposes, Figure 5.7 shows a unit square, where the axes represent the membership values for cluster 1 and cluster 2. In the possibilistic model, the membership in cluster 1 has no influence on the membership in cluster 2 and the data point representing the 2 dimensional membership value , $[\mu_1, \mu_2]$, can exist anywhere within this square as represented by point A. In the probabilistic approach, the two memberships must sum to 1, hence the data point representing this 2D probabilistic membership must exist along the line $\mu_1 + \mu_2 = 1$, as shown by point B. The graded approach allows the membership to exist in a region inside the box bounded by constraints less restrictive than those implemented in the probabilistic approach, but

more restrictive than the possibilistic scenario. This is illustrated by point C, which can lie anywhere between the two dotted lines. This can be extended mathematically to more than 2 dimensions as follows [57].

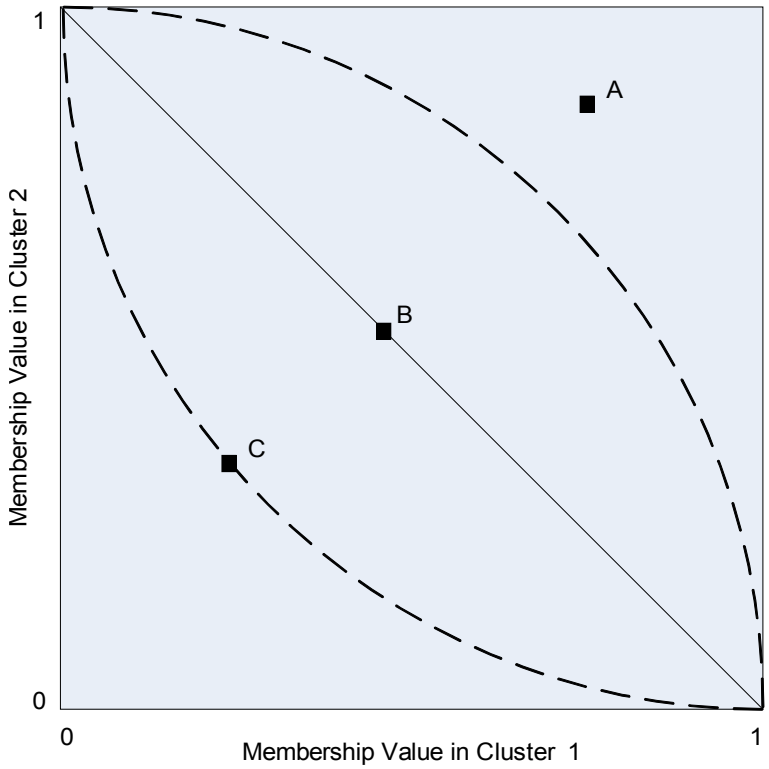


Figure 5.7: Illustration of how membership values can vary for a 2 cluster scenario. Graph modified from information presented in [57].

Let $\xi$ be a variable so that:

$$\underline{\xi} \leq \xi \leq \overline{\xi} \tag{5.17}$$

and allow there to be a particular value $\xi^* \in \left[\underline{\xi}, \overline{\xi}\right]$, so that:

$$\sum_{j=1}^{K} u_{ij}^{\xi^*} = 1 \tag{5.18}$$

In the probabilistic approach $\underline{\xi} = 1$ and $\overline{\xi} = 1$, whilst the possibilistic approach sets limits of $\underline{\xi} = 0$ and $\overline{\xi} = \infty$. In the graded approach, select $\underline{\xi} = \alpha$ and $\overline{\xi} = \frac{1}{\alpha}$ where $\alpha \in [0, 1]$, so the single parameter $\alpha$ controls the level of possibility.

The membership value is now calculated as $u_{ij} = v_{ij}/Z_j$, where $v_{ij}$ is the 'free' member-

ship as calculated in possibilistic clustering and $Z_j$ is calculated as follows [57]:

$$\begin{aligned} Z_j &= \left( \sum_{i=1}^{K} v_{ij}^{1/\alpha} \right)^{\alpha} \quad \text{if } \sum_{i=1}^{K} v_{ij}^{1/\alpha} > 1 \\ Z_j &= \left( \sum_{i=1}^{K} v_{ij}^{\alpha} \right)^{1/\alpha} \quad \text{if } \sum_{i=1}^{K} v_{ij}^{\alpha} < 1 \\ Z_j &= 1 \qquad\qquad\qquad \text{otherwise} \end{aligned} \tag{5.19}$$

Although this approach is still less constraining than the probabilistic approach, consider the scenario that the new data point to be classified is from an unknown cluster. This may be due to a new appliance added to the system, or the detection of a transient, or an error. The distances from the new data point to the existing cluster centres will be very large and consequently, the free membership value, $v_{ij}$, will be very small for all clusters. This means that the value for $Z_j$ will be very small and if one free membership value, $v_{i^*j}$, is considerably larger than the others (which may be the case as they are all very small), the value for $Z_j$ will be dominated by the largest free membership value. This means that the corresponding membership value $u_{i^*j}$ will be close to 1, despite the large distance between the data point and the cluster center.

This can also be seen by reference to Figure 5.7. The membership value generated in graded clustering will never be allowed to approach zero for all clusters (i.e. an unknown data point) unless the value for $\alpha$ is sufficiently small that the allowed region for graded clustering is almost the entire square, with results approaching possibilistic clustering. Therefore, the criteria which states that the membership generation must allow a data point to remain unidentified is not satisfied.

A method for membership generation which allows both the probabilistic and possibilistic values to individually have an input to the final membership value may be more suitable.

### 5.3.6 Fuzzy clustering using a combined approach

This section explores a way in which both the probabilistic and possibilistic membership values may have an input into the fuzzy clustering algorithm. Although the mechanism

considered here is not entirely suitable for the approach taken in this research, there are important ideas which are subsequently used in a novel way (explored in Section 5.3.7).

A fuzzy clustering algorithm is proposed which calculates both the probabilistic memberships, $u_{ij}$, and possibilistic typicalities, $t_{ij}$, and combines the results in order to perform the cluster centre update equation (Step 3 in the fuzzy clustering algorithm outlined in Section 5.3.2) in the following way [58]:

$$\mathbf{c}_j = \frac{\sum_{i=1}^{N} \left( a u_{ij}^m + b t_{ij}^\gamma \right) \mathbf{x}_i}{\sum_{i=1}^{N} \left( a u_{ij}^m + b t_{ij}^\gamma \right)} \tag{5.20}$$

where $1 \leq j \leq \bar{K}$ and $\bar{K}$ is the total number of clusters. As discussed above, the exponent $m$ should be set to 2 and $\gamma$ set to 1.5. By varying the relative weightings of $a$ and $b$, the amount of influence of $u$ and $t$ into the cluster centres can be varied.

This algorithm combines the probabilistic memberships and possibilistic typicalities in order to overcome problems presented when using $u$ and $t$ independently to calculate the cluster centres from a set of unlabelled, unclustered data. Whilst this is inapproporiate for the work considered in this thesis, as the cluster centres are already known, the idea of combining the probabilistic and possibilistic membership values to generate a novel membership is explored in the following section.

### 5.3.7 Generating membership values

This section considers the combination of the probabilistic and possibilistic memberships in order that they can be used to overcome the potential problems when either is considered independently. The problem with the probabilistic membership is the potential misclassification that may occur if a data point is from an unknown source and therefore not typical in any cluster. In this scenario it is desirable that the membership of the data point be negligible in all clusters, but this will not be the case due to the constraint that the membership values across all clusters must sum to 1. However, the possibilistic

typicality will be close to zero for all clusters and may therefore be used to modify the probabilistic membership.

In the case that a data point is highly typical in more than one cluster, for example, if there are two clusters with large spreads of data which are close together, the possibilistic typicality may be significant for both clusters and distinction between these clusters will be hard. However, due to the constraints on the probabilistic membership, this value should be higher for the more typical cluster and lower for the less typical one and therefore in this case the probabilistic membership may be used to modify the possibilistic typicality.

A geometric combination of the probabilistic membership and possibilistic typicality is proposed in order to overcome the potential problems present in either of the individual memberships. Taking the geometric mean allows the possibilistic typicality and the probabilistic membership to provide the bounds for the overall membership, such that they both play a role in the resultant membership value, $m_{ij}$, where:

$$m_{ij} = \sqrt[a+b]{u_{ij}^a t_{ij}^b} \tag{5.21}$$

An unweighted geometric mean in which $a = b = 1$ gives equal weighting to both the probabilistic and possibilistic membership. However, it is possible to give one of the variables more weighting by changing the relative values of $a$ and $b$.

**Varying the relative weightings of probabilistic and possibilistic memberships**

The relative weightings to give to the probabilistic membership and possibilistic typicality may be determined by considering the range of values that these variables may take and determining the bounds on each for which classification should be possible.

This section begins by making a set of assumptions about the data to enable a critical membership value, $m_c$, to be established. The value for $m_c$ must be known so that a

comparative context may be placed on following discussion about the membership values obtained through combination of $u$ and $t$.

Therefore, to allow a critical membership value to be calculated, it is assumed that the clusters to which the AEM has been trained are compact and disjointed so that the probabilistic membership generated for a data point in the cluster to which is actually belongs is close to 1. Lower bounds on the range for this membership function should still lead to classification and are discussed later in this section.

It should also be recalled from Section 5.3.4 that the value of $t$ will be greater than 0.05 for 95% of data points in a particular cluster due to the mechanism by which $t$ is generated. Therefore, to determine a critical membership function, an unweighted geometric combination is performed on the values of $u = 1$ and $t = 0.05$ and this leads to the proposition of a critical membership value of 0.2, i.e. $m_c = 0.2$.

Having determined a value for $m_c$, it is possible to explore the effect of varying the relative weightings of $a$ and $b$ on the final membership value over an appropriate range of $u$ and $t$. The appropriate range of these value is from 1 down to a lower bound which is determined by considering the scenarios under which particular values may be generated.

In order to determine an appropriate lower bound for $u$, it is necessary to determine at what point the value obtained is no longer considered substantial enough for classification to occur. This is done by considering the value that would be deemed substantial in a home with 20 appliances. Let a substantial probabilistic membership be defined as one in which the membership value for a single cluster is at least 10 times bigger than that in any other cluster. For simplicity, let is also be assumed that the values generated in all the other clusters are equal. If $m_1$ denotes the membership value in the single cluster and $m_2$ the membership value in all other clusters, these constraints may be written mathematically:

$$m_1 + 19m_2 = 1 \tag{5.22}$$

$$m_1 = 10m_2 \tag{5.23}$$

These equations can be solved to give $m_1 = 0.34$ to the nearest 2 decimal places. Therefore, if the probabilistic membership is greater than 0.34 it is defined as substantial and it is desirable that data points with $u > 0.34$ and $t > 0.05$ (to include 95% of data points in a cluster) have a final value of $m$ greater than $m_c$.

Table 5.2 shows the variation in $m$ by changing the relative weightings of $a$ and $b$ in the combination of $u = 0.34$ and $t = 0.05$.

| Variation of $a/b$ | Resultant membership $m$ |
|---:|---:|
| 0.1 | 0.06 |
| 0.5 | 0.09 |
| 0.8 | 0.12 |
| 1.0 | 0.13 |
| 1.5 | 0.16 |
| 2.0 | 0.18 |
| 3.0 | 0.21 |
| 4.0 | 0.23 |
| 5.0 | 0.25 |
| 10.0 | 0.29 |

Table 5.2: Variation in $m$ by varying exponent $a$ relative to $b$ for membership values of $u = 0.34$ and $t = 0.05$.

From this table it can be seen that a ratio of $a/b \geq 3$ results in classification when $u = 0.34$ and $t = 0.05$.

If the possibilistic typicality is to be used to modify the probabilistic membership in the scenario that a data point is atypical in a cluster but still has a high probabilistic membership value in that cluster (the reasons for which were discussed in Section 5.3.3), then the combination of the values of $u$ and $t$ generated in this scenario must ensure that $m < 0.2$. To ensure that classification does not occur unless a data point is more than 0.1% typical in a cluster, a value of $t < 0.001$ should ensure that data points, regardless of their probabilistic membership value, do not have a final membership value greater than 0.2.

This relationship is examined for ratios of $a/b$ in the region of 3 and the results are

displayed in Table 5.3.

| Variation of $a/b$ | Resultant membership when $u$ equals | | | | |
|---|---|---|---|---|---|
| | 1.00 | 0.80 | 0.60 | 0.40 | 0.34 |
| 2.5 | 0.14 | 0.12 | 0.10 | 0.07 | 0.06 |
| 3.0 | 0.18 | 0.15 | 0.12 | 0.09 | 0.08 |
| 3.5 | 0.22 | 0.18 | 0.14 | 0.11 | 0.09 |
| 4.0 | 0.25 | 0.21 | 0.17 | 0.12 | 0.11 |

Table 5.3: Variation in $m$ by varying exponent $a$ relative to $b$ for a range of membership values of $u > 0.34$ and $t = 0.001$.

From Table 5.3 it can be seen that the resultant membership value is only reduced to below 0.2 when the ratio of $a/b$ is less than or equal to 3.

Therefore, a ratio of $a/b = 3$ satisfies all the constraints which are placed on the combination of $u$ and $t$ and the geometric combination used in the AEM is:

$$m_{ij} = \left(u_{ij}^3 \times t_{ij}^1\right)^{1/4} \tag{5.24}$$

### 5.3.8 Application of membership generation

This section considers the method by which the cluster prototypes calculated in Section 5.2.4 are used in a novel algorithm to generate membership values for new data points according to the mechanism described in Section 5.3.7.

#### Method

1. For each new data point calculate the Mahalanobis distance to each cluster center. The distance from the $j^{th}$ data point to the $i^{th}$ cluster is calculated as:

$$d_{ij} = (\mathbf{x}_j - \mu_i) \cdot \mathbf{Q}_i^{-1} \cdot (\mathbf{x}_j - \mu_i)^{\mathrm{T}} \tag{5.25}$$

   where $\mu_i$ is the cluster centre of the $i^{th}$ cluster and $\mathbf{Q}_i$ is its covariance.

2. Use the distance measures and the average intracluster distances to calculate the membership values $u_{ij}$ and $t_{ij}$ (where $1 \leq i \leq M$) for the $j^{th}$ data point. The

membership values are calculated according to:

$$u_{ij} = \left( \sum_{k=1}^{M} \left( \frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)} \right)^{-1} \tag{5.26}$$

$$t_{ij} = \frac{1}{1 + \left( \frac{d_{ij}^2}{\eta_i} \right)^{1/(\gamma-1)}} \tag{5.27}$$

3. Calculate the resultant membership value $m_{ij}$ for each data point $j$ in very cluster $i$ according to Equation 5.24.

4. Use the resultant membership to determine which appliances the data point may belong to based on the short term time domain features only. Membership values must be greater than 0.2 for classification.

**Results and analysis thereof**

During the training procedure, the AEM is trained to 'know' the cluster prototypes for a fan heater with 3 modes of operation. The recognition data which is used to test the proposed method consists of a new data set belonging to the fan in all 3 modes in addition to data belonging to an unknown appliance. In this example the unknown appliance is chosen to be a kettle due to is similarity in electrical construction with the fan heater when it is operating in its heating mode.

Figure 5.8 shows a plot of the recognition data for each cycle for the kettle and the fan heater in all 3 modes. The feature data is plotted for each cycle of each appliance, along with the cluster centres obtained during training. Membership values are generated for each data point in each of the known clusters.

Figures 5.9, 5.10 and 5.11 illustrate the variation of the membership values of the new data sets belonging to the fan heater in mode 0 (fan only), mode 1 and mode 2 respectively, in the cluster to which they belong. The membership values generated for each new data set are negligible in the clusters to which they do not belong. Figures 5.12, 5.13 and 5.14

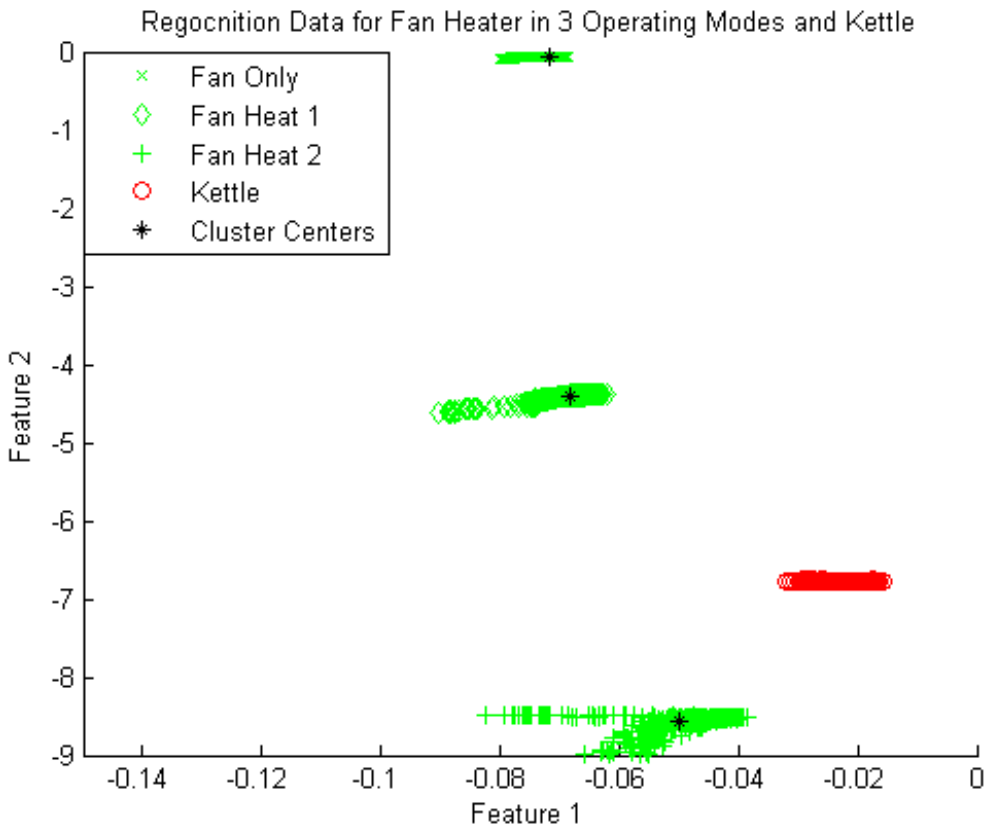


Figure 5.8: Plot of recognition data projected onto 2 dimensions.

illustrate the variation in membership values of the data set generated by the kettle in each of the known clusters from training.

From the figures it can be seen that for the data sets from the appliances to which the AEM has been trained (i.e. the fan heater in all three modes) only 5 data points have a resultant membership value of less than 0.2 in the cluster to which they belong. It has also been stated that the resultant membership value, $m$, must be greater than 0.2 for classification to be possible. This illustrates that more than 99.5% of data points are classified correctly based on their short term membership data alone for the appliances used in this example.

The resultant membership values are negligible for the kettle data points in each of the known clusters. Even though the probabilistic membership lies close to 1 for the kettle data points in the cluster prototype for the fan heater in heat mode 2, the possibilistic

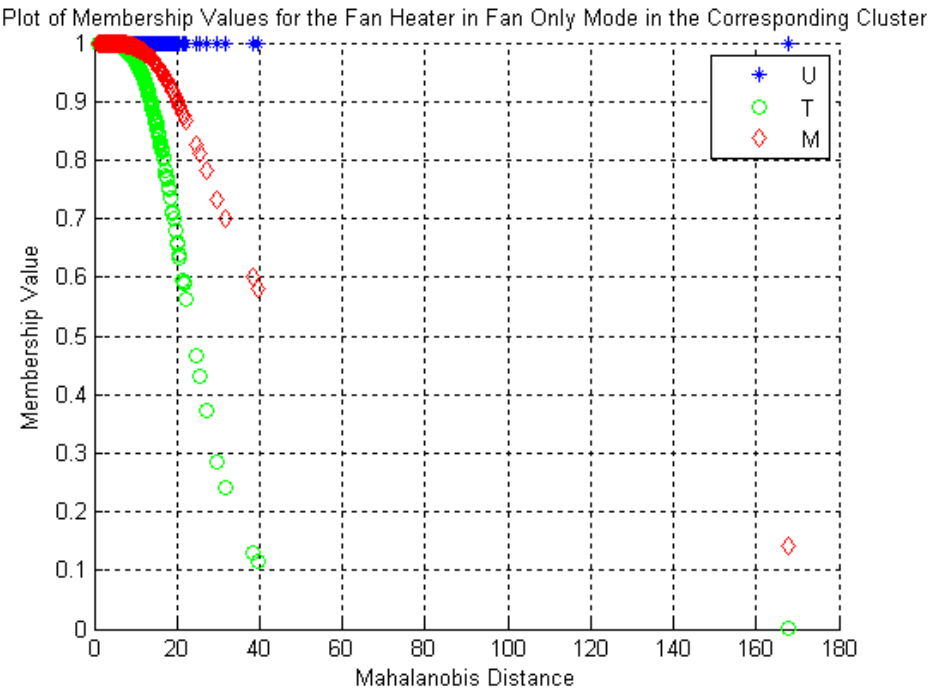


Figure 5.9: Plot of membership values for the fan heater in fan only mode in the corresponding cluster.

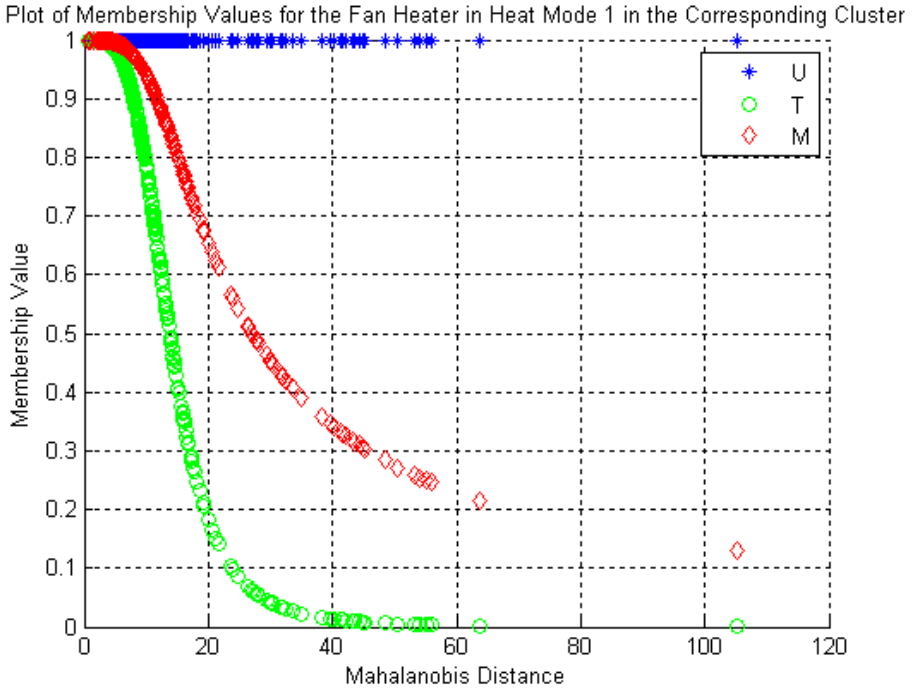


Figure 5.10: Plot of membership values for the fan heater in heat mode 1 in the corresponding cluster.

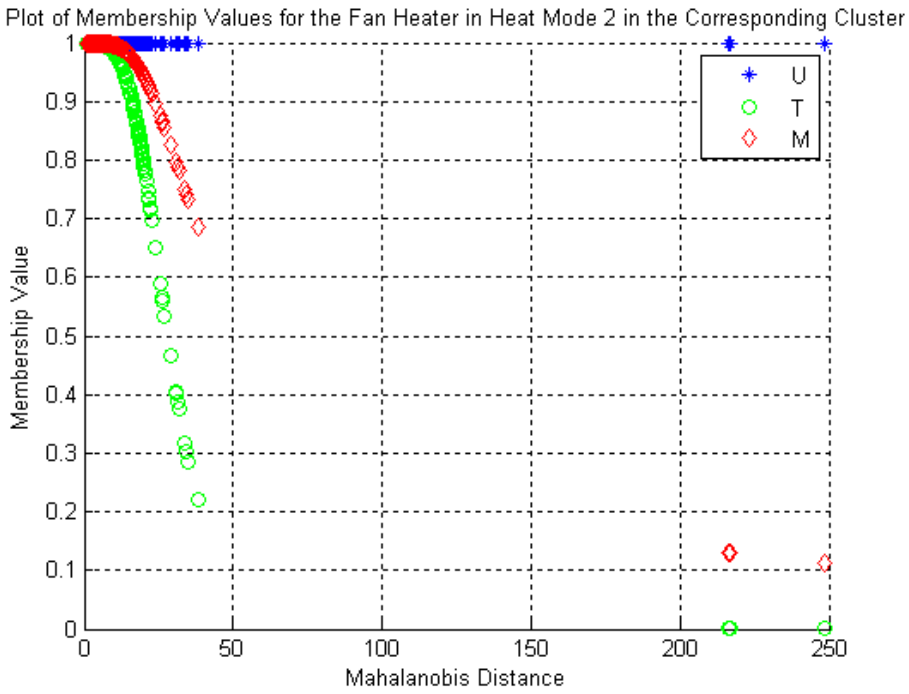


Figure 5.11: Plot of membership values for the fan heater in heat mode 2 in the corresponding cluster.

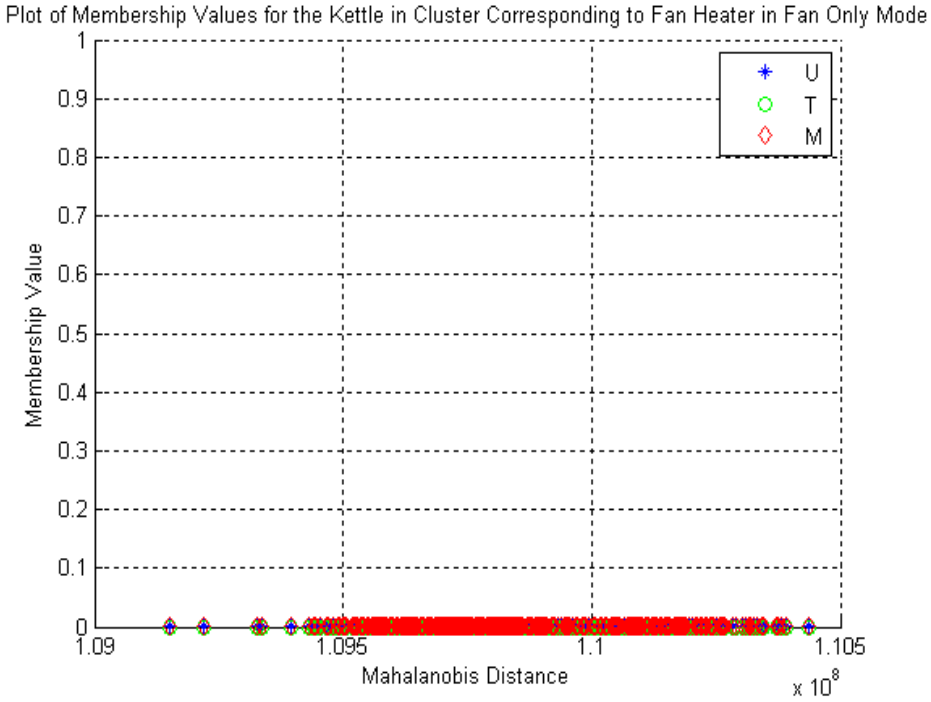


Figure 5.12: Plot of membership values for the kettle in the cluster corresponding to the fan heater in fan only mode.

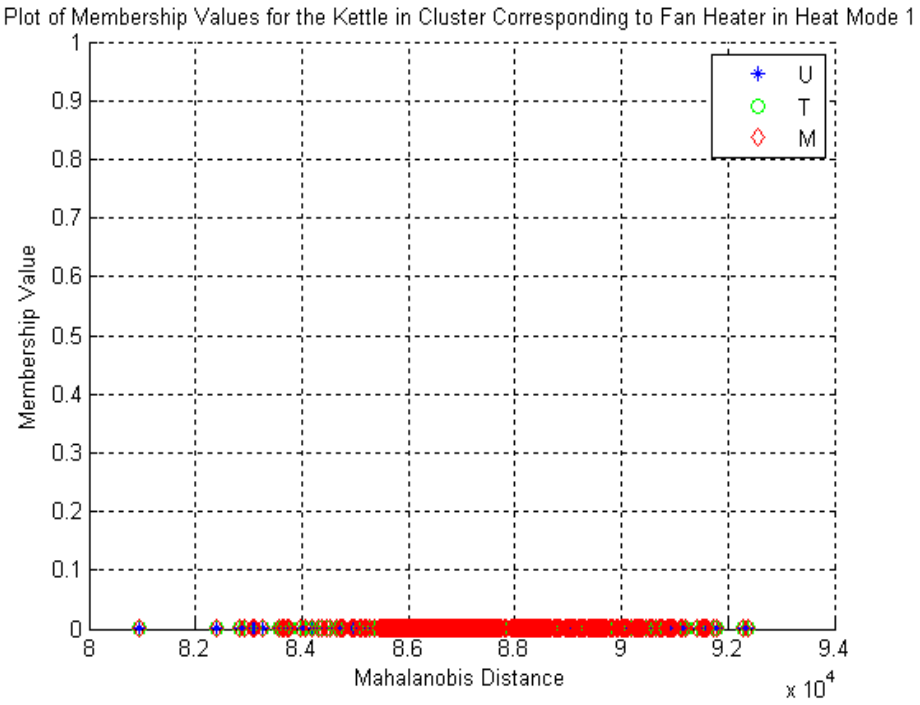


Figure 5.13: Plot of membership values for the kettle in the cluster corresponding to the fan heater in heat mode 1.

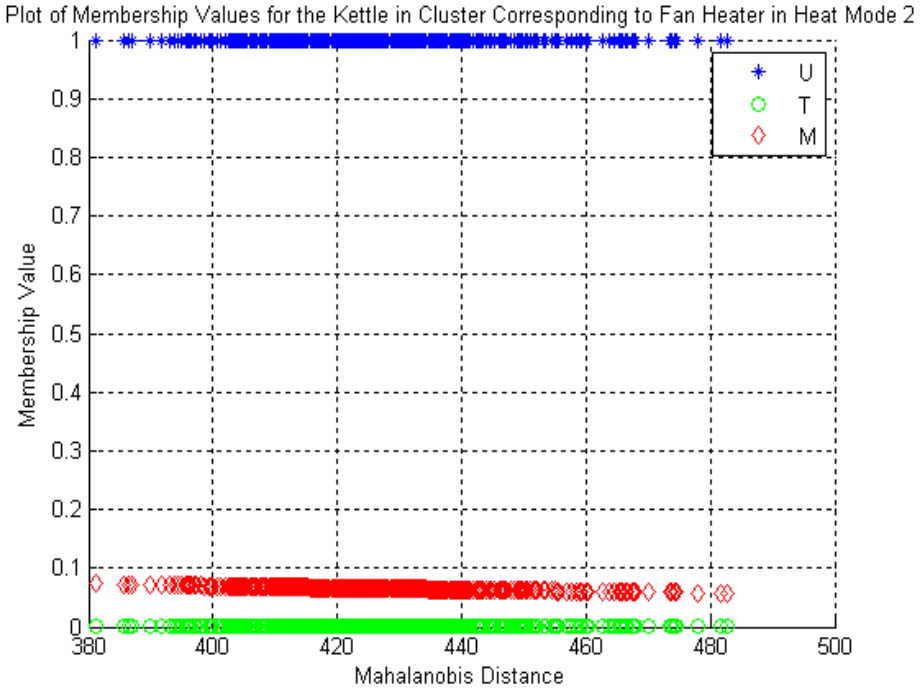


Figure 5.14: Plot of membership values for the kettle in the cluster corresponding to the fan heater in heat mode 2.

typicalities generated are small enough to prevent misclassification

These results illustrate the ability of the membership function to generate correct classifications at an accuracy rate of 99.5% whilst preventing false positive results for appliances which are unknown to the system. These results therefore demonstrate that, for the appliances considered in this chapter, the membership values perform in the way that they were designed to.

## 5.4 Chapter summary

This aim of this chapter was to develop a mechanism for generating the short term membership values for unlabelled feature vectors in a set of existing clusters, which represent the degree to which the unlabelled feature vector belongs to that cluster.

This has been done in a two stage process. The first stage involved training the AEM to 'know' various appliances by storing information relating to the average values of the features obtained, covariance matrices and average intracluster distances for each mode of operation of each appliance.

The second stage was the development of a method for generating memberships for an unlabelled data point in each of the known clusters by considering the proximity and relative proximity of the data point to the known clusters. Probabilistic memberships, $u$, and possibilistic typicalities, $t$, were calculated and used to generate a novel membership $m$, where $m = (u^3 \times t)^{1/4}$.

This membership value is assessed to be 99.5% accurate when used to identify the appliance to which an unknown data point belongs. Therefore the mechanisms proposed in this chapter generate memberships to represent the degree to which an unlabelled feature vector belongs to a set of known clusters to a classification accuracy of 99.5%.

# Chapter 6

# Short Term Features

The aim of this chapter is to identify the feature set $\mathbf{x}$, which is extracted from the observed voltage and delta current signals when an appliance changes state. This extraction is performed by the Feature Extraction Block (FEB) of the AEM system design, as highlighted in Figure 6.1.

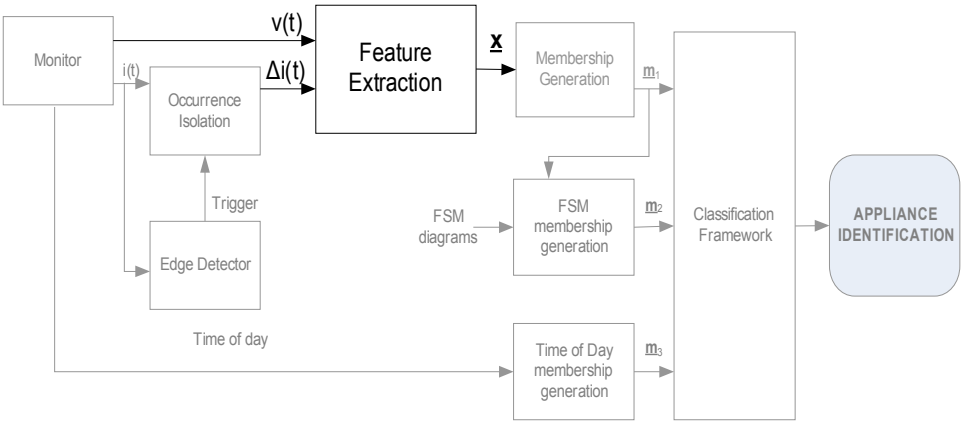


Figure 6.1: Feature Extraction Block.

Because the feature set is extracted from the voltage and current and because the current drawn by an appliance is intrinsically linked to the voltage via the appliance's electrical interface, Section 6.1 begins by exploring variations which may occur over time within the voltage signal. These variations may be carried through to the current and it is important that the features extracted by the MGB are independent of any variations.

Having determined the set of possible voltage variations over time, Section 6.2 examines a selection of typical currents drawn by appliances in the home. This analysis is performed

in order to highlight any potential problems that the AEM will have to overcome when extracting features from these waveforms.

By considering the implications of variations in the voltage and issues highlighted by the current signals observed, a set of criteria to which the selected features must adhere are proposed in Section 6.3. These aim to ensure that when appliances are represented by the selected features, each appliance forms a compact and disjointed cluster within the chosen feature space.

Section 6.4 presents an overview of three possible methods of feature extraction from the voltage and current signals. These three methods are discussed in Sections 6.5, 6.6 and 6.7.

Section 6.5 considers features which are based on a power analysis of the voltage and current monitored. Due to the interdependency of the features established, only the real and apparent power of each half cycle waveform are retained in the 4-dimensional feature set. However, because the positive and negative half cycle power drawn by most appliances is identical, the feature space effectively collapses to 2 dimensions. This results in possible misclassification due to lack of information and therefore this method of feature extraction is not considered appropriate.

Section 6.6 consider how further features may be generated by convolving the current with various test signals. This allows the amount to which the current resembles the current drawn by particular known interfaces to be quantified. However, this method of feature extraction results in features which vary with voltage varations and therefore it is not appropriate.

The final and chosen method of feature extraction attempts to overcome the problems encountered with the previous methods by choosing features which model the underlying electrical interface of individual appliances. Section 6.7 explains how features which correspond to the component parameters of a limited number of appliance interfaces may

be extracted from the voltage and current observed and used to represent appliances. This method of feature extraction is limited by the appliance interfaces proposed, however, the features obtained are appropriate and adhere to the criteria established in Section 6.3.

In Section 6.8 the selected feature set is extracted from real data and an analysis is performed to determine how well the chosen features perform in their ability to represent appliances.

## 6.1 Variation of the voltage signal with time

The feature set considered in this chapter is extracted from the voltage and current waveforms observed for individual appliances. These two waveforms are linked together via the appliance's electrical interface and therefore any changes occurring in the voltage waveform over time may be mirrored in the current waveform. Since it is desirable that the selected features be independent of any voltage variation, this section considers the potential variation in the voltage waveform and the subsequent variation in the current waveform.

Figure 6.2 shows a plot of two cycles of the observed voltage waveform against time. Two data sets are included in the plot. These data sets are captured in the Thom Building (one of the engineering department buildings in Oxford) on different days of the week and at different times during the working the day in order to illustrate the possible variation in the signal over the course of a day in the same location[1].

The first observation to make is that the signal forms a non-ideal sinusoid. The voltage signal is normally taken to be a 50 Hz ideal sinusoid for mathematical calculations, however, from Figure 6.2 it can be seen that higher harmonics are present in the signal.

[1]It is noted that a better illustration would be a comparison of the voltage observed during the peak daytime hours against that seen during the night, however, because the measurements are not used as absolute values and rather as an indication of what may be possible, the two times selected here are sufficient to observe potential voltage variations.

The second observation is that the two curves are slightly different in magnitude and shape, indicating not just a difference in the magnitude of the fundamental harmonic but also in the higher harmonics. This is illustrated further in Figure 6.3, which shows a plot of the relative magnitude (compared to the fundamental) of the harmonic content of these two signals. This highlights the possible variation in both magnitude and relative magnitude of the harmonic content of the voltage signal.

In order to explore the possible impact of this voltage variation on the current, consider 3 fundamental linear components which are likely to be present in a large proportion of appliances. These are a resistor, $R$, an inductor, $L$, and a capacitor, $C$. The following equations describe the way in which current would be drawn for the 3 types of impedance under an applied voltage $V$:

$$\begin{aligned} i_R &= \frac{V}{R} \\ i_L &= \frac{V}{j\omega L} \\ i_C &= j\omega CV \end{aligned} \tag{6.1}$$

Therefore the current will not only be dependent on the magnitude of the fundamental voltage, which has been seen to vary from the curves plotted in Figure 6.2, but also on the magnitude of the harmonics. This harmonic effect is magnified when considering the harmonics in the case of a capacitor due to the multiplication of the magnitude of the voltage harmonic by a factor of $\omega$.

As an example, if there is a 2% change in the magnitude of the fundamental voltage (with the relative magnitudes remaining constant) applied across a capacitive interface there will be a 14% change in magnitude of the $7^{th}$ current harmonic[2] and a 16% change in the the 7th harmonic of the power. Therefore, when considering which feature set to use to represent various domestic appliances, care must be taken in order to ensure that the features are invariant to these changes in the voltage signal over time.

[2]Recall that for a capacitor $i = j\omega Cv$ where $i$ is the current, $\omega$ is the frequency and $C$ is the capacitance. If the relative magnitudes of the harmonics remain the same, for the $7^{th}$ harmonic $\omega = 7\omega_0$ and therefore for a 2% change in the magnitude of the fundamental, there will be a 14% change in the magnitude of the $7^{th}$ harmonic.

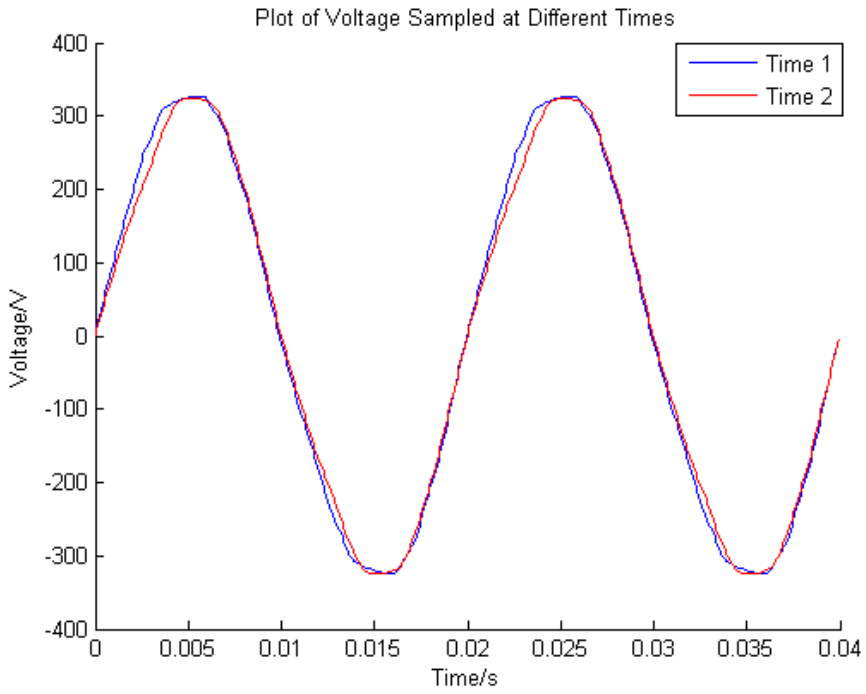


Figure 6.2: A plot of two cycles of the voltage signal for data capture at two different time periods.

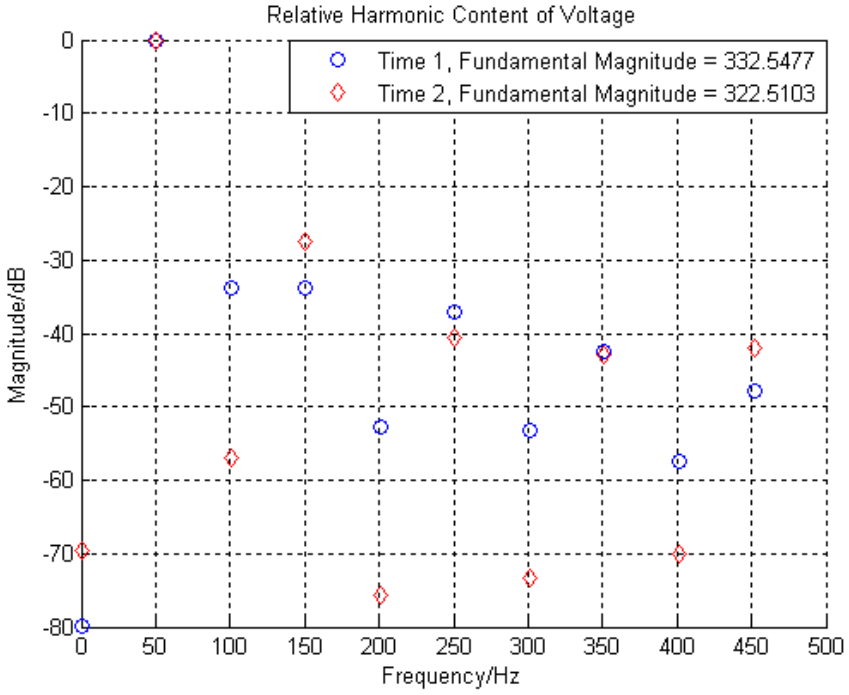


Figure 6.3: A plot of the harmonic content of the voltage signal for two signals captured at two different time periods.

## 6.2 A selection of typical current curves

This section explores the properties of some typical current signals of various domestic appliances in order to highlight any potential problems that the AEM will have to overcome when extracting features from these waveforms. The appliances selected are indicative of the range of the appliances used in the home and include a kettle, a hairdryer, a fan heater, a CD player and a TV. Sections 6.2.1 to 6.2.5 illustrate the typical voltage and current waveforms for the 5 appliances. 2 cycles are chosen from a section of the sampled data just after turn on and 2 cycles are chosen from near the end of the sampled data, so that any underlying trend in the data can be observed. The voltage and current waveforms are plotted in both the time and frequency domains, in order to allow for observation of variations in the the fundamental and higher harmonics between the start of use and end of use data. The implications of these observations are discussed in Section 6.2.6.

### 6.2.1 Kettle

The first device to be considered is a kettle[3]. Figure 6.4 shows a plot of the voltage and current drawn by the kettle against time for 2 cycles of data captured shortly after switch on and 2 cycles worth of data taken shortly before switch off. The time between switch on and switch off in this case is about 3 minutes, which corresponds to the time taken for the kettle to boil. This plot clearly illustrates two points. The first is that the shape of the current curve is very similar to that of the voltage curve. A kettle is mainly resistive and consequently the device is linear with little or no phase shift when compared to the voltage.

The second point to draw is the similarity between the two sets of data which is reflected by the relative magnitude of the harmonics, plotted on a dB scale in Figure 6.5. The

[3]Tesco value cordless kettle, model number JK04W.

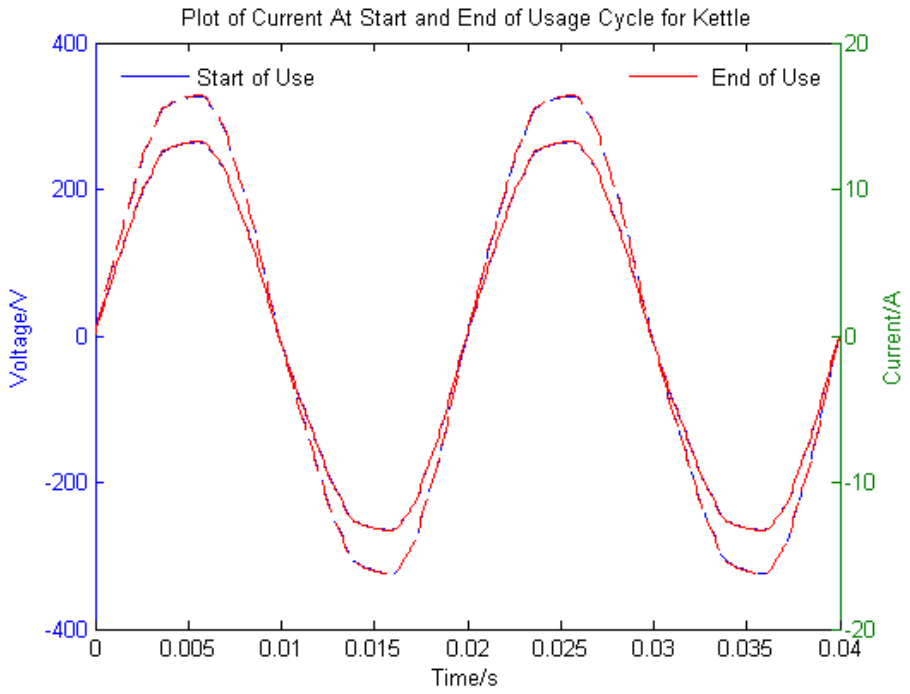


Figure 6.4: Plot of voltage and current signal drawn by a kettle against time. The current curves are represented by solid lines and the voltage curves by dashed lines.

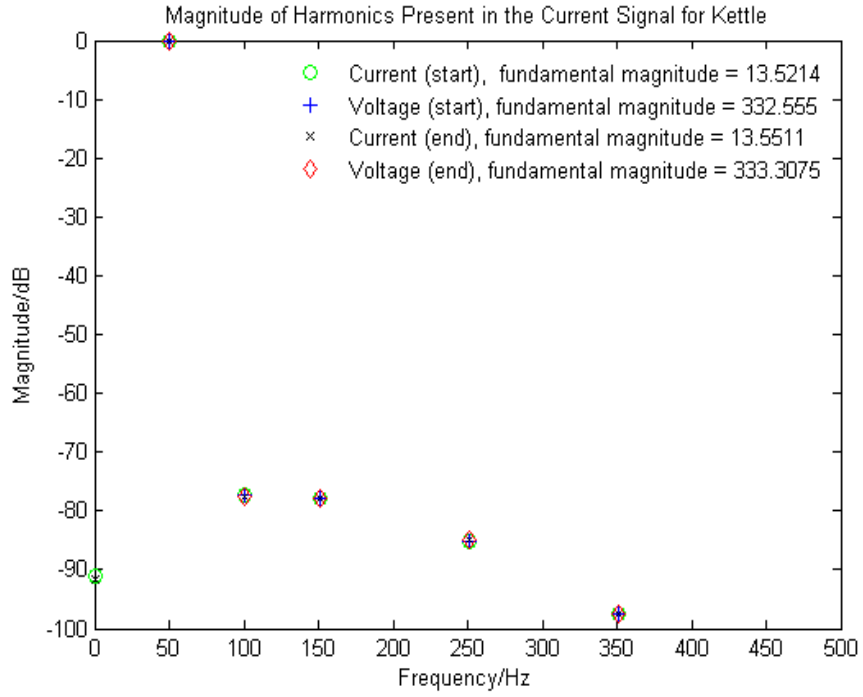


Figure 6.5: Plot of relative magnitude of harmonics (when compared to the fundamental) on a dB scale for both sets of voltage and current data.

magnitude of the fundamental current varies by less than 0.03A (less than 1%) between the two data sets. From this plot it can be seen that all harmonics are at least 75dB smaller than the fundamental and the harmonics are nearly all co-incident for the two current data sets. Any small differences in the current signal is due to the small change occurring in the voltage signal as seen by considering the harmonic at 400Hz.

One factor which has not been considered here is the effect of changing the load by increasing or decreasing the amount of water the kettle has to boil. It shall be assumed for the purposes of this thesis that load variation will just result in an increase in time taken to boil. This is a valid assumption as kettles are simple and cheap appliances and are therefore unlikely to have a mechanism for detecting the load and subsequently adjusting the electrical interface.

### 6.2.2 Fan heater

The second appliance chosen is a fan heater[4] used in the fan only mode. Figure 6.6 shows the plot of the voltage and current drawn by the fan. It can be seen that the current lags the voltage by about 45°. This is due to the inductive nature of the fan motor. Furthermore, the shape of the current is not similar to that of the voltage and seems to exhibit some peaking effect. This indicates that there is some significant amount of higher harmonic content present, probably due to saturating magnetic components.

Figure 6.7 plots the harmonic content of the signals. From this plot it can be seen that the magnitude of the fundamental differs by about 18% between the two sets of data. Furthermore, the magnitude of the 3rd harmonic is significant for both data sets as it is only about 30dB smaller than the fundamental. The relative magnitudes of the 2nd, 4th, 7th, 8th and 9th harmonics are not the same for the 2 data sets. This difference is not due to changes in the voltage but may be caused by changes in the relative values of the underlying electrical components due to a change in the motor speed.

[4]Consort fan heater, model number FHA2.

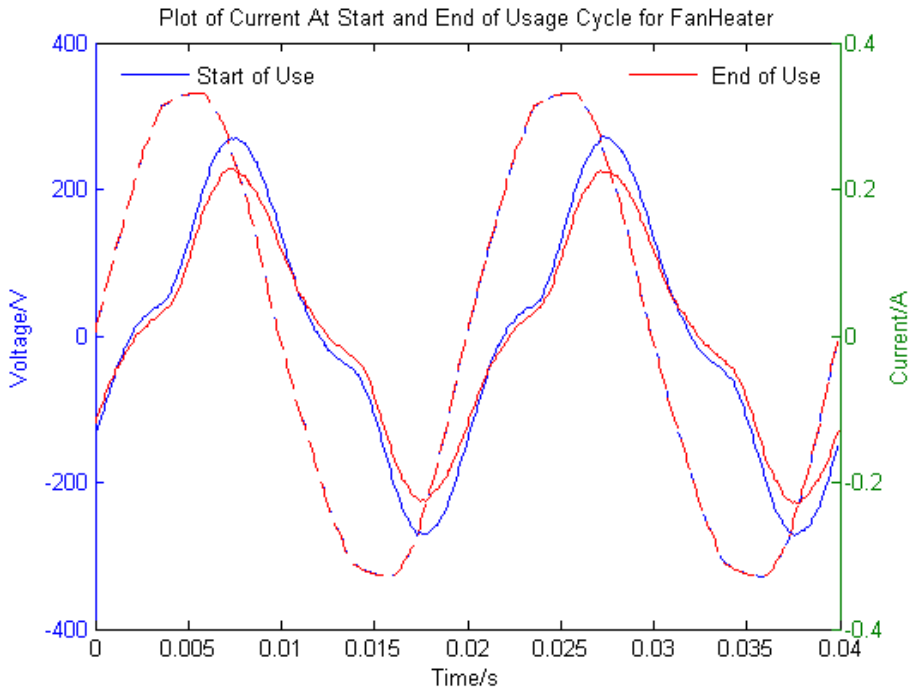


Figure 6.6: Plot of voltage and current signal drawn by a fan heater against time. The current curves are represented by solid lines and the voltage curves by dashed lines.

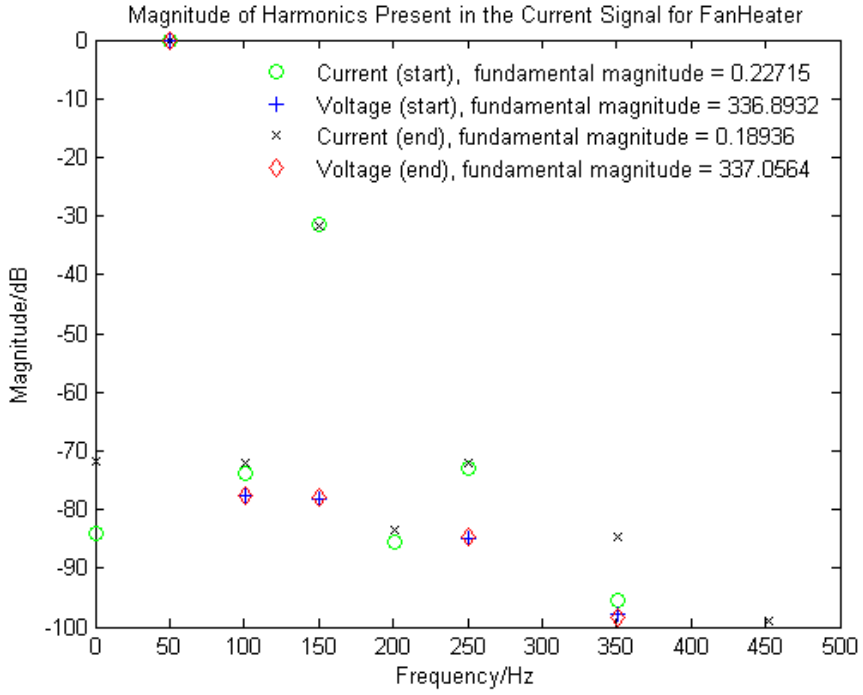


Figure 6.7: Plot of relative magnitude of harmonics (when compared to the fundamental) on a dB scale for both sets of voltage and current data.

### 6.2.3 Hairdryer

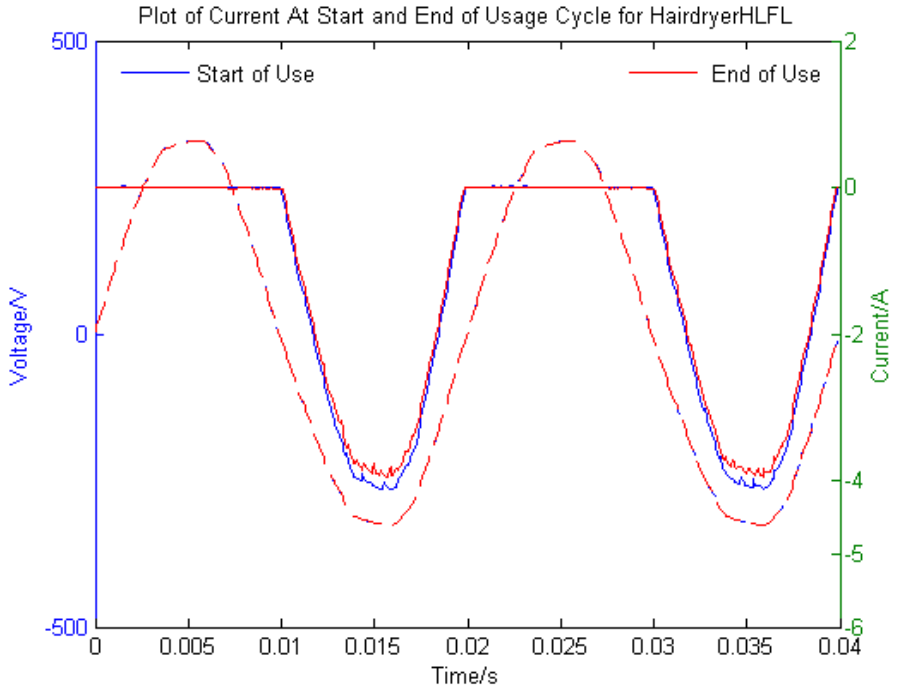


Figure 6.8: Plot of voltage and current signal drawn by a hairdryer against time. The current curves are represented by solid lines and the voltage curves by dashed lines.

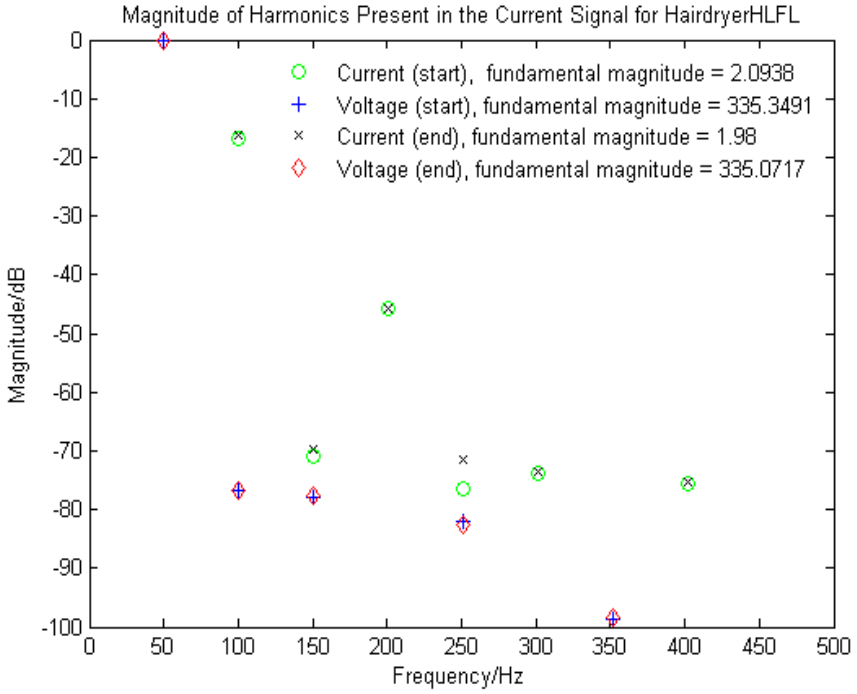


Figure 6.9: Plot of relative magnitude of harmonics (when compared to the fundamental) on a dB scale for both sets of voltage and current data.

The third appliance considered in this Chapter is a hairdryer[5]. Figure 6.8 shows a plot of the voltage and current drawn by the hairdryer for 2 cycles just after switch on and 2 cycles after 20 seconds. From this figure it can be seen that the current is only drawn in the negative half of the cycle. Therefore, there must be a controlled half wave rectifier in the electrical interface to prevent current in the positive direction. The current is in phase with the voltage and appears to be resistive. The current drawn towards the end of usage is smaller than that at the start of usage, but except for the magnitude, the waveforms are similar.

The similarity can be further explored by considering the harmonic content in Figure 6.9. The magnitude of the fundamental of the current varies by 5% between the two data sets. The magnitude of the direct current is substantial (and infact bigger than the fundamental) as the current curve only exists in the negative part of the cycle. The relative magnitudes of the higher harmonics are generally similar between the data sets which suggests that any variation is due to variation in the voltage.

### 6.2.4 CD player

The fourth appliance to be considered is a CD player[6]. The voltage and current drawn by this appliance is illustrated in Figure 6.10 for two sets of data taken at different times during the use of the appliance. From this plot it is seen that the shape of the current and voltage signals are very different to one another. By considering where the current curve crosses the 0V axis, it is seen that the fundamental of the current lags the voltage by about 45°.

For both data sets the shape of the current is similar for the positive and negative parts of the cycle, however there are some differences. The peak that occurs at 0.005s/90° (and also as 0.025s) is larger than the corresponding negative peak at 0.015s/270° (and

[5]Revlon hairdryer, model number 097X.
[6]Technics radio and cd player, model no SCEH790.

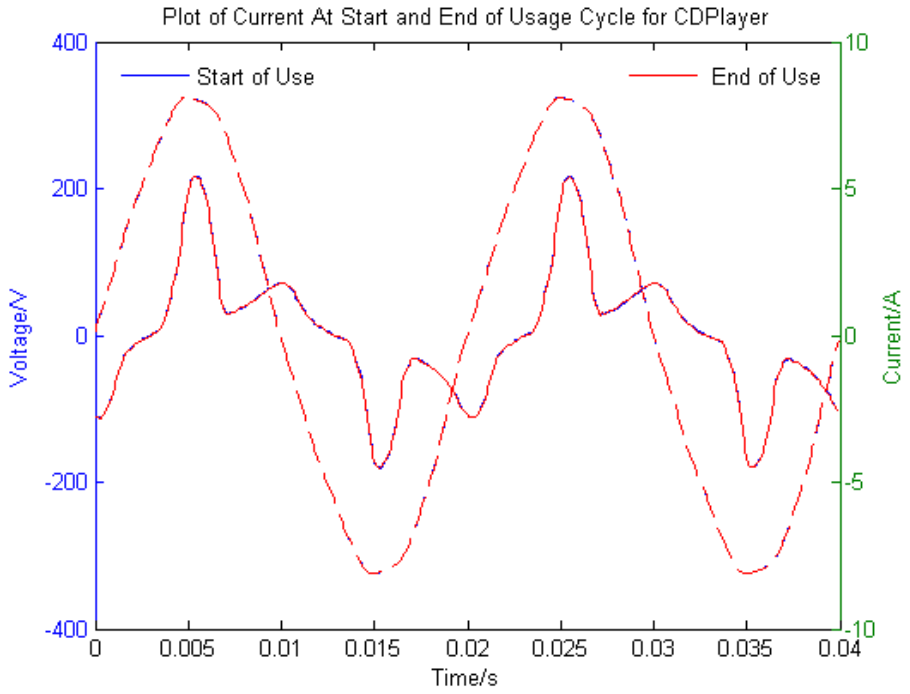


Figure 6.10: Plot of voltage and current signal drawn by a CD player against time. The current curves are represented by solid lines and the voltage curves by dashed lines.

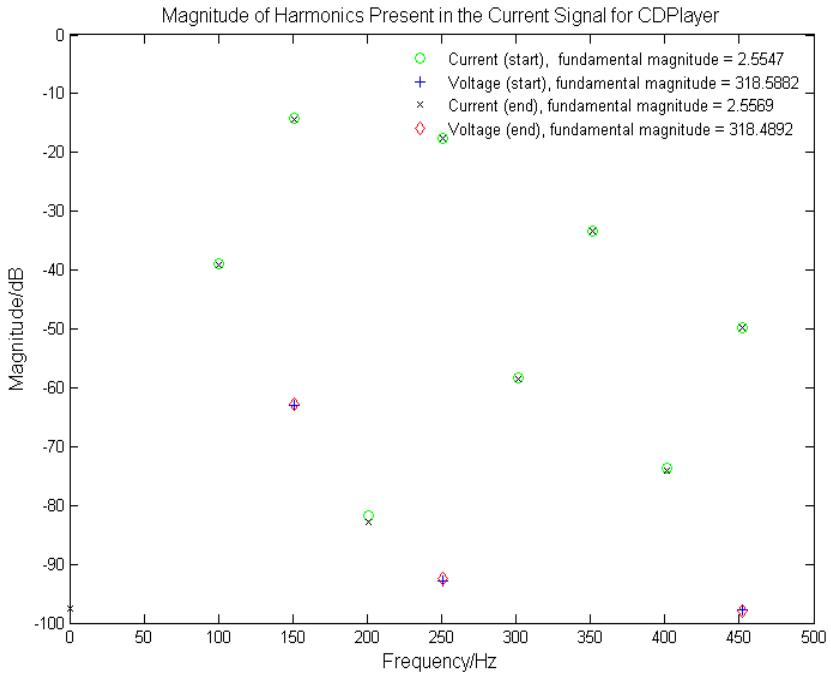


Figure 6.11: Plot of relative magnitude of harmonics (when compared to the fundamental) on a dB scale for both sets of voltage and current data.

0.035s). This peak is approximately co-incidental with the peak of the voltage curve. The current curve for this appliance also exhibits another smaller peak. This second peak occurs in the positive part of the cycle at 0.01s/180° (and 0.03s). This corresponds to where the voltage crosses the voltage axis from positive to negative. This peak is smaller in the positive part of the current cycle than its corresponding negative peak occurring at 0.02s/360° (and 0.04s).

The two data sets are similar to one another when observing a plot of the current against time and this similarity is mirrored by the plot of the relative magnitudes of the harmonics to the fundamental for both data sets. The magnitude of the fundamental differs by less than 0.1% between the two data sets and from Figure 6.11 it can be seen that data points illustrating the relative magnitudes of the various harmonics are co-incident for the two data sets even when the voltage harmonics vary. It can also be seen that the magnitudes of the harmonics up to 450Hz are significant in representing this particular data set. Some of the higher harmonics are only 10dB smaller in magnitude than the fundamental and the odd harmonics up to the 9th (450Hz) are all less than 50dB smaller than the fundamental.

### 6.2.5 TV

The final appliance considered in this chapter is a TV[7]. A plot of the voltage and current drawn by the TV is illustrated in Figure 6.12. The voltage is very similar between the two data sets, however the current curve varies considerably. Although the magnitude is quite similar between the two plots, the phase changes by about 10 degrees between the two curves. The peak of the current at the start of use occurs at about 0.0015 seconds (30°) after the peak of the voltage curve, whilst the peak of the current at the end of use occurs at about 0.002 seconds (40°) after the peak of the voltage. Because of the shape of the current it is likely that the electrical interface of the TV will contain an

[7]Toshiba TV, model number 26WLT66.

ad-dc converter and therefore a small change in the steady state voltage of the output may cause the observed phase changes.

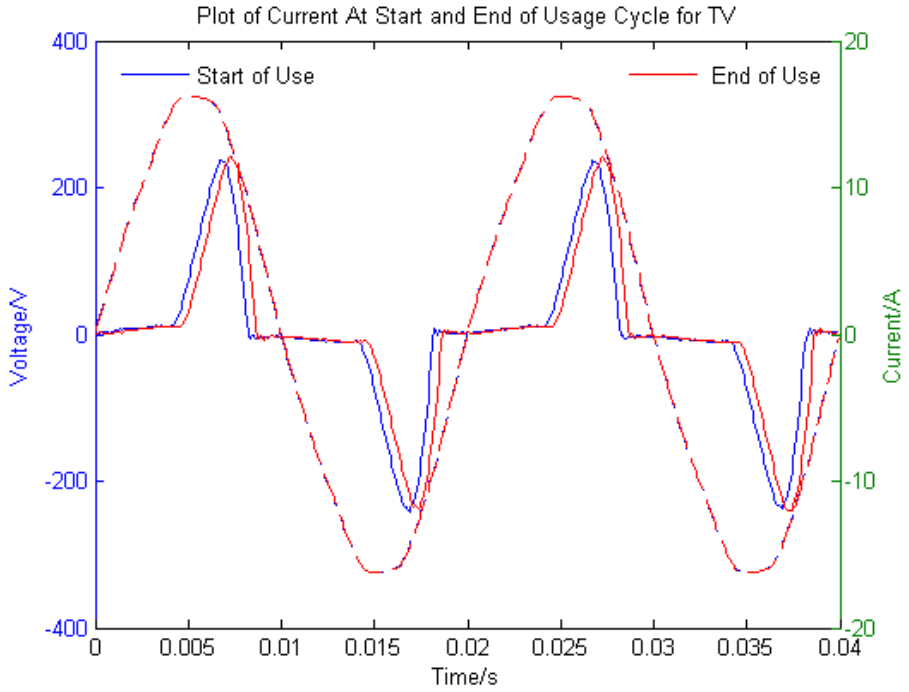


Figure 6.12: Plot of voltage and current signal drawn by a TV against time. The current curves are represented by solid lines and the voltage curves by dashed lines.

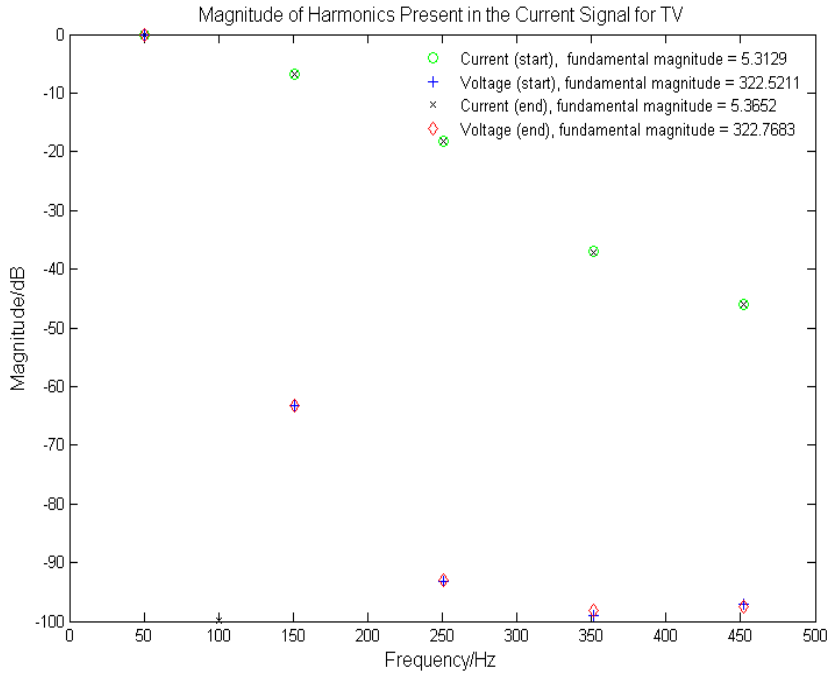


Figure 6.13: Plot of relative magnitude of harmonics (when compared to the fundamental) on a dB scale for both sets of voltage and current data.

More information can be obtained by considering the harmonic content of the signals in Figure 6.13. It can be seen that the magnitude of the fundamental of the current varies by less than 1% between the two sets of data. It can also be seen that the magnitudes of the odd harmonics are significant – the magnitude of the third harmonic is less than 10dB smaller than the fundamental and all odd harmonics up to the 9th are less than 50dB smaller than the fundamental. Furthermore, all the odd harmonics are co-incident in magnitude. The even harmonics are much smaller in magnitude than the odd harmonics and are close to the noise floor of the measuring equipment. Therefore, any variation in the magnitude of the even harmonics is negligible.

### 6.2.6 Issues highlighted by the current waveforms

This section considers particular features observed in the current drawn by the 5 appliances and examines the potential causes and consequences of variation of the signals over time. This allows these variations to be accounted for when considering a feature set with which to represent appliances.

By considering the set of current curves, three observations are made which are useful in distinguishing between appliances. The first observation is to do with the actual shape of the curve. Appliances which have a purely resistive front end, such as the kettle, have a current which has a shape similar to that of the voltage. Other appliances with more complex front ends, such as the CD player or TV, have current waveforms with more complex shapes. This is evident when considering the harmonic content. The kettle has little harmonic content for harmonics higher than the fundamental, whilst the CD player and TV have non-negligible harmonic content up to 450Hz. It can also be seen that the odd harmonics tend to be considerably greater in magnitude than the even harmonics.

The second observation is the difference between the positive and negative half cycles, as seen very clearly when considering the hairdryer. Again, this is reflected in the harmonic

content. There is considerable dc (direct current) content and the odd harmonics are larger than the even harmonics in this case.

The third observation is that the phase of the current with respect to the voltage varies. For the kettle the current is in phase with the voltage, whilst the fan exhibits current which lags the voltage by about 45 degrees.

Each of the appliances considered above show some variation in the voltage and current signals over time. The variation in the voltage waveform has already been considered in Section 6.1. There are three potential causes for changes in the current waveform.

The first cause is due to the changes in the voltage waveform. The current waveform is dependent on the voltage through appliance's electrical interface. Therefore, care must be taken when selecting features to represent the current waveform and thus the underlying appliance in order to ensure that any variation in the voltage is not reflected in the features.

The second cause for observing changes in the current signal over time is due to slow transients present in the data. This has already been illustrated in Figure 3.7 for the heating effect observed in resistors.

The third cause for changes in the current is due to a change in the operating conditions of a particular appliance, e.g. the effect of varying the load or temperature settings in a washing machine. This effect might not be seen during a single use of an appliance, but rather over multiple uses.

To overcome any changes in the current waveform due to causes 2 and 3 it should be ensured that that all operating points are considered when the AEM is trained to obtain the cluster prototypes for each appliance.

The observations and implications considered in this section are useful in establishing the feature set criteria to which selected features are compared.

## 6.3 Criteria for selecting a feature set

Recall that the objective of this chapter is to establish a feature set which allows each appliance to occupy a compact and disjointed region in feature space, ensuring that the selected features are compatible with the classification framework and clustering algorithm proposed in Chapters 4 and 5 respectively. Therefore the following criteria are established:

1. The features selected should allow data points from the same operating mode of the same appliance to lie co-incident in feature space, with data points from different appliance falling a great enough distance away in order to be able to differentiate between the appliance.

2. The features should be as independent as possible of any changes which may occur in the voltage signal.

3. In order that a subset of the features do not become dominant, the features should be approximately similar in magnitude.

4. In order that the features can be used to form a $d$-dimensional feature space, the features should have the same dimensions (i.e. angle's and magnitudes should not be compared within the same feature space).

5. To keep the time complexity of the clustering algorithm to a minimum, it should be ensure that the feature set contain fewer than approximately 20 features. This will also help reduce the memory requirements of the AEM.

For each method of feature extraction proposed, the selected features are considered with reference to the established criteria in order to examine their potential for use in the classification algorithm.

## 6.4 Overview of feature extraction

The following sections explore 3 known mechanisms of feature extraction. For each mechanism, an overview of the technique is considered with reference to the potential benefits of manipulating the raw data. The method of feature extraction is explained and the results are presented and examined with reference to the feature set criteria.

It should be recalled that the AEM provides instant feedback to the user about the total power consumption of the home. Since the AEM is already calculating the voltage, current and power, Section 6.5 considers features based on the power consumption of appliances. The features considered are the RMS[8] current, RMS voltage and the real, reactive and apparent power. However, because of the co-dependence of these variables, it is shown that for most appliances there are only 2 independent features and consequently there is a large scope for the violation of feature set criteria 1.

An extension to this feature set is proposed in Section 6.6, whereby features are produced by convolving the current waveform belonging to each appliance with a particular 'test' signal. By choosing appropriate test signals, linear and non-linear aspects of the current may be identified. However, this analysis is directly dependent on the shape of the current waveform and therefore changes in the voltage signal are carried through to the features, thus violating feature set criteria 2.

Section 6.7 proposes a novel method of feature extraction from the voltage and current signals to model the underlying electrical interface of appliances. This removes the dependence of the features on the voltage signal. There are limitations prevalent in this method as it is only considered for a finite range of electrical interfaces, although extension to a larger range is possible but non-trivial and beyond the scope of this thesis. However, even with the limitations, this method of feature extraction is selected for use

[8]RMS stands for Root Mean Square. The RMS current is the square root of the mean of the squares of the sequence that makes up the current waveform.

in the AEM and Section 6.8 proposes the innovative mechanism by which the feature extraction is performed so that it is usable in the existing classification framework.

## 6.5 Features relating to power consumption

All smartmeters already monitor the voltage and current signals of the home and use these to calculate the overall power consumption. Therefore, if the features are based on these signals, any existing smartmeter can be adapted slightly so that it is able to provide disaggregated feedback about individual appliance usage.

### 6.5.1 Method for extracting features

Features are based on the time domain current and voltage curves. These curves are monitored by the AEM and sampled at a rate of 12.8KHz. This corresponds to 256 samples per cycle. Each cycle is analysed independently in order to allow features corresponding to that particular cycle to be represented. This allows any underlying trends to be observed and accounted for.

By using the voltage and current signals to calculate the apparent, real and reactive powers, appliances are separated based on the magnitude of the current drawn, as well as the proportion of resistive and reactive electrical components contained in the electrical interface of the appliance.

Therefore, the following features are calculated for each cycle:

- RMS current, $i_{rms}$, and RMS voltage, $v_{rms}$. The RMS value of a signal, $r$, is calculated in the following way [59]:

$$r_{rms} = \sqrt{\frac{1}{N}\sum_{j=1}^{N} r_j^2} \tag{6.2}$$

  where $N$ is the number of points used to represent a single cycle of the signal in the time domain.

- Apparent Power, $S$, calculated as [60].

$$S = v_{rms} \times i_{rms} \tag{6.3}$$

- Real Power, $P$:

$$P = \frac{1}{N} \sum_{j=1}^{N} v_j i_j \tag{6.4}$$

- Reactive Power[9], $Q$:

$$Q = \sqrt{S^2 - P^2} \tag{6.5}$$

Of these features, there are only three independent variables. $Q$ is entirely dependent on the $P$ and $S$ and is therefore discarded as a feature. $S$ is further dependent on $v_{rms}$ and $i_{rms}$ and therefore one of these must be discarded as a feature. It is decided not to discard $v_{rms}$ because this is uniform across all appliances and therefore acts as a constant of proportionality between $S$ and $i_{rms}$. Therefore, $S$ is retained because it is more likely to satisfy feature set criteria 3 which requires that the features be similar in magnitude[10]. Of these 3 retained values, only the real and apparent power are used as features. This is because the voltage is independent of the appliance in use.

The feature values are calculated for the positive and negative half cycles separately, because, as illustrated in Section 6.2, the two half cycles may have entirely different parameters and a complete cycle representation will not be an accurate representation of either half cycle.

### 6.5.2 Appliance representation using the half cycle real and apparent powers

The voltage and current data points for a set of $N_c$ cycles are used to generate the feature vector in the following way.

[9]For the purposes of this thesis it is assumed that $Q$ includes the dispersive power associated with higher harmonics.

[10]In previous work Hart [30] uses the real and reactive power, however, because of the feature set criteria established, the apparent power is more suitable that the reactive and therefore is included in this work instead.

**Method**

- Take $N_c$ cycles of voltage and current data for each appliance.
- For each cycle, split the 256 data points into positive and negative half cycle.
- Use the 128 positive half cycle data points to calculate $v_{rms(+)}$ and $i_{rms(+)}$ and the 128 negative half cycle data points to calculate $v_{rms(-)}$ and $i_{rms(-)}$ according to Equation 6.2.
- Use the values for $v_{rms(+)}$, $i_{rms(+)}$, $v_{rms(-)}$ and $i_{rms(-)}$ to calculate the apparent power for the positive and negative half cycles according to Equation 6.3.
- For each cycle calculate the real power in the positive and negative half cycles (see Equation 6.4).
- Plot the apparent power against the real power for each half cycle and for each appliance.

**Results and analysis thereof**

The half cycle real and apparent power is calculated for 100 cycles worth of data for the 5 appliances and the results are plotted in Figures 6.14 and 6.15 for the positive and negative half cycles respectively.

From the results it can be seen that the 5 appliances do form compact and distinct clusters in feature space. Therefore, feature set criteria 1 is satisfied for the appliances considered in this chapter.

Feature set criteria 2 states that the features should be as independent as possible of any changes occurring in the voltage signal. For the features examined here, this is not true, as the power is directly dependent on the voltage. Some amount of this dependence may be removed by dividing by $v_{rms}^2$, although this will only remove variation in the

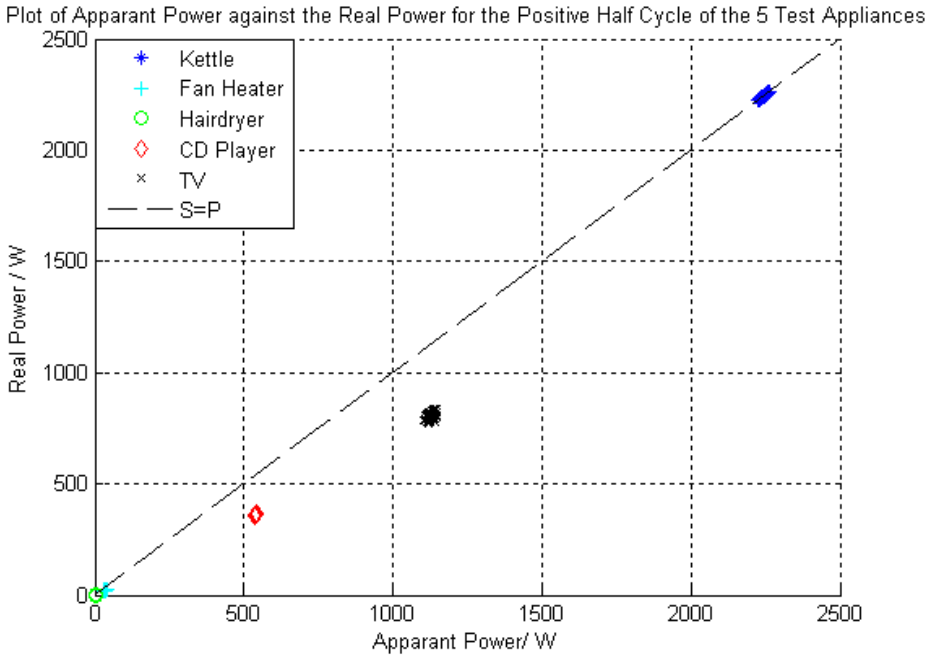


Figure 6.14: Plot of the RMS current against the power factor for the 5 appliances under consideration for the positive half cycle.

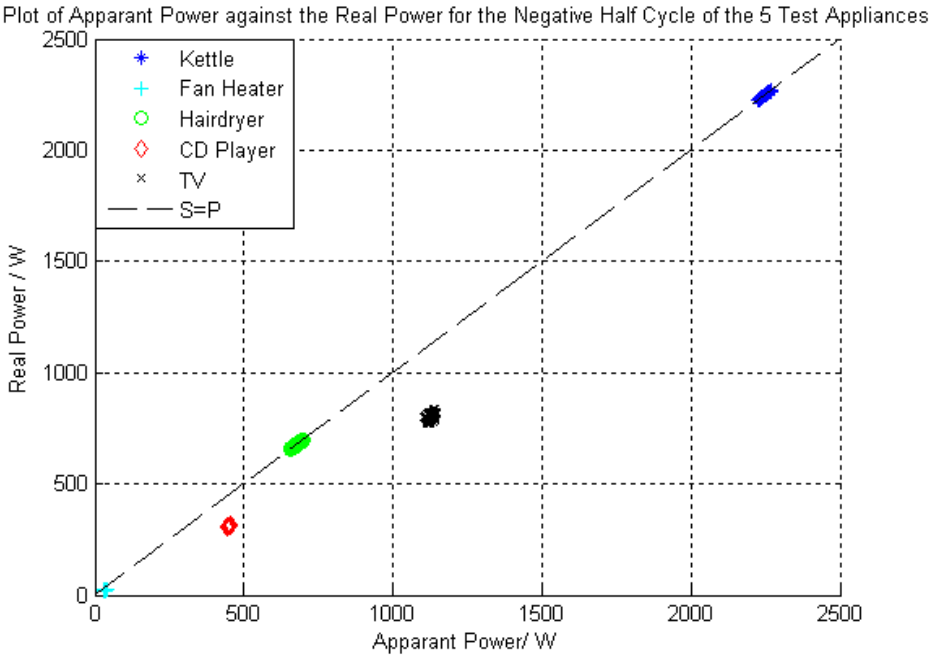


Figure 6.15: Plot of the RMS current against the power factor for the 5 appliances under consideration for the negative half cycle.

magnitude of the fundamental and not in the variation of the relative magnitudes of the harmonics.

From the results it can be seen that the values obtained for the real and apparent power are similar in magnitude and dimensions and therefore feature criteria 3 and 4 are satisfied.

The final feature criteria is that the feature set contain fewer than 20 features. This is satisfied as the feature set contains only 4 features. However, for appliances which have similar positive and negative half cycles (in this case all the appliance aside from the hairdryer), the effective number of features is reduced to 2. This is because the values obtained for the positive half cycle and negative half cycle features will be the same. As more appliances are added to the system there is the possibility that different appliances will overlap in feature space as this space is limited in its dimensionality thus violating feature set criteria 1. Therefore a larger feature set is desirable.

## 6.6 Identifying linear and non-linear content in the current

Further features are generated from the current by identifying particular patterns within the waveform. This is done by convolving the current with a particular 'test' waveform. The current curve is denoted by $x$ and the test signal by $y$. For $N$ pairs of observations of the two variables, the sample correlation coefficient is defined as [61]:

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{6.6}$$

The value of $r$ obtained ($-1 \leq r \leq 1$) reflects to what extent the current is similar to the test signal. The higher in magnitude the value of $r$, the greater the similarity. A negative value indicates a 180 degree phase shift. Therefore, by choosing test signals which reflect the way in which typical electrical components may draw current, the amount that the appliance resembles any particular component can be observed.

### 6.6.1 Identifying linear components

Linear components including resistors, inductors and capacitors all have sinusoidal currents. The difference between these components is the phase between the current and the voltage waveforms. The current through a resistor is in phase with the voltage across it, whilst the current through an inductor lags the voltage by 90° and the current through a capacitor leads the voltage across it by 90°. Therefore, the amount to which a particular signal resembles a resistor, inductor or capacitor can be identified by convolving the current with a sinusoid. This type of analysis is similar to a Fourier representation.

### 6.6.2 Identifying non-linear components

By considering the current waveform obtained for the fan heater, CD player and TV in Figures 6.6, 6.10 and 6.12 respectively, it can be seen that these appliances contain non-linear peaks in their currents. Therefore, in order to identify these non-linear components, the test signal in the convolution should resemble the shape of the underlying non-linear components.

Two typical non-linear components are transformers (in ac to ac converters) and bridge rectifiers feeding capacitors (in ac-dc converters). Transformers exhibit non-linear behavior due to saturation of the B-H curve of the iron core, resulting in a current curve as illustrated in Figure 6.16. This curve is generated based on the assumption that the properties of the transformer core may be modelled as $\mathbf{B} \propto \tan^{-1}(\mathbf{H})$.

Figure 6.17 illustrates the current curve which may be produced when a large rectifying capacitor is connected to an alternating current voltage source. This model is generated by assuming that below a critical voltage no current flows (due to diodes in the circuit) and above this voltage a current flows through a purely resistive circuit which is attached in parallel with the rectifying capacitor.

Therefore the size and location of non-linear peaks may be identified by selecting a test

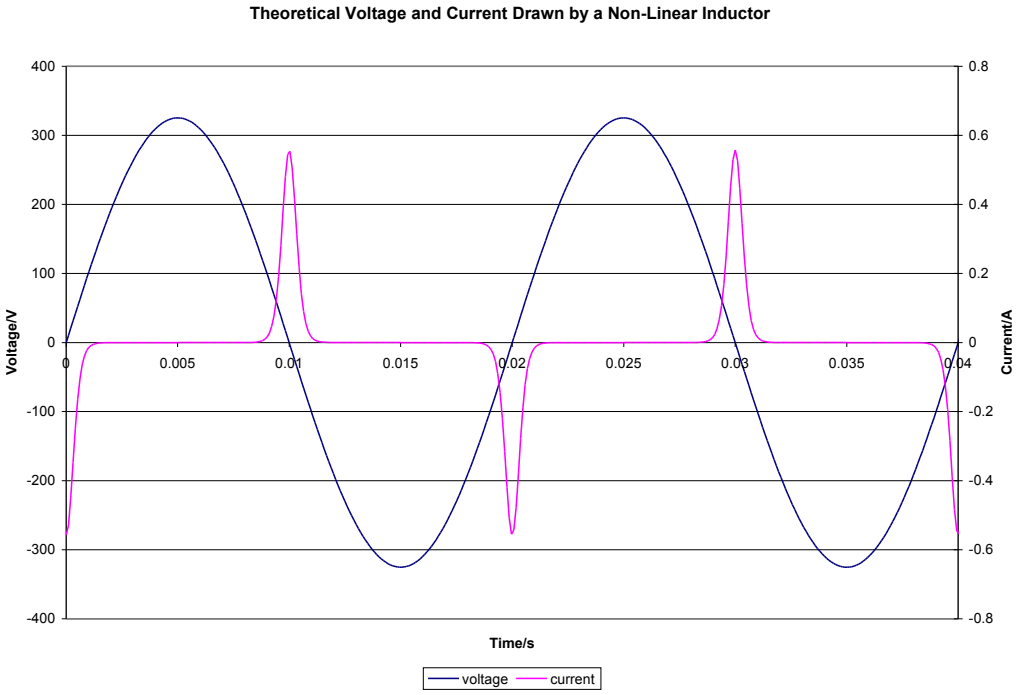


Figure 6.16: Modelled non-linear current drawn by a transformer due to saturation of the iron core.

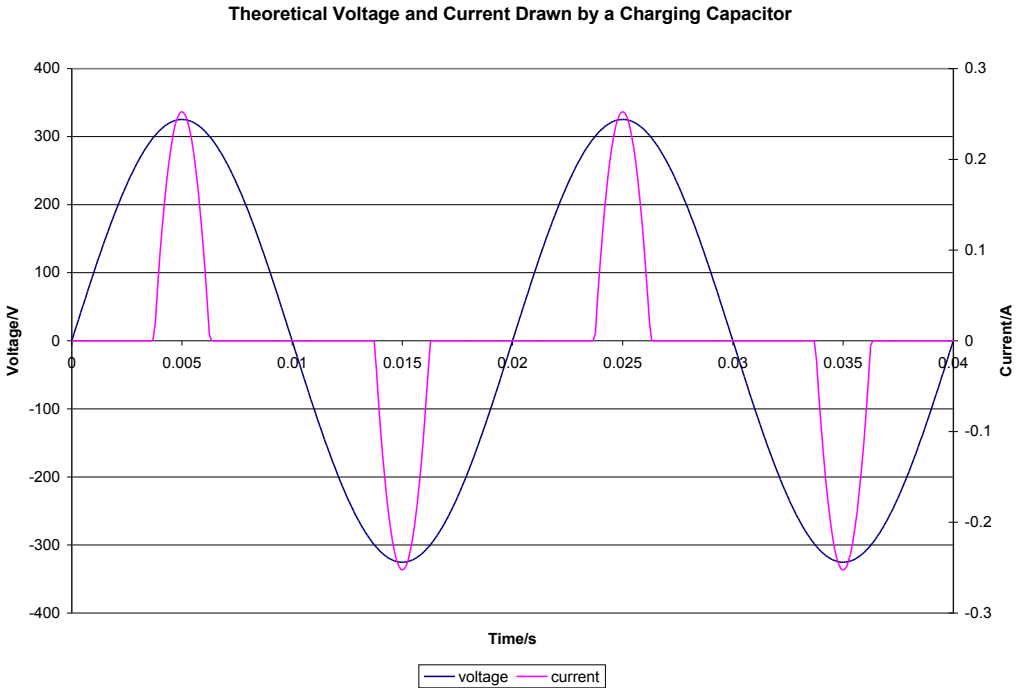


Figure 6.17: Modelled current drawn by an ac-dc or dc-dc converter due to non-linear capacitive effects.

signal which has a similar shape to that of the current curves illustrated in Figures 6.16 and 6.17. This type of analysis is similar to a Wavelet Transform Analysis.

### 6.6.3 Application of the analysis to real data

#### Method

1. Generate a set of test curves which are used to determine the amount of a particular behaviour contained in the current drawn by appliances. For the purposes of this thesis 4 test waveforms are used to observe both the linear and non-linear behaviour of the current. The linear curves include a sinusoid and co-sinusoid and these model the amount or linear resistive, capacitive and inductive behaviour. The non-linear curves include those illustrated in Figures 6.16 and 6.17.
2. Extract one cycle from the current being monitored.
3. Perform the correlation according to Equation 6.6 to generate a correlation coefficient for each of the 4 test signals.
4. Repeat steps 2 and 3 to generate a set of correlation coefficients for a particular current signal over time.
5. Plot the correlation coefficients on a set of graphs.

#### Results and Analysis Thereof

For each appliance 50 cycles of the current over time were used in the algorithm. The results are displayed for the linear correlations in Figure 6.18 and for the non-linear correlations in Figure 6.19. The linear correlation was also performed for the 50 corresponding voltage cycles to observe the effect of voltage variation over time on the features and is illustrated in Figure 6.20.

From Figure 6.18 it can be seen that the kettle and the hairdryer have large linear

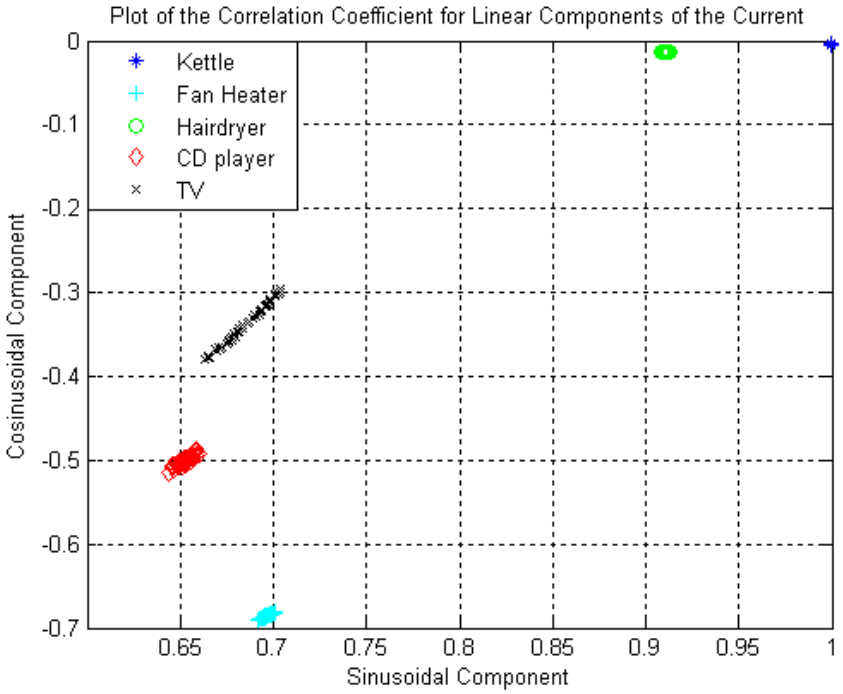


Figure 6.18: Set of linear correlation coefficients for the current.

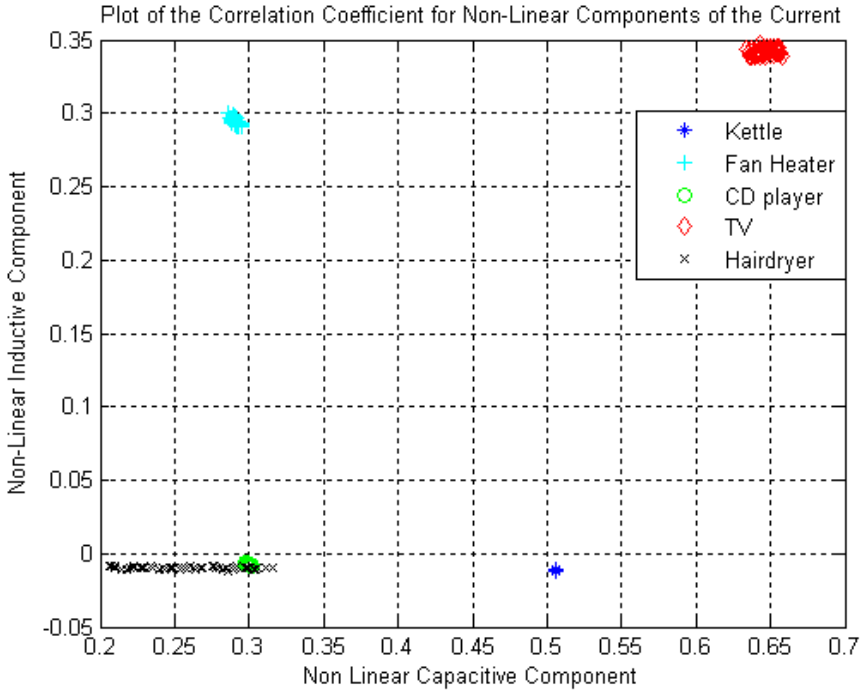


Figure 6.19: Set of non-linear correlation coefficients for the current.

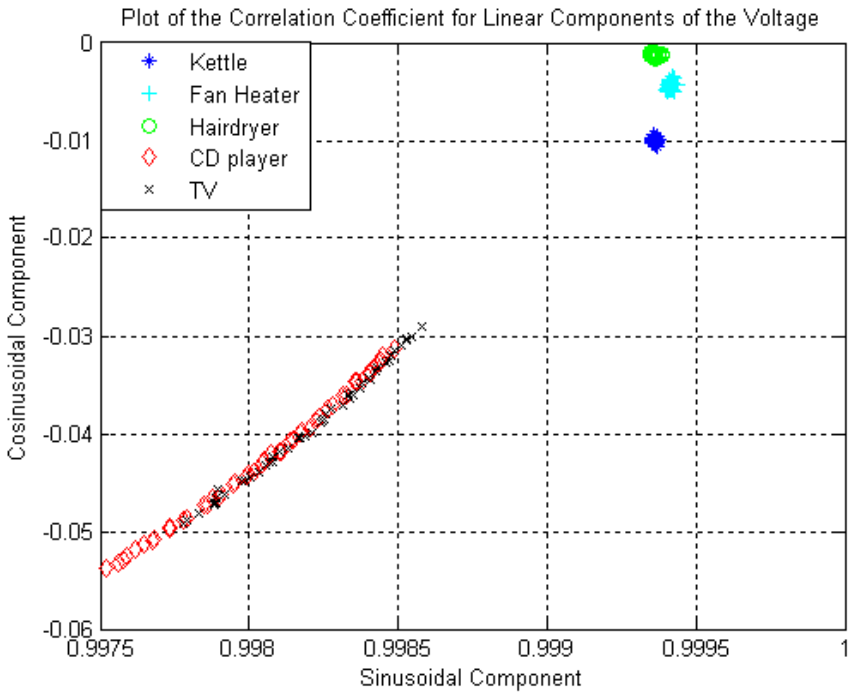


Figure 6.20: Set of linear correlation coefficients for the voltage.

sinusoidal components, and insignificant co-sinusoidal components. This is because these two appliances exhibit linear waveforms which are in phase with the voltage. The reason that the sinusoidal correlation coefficient belonging to the hairdryer is smaller than that of the kettle is due to the lack of current in the positive half cycle for the hairdryer.

The fan heater, TV and CD player all have significant linear correlation coefficients. This indicates the presence of some fundamental content in the current drawn by these appliances, although from Figures 6.6, 6.10 and 6.12 it can be seen that these appliances are clearly not linear. The relative magnitude of the correlation coefficients allows the phase (when compared to the voltage with positive numbers corresponding to a leading current) of the fundamental to be deduced as 45° for the fan heater, 37° for the CD player and 25° for the TV.

It can be seen that whilst the feature presented in Figure 6.18 do fall in different regions of feature space, the clusters into which they fall are by no means compact. This invalidates feature set criteria 1. Furthermore, upon examination of Figure 6.20 which shows the

linear correlation coefficients of the voltage, it becomes evident that the elongation of the clusters in the voltage correlation coefficients are reflected in the current correlation coefficients, which invalidates feature set criteria 2. This dependency may be removed[11] by normalising the features by the magnitude of the fundamental of the voltage, thus solving the two problems discussed in this paragraph, but there are further problems with this feature set which may be seen by considering the non-linear correlation coefficients.

By considering the current drawn by the various appliances, it is predicted that the fan heater should have high non-linear inductive correlation coefficients, and the CD player and TV should have a significant non-linear capacitive correlation coefficients. The kettle and hairdryer are linear appliances, and therefore should not exhibit high non-linear correlation coefficients, however, it can be seen from Figure 6.19 that the kettle has a non-linear capacitive component of 0.5 and the hairdryer of 0.3. They both have negligible non-linear inductive components as the current drawn by these appliances is in phase with the voltage. The reason that these appliances appear to exhibit non-linearity is due to the overlap of positive areas and overlap of negative areas (with no positive/negative overlap) occurring between the current signals and the non-linear capacitive test wave. The correlation coefficient between an ideal sinusoid and this test wave is 0.51, and therefore there are problems when attempting to use this particular test wave to model non-linear behaviour.

More problems with this method are explained by considering the correlation coefficients generated for the TV. From Figure 6.12 it can be seen that the shape of the current drawn by the TV is very similar to that of the non-linear capacitive test wave, however the correlation coefficient between these two data sets is relatively small. This is because the location and width of the test wave and the current signal are not the same, and consequently the identification of the behaviour is missed. Furthermore, the location of

[11] Assuming that the variation in the voltage is due solely to variations in the fundamental, and that the relative magnitude of the harmonics remain constant.

the peak in the current drawn by the TV shifts over time. This gives rise to the elongation of the cluster generated for the TV, as illustrated in Figure 6.19. Therefore, the non-linear capacitive correlation coefficients do not satisfy feature set criteria 1, because the clusters formed when they are included in the analysis are not compact and disjointed.

A solution for this would be to use a Wavelet Packet Transform (WPT) analysis [62] to account for the varying location of the peak with time, however, this type of analysis produces a large number of features as the convolution between the sample and test wave is performed on multiple scales and locations of the test wave along the sample wave. WPT analysis was investigated in great detail but, because the features generated do not satisfy the criteria and because of space constraints, it is not presented in this thesis.

Furthermore, whilst the dependency of the features on the voltage variations may be largely removed for the linear analysis, this is not possible for non-linear appliances and therefore the feature set examined in this section will continue to violate feature set criteria 2. Whilst the other feature set criteria may be satisfied, the two most important are not and convolution techniques are not explored in further detail.

## 6.7 Modelling appliance parameters

This section considers a feature extraction mechanism which attempts to remove any dependency of the features on the voltage signal by modelling the parameters of an appliance's electrical interface. In general it is not possible to remove the voltage variation due to the fact that different electrical interfaces react to the variation in different ways. However, if the electrical interface is assumed, a set of equations may be developed which govern the way in which that particular set of electrical components draw current when a particular voltage is applied. Therefore the voltage variation is accounted for and removed for that electrical interface only.

This thesis considers 3 different types of electrical interface. The list of electrical interfaces

modelled in this section is by no means exhaustive, but extension of this analysis beyond those models considered here is non-trivial and beyond the scope of this thesis. The interfaces which are modelled are illustrated in Figure 6.21. Models (a) and (b) consider a large range of linear devices as $Z$ may model an inductor, $Z = j\omega L$, or a capacitor, $Z = 1/j\omega C$. This covers electrical appliances such as lights, motors whose inductors do not saturate (those which do saturate are complicated models and beyond the scope of this thesis) and those containing heating elements including fan heaters and hairdryers. It should be noted that under a purely sinusoidal voltage source these two models are identical. However, the actual voltage is not ideal and does contain higher harmonics, as seen in Figure 6.2, and consequently both these models are included in the analysis.

Model (c) illustrates a simple full bridge rectifier. Although the resistive, inductive and capacitive components in this model are all linear, because of the presence of diodes, the mode of operation is non-linear. This model may do well at representing appliances which contain ac-dc converters, for example, a TV with a cathode ray tube.

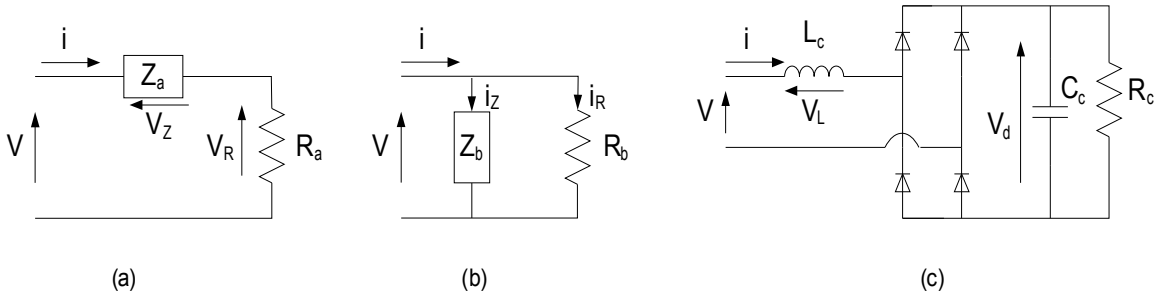


Figure 6.21: Three possible electrical interfaces modelled in this thesis.

This section proposes a feature set which corresponds to the appliance parameters of the model which most accurately represents the underlying appliance. If it is assumed that the selected model does accurately represent the appliance, then the features selected are independent of the voltage and will therefore satisfy feature set criteria 2.

Mathematical models are developed in order to determine the component parameters by considering the circuit analysis for each of the proposed models and constructing simultaneous equations to solve. This analysis is based on Kirchoff's current and voltage

laws [63] which state:

- The sum of the currents into a junction is zero, i.e. $\sum_{k=1}^{n} I_k = 0$ where $n$ is the total number of currents flowing into that point.
- The sum of the voltages in any circuit loop is zero, i.e. $\sum_{k=1}^{n} V_k = 0$ where $n$ is the total number of individual voltages (e.g. sources or impedances) present in the loop.

### 6.7.1 Determining the parameters for models (a) and (b)

Kirchoff's laws may be applied to circuits (a) and (b) by considering the harmonic content. Circuit (c) must be considered separately as harmonic content analysis is not appropriate due to the presence of the diodes. For circuit (a) the analysis proceeds as:

$$v = v_R + v_Z \tag{6.7}$$

$$v_R = i \times R_a \tag{6.8}$$

$$v_Z = i \times Z_a \tag{6.9}$$

Substituting these values for $v_R$ and $v_Z$ into Equation 6.7:

$$v = i \times (R_a + Z_a) \tag{6.10}$$

The same analysis can be done for circuit $b$) to produce circuit equation 6.11:

$$v = i \times \left( \frac{Z_b R_b}{R_b + Z_b} \right) \tag{6.11}$$

Once these equations are known, the parameters $R$ and $Z$ for models (a) and (b), along with the suitability of each model, is determined in the following way.

Models (a) and (b) assume linear currents and therefore the total harmonic distortion (THD) of the current must be assessed to determine whether these models are appropriate. The THD is calculated according to [64]:

$$\text{THD} = \frac{\sqrt{(I^2 - I_1^2)}}{I_1} \tag{6.12}$$

where $I^2$ is the power in the current signal and $I_1^2$ is the power contained in the fundamental. By considering the relative harmonic content of the appliances tested in this chapter, it can be seen that those appliances which are non-linear have harmonics with relative magnitude of around 20dB smaller than the fundamental. Therefore, it is assumed that if the relative harmonic content is greater than -20dB that the appliance is non linear. This corresponds to a THD of 0.1.

If the appliance is deemed linear, i.e. THD$<$ 0.1, Equations 6.10 and 6.11 are used to determine the magnitude of $R$ and $Z$. These equations contain two unknowns and therefore 2 simultaneous equations are required for a solution. These two simultaneous equations are obtained by considering the sinusoidal and co-sinusoidal components of the fundamental of the voltage and current separately[12]. For model (a) this is:

$$\begin{aligned} v_{1(\sin)} &= i_{1(\sin)}R_a - i_{1(\cos)}Z_a \\ v_{1(\cos)} &= i_{1(\sin)}Z_a + i_{1(\cos)}R_a \end{aligned} \tag{6.13}$$

where the subscript $_1$ denotes the fundamental and the subscripts $_{\sin}$ and $_{\cos}$ denote the sinusoidal and co-sinusoidal component of the signals respectively. For model (b) the simulations equations produced are:

$$\begin{aligned} v_{1(\sin)} &= \tfrac{Z_b R_b}{Z_b + R_b}\left(i_{1(\sin)}Z_b - i_{1(\cos)}R_b\right) \\ v_{1(\cos)} &= \tfrac{Z_b R_b}{Z_b + R_b}\left(i_{1(\sin)}R_b + i_{1(\cos)}Z_b\right) \end{aligned} \tag{6.14}$$

If a positive value is obtained for $Z$ then the imaginary impedance corresponds to an inductor, whereby $|Z| = \omega L$. However, if the value returned is negative, this indicates the presence of a capacitive impedance in which case $|Z| = 1/\omega C$. Therefore, the features returned consist of the parameter values in both half cycles (i.e. $R_{pos}$, $R_{neg}$, $Z_{pos}$ and $Z_{neg}$ ).

[12]The sinusoidal and co-sinusoidal components may be isolated from the signal by performing an FFT on the signal and deducing the respective components from the real and imaginary amplitudes returned.

### 6.7.2 Determining the parameters for model (c)

If the THD returned is larger than 0.1, indicating that neither models (a) nor (b) are appropriate, the parameters for model (c) are calculated. The equations for model (c) are somewhat more complex than for models (a) and (b) and must be considered in the two regimes of operation. The first regime is when the diodes are conducting and the second regime is when they are not. These two regimes are illustrated in Figure 6.22. Harmonic analysis may not be used in this scenario as the voltage across the resistance and capacitance is not known for the entire cycle. Therefore, an innovative analysis, which deduces the parameter values by considering particular aspects of the current waveform, is proposed as follows.

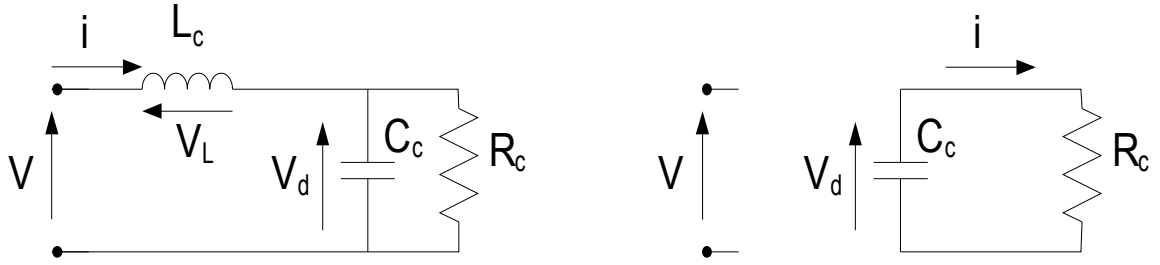


Figure 6.22: Illustration of the two modes of operation of model (c). The regime in which the diodes conduct is illustrated on the left and the regime in which they are turned off is on the right.

The voltage $v_d$ is not constant, however it is known at the point at which the diodes switch on, i.e. at $t = 0$, as $v = v_d$. If $v_d$ is assumed to be constant just after switch on the equation $v_L = v - v_d$ may be solved at time $t = \tau$ where $\tau < 10^{-4}$s. This assumption is reasonable due to the time constant introduced by the resistor and capacitor connected in parallel. Therefore, the magnitude of the inductance may be estimated as:

$$L \approx \frac{v_{L(t=\tau)}}{(di/dt)_{t=\tau}} \tag{6.15}$$

Once the inductance is established, the voltage across the inductor for the remainder of the period during which the diode is conducting is calculated as $v_L = L{}^{di}/_{dt}$ and therefore the voltage across the capacitor and resistor is estimated whilst the diodes are

conducting.

The value $R_c$ is obtained by considering the power consumption of the appliance. This power is due to the voltage across the resistor, $R_c$, and although this voltage does vary throughout the cycle, it is estimated as $v_{t=\tau}$. Therefore:

$$R_c = \frac{(v_{t=\tau})^2}{\frac{1}{\mathrm{N}} \sum_{k=1}^{N} v_k i_k} \tag{6.16}$$

Finally, the value of the capacitance is obtained by considering the current consumption of the capacitor whilst the diodes are conducting according to the following equations.

$$i_c = i - {}^{v_d}/_{R_c} \tag{6.17}$$

$$C_c = \frac{Q}{V} = \frac{\int_{t=t_1}^{t=t_2} i_c.dt}{v_{d(t=t_2)} - v_{d(t=t_1)}} \tag{6.18}$$

The feature vector returned for this model consists of the resistance, inductance and capacitance values, $R$, $L$ and $C$.

## 6.8 Application of the analysis to real data

This section proposes a mechanism by which the analysis presented in Section 6.7 may be used in order to determine a set of features for each appliance. The feature set should contain an indication of the correct model and the corresponding parameters.

**Method**

1. For each appliance take $N_c$ cycles worth of data. Consider each cycle separately.
2. For each cycle calculate the THD in the positive and negative half cycle, assuming that the half cycle RMS current is greater than 0.1A. Note - if the half cycle current is less than 0.1A then set the corresponding half cycle parameters to zero in the rest of the analysis.

3. If the THD in both half cycles is less than 0.1 then assume that the appliance is linear and calculate the model parameters for models (a) and (b). If the THD is greater than 0.1 the appliance is non-linear and the model (c) parameters are calculated.

4. Determine the suitability of each model proposed. This is done by considering the error, $\varepsilon$, between the current observed and the current predicted using the observed voltage and the model parameters obtained, where, for models (a) and (b),

$$\varepsilon = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(i_{(k)} - i_{predict(k)}\right)^2} \Big/ i_{RMS} \tag{6.19}$$

and for model (c),

$$\varepsilon = \frac{\left|\int_{t=\tau_{on}}^{t=\tau_{off}} i_{(t)}.dt - \int_{t=\tau_{on}}^{t=\tau_{off}} i_{predict(t)}.dt\right|}{\int_{t=\tau_{on}}^{t=\tau_{off}} i_{(t)}.dt} \tag{6.20}$$

where $\tau_{on}$ is the moment at which the diodes begin conducting, and $\tau_{off}$ is when the diodes switch off and no further current is drawn. This happens twice per cycle, as seen in Figure 6.12, so both half cycle errors should contribute to the total error.

5. Consider the error. If $\varepsilon \leq 0.1$, select the model with the smallest error and features corresponding to the appropriate component parameters.

6. If $\varepsilon > 0.1$ then none of the models proposed in this chapter are suitable. This scenario may occur for the appliances considered in this thesis due to the limited range of electrical interface models proposed, however, as the number of models expands, the chance of this happening reduces. However, for the purposes of this thesis, when this situation does occur, use a generic feature set containing the normalised real and reactive power of the first, third and fifth current harmonics[13]. Although this feature set does not satisfy all the feature set criteria (see Section 6.6) it prevents the feature vector remaining empty for appliances who electrical interface do not match the proposed models.

[13]The harmonic content is obtained by performing the FFT Matlab function on the current signal. This is then normalised by the ratio of 230V to the magnitude of the fundamental of the voltage signal.

**Results and analysis thereof**

The above analysis is performed for the kettle, fan, hairdryer, CD player and TV. 50 cycles are examined for each appliance, the model type (if applicable) of each cycle is determined and the corresponding model parameters are calculated. For each appliance, the number of cycles which fall under each model type is listed Table 6.1. Models A and B consider the positive half cycle (PHC) and negative half cycle (NHC) separately whilst model C considers the total cycle (TC).

Number of cycles of data belonging to each model

| Appliance | Model A | | Model B | | Model C | Comment |
|---|---|---|---|---|---|---|
| | PHC | NHC | PHC | NHC | TC | |
| Kettle | 0 | 0 | 50 | 50 | 0 | - |
| Fan | 0 | 0 | 0 | 0 | 0 | THD>0.1, (c) not suitable |
| Hairdryer | 0 | 33 | 0 | 17 | 0 | (a) and (b) suitable |
| CD Player | 0 | 0 | 0 | 0 | 0 | THD>0.1, (c) not suitable |
| TV | 0 | 0 | 0 | 0 | 40 | $\varepsilon > 0.1$ in 10 data cycles |

Table 6.1: Number of cycles of data (out of 50 for each appliance) assigned to each model type.

From Table 6.1 it can be seen that the kettle, hairdryer and TV may be represented by the proposed models A, B or C, though the fan and the CD player are not covered by any of the model types. This is because the current drawn by the fan and CD player appliances is non-linear (therefore neither model A nor B are appropriate) and does not correspond to a rectifier (therefore model C is not appropriate). In this thesis models for the type of electrical interfaces assumed by the fan and CD player are not considered. They both contain non-linear saturating inductors and extension of circuit analysis to cover this type of model is non-trivial and beyond the scope of this research.

The kettle is linear in both positive and negative half cycles and every cycle considered is best represented by model B. The hairdryer is also linear in the negative half cycle (no current flows in the positive half due to a half bridge rectifier) and may be represented by both models A and B. As seen in Table 6.1, model A is more accurate 66% of the time and model B 34% of the time.

The TV is non linear and is represented best by model C. By considering the electrical interface of the TV is it known that model C should produce an accurate representation of the current drawn. However, the model is only accurate to within 10% 40 times out of 50. Figure 6.23 shows the rectified predicted current compared to the rectified actual current when Model C is deemed a good match, and Figure 6.24 shows the comparison when the error is greater than 10%. From these figures it can be seen that the general shape of the curve is correct in both cases, but the ability of the model to make an accurate prediction is dependent on the parameter values returned.

The reason that the parameters obtained during circuit analysis are inaccurate at modelling the current 20% of the time is because the parameters are based on the derivative of the current (see Section 6.7.2). The derivative of the current is calculated by considering the changes in the value of the current between consecutive sampled data points. This paragraph considers the potential for error in the derivative signal by considering the equipment used to monitor the current. The current is monitored using an 12 bit ADC (analogue to digital converter). The most significant bit is used for sign determination and this reduces the effective number of bits to 11. The least significant bit is not reliably accurate and therefore there is a potential error of up to approximately 0.1% of the maximum value measured. This error is denoted by $\varepsilon_{\max}$.

In order to obtain the maximum possible error in the derivative curve over the entire cycle, it is assumed that the first data point is smaller than it should be by $\varepsilon_{\max}$, the next current value is larger than it should be $\varepsilon_{\max}$ and so on. Whilst this is unlikely to occur, it places the bounds on the maximum possible error, averaged over a cycle, that may be observed in the derivative curve. By varying the error in the current as described above, every point in the derivative curve will have an error of $2\varepsilon_{\max}$. This error may be represented as a percentage of the RMS value of the derivative curve and this maximum potential error is found to be as high as 87% if the current is assumed to be sinusoidal.

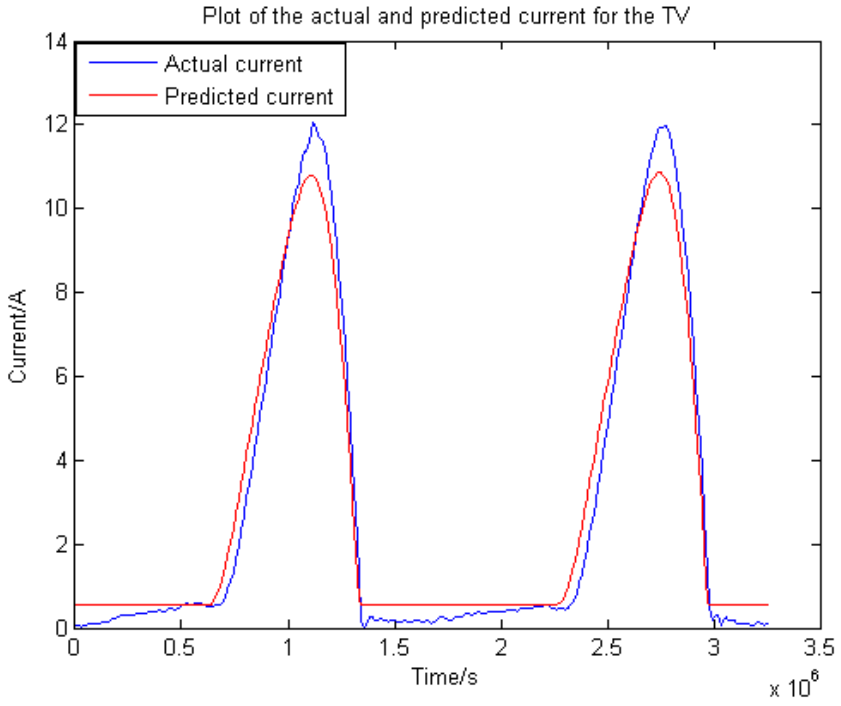


Figure 6.23: Plot of the rectified real and predicted currents drawn by the TV under model (c) when the prediction is accurate.

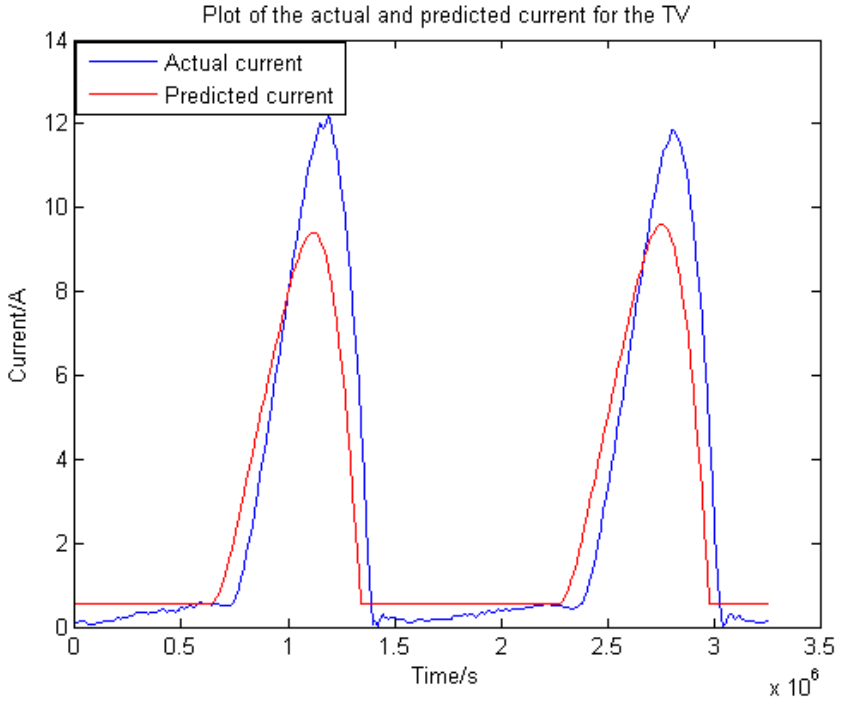


Figure 6.24: Plot of the rectified real and predicted currents drawn by the TV under model (c) when the prediction is inaccurate.

Therefore, in order to improve the performance of model C so that it generates parameters which more accurately model the TV a greater percentage of the time, a larger ADC is required. By using a 16 bit ADC the average maximum potential error in the derivative curve is reduced to 5%.

Further improvements will be made to the results obtained for model C by increasing the sampling rate of the ADC. In the experiments described in this thesis, the sampling rate is limited to 256 samples per cycle due to restraints on the equipment available. However, due to the assumptions on which Equation 6.15 is based, a faster sampling rate will enable a more accurate pinpoint of the moment at which the diodes switch on and a more accurate representation of the inductance. It should be noted, however, that this may increase the error in the derivative curve and must be considered with respect to the accuracy of the ADC.

Figures 6.25 to 6.28 show a plot of the features returned for each of the appliances in each of the applicable models. From these plots the extent to which the proposed feature set meets the feature set criteria is assessed. From these figures it can be seen that most of the appliances lie in different feature spaces due to their different electrical interfaces. This improves the ability of the features to differentiate between appliances with different electrical construction.

Feature set criteria 1 states that data points from the same appliance should lie co-incident in feature space, with data points from different appliances falling a great distance away. From the figures it can be seen that data points from different appliances do lie far away from each other, and in many cases fall in separate feature spaces.

From Figures 6.25 to 6.27 it can be seen that the data points from the same appliance do lie co-incident along the $R$ axis, but have a larger spread in general along the $Z$ axis. This may be explained by considering the relative magnitudes of the observed real and imaginary impedances. In Model A where the real and imaginary impedances are in

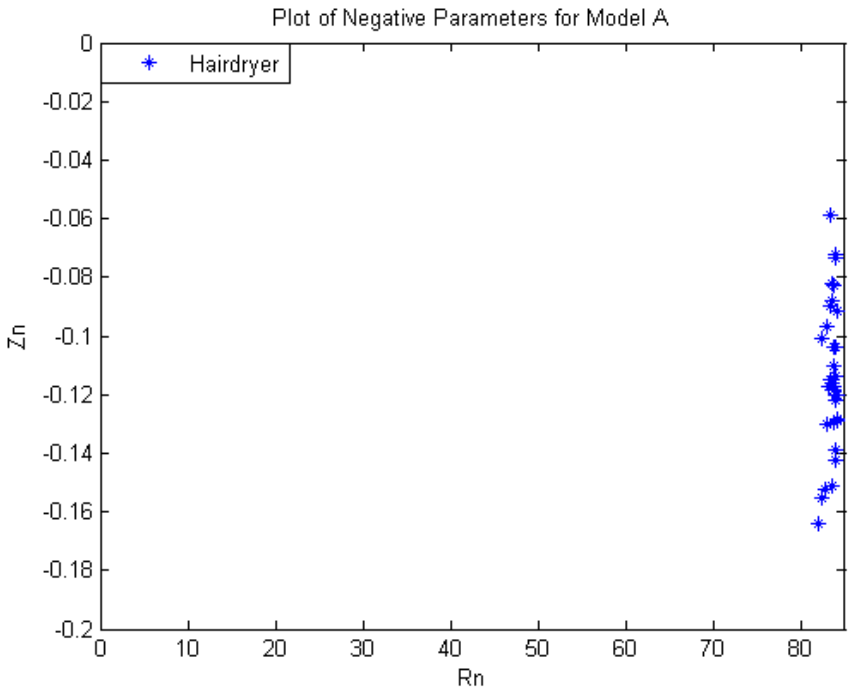


Figure 6.25: Plot of the negative parameters for model A.

series, the magnitude of the real impedance in the hairdryer is over 500 times bigger than the magnitude of the imaginary impedance and therefore the imaginary impedance is negligible and most likely to be caused by errors in measurement or stray capacitance.

In Model B, where the real and imaginary impedances are connected in parallel, the ratio of the imaginary impedance to the real impedance is over 500 for the hairdryer and over 100 for the kettle. Therefore, these imaginary impedances may be likened to an open circuit and again present due to measurement errors. In both these scenarios is may be more applicable to model the kettle and the hairdryer as pure resistances. Therefore, under the scenario that the data belongs to model A whereby $R/Z \geq 50$, or to model B whereby $Z/R \geq 50$, both models are deemed inappropriate and the data is actually assigned to model R, which has 2 features; $R_{pos}$ and $R_{neg}$. These are calculated with respect to the power consumption of the appliance, i.e. $R = V_{rms}^2/P_{rms}$. A value of 50 is chosen because this represents the scenario in which the phase measurement error is maximum, i.e. when there is some appearance of phase difference between the voltage and current signals due only to quantization errors.

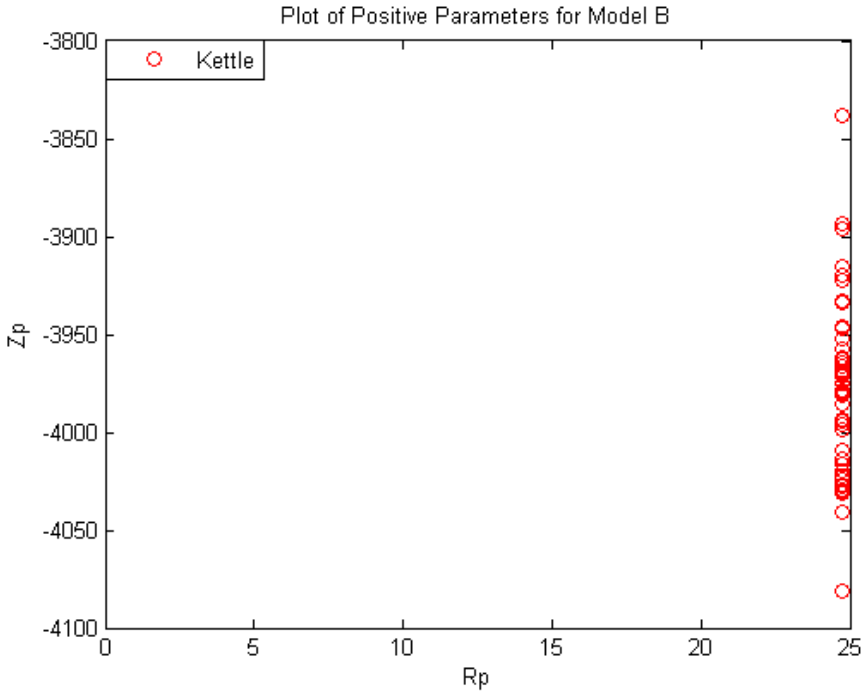


Figure 6.26: Plot of the positive parameters for model B.

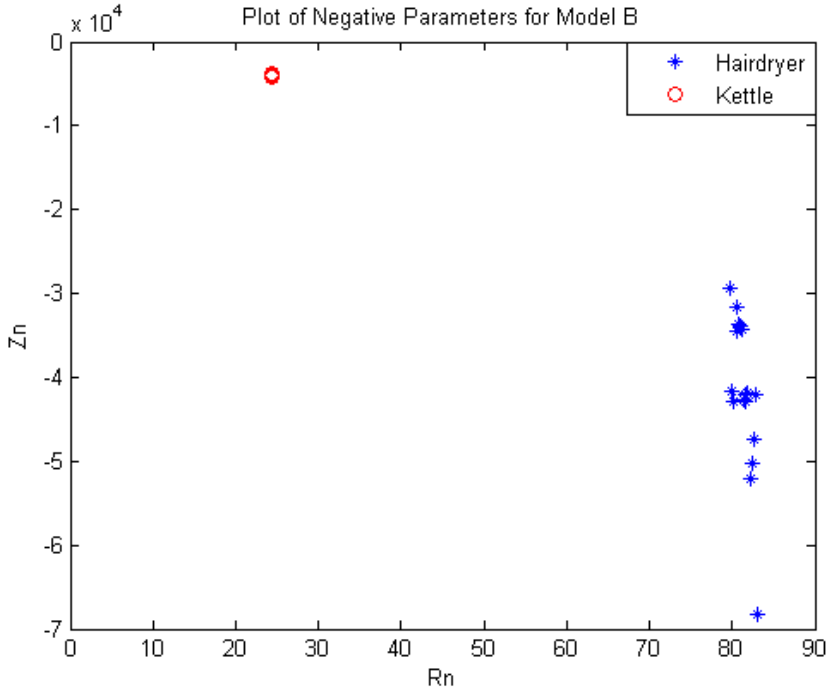


Figure 6.27: Plot of the negative parameters for model B.

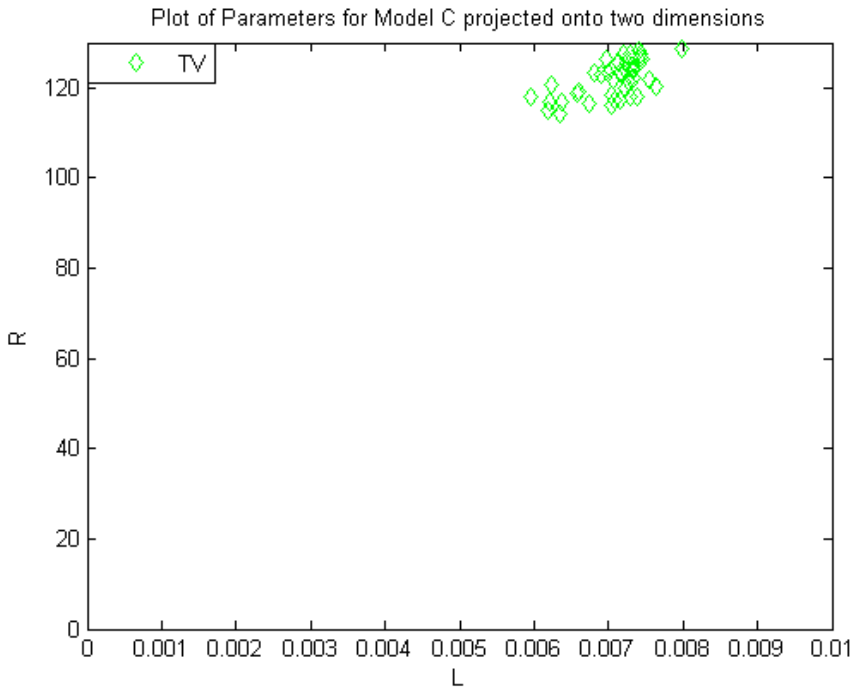


Figure 6.28: Plot of the parameters for model C.

Figure 6.29 shows a plot of Model R feature space for the kettle and hairdryer. It is seen in this figure that the hairdryer appears to no resistance in the positive half cycle, i.e. $R_{pos} = 0$ for all data points. This is caused by the lack of current in the positive half cycle, which results in an infinite resistance. Because an infinite resistance is impossible to plot, is it plotted instead as exactly zero resistance.

From Figure 6.29 it can be seen that the data points from the same appliance do lie near enough co-incident in feature space so that the clusters generated do form compact and disjointed clusters. This satisfies feature set criteria 1. Furthermore, by modelling these appliances as resistances, the potential problems caused by having the same appliance fall under two different models (the hairdryer is modelled sometimes under model A and other times under model B) is removed[14].

Feature set criteria 2 states that the features should be as independent as possible of any changes in the voltage. Because the features are extracted by considering the underling

[14]It should be noted that the sole reason that the hairdryer does fall under two different models is because the observed capacitance is due to measurement errors and stray capacitance and therefore not an indication as to the underlying electrical construction of the appliance.

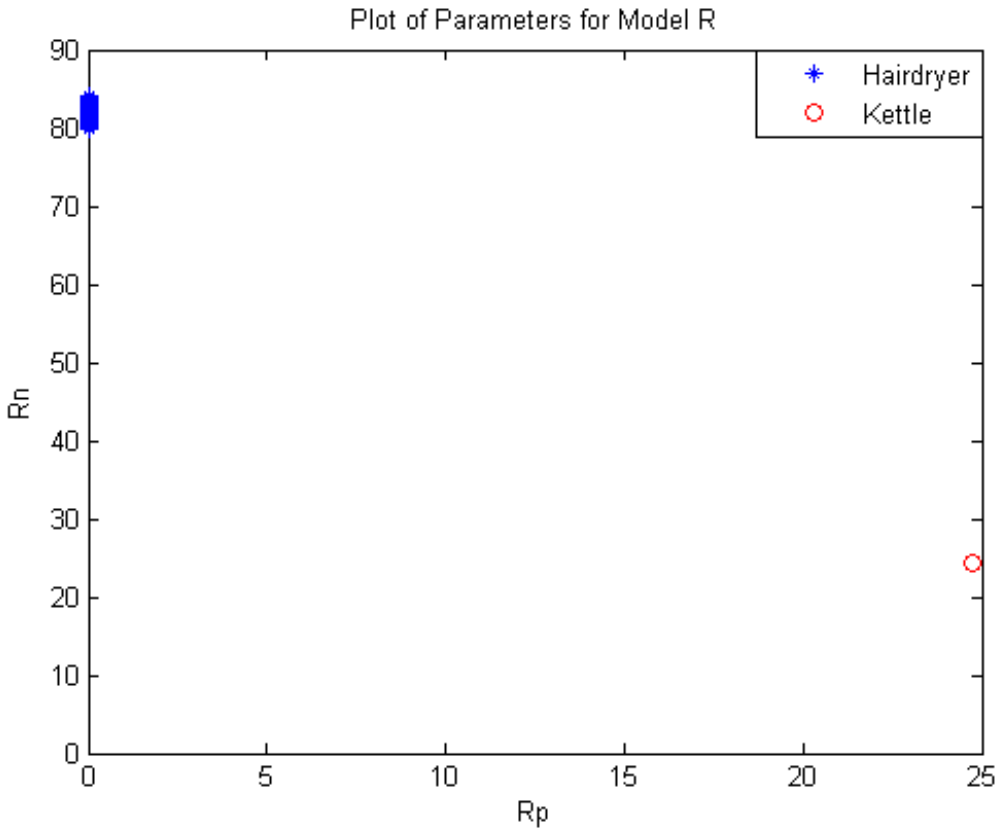


Figure 6.29: Plot of the parameters for model R.

electrical construction, if the selected model is accurate, any changes occurring in the voltage signal are accounted for in the features and therefore the features are indeed independent of these changes. If the model is not accurate, then the changes in the voltage will not be removed when calculating the features and they will be carried through. However, because the model is only applied in the case when the error between the modelled current and the observed current is less than 10%, any inaccuracies in the model should be small and therefore any voltage dependencies should be minimal, thus satisfying feature set criteria 2.

The third feature set criteria states that the magnitudes of the features should be approximately equal. In models A and B the maximum of the real to imaginary impedances is 50. If the ratio exceeds this then the data is assigned to model R. In model R the ratio between the resistance in the positive half cycle and negative half cycle is determined by

the appliance data, however, the limiting ratio is 1000[15]. In model C the ratio of the real to imaginary impedances is 50 for the inductive impedance and 1000 for the capacitive impedance. Therefore, the largest observed magnitude difference is in the region of 1000, which is easily accountable for if a 16 bit ADC is used (where the ratio of the MSB to the LSB for signed numbers is 32768). Therefore, the third feature set criteria is satisfied.

The forth feature set criteria stipulates that the features must be of the same dimension in order that they may be plotted in the same feature set criteria. All the features considered in this section are impedances and therefore this feature set criteria is satisfied.

The final feature set criteria is that the feature set must contain 20 or fewer features. The feature set proposed in this section contains one feature corresponding to the model type (which is not actually used in the clustering algorithm) and a further set of features corresponding to the appropriate appliance parameters. For models A and B this may be up to 4 features and for model C there will be 3 appliance parameter features. Therefore, this final feature set criteria is satisfied for the models proposed in this section. Whilst an extension to further models is deemed necessary to allow the proposed feature set to identify a greater range of appliances, by keeping the models relatively simple and limiting the number of electrical components, this feature set criteria should continue to be satisfied.

Because all 5 feature set criteria are satisfied, the method of feature extraction proposed in this section is selected for use in the FEB of the AEM system design. Whilst it is acknowledged that the range of appliances which may be identified is constrained due to the limited number of models developed in this thesis, this feature set is still proposed due to it suitability with regards to the feature set criteria. It is also acknowledged that the range of models proposed may be extended to cover a greater spectrum of appliances, however, this is non-trivial and beyond the scope of this thesis. Furthermore, without

[15]If the ratio is greater than 1000 then the current drawn in one half cycle is less than 0.1% of the current drawn in the other half cycle and therefore cannot be measured due to the resolution of the 12 bit ADC.

knowing the actual appliances connected to the system, this method of analysis enables the power consumption of particular types of appliances (i.e. an 'untrained' system might know that a heating type appliance is drawing power whilst the 'trained' system would recognise this as a particular kettle known to be used in the home) to be determined.

## 6.9 Chapter summary

The aim of this chapter was to identify the feature set $\mathbf{x}$, extracted from the observed voltage and delta current signals when an appliance changes state. This chapter set about identifying the feature set $\mathbf{x}$ by first considering the nature of the signals and highlighting any observed issues in order to establish a set of criteria to which the features should adhere.

The objective of the feature set is to represent appliances so that they are separated into compact and disjointed clusters in the feature space identified. To ensure that the features do form compact clusters they should be independent of changes occurring in the voltage signal. They should also be of the same dimensions so they may be represented in the same feature space and of similar magnitude to avoid over-domination of any one feature. These objectives formed an array of 5 feature set criteria to which any features proposed are measured.

Initially a set of features based on a power analysis was proposed. The apparent, real and reactive powers are calculated from the waveforms monitored and used as features in $\mathbf{x}$. Of these features only two are independent and therefore the features retained in the feature set correspond to the real and apparent power in the positive and negative half cycles respectively. The features are based to some extent on the underlying electrical construction, as the real and imaginary impedances are both represented and therefore the data points do form compact and distinct clusters. However, the features are not independent of changes in the voltage signal and furthermore, because the feature set is

so limited (especially when the negative and positive half cycles are the same), there is a high possibility that when a greater number of appliances are considered, the clusters formed may overlap in feature space.

In order to enable an expansion of the feature set to improve the ability of the features to distinguish between appliances, a method of feature extraction which allows linear and non-linear components to be identified from the current was proposed. This method of feature extraction proposes a convolution of the current with different test signals. Various linear and non-linear components are identified with careful consideration of the test signals proposed, with linear convolution analogous to a Fourier analysis and non-linear convolution analogous to a Wavelet Transform Analysis. However, this method of feature extraction results in features which are entirely dependent on the shape of the current, which varies with any voltage variation and therefore the features are not independent of the voltage. Because of this, convolution is not discussed in greater detail here.

To overcome the voltage dependency problems of the previous features, an innovative feature set is proposed which attempts to model the underlying electrical interface of the various appliances by identifying particular aspects of the current drawn by the appliance. The features in this set correspond to the identified model which best represents each appliance (assuming that the representation is reasonably accurate) and the corresponding components contained in the electrical interface. These features satisfy all the feature set criteria for the set of models proposed in this chapter and whilst some appliances remain unidentified, this is due to the limited range of models proposed.

Having identified a suitable feature set, a novel mechanism of feature extraction was proposed. This mechanism first determines the linearity of the underlying appliance and then establishes the parameter values for the set of models which may be appropriate. Once the parameter values have been determined, they are used to generate a set of

expected currents, which are compared to the observed currents in order to determine the accuracy of the model. The most accurate model (assuming that the accuracy is above a certain threshold) is then used to represent that appliance.

Therefore, this chapter has satisfied its aim in identifying a feature set $\mathbf{x}$ which may be used to represent appliances in a manner which satisfies all the established feature set criteria.

# Chapter 7

# A Worked Example

This chapter uses the theory and ideas presented in Chapters 4, 5 and 6 and applies it to a real example in order to demonstrate the ability of the proposed classification scheme to correctly identify appliances based on their observed current and voltage waveforms.

Before the AEM may be used to identify the energy consumption of appliances, it must first be trained to know the appliances in use in the home[1]. Therefore, Section 7.1 begins by considering the mechanism of data collection to train the AEM.

The data collected is analysed separately for each appliance and used to determine the model type and corresponding features which may be used to represent each appliance considered in this chapter. The model type of each cycle is first determined, as outlined in Section 7.2.1. Each model type has a different set of parameters which are used to represent it and therefore each model type is considered separately. For each model type, the corresponding appliance parameters (which form the feature set) are calculated and from the set of features generated, the number of modes of operation of the appliance is determined, as described in Section 7.2.2.

Each mode of operation is characterised by a set of cluster parameters, which include the cluster centre (already determined in the previous part of the analysis), the covariance

[1]There is the possibility of extending the analysis proposed in this thesis to incorporate an AEM which does not require manual training, see Chapter 3 for more details, however, this is beyond the scope of this research and not considered here.

matrix, average intracluster distance and cluster label. The covariance matrix varies with the number of data points used to calculate it and Section 7.2.3 describes how the number of required data points is determined. Once this value is determined, the cluster prototypes are evaluated and stored to memory, as described in Section 7.2.4.

Having established a set of cluster prototypes for all the appliances considered, Section 7.3 explores the mechanism by which a set of membership values are generated for a new set of data points. These membership values represent the level to which each new data point belongs to each of the known clusters, in order that they may be used to identify the appliance to which the new data point belongs. The results are compared to the known identity of the data points and this enables the accuracy of the techniques developed in this research to be determined.

In Section 7.4 confidence levels are generated for each of the classifications made in the previous section. Therefore, this chapter illustrates the workings of the AEM with reference to the short term time domain features.

## 7.1 Generating training data

This section examines the set of training data which is used to train the AEM. The data comes from a total of 4 different appliances but covers 9 user controlled modes of operation. For the rest of this chapter it is assumed that there are 9 separate appliances, as each user controlled mode may be operated separately. These appliances include a kettle[2], a fan heater[3] in 3 modes (fan only, heat 1 and heat 2), a soldering iron[4] and a hairdryer[5] in 4 modes (heat-low-fan-low, heat-low-fan-high, heat-high-fan-low and heat-high-fan-high).

The data collected corresponds to the voltage and current signals observed when each

[2]Tesco value cordless kettle, model number JK04W.
[3]Consort fan heater, model number FHA2.
[4]Weller soldering iron, model number WTCP 50.
[5]Revlon hairdryer, model number 097X.

appliance draws power from the supply. These signals are sampled using an 12 bit ADC at a rate of 256 samples per cycle. The number of cycles collected (see Table 7.2) are sufficient to cover all operating modes of each appliance (see Section 5.2 in Chapter 5). For each cycle of data sampling is begun as the voltage crosses the zero point with a positive gradient. Three cycles of the current for each appliance are illustrated in Figures 7.1 to 7.9. The data collected is then used to generate cluster representations for each appliance.

From Figures 7.1 to 7.9 it can be seen that the kettle, the fan heater in heat modes 1 and 2 and the soldering iron have currents which are sinusoidal with similar parameters in both the positive and negative half cycles. This implies that these appliances are linear. Of these 4, the kettle and the fan heater in both modes are in phase with the voltage and are therefore mainly resistive, whilst the soldering iron exhibits a significant current phase lag indicating the presence of an inductor.

The hairdryer in all 4 modes has a current in phase with the voltage and therefore indicates the presence of a resistive electrical interface. However, there are differences in the magnitude of the current drawn in the positive and negative half cycles which is particularly prevalent in the low-heat-low-fan mode in which no current at all is drawn in the positive half cycle. This implies a linear resistive behaviour within a non-linear interface.

The only truly non-linear appliance considered in this chapter is the fan heater in fan only mode. The current is similar to that drawn by an electric motor with a saturating inductor (indicative of the cheap 2 pole motors found in this type of apparatus).

## 7.2 Generating cluster prototypes

Having collected a set of training data for each appliance, this data is then used to determine a set of cluster prototypes. Cluster prototype generation is performed for each

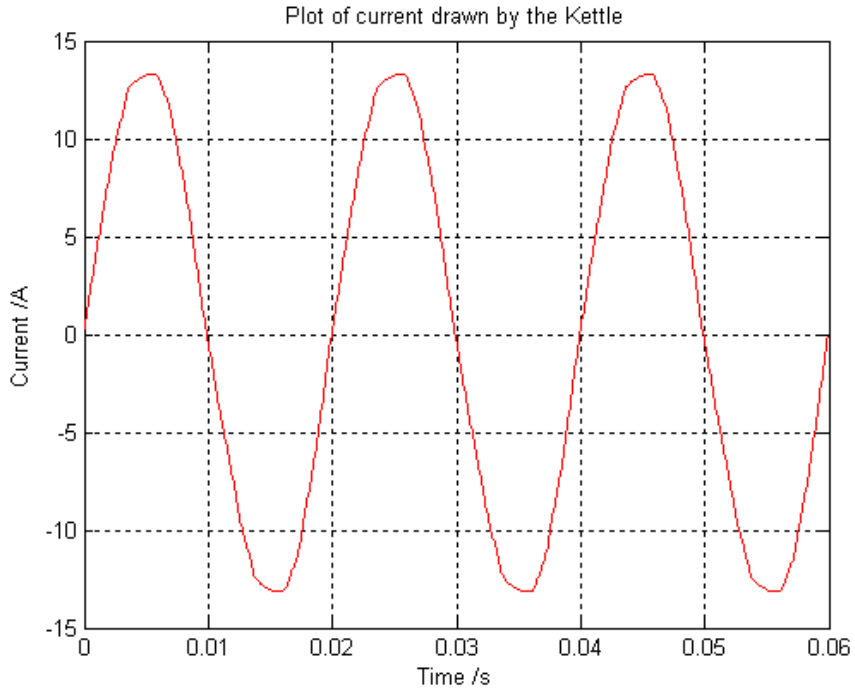


Figure 7.1: Plot of 3 cycles of the current curve for a kettle

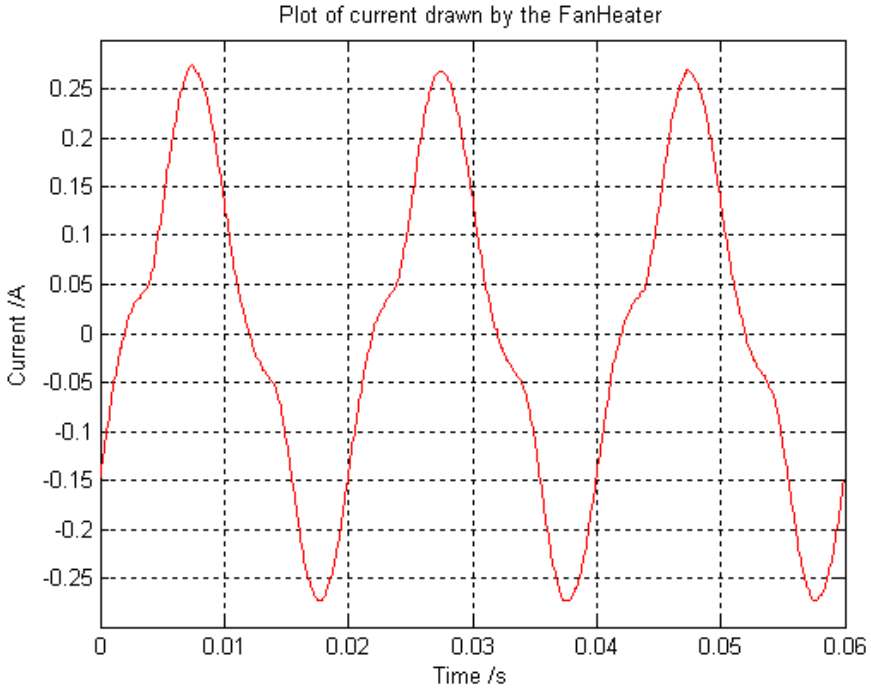


Figure 7.2: Plot of 3 cycles of the current curve for a Fan Heater in fan only mode

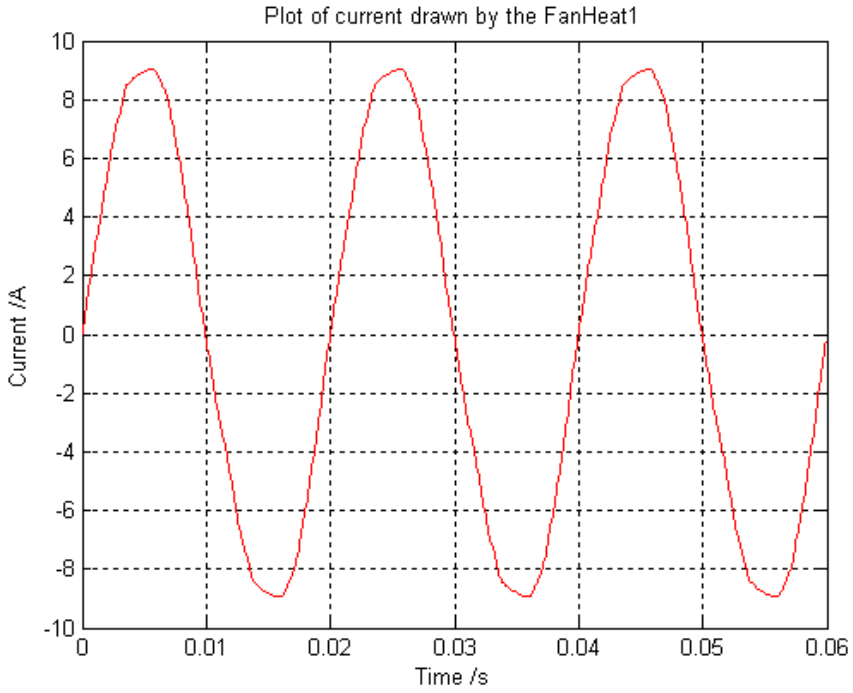


Figure 7.3: Plot of 3 cycles of the current curve for a Fan Heater in heat 1 mode

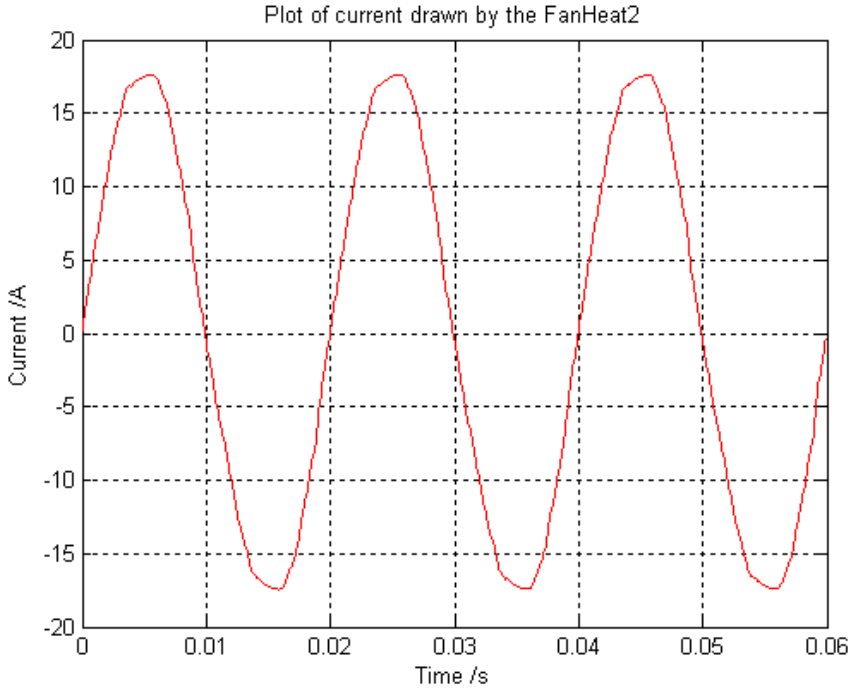


Figure 7.4: Plot of 3 cycles of the current curve for a Fan Heater in heat 2 mode

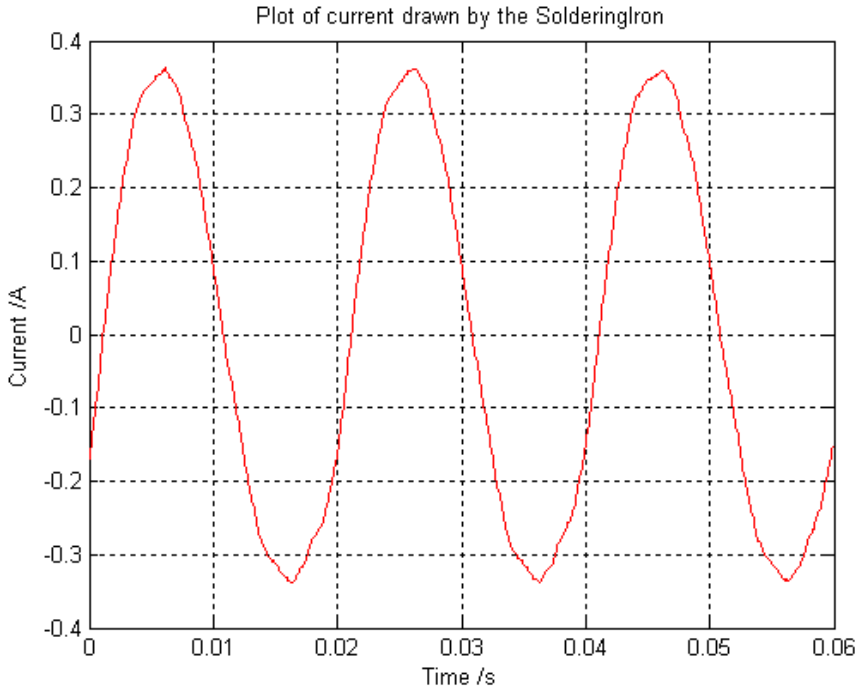


Figure 7.5: Plot of 3 cycles of the current curve for a soldering iron

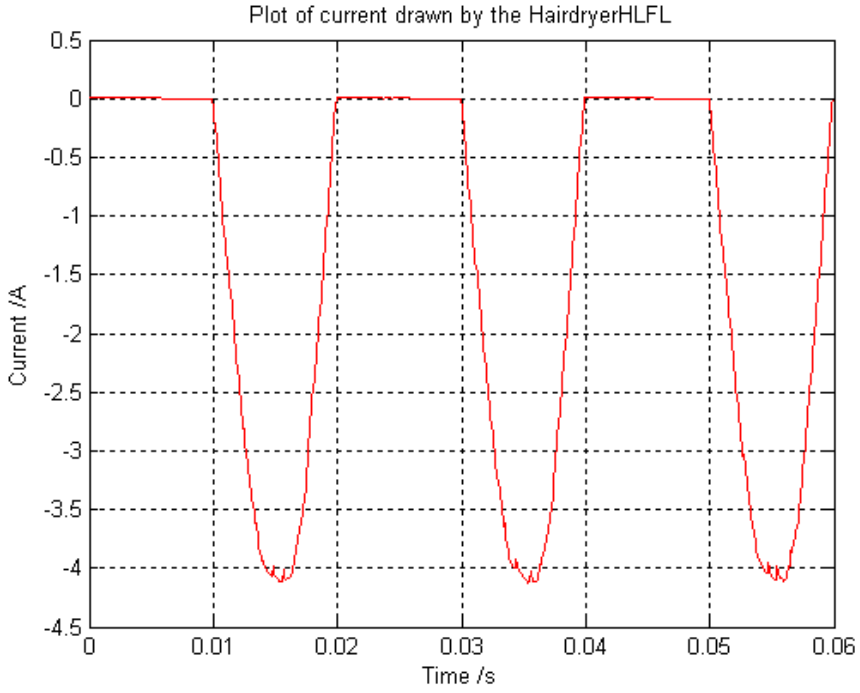


Figure 7.6: Plot of 3 cycles of the current curve for a hairdryer with low fan and low heat

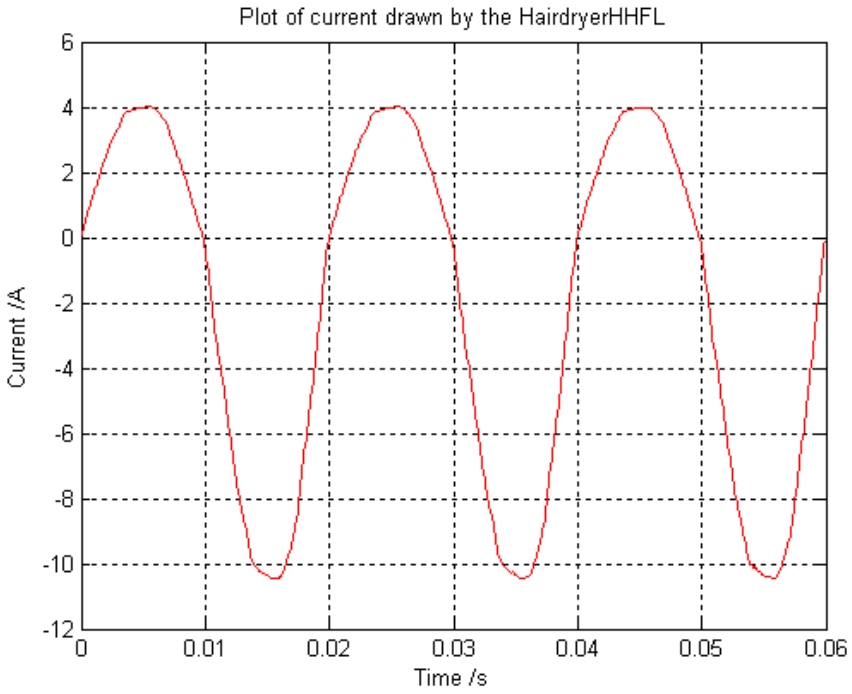


Figure 7.7: Plot of 3 cycles of the current curve for a hairdryer with low fan and high heat

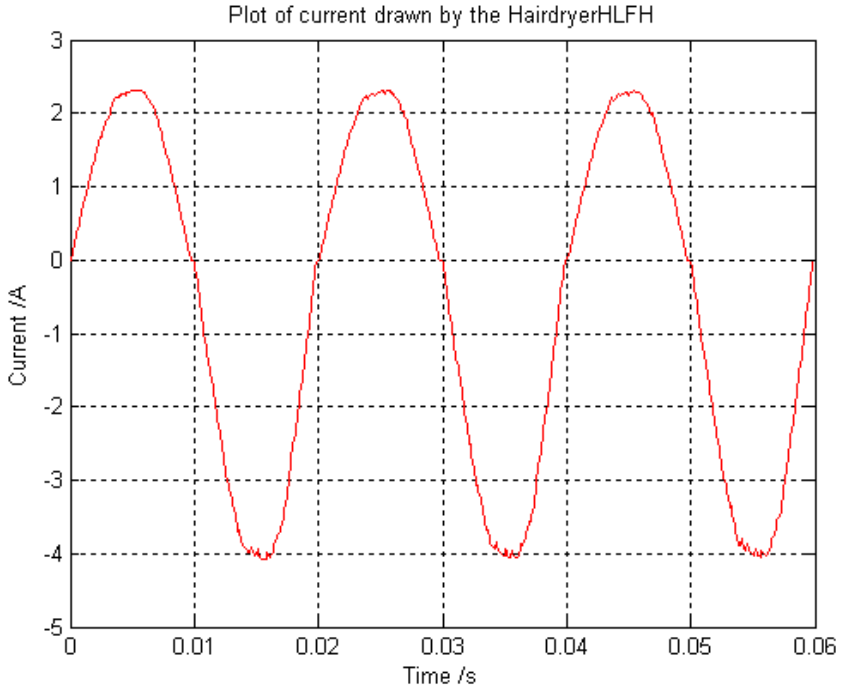


Figure 7.8: Plot of 3 cycles of the current curve for a hairdryer with high fan and low heat

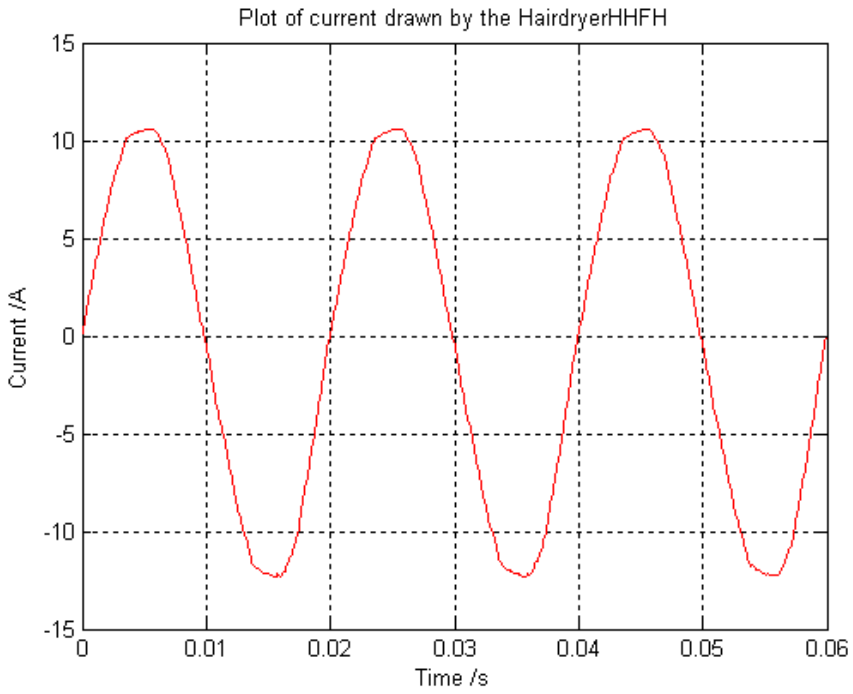


Figure 7.9: Plot of 3 cycles of the current curve for a hairdryer with high fan and high heat

appliance in the following way:

1. For each cycle of data collected determine the feature vector used to represent the appliance. This contains the model type and corresponding parameters (or if none of the models are appropriate, the generic feature set containing the normalised real and reactive power of the first third and fifth current harmonics).

2. Consider each model type separately. For each model type use the feature vectors generated to determine the number of clusters in the data and corresponding cluster centers. This is done according to the method described in Section 5.2.1.

3. Once the number of cluster centres is known, determine the number of data points needed to characterise each cluster according to the method described in Section 5.2.2.

4. Use the pre-determined number of data points to generate cluster prototypes, including the cluster center, covariance matrix and average intracluster distance for

each cluster found in the data (see Section 5.2.3).

### 7.2.1 Determining the model type and parameters

Each appliance may operate in a number of uncontrollable modes and each mode may correspond to a totally different electrical interface and parameter set. Because different appliance models cannot be considered in the same feature space, it is necessary to separate the data by model type before the number of clusters within each model type is determined.

#### Method

Each cycle of data is considered separately to allow trends or changes in the current to be observed. For each cycle of data the model type and corresponding appliance parameters are calculated as explained in Section 6.8. The model types and parameters include: The

| Model | Parameters | Comment |
|---|---|---|
| A | $R_p, R_n, Z_p, Z_n$ | Developed in Chapter 6 |
| B | $R_p, R_n, Z_p, Z_n$ | Developed in Chapter 6 |
| R | $R_p, R_n$ | If A/B appropriate but dominated by $R$ |
| C | $R, L, C$ | Developed in Chapter 6 |
| G | $H_{1re}, H_{1im}, H_{3re}, H_{3im}, H_{5re}, H_{5im}$ [a] | Generic model when IMs[b] invalid |

[a] $H_{Nre}, H_{Nim}$ denotes the power of the $N^{th}$ real and reactive harmonics in the current.
[b] IM stands for Interface Model, i.e. Models A, B, R and C which are based on the appliance interface.

Table 7.1: List of possible model types and their corresponding parameter values.

model type and corresponding parameter values which are considered most appropriate to represent each cycle of data form the feature vector which is subsequently used in the clustering procedures.

#### Results and Analysis Thereof

Table 7.2 lists the number of cycles used to describe each appliance and the corresponding number of cycles which fall under each model type considered.

| Appliance | Number of Cycles | | | | | |
|---|---|---|---|---|---|---|
| | Model A | Model B | Model R | Model C | Model G | Total |
| Kettle | 0 | 0 | 4768 | 0 | 1 | 4769 |
| Fan Only | 0 | 0 | 0 | 0 | 903 | 903 |
| Fan Heat 1 | 0 | 0 | 1860 | 0 | 0 | 1860 |
| Fan Heat 2 | 1 | 0 | 2024 | 0 | 0 | 2025 |
| Soldering Iron | 0 | 3535 | 0 | 0 | 4 | 3539 |
| Hairdryer HLFL | 0 | 0 | 976 | 0 | 0 | 976 |
| Hairdryer HLFH | 0 | 0 | 722 | 0 | 1 | 723 |
| Hairdryer HHFL | 0 | 0 | 963 | 0 | 1 | 964 |
| Hairdryer HHFH | 0 | 1 | 915 | 0 | 0 | 916 |

Table 7.2: Number of cycles which fall under each model type considered, listed for each appliance.

From Table 7.2 it can be seen that most cycles for all appliances bar the fan heater in fan only mode fall under one of the IM types. Every cycle for the fan heater in heat mode 1 and the hairdryer in heat-low-fan-low mode fall under a single IM (model R in both cases), although it should be noted that this does not automatically imply that these appliances will have just a single mode of use.

The fan heater in heat mode 2 and the hairdryer in heat-high-fan-high mode are appliances which have training data corresponding to two different IMs. However, in both scenarios there is a dominant type which accounts for over 99% of the cycles monitored, indicating that the second model type is probably an anomaly due to noise on the system or transient behaviour.

The kettle, fan in heat 1 mode and hairdryer in heat-low-fan-high and heat-high-fan-low mode have nearly all cycles of data falling under a single IM, with the remaining few belonging to the generic feature set. However, in all these cases less than 0.1% of cycles belong to model G and this can again be assumed to be due to noise or transient behaviour.

The fan heater in fan only mode is the only appliance whose data is not represented by any of the IMs. This may be explained by considering the current drawn by this appliance in the context of the models considered. It can be seen from Figure 7.2 that

the current drawn is non-linear. This is confirmed by the value of the total harmonic distortion (see Section 6.7) which has an average value of 0.1954 in the positive half cycle and 0.2243 in the negative half cycle. Therefore models A, B and R will not accurately represent the fan heater in fan only mode. Model C is representative of a rectifier which draws current in a manner indicated in Figure 6.12 in the previous chapter. Again, this is not representative of the fan and consequently model C will not accurately represent this appliance. For the purposes of this thesis this appliance is therefore represented by a generic feature set consisting of the the real and reactive power of the first, third and fifth current harmonics (normalised by the magnitude of the voltage). All the cycles of data considered for the fan in this mode are represented by the generic model type.

Figures 7.10 through to 7.12 show a plot of the parameters obtained for each cycle, projected onto 2 dimensions, for each of the appliances considered. From these figures it can be seen that most of the appliances considered in this chapter have purely resistive electrical interfaces.

From Figure 7.10 it can be seen that, despite the limited features used to represent appliances, the clusters formed are disjoint. Most of the clusters formed are also elongated and this elongation is due to the heating effect of the resistors. It can also be seen that the data points for the kettle and the fan heater lie along the line $R_{pos} = R_{neg}$, whilst those data points from the hairdryer lie off this line, due to differing resistances in the positive and negative half cycle. Furthermore, in a low heat and low fan mode, the hairdryer has no current in the positive half cycle and this is represented by a resistance of exactly 0 (explained in Chapter 6).

The soldering iron is the only appliance to be represented by model B (see Figure 7.11) and the fan heater in fan only mode is the only appliance to be represented by model G (see Figure 7.12). The soldering iron forms a compact cluster, though there are some points which lie slightly away from the bulk. This is probably due to transients caused

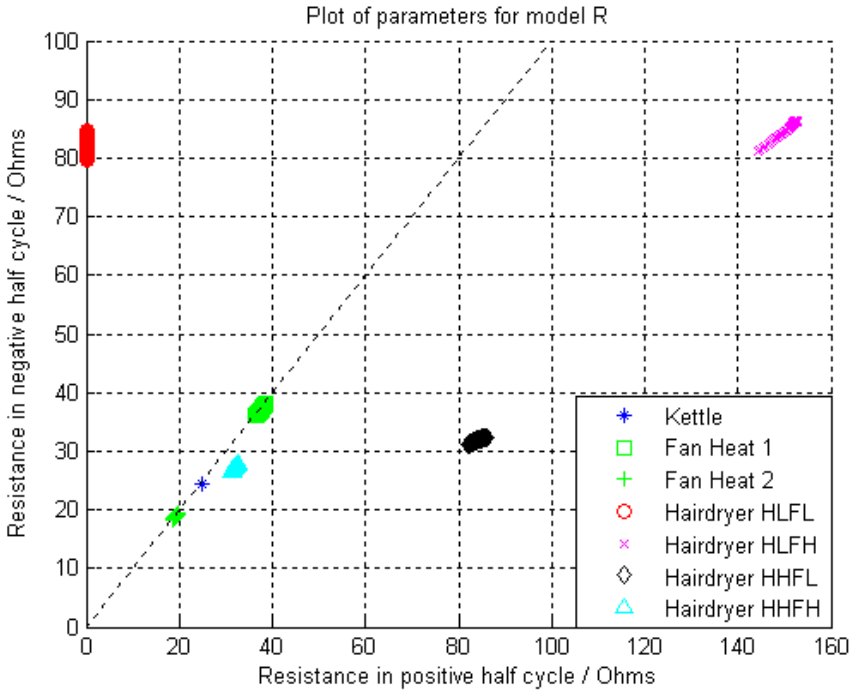


Figure 7.10: Plot of data, projected onto 2 dimensions, which forms the feature vector for appliances with model type R.

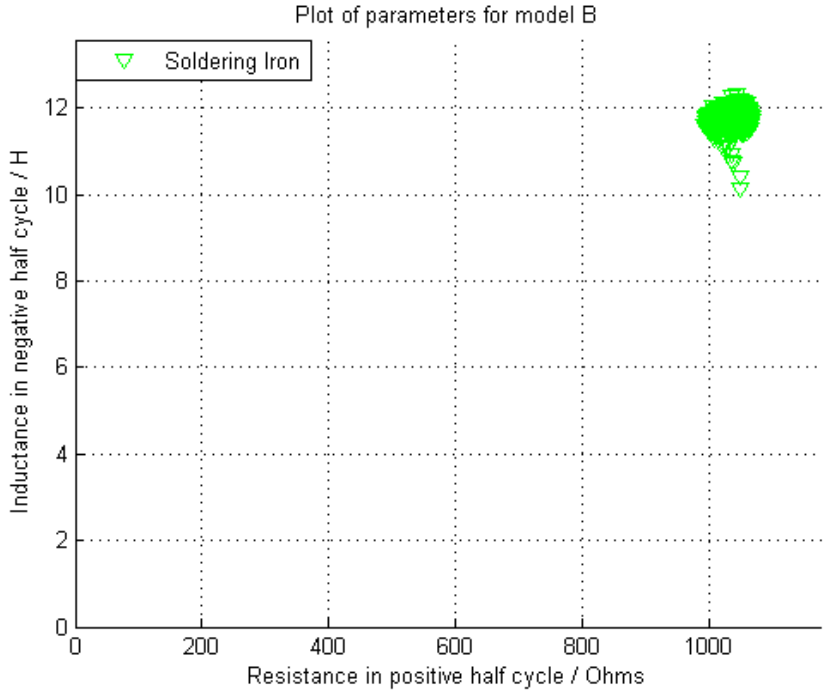


Figure 7.11: Plot of data, projected onto 2 dimensions, which forms the feature vector for appliances with model type B.

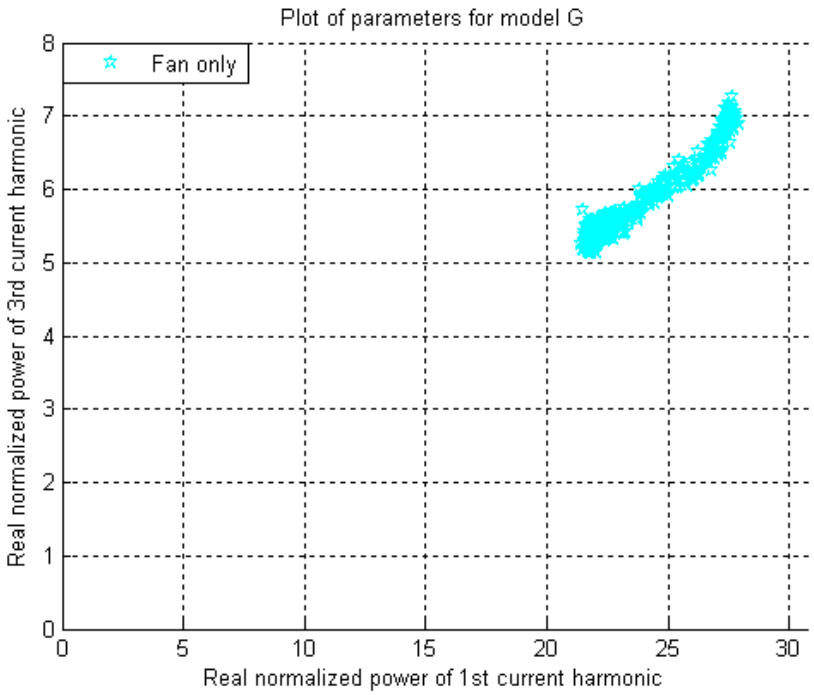


Figure 7.12: Plot of data, projected onto 2 dimensions, which forms the feature vector for appliances with model type G.

by changes in the voltage signal[6]. The data points from the fan heater form an elongated cluster, as these features are correlated. If the magnitude of the current increases for some reason (due to the heating effect of resistors, or due to changes in the voltage signal) this will be reflected by all the harmonics so long as the shape of the current waveform remains the same.

Although it is seen from Figures 7.10 to 7.12 that the data from each appliance falls in a single cluster, the natural number of clusters in the data is determined mathematically via the hierarchical clustering algorithm.

### 7.2.2 Determining the number of clusters

Having determined the model type of all the cycles of data considered for each appliance, each model is assessed individually in order to determine the number of clusters into which the feature vectors form.

[6]Whilst the selected features do not reflect steady state changes in the voltage signal, they will respond to transient behaviour.

### Method

Each feature vector contains the model type (A, B, R, C or G) and the corresponding appliance parameters. For each appliance each model type is considered separately. The appliance parameters are used in the clustering algorithm described in Section 5.2.1 to evaluate the natural number of clusters in the data and the corresponding cluster centers.

### Results and analysis thereof

For each appliance only one model type is applicable. Section 5.2.1 states that the hierarchical clustering algorithm is performed multiple times on different groups of data so that each group contains $y$ feature vectors, where $y$ is determined according to Equation 5.1. Table 7.3 shows how many clusters are formed each time the hierarchical clustering algorithm is performed[7].

From Table 7.3 it can be seen that the resultant number of clusters for each appliance is 1 for all the appliances considered in this chapter. This is the correct value in all cases, as none of the appliances considered in this chapter are multi-modal, aside from the fan heater and hairdryer where the different modes are user controlled and considered as separate appliances.

From Table 7.3 it can also be seen that the hierarchical clustering of data subsets often produces more than one cluster. This is due to the mechanism by which the data is clustered. Ward's clustering mechanism is used to determine the number of clusters in each group of $y$ data points. In this clustering mechanism the nested set of clusters produced is done so according to a function which minimizes the Euclidean error at each step (see Appendix C for more information). However, because of the presence of slow transients caused by the heating effect of resistors, or slight changes in operating parameters during use of appliances, often the underlying data sets may be elongated.

[7]Recall that hierarchical clustering is performed multiple times in order to ensure that all possible modes of operation are included in the training algorithm.

| Appliance | Number of Clusters in Data Set: | | | | | | | | | | | | | | | | | | Final Number of Clusters |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | |
| Kettle | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | - | - | - | - | - | - | - | - | - | 1 |
| Fan Only | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - | - | - | - | - | - | - | - | - | 1 |
| Fan Heat 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Fan Heat 2 | 3 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | - | - | - | - | - | - | - | - | 1 |
| Soldering Iron | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | - | 1 |
| Hairdryer HLFL | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - | - | - | - | - | - | - | - | - | 1 |
| Hairdryer HLFH | 2 | 1 | 2 | 1 | 2 | 2 | 2 | - | - | - | - | - | - | - | - | - | - | - | 1 |
| Hairdryer HHFL | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | - | - | - | - | - | - | - | - | - | 1 |
| Hairdryer HHFH | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | - | - | - | - | - | - | - | - | - | 1 |

Table 7.3: Number of clusters formed in each data following the hierarchical clustering algorithm.

This elongation of the data is not accounted for in Ward's clustering algorithm and subsequently results in the apparent presence of more than one cluster.

Because of the nature of the data, the final set of cluster centres produced are also likely to form an elongated cluster. Therefore, a single linkage algorithm (see Appendix C) is used to cluster this data set to account for the underlying trend and consequently the apparent existence of extra clusters is removed. It should be noted that the single linkage algorithm is not used at every step in the clustering procedure because is it corruptible by noisy data which may to be present in the initial data set.

At the end of this stage of the training algorithm, the number of clusters in each data set and the corresponding cluster centres have been identified. Table 7.4 shows the cluster centres produced and Table 7.5 shows the error between the true cluster centres and the cluster centres returned from the algorithm. This error is present because in the final clustering procedure it is assumed that each of the existing cluster centres (from which the final set of clusters and corresponding cluster centres are determined) should have the same weighting. This is not the case, because in the previous clustering steps, the different clusters produced contain a varying number of data points. However, from Table 7.5 it can be seen that the error between the true cluster centres and those obtained from the algorithm is less than 0.5% in all dimensions for each appliance. The limitations on the ADC used to measure the current and voltage produces an error of 0.2% on each signal and therefore an error of this magnitude may be assumed negligible.

### 7.2.3 Determining the number of data points needed to characterise the clusters

Having determined the number of clusters in the data and the cluster centers, this section explores the mechanism for determining the number of data points needed to characterise each cluster prototype.

Cluster centres Returned for all Appliance

| Appliance | Model | Cluster Centre in Dimension | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Kettle | R | 24.76 | 24.42 | - | - | - | - |
| Fan Only | G | 22.90 | -24.40 | 5.60 | 3.97 | -0.63 | -0.68 |
| Fan Heat 1 | R | 38.27 | 37.79 | - | - | - | - |
| Fan Heat 2 | R | 19.59 | 19.30 | - | - | - | - |
| Soldering Iron | B | 1038.15 | 1025.33 | 11.79 | 11.91 | - | - |
| Hairdryer HLFL | R | 84.22 | - | - | - | - | - |
| Hairdryer HLFH | R | 152.03 | 86.05 | - | - | - | - |
| Hairdryer HHFL | R | 85.66 | 32.32 | - | - | - | - |
| Hairdryer HHFH | R | 32.58 | 27.69 | - | - | - | - |

Table 7.4: Cluster centres returned by the hierarchical clustering algorithm. Note that units for the Fan Only are Watts in all dimensions, for all other appliances dimensions 1 and 2 are Ohms and 3 and 4 are Henrys.

Errors in Cluster centres for all Appliances

| Appliance | Model | Percentage Error in Dimension | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Kettle | R | 0.011 | 0.006 | - | - | - | - |
| Fan Only | G | 0.050 | 0.000 | 0.125 | 0.096 | 0.000 | 0.000 |
| Fan Heat 1 | R | 0.213 | 0.211 | - | - | - | - |
| Fan Heat 2 | R | 0.312 | 0.310 | - | - | - | - |
| Soldering Iron | B | 0.031 | 0.015 | 0.010 | 0.001 | - | - |
| Hairdryer HLFL | R | 0.000 | 0.007 | - | - | - | - |
| Hairdryer HLFH | R | 0.186 | 0.210 | - | - | - | - |
| Hairdryer HHFL | R | 0.042 | 0.029 | - | - | - | - |
| Hairdryer HHFH | R | 0.191 | 0.210 | - | - | - | - |

Table 7.5: Error between the true cluster centres and the cluster centres returned by the hierarchical clustering algorithm. Error is expressed as a percentage of the true cluster centre value in each dimension considered.

**Method**

The number of data points, $n_p$, needed to characterise each cluster are determined with consideration to the cluster mean. As the appliance is in use there may be some trend in the data set, be it an underlying ripple (for example, motor variation in a washing machine) or slow transient (for example, the increase in resistance caused by the heating effect in lights or heaters). This variation will cause a change in the cluster centre calculated as the number of data points used to calculate the cluster centre increases. The number of data points needed to characterise a cluster is therefore determined with

Number of cycles needed to characterise each cluster

| Appliance | Number of cycles | |
|---|---|---|
| | Case 1 | Case 2 |
| Kettle | 1670 | 1820 |
| Fan Only | 250 | 260 |
| Fan Heat 1 | 200 | 220 |
| Fan Heat 2 | 200 | 230 |
| Soldering Iron | 870 | 890 |
| Hairdryer HLFL | 50 | 50 |
| Hairdryer HLFH | 30 | 30 |
| Hairdryer HHFL | 110 | 110 |
| Hairdryer HHFH | 80 | 90 |

Table 7.6: Number of cycles needed to characterise each cluster. Case 1 is the instance that the cluster centres returned by the clustering algorithm are used whilst Case 2 is when the true cluster centres are used.

consideration to the changing cluster centre as more data points are added. The method used is described in detail in Section 5.2.2.

**Results**

For each appliance the number of data points needed to characterise each cluster is expressed in Table 7.6. From this table it can be seen that the number of cycles needed to represent the clusters varies considerably, from a minimum of 30 for the hairdryer in heat low fan low mode, to a maximum of 1670 for the kettle. The reason for this variation is due to the variance of the data. Recall that the t-test is performed according to the equation:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \tag{7.1}$$

Therefore, a cluster with a small variance will produce a larger t-value and consequently the null hypothesis, which is that the data mean and the cluster centres are identical, is more likely to be rejected, thus requiring a larger number of data points to be considered.

Having determined the number of data points which are needed to characterise each cluster and the number of clusters and cluster centers, the next step is to determine the cluster prototypes.

### 7.2.4 Determining the cluster prototypes

This section explores the generation of cluster prototypes for each appliance according to the method outlined in Section 5.2.3. The cluster prototype is composed of the cluster centers, covariance matrix, average intracluster distance and cluster label. The cluster centres are those calculated from the clustering algorithm (although it should be noted that those calculated from the set of $n_p$ data points needed to represent the cluster are not statistically different according to the t-test). The covariance matrices and average intracluster distances are calculated from the set of $n_p$ data points and these are written to hard memory so that they may be used in the recognition side of the algorithm. The results are not displayed here as the actual covariance matrices would prove to be somewhat meaningless at this stage. Rather, the discussion as to the cluster prototypes developed is performed when considering the membership generation of a new set of data points in the set of known clusters.

## 7.3 Membership generation for new feature vectors

This section considers the way in which memberships are generated for a new set of data corresponding to each of the known appliances in order to demonstrate the ability of the classification scheme to correctly identify appliances based on the features selected to represent them.

A set of data is gathered for each of the 9 appliances considered in this chapter to be used in the recognition side of the AEM. For each of the 9 appliances, $N_R$ cycles of data are gathered over a period of time. The applicable model if first established for each of the $N_R$ cycles. Once this has been determined the corresponding feature vector of parameter values is calculated. Both the model type and the corresponding feature vector of the new data point are then used along with the existing cluster prototypes in order to generate a set of membership values for that data point. These membership values indicate the

level of belongingness of that data point in each of the known clusters.

**Method**

- For each appliance generate $N_R$ cycles of recognition data. In the actual implementation of the AEM in the domestic environment, the recognition data would be generated when a step change in the power signal is observed. This recognition data is denoted $S_\delta$ and corresponds to the voltage seen by and the current drawn by the individual appliance which has just changed state. For the purposes of this research $S_\delta$ is approximated by the individual current cycles generated and the voltage measured across the appliance terminals.
- For $S_\delta$ determine which model (A, B, C, R or G) is appropriate to model the appliance and calculate the corresponding appliance parameters.
- Having determined the appropriate model and feature space, use the appliance parameters to calculate the probabilistic membership and possibilistic typicality for the new data point in each of the existing clusters in that feature space. This is done according to the method described in Section 5.3.8.
- Use values obtained for $u_{ij}$ and $t_{ij}$ to calculate $m_{ij}$ according to Equation 5.24.
- Use $m_{ij}$ to determine the appliance (if any) to which the unknown signal, $S_\delta$, corresponds. In the fully developed AEM this involves combining the short term membership values generated with a set of long term memberships according to the method developed in Section 4.2.3. However, for the purposes of this thesis it will be assumed that the long term memberships add no beneficial information and therefore classification is performed using the short term membership values only. This assumption is valid for the purposes of this research because the objective of this work is to determine the classification ability of the short term time domain features and the membership generation algorithm to correctly identify appliances from which they were drawn.

- Calculate the confidence level for the classification using the set of $m$ membership values generated according to the method described in Section 4.2.4.

**Results and Analysis Thereof**

This method is performed for 100 cycles worth of data which have not been used in training from each of the 9 appliances considered in this chapter. For each appliance the number of cycles which fall under each model type is illustrated in Table 7.7. From these results it can be seen that all of the cycles considered for each appliance type fall under the correct model type (i.e. the model type to which the underlying appliance actually belongs to), however, this does not necessarily mean that each cycle will be correctly identified.

| Appliance | Number of Cycles | | | | | |
|---|---|---|---|---|---|---|
| | Model A | Model B | Model R | Model C | Model G | Total |
| Kettle | 0 | 0 | 100 | 0 | 0 | 100 |
| Fan Only | 0 | 0 | 0 | 0 | 100 | 100 |
| Fan Heat 1 | 0 | 0 | 100 | 0 | 0 | 100 |
| Fan Heat 2 | 0 | 0 | 100 | 0 | 0 | 100 |
| Soldering Iron | 0 | 100 | 0 | 0 | 0 | 100 |
| Hairdryer HLFL | 0 | 0 | 100 | 0 | 0 | 100 |
| Hairdryer HLFH | 0 | 0 | 100 | 0 | 0 | 100 |
| Hairdryer HHFL | 0 | 0 | 100 | 0 | 0 | 100 |
| Hairdryer HHFH | 0 | 0 | 100 | 0 | 0 | 100 |

Table 7.7: Number of cycles which fall under each model type considered, listed for each appliance.

Once the model type has been determined, the corresponding feature vector is calculated for each cycle of data. These feature vectors are plotted in 2-D feature space for each of the appliances in Figures 7.13, 7.14 and 7.15 for models R, B and G respectively.

From these figures it can be seen that the set of recognition data does not fall equidistant around the existing cluster prototype of the cluster to which is belong. In the case of Models B and R this may occur because the recognition data is captured at the start of use of the appliance (in order to ensure it is as similar as possible as that data which may be observed in the signal $S_\delta$) which, due to the heating effect of resistors, means that it is

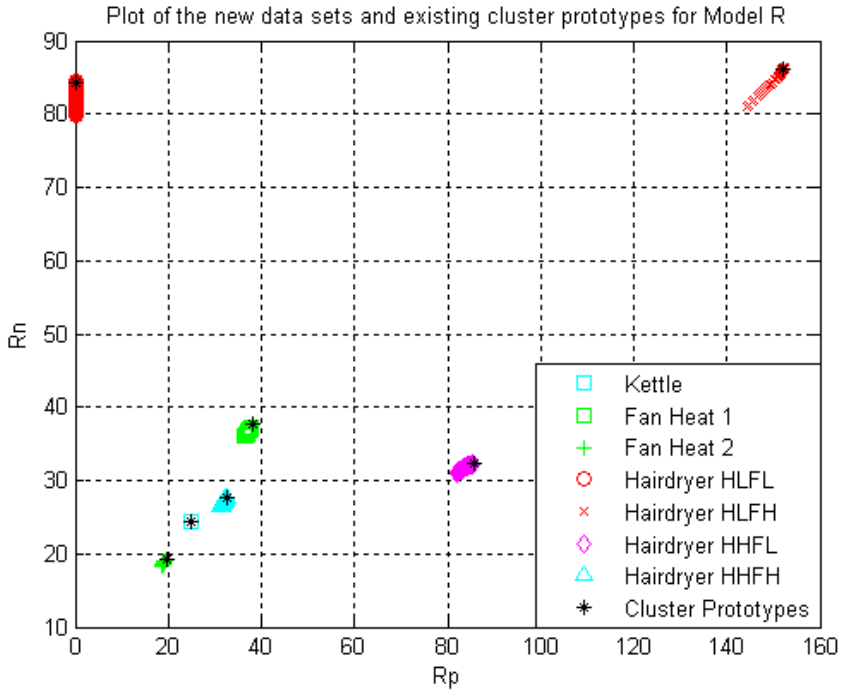


Figure 7.13: Plot of the feature values generated for the recognition data which falls under Model R and the existing cluster prototypes which fall under the same model.


Figure 7.14: Plot of the feature values, projected onto two dimensions, generated for the recognition data which falls under Model B and the existing cluster prototypes which fall under the same model.

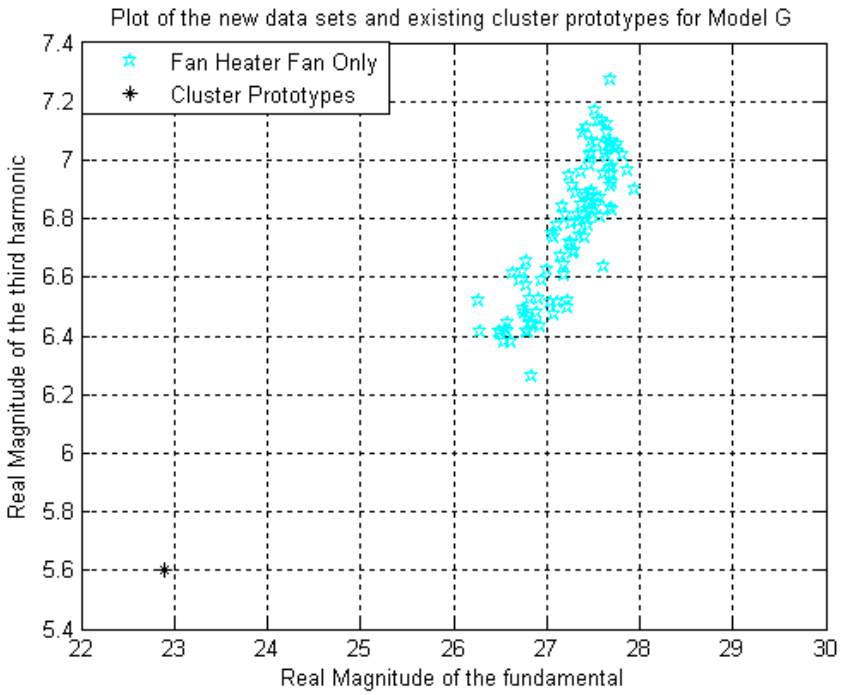


Figure 7.15: Plot of the feature values, projected onto two dimensions, generated for the recognition data which falls under Model G and the existing cluster prototypes which fall under the same model.

not necessarily identical to the mean of the training data which covers the entire mode of use of the appliance. In Model G the difference in data value may be cause by variations in the voltage between capture of the training data and capture of the recognition data, because the parameters in the feature vector in this model are dependent on voltage variations.

The differences observed between the recognition data and the cluster prototypes is particularly important when considering the set of membership values which are subsequently generated. These membership values are plotted for each set of recognition data in the cluster to which they actually belong in Figures 7.16 to 7.24. The membership values for each cycle of recognition data in the clusters to which they do not belong is negligible.

From these figures it can be seen that for all recognition data points bar one, the resultant membership value is greater than 0.2 and consequently 899 of the 900 data points in the

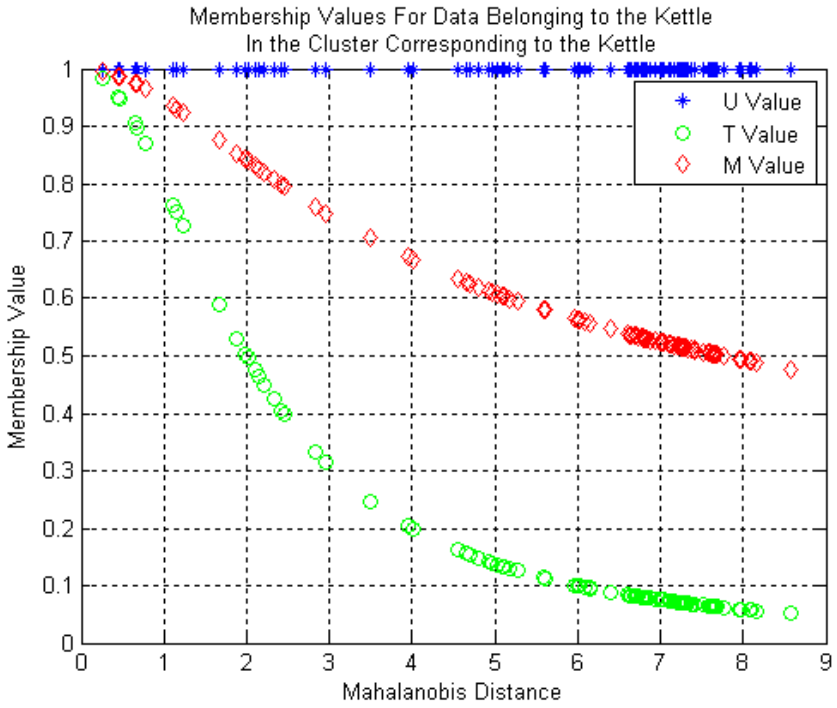


Figure 7.16: Plot of the membership values generated for the recognition data for the kettle in the cluster generated by the corresponding training data.

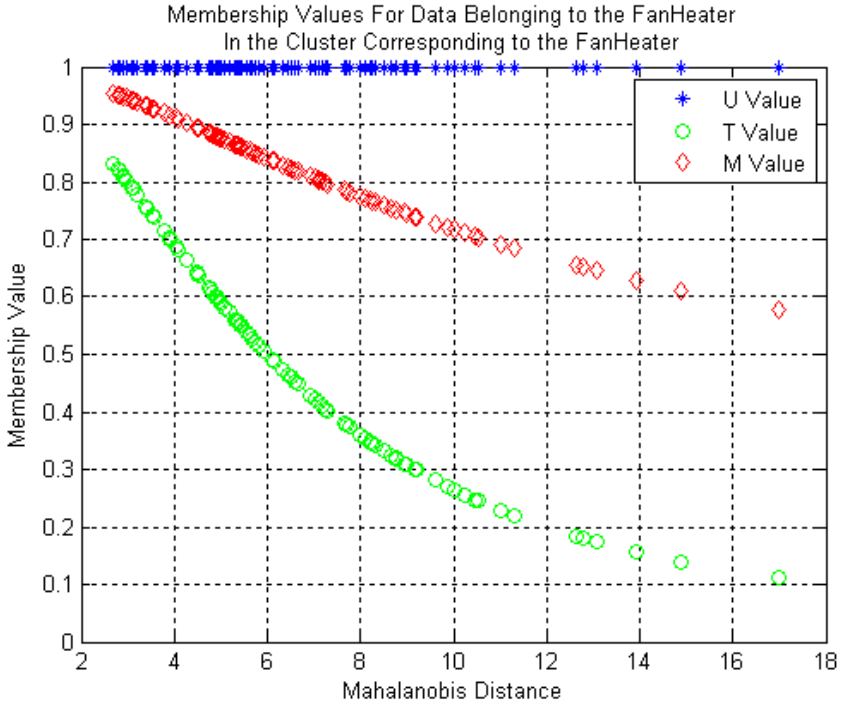


Figure 7.17: Plot of the membership values generated for the recognition data for the fan heater, fan only mode, in the cluster generated by the corresponding training data.

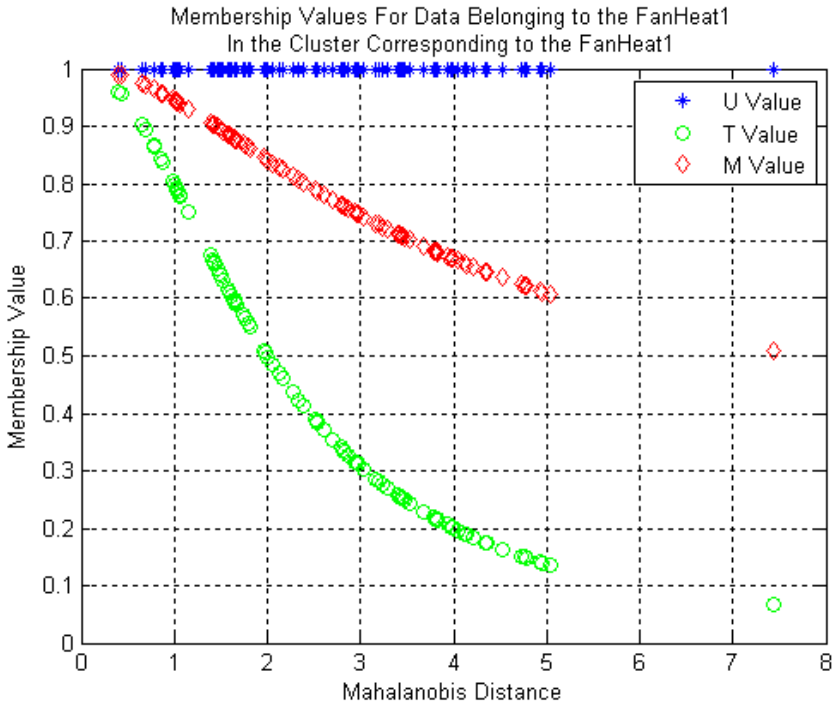


Figure 7.18: Plot of the membership values generated for the recognition data for the fan heater, heat mode 1, in the cluster generated by the corresponding training data.

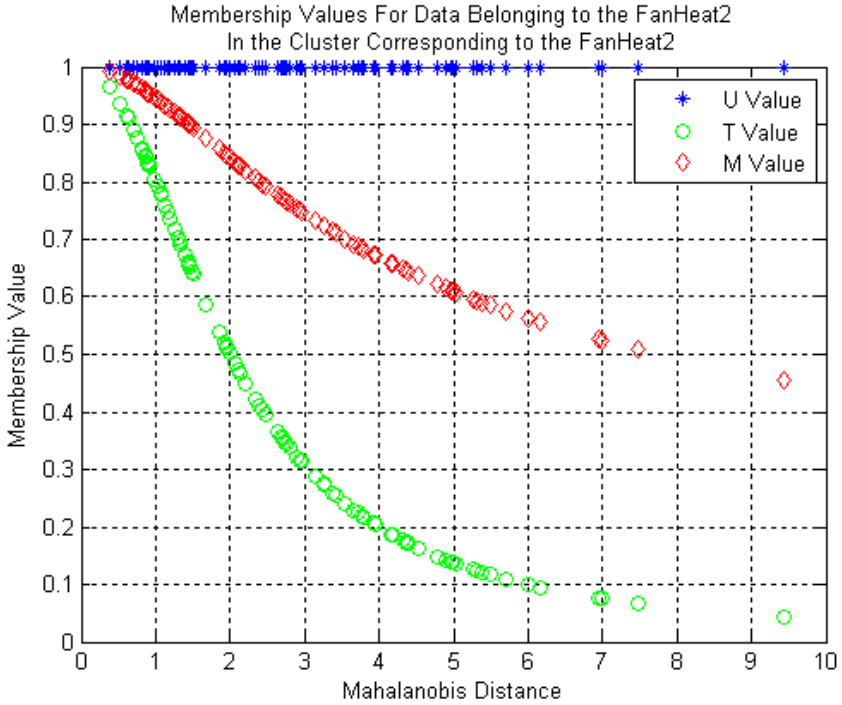


Figure 7.19: Plot of the membership values generated for the recognition data for the fan heater, heat mode 2, in the cluster generated by the corresponding training data.

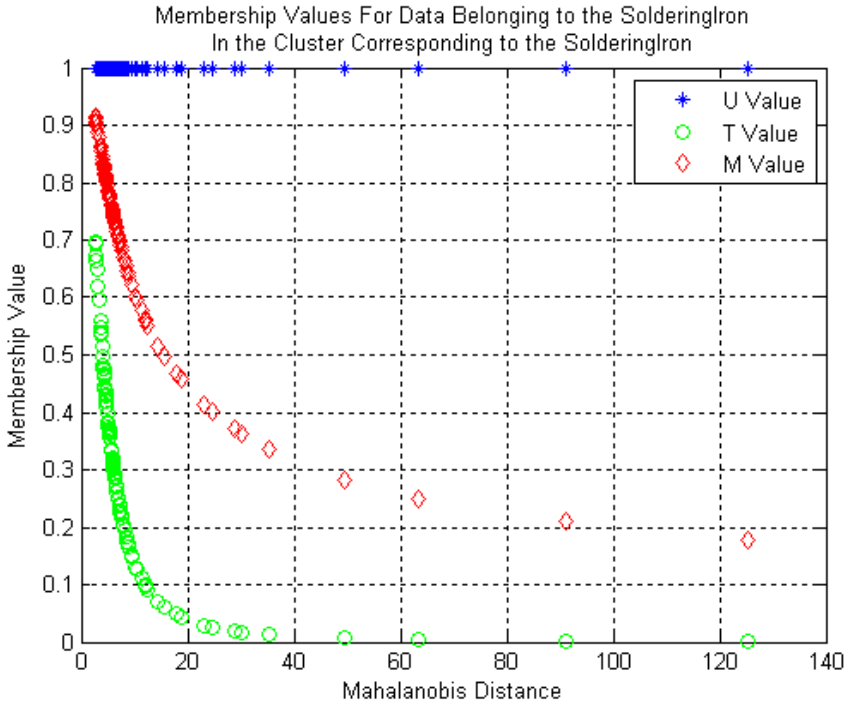


Figure 7.20: Plot of the membership values generated for the recognition data for the soldering iron in the cluster generated by the corresponding training data.

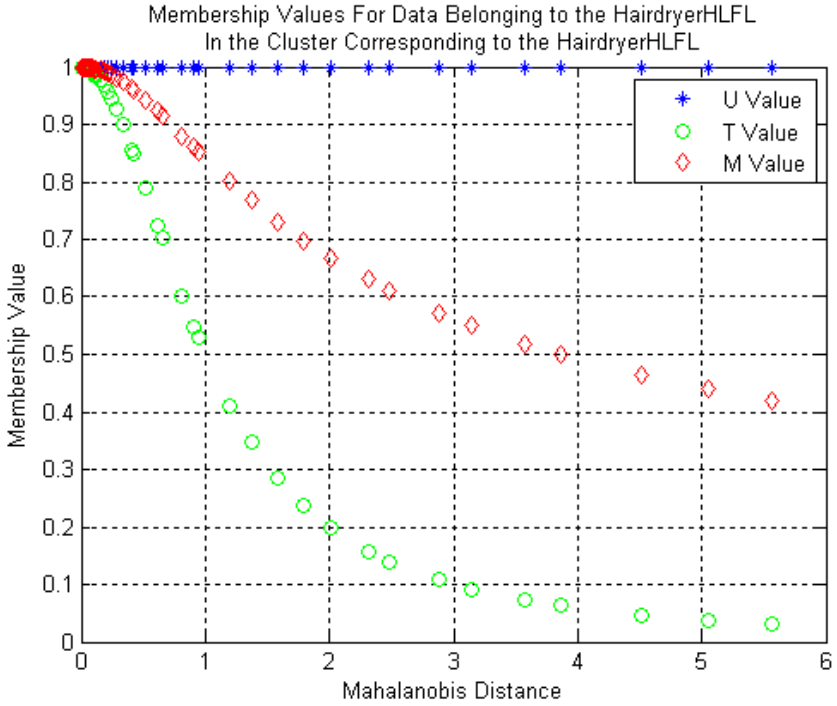


Figure 7.21: Plot of the membership values generated for the recognition data for the hairdryer in low-heat-low-fan mode in the cluster generated by the corresponding training data.

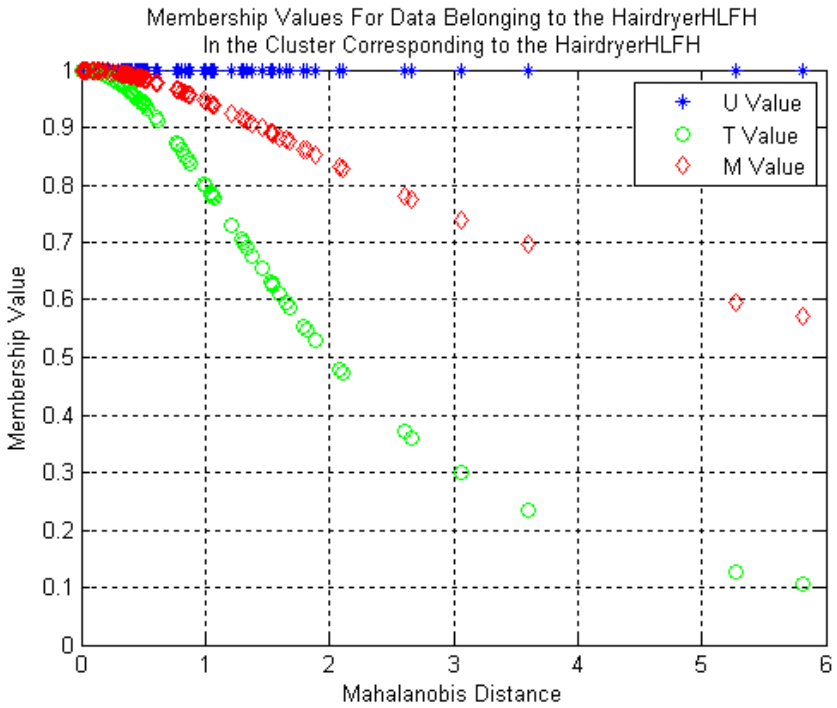


Figure 7.22: Plot of the membership values generated for the recognition data for the hairdryer in low-heat-high-fan mode in the cluster generated by the corresponding training data.

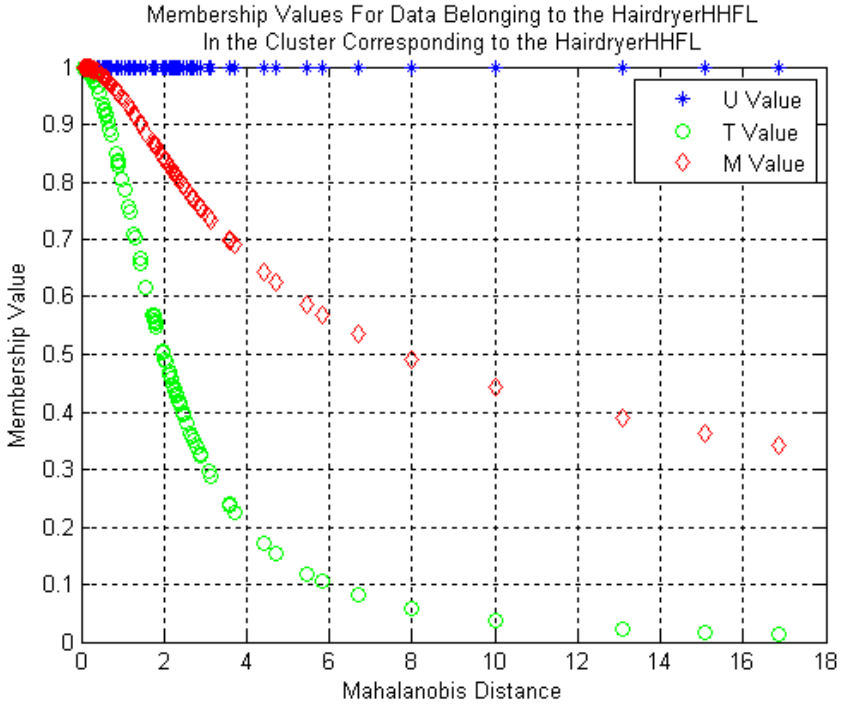


Figure 7.23: Plot of the membership values generated for the recognition data for the hairdryer in high-heat-low-fan mode in the cluster generated by the corresponding training data.

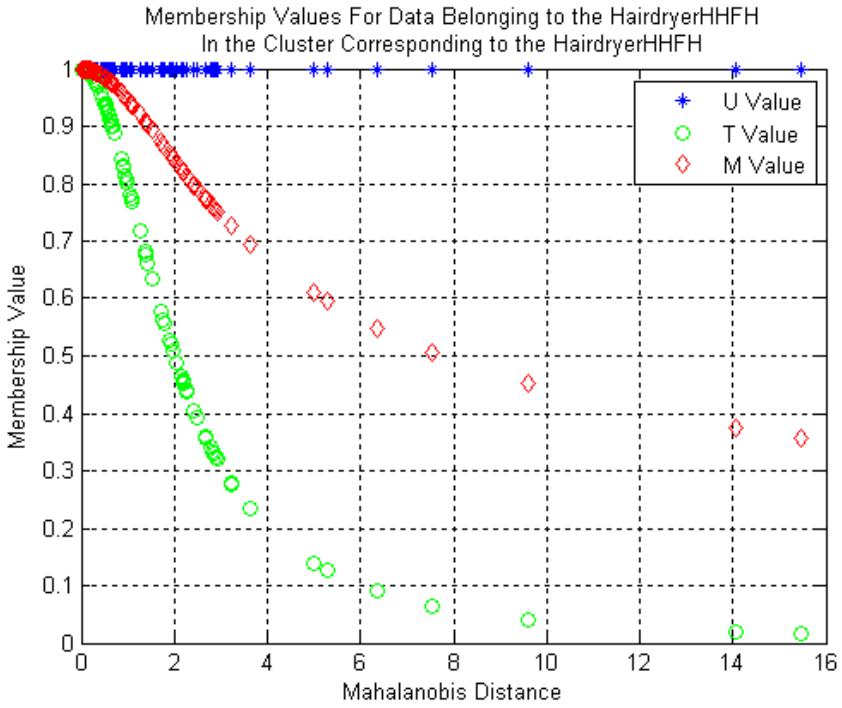


Figure 7.24: Plot of the membership values generated for the recognition data for the hairdryer in high-heat-high-fan mode in the cluster generated by the corresponding training data.

recognition data set are classified correctly according to their short term time domain features. This, along with the observations made regarding Figures 7.13, 7.14 and 7.15, illustrates that the cluster prototypes do account for data points which are typical in that cluster but which are not coincident with the cluster mean, whilst rejecting data points which do not belong in that cluster. This indicates well designed cluster prototypes which are able to accurately identify those data points which exist some distance from the cluster prototype.

From Figures 7.16 to 7.24 it can be seen that in general the performance of the membership generation algorithm is better for those appliances whose features correspond to the IMs considered in this chapter. The only appliance which has features corresponding to Model G is the soldering iron and it can be seen that whilst the possibilistic memberships generated for the recognition data points in this cluster are close to 1, the possibilistic typicalities are significantly smaller than those generated for the other recognition data sets. Furthermore, the data point which has a resultant membership of less than 0.2 does

so because of the low possibilistic typicality generated for that data point. This indicates the improved performance of the IM feature sets, which are independent of the voltage variations, compared to the generic feature set.

Whilst the accuracy of classification of the 100 data points per appliance considered is very high (99.8%) there may be scenarios in which the classification is not accurate. This is demonstrated by considering the initial set of training data. From Table 7.2 it can be seen that a total of 8 data points of the 16675 considered are identified as belonging to the incorrect model. Therefore, had these data points been observed during the recognition part of the algorithm, they would have been incorrectly classified or not classified at all. However, this represents less than 0.1% of data points. Therefore, the classification algorithm is deemed to be accurate to greater than 99.8% for the data considered in this chapter.

## 7.4 Generating confidence levels for the classifications made

Recall from Chapter 4 that the objective of the confidence level is to determine the strength of the classification made. The confidence level, $CL$, is calculated according to

$$CL = \sqrt{M_1 \left(M_1 - M_2\right)} \tag{7.2}$$

where the final membership value be denoted $M_1$ for the highest membership and $M_2$ for the second highest.

These membership values, $M$, are obtained through combining the short term memberships $m$, calculated as described above, with long term membership values calculated according to the methods described in Chapter 4. However, the version of the AEM presented in this thesis does not consider long term membership values and therefore, for the purposes of this work $M$ is replaced by $m$, the short term membership values.

The confidence levels are calculated for each of the classifications made and these are presented in Figures 7.25 to 7.33.

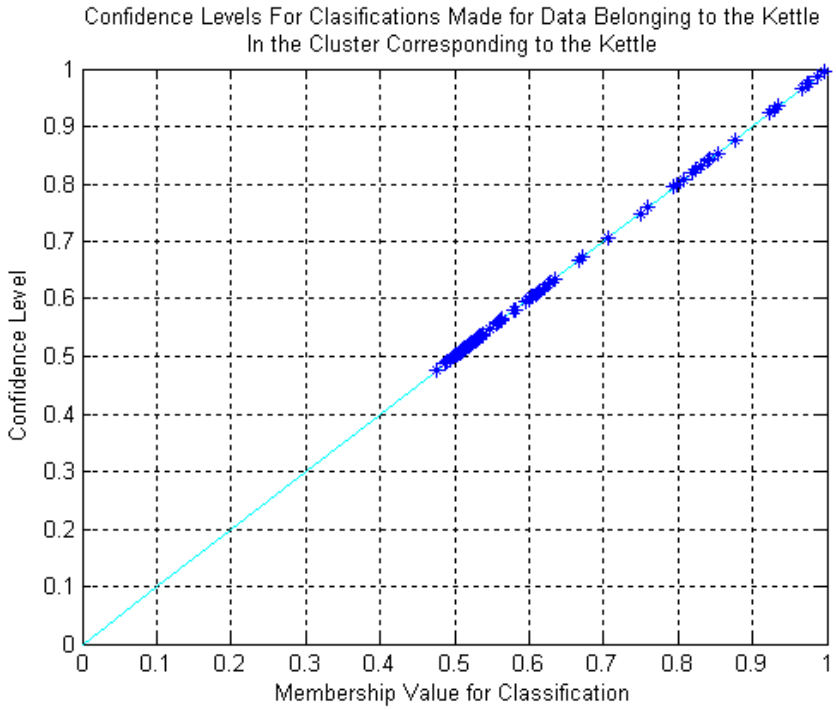


Figure 7.25: Plot of the confidence levels generated for the recognition data for the kettle in the cluster generated by the corresponding training data.

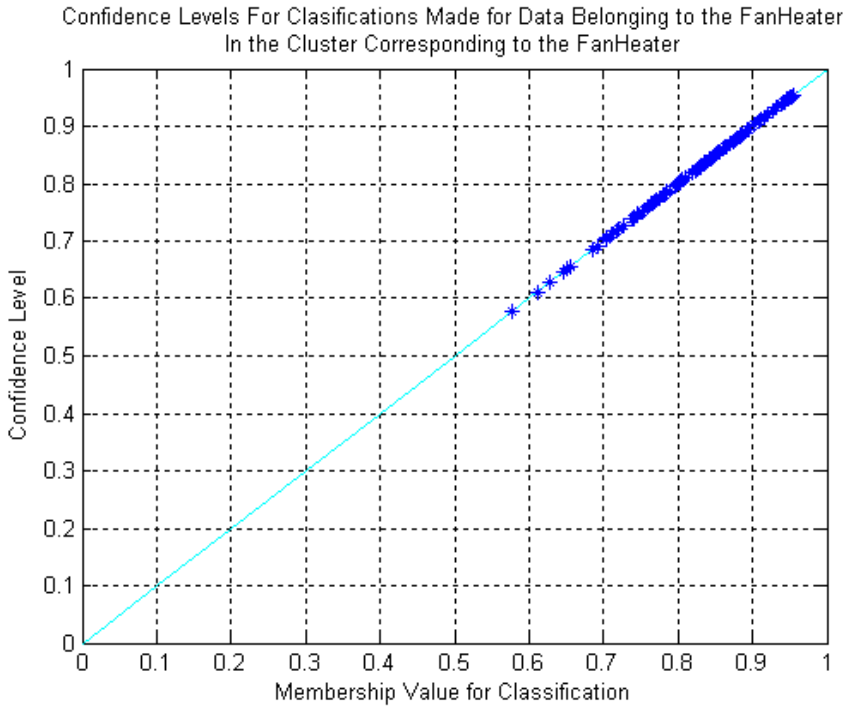


Figure 7.26: Plot of the confidence levels generated for the recognition data for the fan heater, fan only mode, in the cluster generated by the corresponding training data.

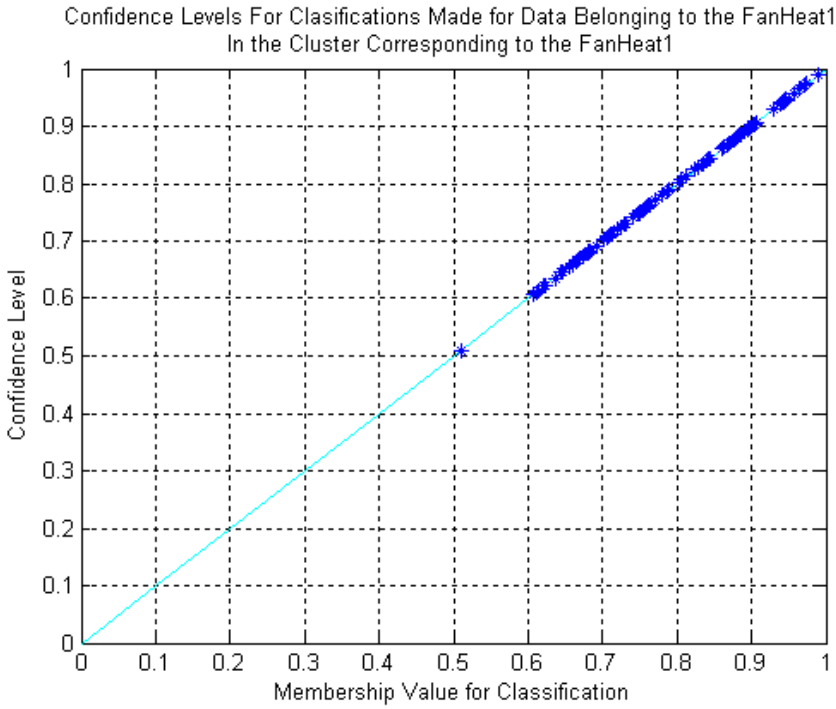


Figure 7.27: Plot of the confidence levels generated for the recognition data for the fan heater, heat mode 1, in the cluster generated by the corresponding training data.

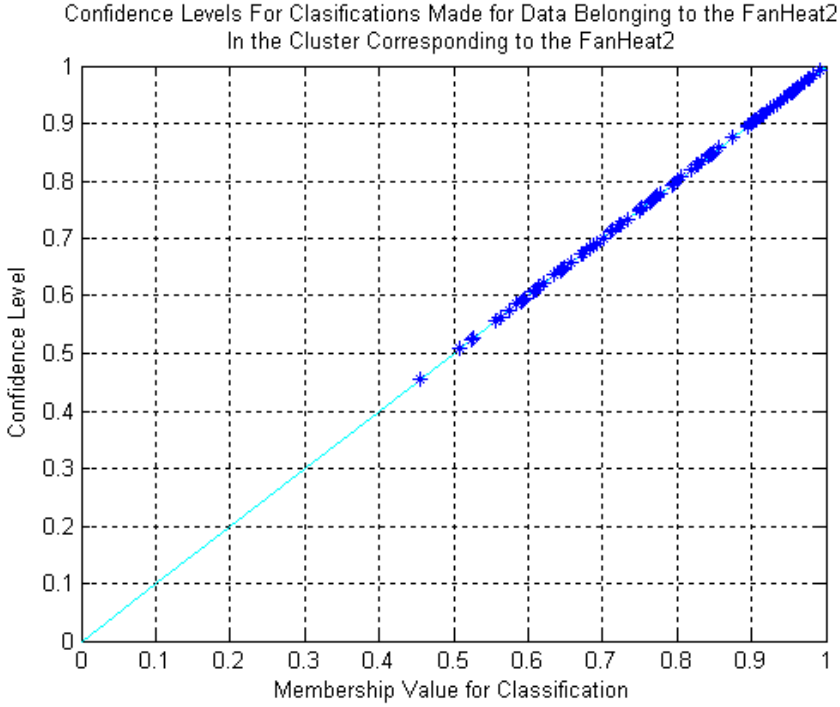


Figure 7.28: Plot of the confidence levels generated for the recognition data for the fan heater, heat mode 2, in the cluster generated by the corresponding training data.

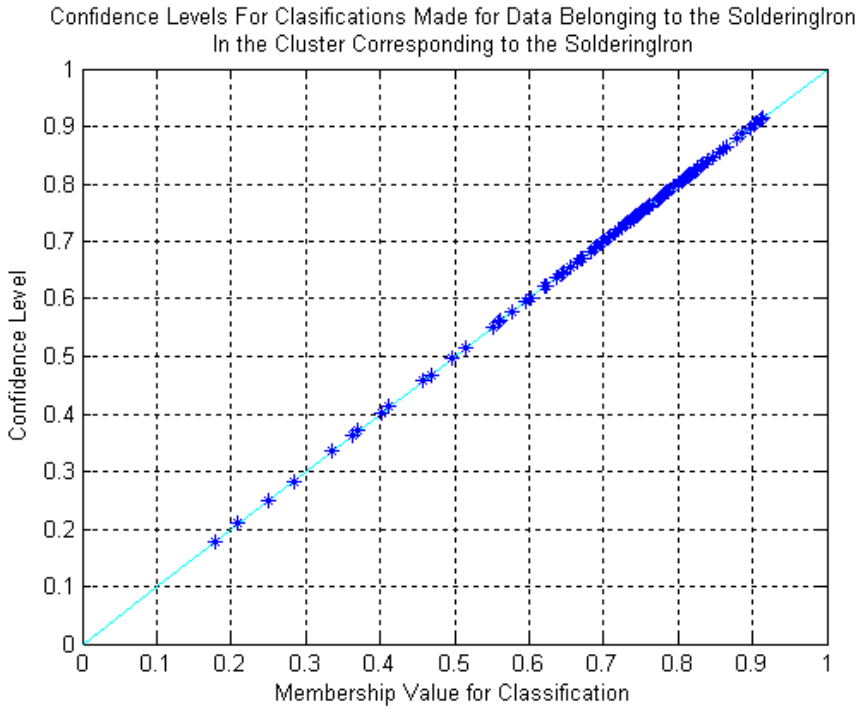


Figure 7.29: Plot of the confidence levels generated for the recognition data for the soldering iron in the cluster generated by the corresponding training data.

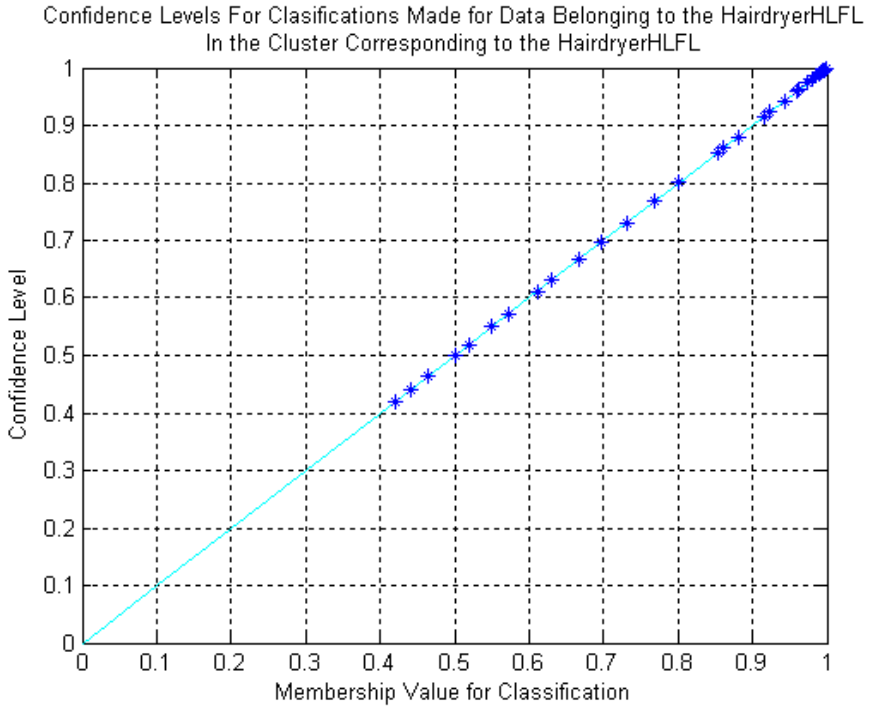


Figure 7.30: Plot of the confidence levels generated for the recognition data for the hairdryer in low heat and low fan mode in the cluster generated by the corresponding training data.

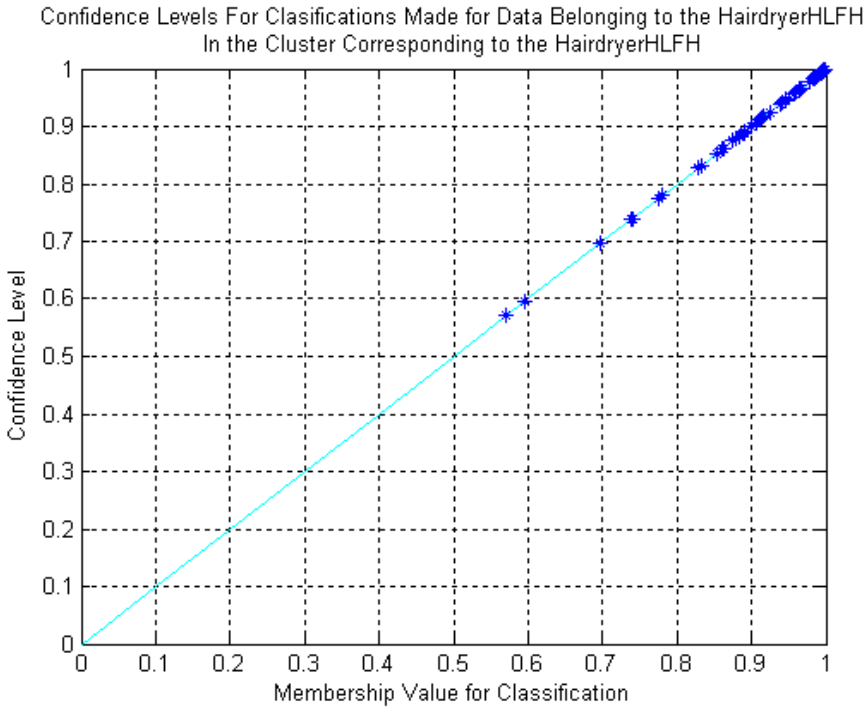


Figure 7.31: Plot of the confidence levels generated for the recognition data for the hairdryer in low heat high fan mode in the cluster generated by the corresponding training data.

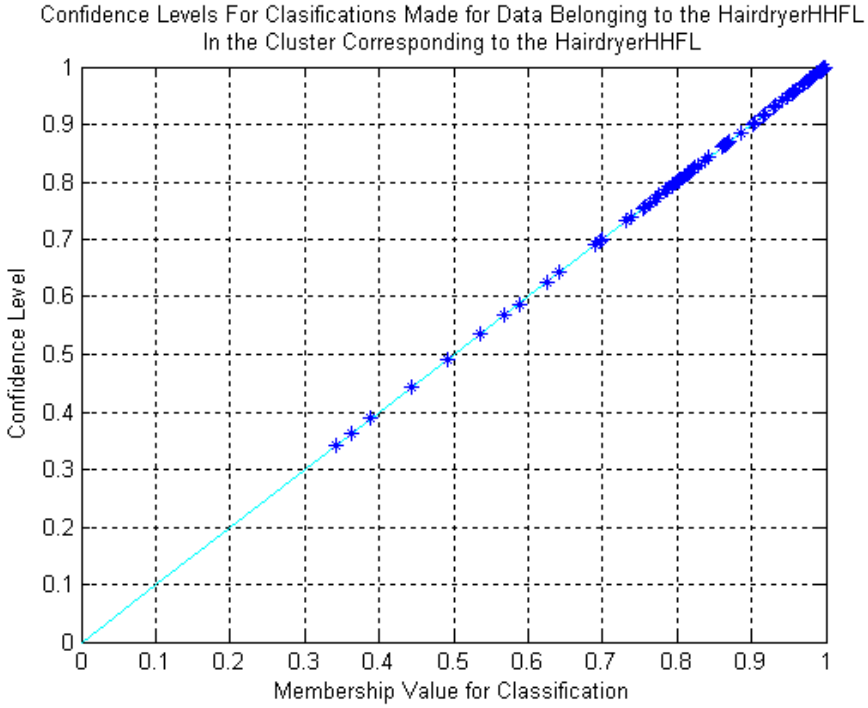


Figure 7.32: Plot of the confidence levels generated for the recognition data for the hairdryer in high heat low fan mode in the cluster generated by the corresponding training data.

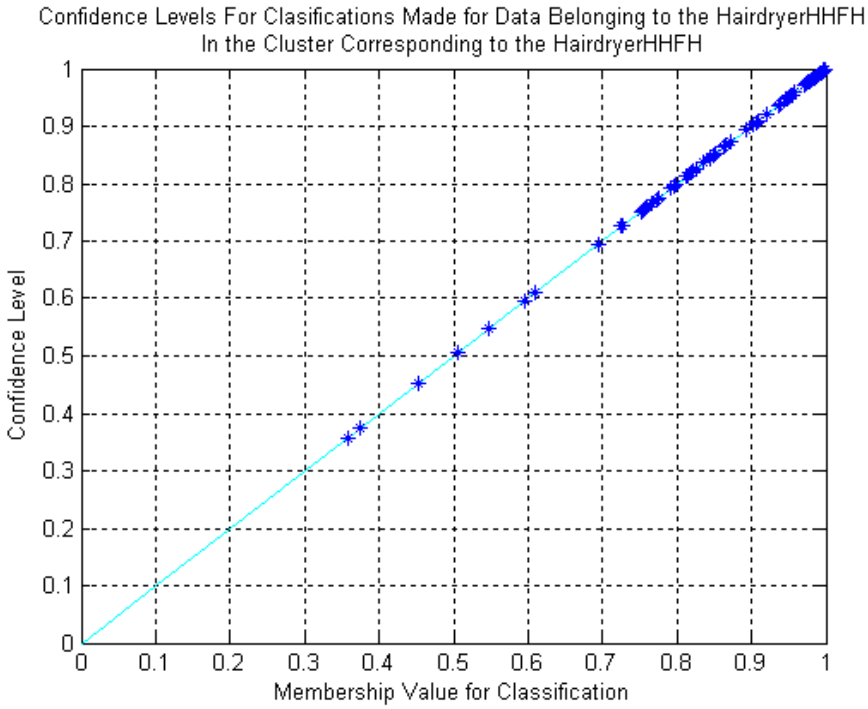


Figure 7.33: Plot of the confidence levels generated for the recognition data for the hairdryer in high heat high fan mode in the cluster generated by the corresponding training data.

From these figures it can be seen that the confidence levels lie on the line $CL = m_1$ and therefore the confidence in the classification is as strong as the classification itself. This is because in all cases the value of $m_2$ is negligible. Therefore the confidence levels lie in the range $0.2 < CL < 1.0$[8].

## 7.5 Chapter summary

This chapter has shown how the classification techniques developed in Chapters 4 and 5 are implemented in an example using real world data for 9 different appliances. Initially a training algorithm is performed on a set of training data for each of the appliances. This first determines the natural number of clusters in each appliance's data set, before determining the cluster center, covariance matrix and average intracluster distance for each of the clusters considered.

[8]The confidence level has been generated for the data point which belongs to the soldering iron which has $m < 0.2$ and therefore $CL < 0.2$, however, in the working regime of the AEM this data point is not classified and therefore no confidence level would be generated.

Once the cluster prototypes are determined, a set of labelled recognition data is taken for each of the appliances for which training data is available. For each appliance in the recognition data the correct model is first identified in order to determine the feature space within which to perform the recognition algorithm. For each cycle of the recognition data a set of membership values are generated in each of the existing cluster prototypes in the correct feature space. The membership values generated are those for the probabilistic membership, $u$, the possibilistic typicality, $t$, and the resultant membership value, $m = \sqrt[4]{(u^3 \times t)}$. Classification is performed if the resultant membership value is greater than 0.2 and for the data used in this chapter, over 99.8% of the recognition data is classified correctly.

A confidence level is developed for each of the classifications, which shows that the confidence level of the classification is as strong as the actual classification itself, as the second highest membership value generated for each data point is negligible.

Therefore, the data considered in this chapter illustrated how the current and voltage signals from an individual appliance are used in order to identify the correct appliance to a 99.8% degree accuracy using only the short term data.

# Chapter 8

# Conclusion

The aim of this thesis was to develop a set of techniques to be implemented in an Advanced Electricity Meter which provides homeowners with useful information about the way in which they use electrical energy in the home to give them the means and motivation to reduce their domestic energy consumption. So that these techniques were developed in an appropriate manner, a set of key questions were posed in order to understand the mechanism by which such a device may satisfy its objective. The work presented in this chapter recaps on how these key questions were answered and how the set of objectives was satisfied in order to enable the development of an appropriate set of techniques and on the novelty introduced in this development.

## 8.1 Key steps in the research

**How can individuals be encouraged and enabled to reduce their domestic energy consumption?**

In order to establish how to aid individuals most effectively to reduce their domestic energy consumption Chapter 2 first explored factors which influence an individuals motivation towards a task. It was found that motivation varies along a scale from extrinsic externally controlled motivation (whereby a task is undertaken in order to avoid external punishment or to receive some external reward) to intrinsic internally controlled moti-

vation which is due to the task in itself. Task performance is affected by the type of motivation an individual possesses towards the task and it is found that intrinsically motivated individuals undertake tasks with more effort and persistence than extrinsically motivated individuals. Therefore, in order to cultivate an intrinsic interest, the task should be made challenging, novel and interesting.

Levels of motivation are also affected by perceived task ability and therefore to increase an individuals coping efforts towards a task, which will in turn affect their effort and persistence, perceived self efficacy and competence levels should be heightened. This can be done by allowing individuals to set themselves goals and intermediate goals and providing them with feedback so they can see the results of their efforts.

Having established a set of factors which improve general task performance, Chapter 2 examined these factors with reference to the specific task of reducing domestic energy consumption. In order to cultivate an intrinsic interest, the method by which energy information is captured and presented to the user should not be considered a burden, but rather be challenging, novel and encourage user interaction. This implied an automated method of calculating the energy consumption and a method of feeding back this information to individuals whereby they can set themselves goals and monitor their performance in a self controlled manner.

In order to heighten perceived task ability, the feedback should allow individuals to clearly observe their energy saving endeavours. This has two implications. The first is that the feedback should be available immediately and the second is that they should be able to see the energy consumption of individual appliances.

Therefore, individuals can be encouraged and enabled to reduce their domestic energy consumption by providing them with an automated feedback system which presents them with instantaneous and disaggregated information relating to their domestic energy consumption.

**What are the design implications of such a feedback system?**

The design implications of the proposed feedback system were explored in Chapter 3 and were based on the set of design criteria established as a result of the implications on motivation and task performance. These design criteria also take into account the feasibility of such a system and it was proposed that the developed system be easy to install into existing homes and cost no more than £70 to purchase. This implies that the system must utilise simple hardware and monitor the electricity consumption of the home from a single point.

The proposed design should also allow the user to observe the energy consumption of individual appliances, which, along with the criteria discussed above, implies that the system must use complex software in order to track the use of individual appliances from the total electricity consumption. To do this, it is proposed that the AEM examine step changes occurring in the whole house power consumption (corresponding to appliances switching on and off) and from this extract an appliance signature which may be used to identify the individual appliance causing the step change in power.

**Is there any existing state of the art?**

In Chapter 3 a critical analysis was performed on other systems already in existence which satisfy the design implications of the AEM. The most fitting existing systems are based on non-intrusive appliance load monitoring (NALM), as intrusive load monitoring requires more complex hardware than is appropriate for the AEM.

The NALM system works by monitoring the total power consumption, identifying step changes in power consumption, extracting features from the step changes which correspond to the real and imaginary normalized power and uses these features in order to build state diagrams for individual appliances to subsequently track the energy consumption of these appliances.

The advantages and limitations of such a system are discussed and it is established

that one of the largest problems is due to the lack of features which may be used in order to identify appliances. Existing systems utilize only two features (the real and reactive power) from which to identify appliances, so a question was posed as to a possible extension of the feature set which may be extracted to form the appliance signature.

**What information is used to form the appliance signature?**

An appliance signature is proposed in Chapter 4 which is made up from 3 different sets of features. These feature sets correspond to the short term time domain behaviour (which relates to the shape of the current drawn and therefore to the appliance's electrical interface), the FSM behaviour observed (which may or may not be time dependent) and influences on appliance use due to the time of day.

The features contained in these different feature sets allow more information to influence the identification of appliances and therefore the accuracy with which appliances may be identified following step changes in power is improved. However, because these features exist in different domains, a framework within which the classification can take place is deemed necessary.

**How can these features be used in a framework to identify the underlying appliance?**

The mechanism by which the different feature sets may be used together in a framework to enable appliance identification was explored in Chapter 4. The set of available information (which corresponds to the three feature sets discussed above) is noted and the desired capabilities of the framework are identified. With consideration to the desired capabilities and the inter-dependency between feature sets, a Bayesian Belief Network is proposed. However, problems with this network relating to the inability to account for unknown appliances leads to the proposition of a novel membership model.

A novel framework is proposed with similar characteristic to the BBN, but which analyses memberships, rather than probabilities, for each feature set. These memberships deter-

mine the level to which each feature set belongs to each of the known appliances and these memberships may sum to zero across all appliances if the feature set is generated from an unknown appliance.

Having identified the membership values for each set of features in each applicable appliance group, the memberships are combined in a weighted geometric mean in order to allow the different feature sets to have a differing level of input to the final classification and to ensure that the addition of information does not incorrectly weaken the result.

The actual values chosen for the weightings are subject to further research and are not considered in this report. This is because the research contained in this report is concerned with the method by which the set of short term features extracted from the appliance signature are used to determine the membership values in each appliance which is 'known' to the AEM. This poses three questions. The first is concerned with how the short term features are extracted and is dealt with at a later stage. The second considers the mechanism by which the AEM can be trained to 'know' various appliances and the third explores the way in which these 'known' appliances are subsequently used in order to generate memberships for an unlabelled feature vector in order to identify the appliance to which it belongs.

**How can the AEM be trained to 'know' individual appliances?**

Chapter 5 explored a novel mechanism by which the AEM is trained to 'know' the individual appliances connected to the system. It is assumed that the features generated by an appliance in a particular mode of operation form a normal data distribution and fall into compact and disjoint clusters within the selected feature space. These clusters may be represented by a set of parameters which characterise the data set and the AEM may therefore be trained by determining the set of parameters.

However, appliances may exhibit more than one mode of operation and the features generated by each mode may be different to one another. Therefore, whilst the feature

sets generated while the appliance is in use are known to come from that appliance, the mode of operation is unknown and furthermore the total number of modes are unknown.

The novel mechanism by which the AEM may be trained is twofold and utilises two standard clustering techniques. The novelty is in the particular implementation of the established techniques to overcome the problems caused by the unknown number of modes in the data. Because the modes of operation are uncontrollable, a hierarchical clustering procedure is used to determine the natural number of clusters (i.e. the number of modes of operation) into which the data for each appliance falls. Once this value has been deduced, nearest neighbour clustering is performed whereby data points are assigned to their corresponding clusters. This allows the data from each appliance to be separated into individual modes so that each mode may be characterised separately.

Two important questions must be answered before the clusters can be characterised:

1. What information should be used to characterise each cluster?

2. How many data points are needed to characterise each cluster?

In Chapter 5 it was identified that as well as representing each mode of operation separately, each mode should be represented with as few parameters as possible, whilst still ensuring that the representation encompasses all operating points and accounts for the spread of the data. The mean, covariance and average intracluster distance are therefore chosen as parameters with which to characterise the data.

Therefore, to train the AEM to individual appliances, a set of data points is used in order to generate the parameters which represent each mode of operation. The mean of the entire data set is calculated during the hierarchical clustering procedure, however, because of the presence of slow transients in the data set, the covariance matrix is dependent on the number of data points used to characterise each cluster. To prevent generating misrepresentative covariance matrices, a Students t-test is proposed to determine the number of required data points.

Therefore, Chapter 5 illustrated the mechanism by which the features from the short term time domain are used in a novel algorithm which determines a set of representations for each mode of operation for each appliance. Storing these cluster representations enables the AEM to 'know' these appliances in order that they may subsequently be used to generate membership values for unknown data points.

**How can these cluster prototypes be used to generate membership for an unlabelled data point in order to identify the appliance to which it belongs?** Having established a set of cluster prototypes for each mode of each appliance to which the AEM has been trained to recognise, these prototypes are then used in order to generate a set of membership values for an unlabelled feature vector $\mathbf{x}$. These membership values represent the degree to which $\mathbf{x}$ belongs in each of the existing clusters and are subsequently used in order to determine the identity of the appliance from which $\mathbf{x}$ was drawn. The method by which this is done was explored in Chapter 5, whereby a novel method of membership generation based on both probability and possibility theory was proposed.

When a step change in power is observed and feature vector $\mathbf{x}$ subsequently identified, this feature vector corresponds to a single appliance changing states. Because of the assumption that only a single appliance changes states between two inspections of the system load, this implies that the membership values should be representative of a set of mutually exclusive events. Therefore a membership value based on the probability of the feature vector belonging to one appliance group over another is generated. This membership value is denoted $u_i$ for the membership of the data point in the $i^{th}$ cluster and it is calculated by considering the relative distance between the data point given by $\mathbf{x}$ and the existing cluster prototypes. Because the memberships represent mutually exclusive events, the membership values across all the clusters must sum to 1. This constraint causes problems if the data point is representative of noise or transients on the system, or of a new appliance to which the system has not been trained.

Therefore, a method of generating memberships based on the extent to which a data point is typical in each cluster without consideration of any of the surrounding clusters is proposed. This allows the membership values to be negligible in all the existing clusters if the data point is not representative of any existing appliances. This membership value is denoted $t_i$ and is function of the absolute distance from the data point to the cluster center.

From these two well documented memberships, a novel membership, $m$, was proposed which combines the probabilistic and possibilistic memberships geometrically. In the domestic model an observed appliance state change is a mutually exclusive event due to the assumption that only a single appliance changes state between subsequent observations of the system. However, the model should also account for the observed state change to belong to none of the known appliances due to the ability of an individual to update their domestic appliances at any time. Because this new membership $m$ combines $u$ and $t$ geometrically it allows both mutual exclusivity and typicality within a particular grouping to play a part in the resultant membership and this makes it appropriate for use in the domestic model.

Chapter 5 also considered the relative weightings of $u$ and $t$ in determining $m$. A constraint imposed in Chapter 4 stated that a classification will not be made unless the final membership value is above 0.2. This, along with a consideration of the distribution of $u$ and $t$ values expected for positive and negative results, allowed the relative weightings to be determined. Hence $m$ is calculated according to $m = \sqrt[4]{u^3 \times t}$. These membership values can then be used to identify the appliance based on just the short term data (by classifying the data point to the appliance in which its membership value is the largest, assuming that it is above 0.2) or can be fed into the classification network proposed in Chapter 4.

Therefore, Chapter 5 identified a novel mechanism of generating memberships for an

unlabelled data point $\mathbf{x}$ by considering the extent to which the data point belonged in each existing cluster prototype. These membership values and corresponding cluster prototypes are subsequently used to identify the appliance from which the data point was drawn.

The mechanism by which the membership values are generated and subsequently used in the proposed classification network has so far developed independently to the features which constitute the feature vector $\mathbf{x}$. However, in order to implement the proposed classification scheme, a mechanism of generating features which are compatible with the developed algorithms needed to be considered.

**How are features extracted from the current and voltage signals?**

It is established in Chapter 6 that the aim of the feature set $\mathbf{x}$ is to represent individual appliances so that data points from the same appliance are grouped together and data points from different appliances are separated within the feature space identified. In order that the selected features allow the clustering algorithm to successfully identify appliances in the feature space, the data points identified from each appliance should adhere to the following criteria.

1. The features selected should allow data points from the same operating mode of appliance to lie co-incident in the feature space, with data points from different appliance falling a great enough distance away in order to be able to differentiate between the appliance.

2. The features should be as independent as possible of any changes which may occur in the voltage signal.

3. In order that a subset of the features do not become dominant, the features should be approximately similar in magnitude.

4. In order that the features can be used to form a $d$-dimensional feature space, the features should have the same dimensions (i.e. angles and magnitudes should not

be compared within the same feature space).

5. To keep the time complexity of the clustering algorithm to a minimum, ensure that the feature set contains fewer than approximately 20 features. This will also help reduce the memory requirements of the AEM.

Three methods of feature extraction are explored in Chapter 6 and are assessed against the feature set criteria. The selected method proposes a novel mechanism which extracts features based on the observation of particular characteristics in the current and voltage signals which give rise to a model of the appliance's electrical interface. In this thesis 3 different models are considered[1]. The first two are the linear combination of real and imaginary impedances connected in series and in parallel. The third is the non-linear combination of linear components which constitute a rectifier.

For each set of voltage and current signals generated by each appliance, the three models are considered and used to calculate the corresponding parameter values. Current signals are then generated based on the observed voltage and deduced electrical interface and the errors between the generated current and actual current is then used to determine whether or not the model is correct. The feature set therefore contains an indication as to the accurate model as well as the corresponding parameter values.

This innovative method of feature extraction satisfies all of the feature set criteria, as long as the appliance falls under one of the proposed model types. If it does not a generic feature set is used which contains the power of the real and reactive 1st, 3rd and 5th current harmonics. This feature set does not satisfy all feature set criteria but it is used to prevent the feature set remaining empty when none of the interface models developed to data are appropriate. As more models are developed, the generic feature set is used

[1]The number of interface models is extended to 4 when the results produced in this chapter are analysed. The fourth model is a purely resistive electrical interface, and is proposed if either model (a) or (b) is selected whereby the voltage drop across the imaginary impedance or the current drawn by the imaginary impedance is significantly smaller than the voltage drop across the resistor or the current through the resistor respectively.

less and less often and therefore the non-satisfaction of all feature set criteria for this generic feature set is not problematic.

Chapter 6 has thus established a novel mechanism by which features are extracted from the voltage and current signals of each appliance considered, in order that they allow data points from different appliances to be separated in feature space and allow data points from the same operating mode of the same appliance to lie close together.

**How well do the techniques developed in Chapters 5 and 6 cope with the task of identifying appliances based on their short term time domain data?**

In order to demonstrate the ability of the techniques developed in Chapters 5 and 6 to identify appliances based on their short term time domain features, these techniques are applied to a set of real world data in Chapter 7.

The proposed clustering algorithm generates memberships for unlabelled feature vectors based on their probabilistic membership and possibilistic typicality in the set of existing appliance groups. Therefore, before the algorithm is applied to a set of test data, a set of training data is used to generate the cluster prototypes. Chapter 7 explores the mechanism by which the AEM is trained to 'know' a set of 9 different appliances. A set of training data is drawn from each of the appliances considered, features are extracted from each cycle of training data and for each appliance the feature sets are used to first determine the number of clusters and then to determine the cluster representations. Each appliance exhibits one mode of operation and for each mode the cluster representation (including the model to which the data belongs and the corresponding cluster center, covariance matrix and average intracluster distance) and cluster label is stored in a cluster prototype.

Having established a set of cluster prototypes, features are extracted from a set of test data which consists of 100 cycles worth of data per appliance considered. For each feature vector the correct model type is established and a set of membership values are generated

for the feature vector in each of the existing appliance groups within the identified feature space. These membership values are calculated according to the method proposed in Chapter 5.

The membership values are subsequently used to identify the appliance and to calculate the confidence level of the classification. From the data used in Chapter 7 it was shown that the proposed clustering algorithm produced the correct appliance classification for over 99.8% of the data points. The reason that the classification accuracy is less than 100% is due to the final membership value generated being below 0.2, the value deemed necessary for classification, for one of the test data points generated by the soldering iron.

Confidence levels were generated for each classification made and it was shown that the confidence levels for every classification are equal to the strength of the classification itself. Therefore, in this thesis, there are no circumstances under which a false positive classification may occur when these results are used in the classification framework.

Therefore, a mechanism by which the techniques developed in Chapters 5 and 6 are implemented on a set of real world data has been illustrated in Chapter 7. This implementation demonstrated that the ability of the classification scheme proposed to correctly identify appliances based on their short term time domain data alone is greater than 99.8%, whereby the errors are due to false negative classifications and no false positives.

## 8.2 What novel work is produced in this research?

The novelty of the work presented in this thesis is in the specific application of existing techniques to solve a unique problem.

The first novel aspect of this work is the development of the classification framework which enables each feature domain[2]to affect the determination and strength of the fi-

[2]These include short term time domain, time dependent FSM model domain and time of day domain.

nal classification. This framework proposes an innovative use of memberships (which are based on the degree to which the particular feature set belongs within a particular appliance group) combined in a weighted geometric mean to allow more confident and rigorous features to have a greater contribution to the final classification than weaker and more flexible features. This enables more information to be incorporated into appliance identification than any previous similar systems.

The second novel aspect identified in this research is in the particular application of known clustering techniques in order to characterise existing appliances in the home. Supervised and unsupervised clustering techniques are combined in an innovative manner so that a set of labelled training data with an unknown number of user modes may be characterised by a compact set of parameters which represent the distribution of the features generated by each mode.

The third contribution to knowledge is the mechanism by which short term time domain memberships, which represent the degree to which a new data point belongs to each of the known appliances, are generated. These new memberships are produced by a weighted geometric combination of two known fuzzy memberships, the probabilistic membership $u$ and the possibilistic typicality $t$, and enable both mutual exclusivity and typicality to be incorporated into the final result. The particular manner in which the memberships $u$ and $t$ are combined is such that it is appropriate for use in a domestic model.

The final innovation presented in this thesis concerns the mechanism by which features are extracted from the short term time domain signals to form the short term time domain feature set. It is not new to use models of electrical interfaces to determine the current which is drawn by that interface, however, this research uses particular characteristics in the current signal to determine the underlying interface. The new feature set proposed thus contains an indication of the electrical interface model and the corresponding set of parameter values. This feature set therefore first separates appliances by their interface

model type, reducing the number of appliances which exist in each feature space[3], before representing appliances with features which satisfy all the proposed criteria.

This section has highlighted the four novel contributions of the work presented in this thesis to the scientific community.

## 8.3 Areas of further research

There are four areas of future work which are evident in order to develop the AEM further and which are discussed in this section.

The first extension of work to this thesis is further development of the electrical interface models which are used to determine the short term time domain features. More complex models which include the operation of non-linear components within both linear and non-linear interfaces should be included. This allows interfaces to be included in the analysis which are particularly prevalent in the domestic environment (e.g. saturating transformers which are present in many cheap motors including the one contained in the electrical interface of the fan heater of Chapters 6 and 7).

The second area of further work is the further development of the classification algorithm to cope with multi-modal appliances. For and appliance with two on modes, x and y, the state transitions may be x↔y as well as off↔x and off↔y. If both modes have the same electrical interface, for example, a heater with two different heat settings, both resistive, they will have the same model type and the transition x→y may be modelled as a linear combination of x→off and off→y. However, if the different modes of operation have different electrical interfaces and thus different model types, for example a washing machine, the analysis becomes more complex as the combination of the two transitions will most likely not belong to any of the IMs developed. By considering the previous changes on the system and the known FSM models it may be possible to predict these

[3]Previous work has only considered features generic to all appliances and therefore, within each feature space there exist clusters representing all the appliances connected to the system.

more complex occurrences and enable an limited search through various combinations of events, but further research is needed to justify this.

The third extension includes the development of the FSM features and the Time of Day features within the classification framework. Research should be carried out into the effect of both these feature domains on the probability of observing particular appliance state changes. For example, a washing machine may have a number of wash cycles, each of which may be represented as a time dependent finite state machine. Once the machine has been observed to enter the first state, the probability of observing the subsequent state transitions increases at particular time intervals after the previous state transition. These state/time dependencies should be incorporated in the final membership through the classification framework. Therefore, further work includes modelling the various domestic appliances to deduce the membership generation mechanism for the FSM and time of day nodes of the classification framework, as well as deducing the particular combination of the memberships within the framework.

The forth area of further work is in the ability of the AEM to become self training. As seen in the NALM analysis, training regimes may be replaced by an algorithm which builds usage patterns of particular appliances from repetitive observations, without knowing the actual appliances. Because the AEM uses multiple feature domains, this algorithm should know the type of electrical interface possessed by the appliance, as well as typical state diagrams and time of day usage patterns. With this large amount of information it may be possible for the AEM to deduce the appliance from the database of information present. Other possibilities include on-line cataloguing of information, much like that present in music systems such as iTunes, so that the AEM may use the feature patterns built up about appliances along with some on-line system to deduce the specific underlying appliance.

This section has outlined four areas of future work which are needed in order to further

develop the AEM in the hope that the techniques developed in this thesis may contribute to the development of a smart system which may accurately identify the electrical energy consumption of individual appliances in the home.

## 8.4 Publications

Early in the development of this research a patent application was made but subsequently dropped due to it's similarity to the work done in this area by Hart. However, as the research presented in this thesis has progressed it has become evident that the reason for the apparent similarity was due to the lack of detail in the original patent. Consequently, the option to re-apply is currently being considered.

At the time of submission of this thesis there are no outstanding publications of this work. This is due to reasons relating to IP issues. However, a series of papers are planned for submission to Energy and Buildings, an international journal by Elsevier which investigates the energy use and energy efficiency in buildings.

# Appendix A

# Apparatus for Data Collection

This appendix provides details into the apparatus used to collect the data. The schematics are provided in Figure A.1.

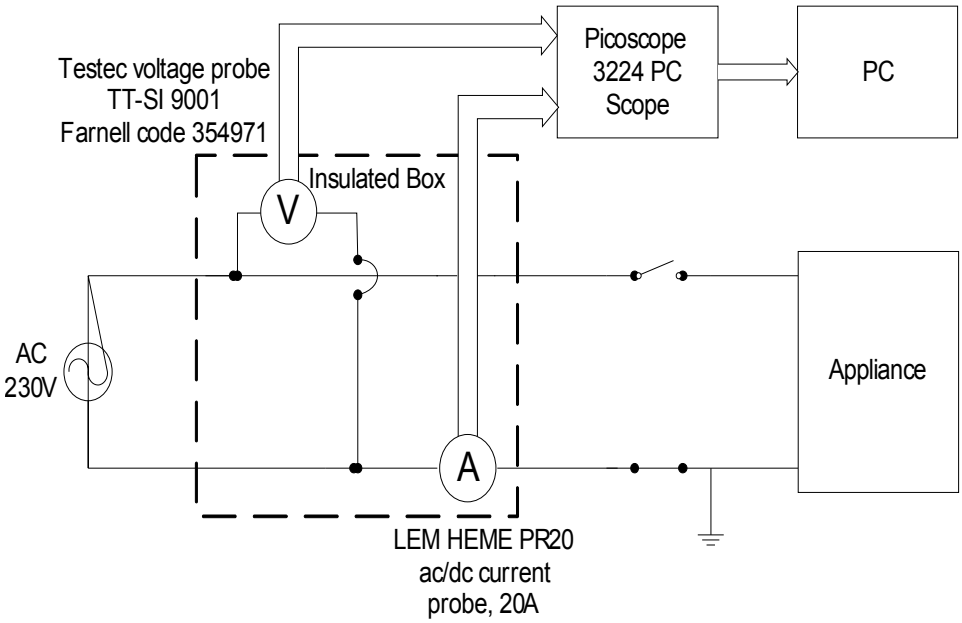


Figure A.1: Illustration of the apparatus used to collect the data relating to the voltage across and the current drawn by individual appliances.

An appliance is connected to the mains voltage supply across the live and neutral terminals. Between the mains point and the appliance is an insulated box (to protect the user from the voltage which lies in the region of 230V rms) in which the voltage and current measurements take place.

The voltage is measured using a Testec voltage probe connected between the live and

neutral terminals. This probe measures the differential voltage between these terminals. To account for the differential magnitude (i.e. up to approximately ±350V maximum differential amplitude) the attenuation factor is set to 100 at which the cut off frequency is 25MHz.

The current is measured using a LEM current probe connected around the neutral wire. This probe measures current up to a maximum value of ±30A at a resolution of 1mA. The output sensitivity of the probe is 100mV/A with a cut off frequency of 20kHz. The fundamental frequency is 50Hz, and therefore up to 400 harmonics may be accounted for without any attenuation, which is more than enough to correctly represent the current drawn by the types of appliances considered in this research.

These measurements are fed into a Picotech oscilloscope. This scope has an input voltage range of ±20mV up to ±20V in 10 different ranges. The software may be set to automatically select the desired range for the most accurate readings. The bandwidth of the scope is 10MHz and the data is sampled at 256 samples/cycle (i.e. 12.8kHz) using a 12 bit ADC. The sampling of each cycle is begun when the voltage crosses the 0V axis from positive to negative. The readings are then stored to a file on the PC.

# Appendix B

# Hidden Markov Models

In certain scenarios in the domestic environment a Hidden Markov Model (HMM) may be appropriate for generating probabilities of observing a particular sequence of events. A HMM becomes very complex as the number of possible states is increased (for a system with N linked appliances or modes of operations there are $2^N$ possible stated) and therefore this model should only be used for FSM models which link together different operating modes within an appliance, or 2 appliances which are used together, for example a computer and a monitor.

When a state transition is observed, the actual state of affairs is in itself unobservable and the actual observation relates to the short term features seen in the current waveform. A HMM is one in which the state of the system itself cannot be seen, but the output relating to a particular state can be observed. This appendix therefore explores the development of an HMM appropriate to the domestic system.

The state of the system relates to which appliances are in which state at any one time. Each state of the system is influenced only by the previous state, which is a phenomenon observed with some appliances in the home. For example, once a computer has been switched on the likelihood of a monitor being switched on increases from what it would have been if the PC had not been switched on. This process is illustrated in Figure B.1 with the total number of appliances limited to just two, the monitor and the PC. There

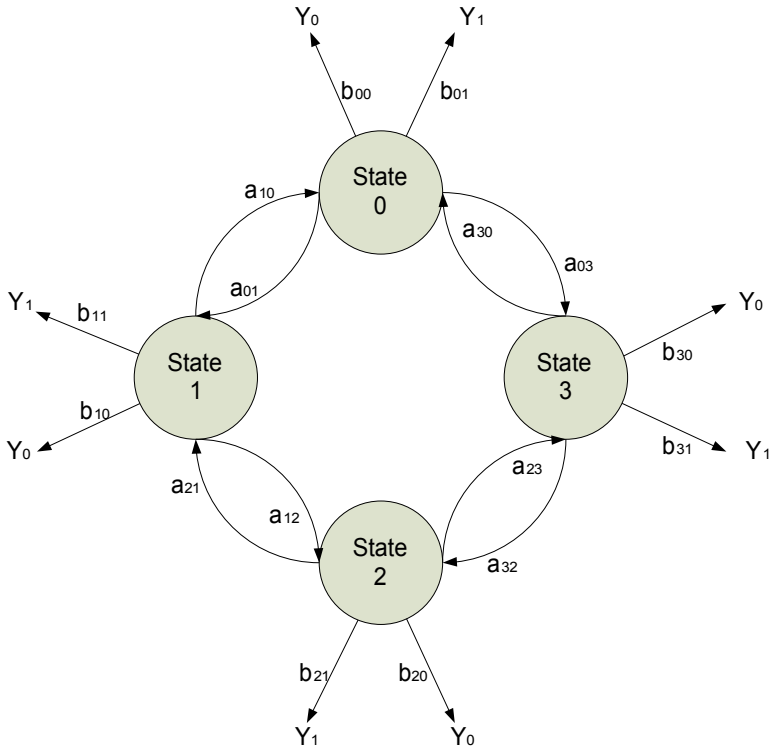


Figure B.1: Transition diagram for a Hidden Markov Model.

are 4 possible options for the states:

- State 0 — PC off and monitor off
- State 1 — PC off and monitor on
- State 2 — PC on and monitor on
- State 3 — PC on and monitor off

There are two potential observations in this example if the sign of the observation is ignored (i.e. the features corresponding to a switch off are identical to those for a switch on aside from the direction of the change). These observations relate to the feature set obtained when the PC is switched on or off, $(Y_0)$, and the feature set from switching the monitor on or off, $(Y_1)$. The probability of observing event $Y_k$ when in State $j$ is given by $b_{jk}$. The $a's$ are also probabilities, but these relate to state transitions, where $a_{ij}$ is the probability of the next state being $j$ given that the current state is $i$.

Given a set of training data it is possible for the model to learn the parameters $a_{ij}$ and $b_{jk}$ for all $i, j$ and $k$ using the Forward-Backward algorithm [65]. Once these parameters

have been identified and the model completed it is then possible to predict the state changes following the observable outputs by calculating the likelihood of particular paths and concluding on the most likely path [65].

This model exhibits some behaviour which is appropriate to the particular system under consideration, but there are also problems associated with it. One problem is to do with the observable outputs, Y, and the probabilities of these observations given a particular state. This is demonstrated by considering the simple HMM above.

If the system undergoes the state transition $0 \rightarrow 1$, the underlying transition is that of the monitor switching on, so the probability of observing the feature set $Y_1$ is very high whilst the probability of observing the feature set $Y_0$ is low. This implies that $b_{11}$ is large and $b_{10}$ is small. However, now consider the transition $2 \rightarrow 1$, i.e. the PC switching off and monitor remaining the same. This implies a small value for $b_{11}$ and a large one for $b_{10}$. These 2 situations cannot both be true and the error is because the observable outputs are indicative of the change in the system and not of the state of the system. It may be possible to overcome this problem by adapting the model slightly so that the observable outputs of the system are related to the changes rather than the states as illustrated in Figure B.2.

This model also has a number of problems which can be observed by considering the transition of the PC remaining the same and the monitor switching on. This would result in the observation of event $Y_1$, although it would not be clear from this information alone whether the state transition occurring is from State 3 to State 2 or State 0 to State 1. There are therefore two possible options for the current state, 1 or 2. This becomes more complex as more interconnected appliances or interconnected states within appliances are added to the diagram and there are many options for the states of other appliances to remain the same whilst a single appliances switches state.

Another problem with the Hidden Markov Model is its inability to account for time.

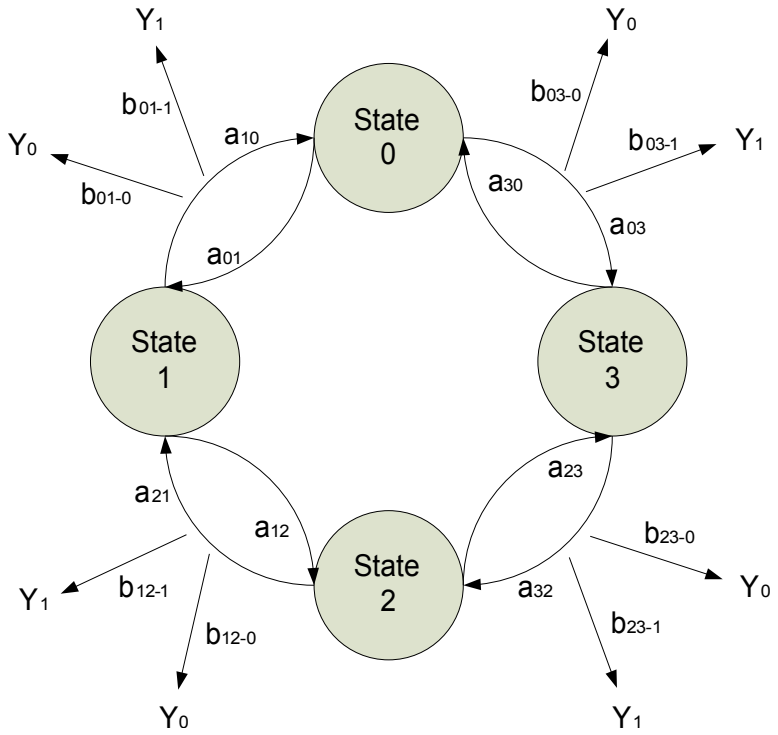


Figure B.2: Transition diagram for a Hidden Markov Model where the observable outputs relate to the transitions between states rather than actual states.

The state transitions are solely dependent on the previous state of the system and not on the time lapse between state changes. Consider the simple example used in this section of the computer and monitor. If the initial state of the system is 0 so that both the appliances are off, the probability of either the monitor or computer being switched on is independent of time. This means that the probability of transitions 0 to 1 or 0 to 3 ($a_{01}$ and $a_{03}$ respectively) are independent of the time lapse between the state changes. However, once one of these appliances has been switched on, the probability of observing the other appliance being switched on becomes dependent on the time lapse between the state changes. In this example it is likely that the second appliance will be switched on within a short time period of the first being switched on and the probability of the observation will decrease with time. This problem can be overcome by making the probabilities $a_{12}$ and $a_{32}$ functions of time.

Therefore, if a particular FSM type behaviour is observed, a HMM in which the observable outputs are indicative of a state transition and where the probabilities are functions of

time may prove useful in generating a probability of observing a future transition due to this state like behaviour. This will feed into the respective node in the classification framework.

# Appendix C

# Hierarchical Clustering

This appendix explores the way in which a hierarchical clustering algorithm is used in order to determine the number of natural groupings within a data set.

## C.1 Producing a nested set of clusters

Consider a set of $n$ data points which exist in $d$-dimensional feature space. The data points exist in a number of natural groupings $M$ but this number is unknown. The following algorithm [53] is implemented in order to produce a set of nested clusters from which the natural number of clusters can subsequently be identified.

1. Begin by assuming that each data point is a separate cluster in it's own right and therefore the number of clusters $K$ is equal to $n$. The cluster centres are therefore the locations of the individual data points.
2. Reduce the value of $K$ to $n-1$ and determine the new cluster centers. This is done by merging the two closest clusters and recalculating the new cluster center.
3. Repeat Step 2) until only a single cluster remains.

There are a number of ways in which the clusters can be merged at each step which depends on the definition of closeness used to define the distance between the two 'closest' clusters. The following paragraphs explore a few different standard techniques.

### Single linkage

In a single linkage clustering algorithm the closeness between a pair of clusters is defined as a function of the distance between the closest two data points in the clusters [66]. This can be defined mathematically as:

$$d(c_r, c_s) = \min\left(dist\left(x_{r,i}, x_{s,j}\right)\right) \quad i \in (1, 2, \ldots, n_r) \quad j \in (1, 2, \ldots, n_s) \tag{C.1}$$

Where $c_r$ and $c_s$ are clusters $r$ and $s$ respectively, with corresponding number of data points $n_r$ and $n_s$. This mechanism of clustering will join two clusters if just two points are similar, regardless of the similarity between the other points in the clusters. This tends to produce clusters that are elongated and it does not cope well with noisy data [53].

### Complete linkage

By contrast with the single linkage algorithm, the complete linkage defines the distance between the two closest clusters as the maximum distance between any of the points in the two clusters [67].

$$d(c_r, c_s) = \max\left(dist\left(x_{r,i}, x_{s,j}\right)\right) \quad i \in (1, 2, \ldots, n_r) \quad j \in (1, 2, \ldots, n_s) \tag{C.2}$$

This tends to favour clusters which are compact and disjointed in feature space, which is not entirely suitable for this application, as the clusters produced do tend to be slightly elongated due to the correlation of some features.

### Average linkage

This computes the average distance between cluster $r$ and $s$ by defining the distance as the average of all pairwise combinations between the two clusters $r$ and $s$ [66] such that:

$$d(c_r, c_s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{i=1}^{n_s} dist\left(x_{r,i}, x_{s,j}\right) \tag{C.3}$$

This method allows all the data points to have some input into the closeness measure between the two clusters so should avoid the problem in single linkage of merging two clusters if two data points are similar even if the others are not. However, each time two clusters are merged to form a reduced cluster set, the distance between the new cluster and the existing ones needs to be recomputed to include all the data points in the new cluster.

**Ward's linkage**

Ward's clustering mechanism [68] reduces the number of clusters from $n$ to $n-1$ by joining together the two clusters such that an objective function is minimised. This objective function is related to the error increased caused by merging two clusters. Define the error in present in each cluster as the sum of squares distance between each data point and cluster center, so for cluster $j$ which contains $n_j$ data points, the error is defined as:

$$e\left(j\right) = \sum_{i=1}^{n_j} x_i^2 - \frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_i\right)^2 \tag{C.4}$$

When two clusters are merged the error always increases, so the definition of the two closest clusters are those which result in the smallest increase of the total sum of squares error over all clusters, giving a distance measure:

$$d^2(c_r, c_s) = \frac{n_r n_s}{n_r + n_s} \left\|\bar{x}_r - \bar{x}_s\right\|_2^2 \tag{C.5}$$

It has been chosen to use Ward's Linkage for the algorithm due to its direct link to the error introduced by merging two clusters, rather than just merging the two apparently nearest in Euclidean distance to each other.

Having established the way in which to measure the distance between clusters, a hierarchical clustering algorithm is applied to the data. For illustration purposes this is performed on a data set consisting of 4 natural groupings and the dendrogram as illustrated in Figure C.1 is generated. As the distance measure decreases, the number of clusters increases one at a time, though the distance between these events is varying.

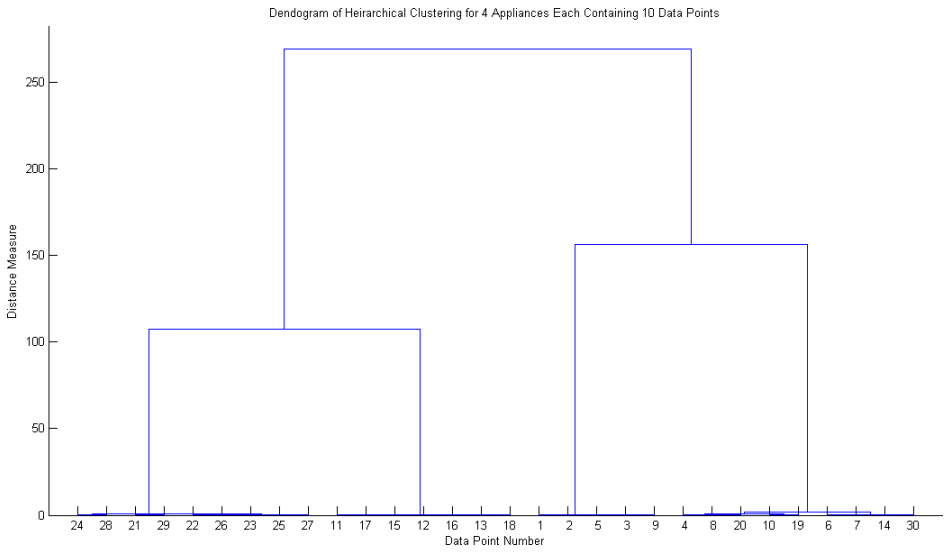


Figure C.1: Dendrogram illustrating hierarchical clustering for 4 appliances

From the dendrogram in Figure C.1 it seems fairly obvious that the natural grouping of data is into 4 clusters, however, by considering the dendrogram at a different scale as shown in Figure C.2, it becomes less obvious to see the actual natural number of clusters. In order to fully determine the true number of natural clusters, it is important to have some means of testing the data to see if further splits in the number of clusters results in a better natural grouping.

## C.2 Determining the number of clusters

This section considers a mechanism by which the natural number of clusters can be determined from the obtained nested set of clusters. In order to provide some mathematical grounding to the mechanism, an error measurement is used to determine whether or not a given cluster should be divided into two separate smaller clusters.

A common error measurement is the sum of squares error:

$$e^2(\mathscr{X}, \mathscr{L}, K) = \sum_{j=1}^{K} \sum_{i=1}^{n_j} \left\| \mathbf{x}_i^{(j)} - \mathbf{c}_j \right\|^2 \tag{C.6}$$

where there are $K$ clusters in a cluster set $\mathscr{L}$ in the feature set $\mathscr{X}$. This error will reduce

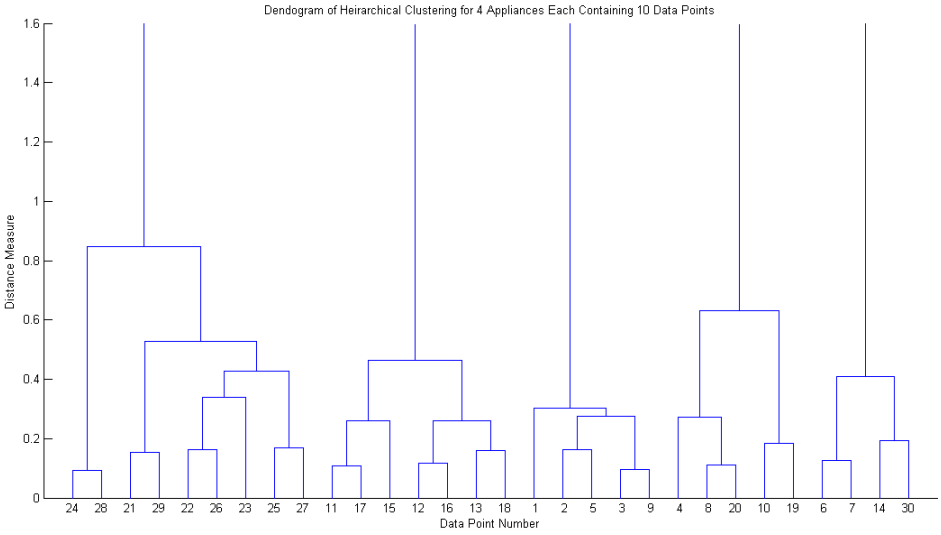


Figure C.2: Dendrogram illustrating hierarchical clustering for 4 appliances at a lower scale

as the cluster set $\mathscr{L}$, selected from the feature set $\mathscr{X}$, becomes a better representation of the actual grouping of the data.

However, this error alone cannot be used as an objective function, otherwise the trivial solution of $n$ clusters ($n$ is the total number of data points) will result, corresponding to a zero error! The error will always reduce as the number of clusters increases.

Assuming there are a natural number of $M$ clusters, the error should decrease rapidly as the number of clusters are increased from 1 to $M$ and then the error should decrease slowly as the number of clusters is further increased. It is therefore necessary to understand what constitutes a statistically significant reduction of the error. This can be done by the following hypothesis testing taken from 'Pattern Classification' by R. Duda, P. Hart and D. Stork [65].

Postulate a null hypothesis of the data containing $K$ clusters. Begin with $K = 1$. The aim of the test is to determine whether there is any reason to assume that the data comes from more than just a single cluster.

**Null hypothesis** All $n$ samples come from a normally distributed population of mean $\mu$ and covariance $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

If the null hypothesis is true then any multiple clusters found would be formed by chance and therefore any reduction in the error as a result of multiple clusters would have to be considered insignificant. Considering the error calculation for a single cluster (and dropping the $\mathscr{X}$ and $\mathscr{L}$ as the error will be assumed to be minimised for the particular number of clusters under investigation) the error resulting from a single cluster will be:

$$e^2(1) = \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{c}\|^2 \tag{C.7}$$

Under the null hypothesis, this error should be normal with mean $nd\sigma^2$ and variance $2nd\sigma^2$, where $d$ is the dimension number of the data. If the data is now partitioned into two subsets, there error will be:

$$e^2(2) = \sum_{j=1}^{2} \sum_{i=1}^{n_j} \left\|\mathbf{x}_i^{(j)} - \mathbf{c}_j\right\|^2 \tag{C.8}$$

As the optimal partitioning is not known the answer cannot be exact, however, an approximation can be made by considering the sub-optimal partitioning of the data through the sample mean. The error resulting in this partition should be approximately normal with mean $n\left(d - \frac{2}{\pi}\right)\sigma^2$ and variance $2n\left(d - \frac{8}{\pi^2}\right)\sigma^4$, demonstrating that the error with two clusters is indeed smaller than that for a single cluster. By considering this sub-optimal partition as nearly optimal and assuming a normal approximation for the sampling distribution, the critical value for this $e^2(2)$ can be approximated. Estimate $\sigma^2$ as:

$$\hat{\sigma}^2 = \frac{1}{nd} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{c}\|^2 = \frac{1}{nd} e^2(1) \tag{C.9}$$

Therefore, reject the null hypothesis at the $p$-percent significance level if:

$$\frac{e^2(2)}{e^2(1)} < 1 - \frac{2}{\pi d} - \alpha\sqrt{\frac{2\left(1 - 8/\pi^2 d\right)}{nd}} \tag{C.10}$$

Where $\alpha$ can be determined by:

$$p = 100 \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = 50\left(1 - \mathbf{erf}\left(\alpha/\sqrt{2}\right)\right) \tag{C.11}$$

Recalling that $p$ is the percentage significance level and **erf** is the standard error function.

Therefore, by considering the error at each step and the critical value as determined from the data set, the natural number of clusters, $M$, is determined.

# Bibliography

[1] IPCC, *Climate Change 2007: Impacts, Adaptation and Vunerability. Working Group II Contribution to the Intergovernmental Panel on Climate Change Forth Assesment Report*, 2007.

[2] P. Brohan, J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, "Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850," *Journal of Geophysical Research*, vol. 111, June 2006.

[3] J. R. Petit, J. Jouzel, D. Raynaud, N. I. Barkov, J. Barnola, I. Basile, M. Bender, J. Chappellaz, M. Davisk, G. Delaygue, M. Delmotte, V. M. Kotlyakov, M. Legrand, V. Y. Lipenkov, C. Lorius, L. Pepin, C. Ritz, E. Saltzmank, and M. Stievenard, "Climate and atmospheric history of the past 420,000 years from the vostok ice core, antarctica," *Nature*, vol. 399, pp. 429–436, June 1999.

[4] B. Bolin, E. T. Degens, P. Duvigneaud, and S. Kempe, *The Global Carbon Cycle*, ch. The Global Biogeochemical Carbon Cycle. Wiley on behalf of the Scientific Committee On Problems of the Environment (SCOPE), 1979.

[5] W. F. Ruddiman, "The anthropogenic greenhouse era began thousands of years ago," *Climatic Change*, vol. 61, pp. 261–293, 2003.

[6] B. Bolin and E. Eriksson, *The Atmosphere and the Sea in Motion*, ch. Changes in the Carbon Dioxide Content of the Atmosphere and Sea due to Fossil Fuel Combustion, pp. 130–142. Rockefeller Institute Press, New York, 1959.

[7] International Energy Agency, *World Energy Outlook 2006*. OECD/IEA, Paris, 2006.

[8] Department of Trade and Industry, *Meeting the Energy Challenge: A White Paper on Energy*. TSO (The Stationary Office), May 2007.

[9] Department of Trade and Industry, *The Energy Challenge Energy Review Report*, July 2006.

[10] G. Wood and M. Newborough, "Dynamic energy consumption indicators for domestic appliances: Environment, behaviour and design," *Energy and Buildings*, vol. 35, pp. 821–841, 2003.

[11] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemparary Educational Psychology*, vol. 25, pp. 54–67, 2000.

[12] R. M. Ryan and E. L. Deci, "Self–determination theory and the facilitation of intrinsic motivation, social development and well–being," *American Psychologist*, vol. 55, pp. 68–78, 2000.

[13] M. Kent, *The Oxford Dictionary of Sports Science*, ch. Locus of Causality. Oxford University Press, 2007.

[14] A. Bandura, "Self–efficacy mechanism in human agency," *American Psychologist*, vol. 37, pp. 122–147, 1982.

[15] A. Wigfield, "Expectancy–value theory of motivation.," *Contemparary Educational Psychology*, vol. 25, pp. 68–81, 2000.

[16] A. Bandura and N. E. Adams, "Analysis of self–efficacy theory of behavioural change," *Cognitive Therapy and Research*, vol. 1, pp. 287–310, 1977.

[17] A. Bandura, "Human agency in social cognitive theory," *American Psychologist*, vol. 44, pp. 1175–1184, 1989.

[18] G. Gaskell and R. Pike, "Residential energy use: An investigation of consumers and conservation strategies," *Journal of Consumer Policy*, vol. 6, pp. 285–302, 1983.

[19] J. H. V. Houwelingen and W. F. V. Raaij, "The effect of goal–setting and daily electronic feedback on in–home energy use," *Journal of Consumer Research*, vol. 16, pp. 98–105, 1989.

[20] G. Brandon and A. Lewis, "Reducing household energy consumption; a qualitative and quantitative field study," *Journal of Environmental Psychology*, vol. 19, pp. 75–85, 1999.

[21] H. Wilhite and R. Ling, "Measured energy savings from a more informative energy bill," *Energy and Buildings*, vol. 22, pp. 145–155, 1995.

[22] M. L. Dennis, E. J. Soderstrom, W. S. K. Jr., and B. Cavanaugh, "Effective dissemination of energy–related information," *American Psychologist*, vol. 45, pp. 1109–1117, 1990.

[23] T. Ueno, F. Sano, O. Saeki, and K. Tsuki, "Effectiveness of an energy–consumption information system on energy savings in residential houses based on monitored data," *Applied Energy*, vol. 83, pp. 166–183, 2006.

[24] T. Ueno, R. Inada, O. Saeki, and K. Tsuji, "Effectiveness of an energy–consumption information system for residential buildings," *Applied Energy*, vol. 83, pp. 868–883, 2006.

[25] BERR: Department for Business Enterprise and Regulatory Reform, *Quarterly Energy Prices*, June 2008.

[26] G. Barnwell, "Your contribution to global warming," *National Wildlife*, p. 53, February – March 1990.

[27] G. Owen and J. Ward, *Smart Meters: Commercial, Policy and Regulatory Drivers*. Sustainability First, March 2006.

[28] G. Owen and J. Ward, *Smart Meters in Great Briton: the next steps?* Sustainability First, July 2006.

[29] S. Marvin, H. Chappells, and S. Guy, "Pathways of smart metering development: shaping environmental innovation," *Computers, Environment and Urban Systems*, vol. 23, pp. 109–126, 1999.

[30] G. W. Hart, "Non intrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[31] R. McWilliam and A. Purvis, "Potential uses of embedded RFID in appliance identification," *International Appliance Manufacturing*, vol. 10, pp. 100–107, 2006.

[32] Standards and Technical Regulations Directorate, *Electrical Supply Tolerances and Electrical Appliance Safety*, July 2005.

[33] M. Baranski and J. Voss, "Genetic algorithm for pattern detection in NIALM systems," in *2004 IEEE International Conference on Systems, Man and Cybernetics*, pp. 3462–3468, 2004.

[34] L. K. Norford and S. B. Leeb, "Non–intrusive electrical load monitoring in commercial buildings based on steady–state and transient load–detection algorithms," *Energy and Buildings*, vol. 24, pp. 51–64, 1996.

[35] M. R. Durling, Z. Ren, N. Visnevski, and L. E. Ray, "Cognitive electric power meter." UNITED States Patent Application Publication, February 2009.

[36] K. Yoshimoto, "Development of non-intrusive appliance monitoring system (2): Inference of load consumption of household appliances," *CRIEPI Report T00010*, pp. 68–69, February 2001.

[37] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[38] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. London Academic Press, 2003.

[39] C. N. Jardine, "Synthesis of high resolution domestic electricity load profiles," in *First International Conference and Workshop on Micro–Cogeneration Technologies and Applications National Arts Centre*, 2008.

[40] I. Ben-Gal, *Encyclopedia of Statistics in Quality and Reliability*. John Wiley and Sons, 2007.

[41] D. Heckerman and M. P. Wellman, "Bayesian networks," *Communications of the ACM*, vol. 38, pp. 27–30, 1995.

[42] L. A. Alexandre, A. C. Campilho, and M. Kamel, "On combining classifiers using sum and product rules," *Patter Recognition Letters*, vol. 22, pp. 1283–1289, 2001.

[43] Y. Lu and F. Yamaoka, "Fuzzy integration of classification results," *Pattern Recognition*, vol. 30, pp. 1877–1891, 1997.

[44] L. Devroye, L. Gyrfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. Springer, 1996.

[45] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical Methods for Physics and Engineering*. Cambridge University Press, 2006.

[46] A. Howatson, P. Lund, and J. Todd, *Engineering Tables and Data*. Chapman and Hall, 1991.

[47] A. J. Griffiths, W. M. Gelbart, R. C. Lewontin, and J. H. Miller, *Modern Genetic Analysis, 2nd Edition*. W. H. Freeman, 2002.

[48] H. Timm, C. Borgelt, C. Doring, and R. Kruse, "An extension to possibilistic fuzzy cluster analysis," *Fuzzy Sets and Systems*, vol. 147, pp. 3–16, 2004.

[49] J. Orwant, J. Hietaniemi, and J. Macdonald, *Mastering algorithms with Perl*. O'Reilly and Associates, Inc., 1999.

[50] A. Webb, *Statistical pattern recognition*. Arnold, 1999.

[51] R. D. Maesschalck, D. Jouan-Rimbaud, and D. Massart, "The mahalanobis distance," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, pp. 1–18, 2000.

[52] L. A. Zadeh, "Fuzzy logic and approximate reasoning (in memory of Grigore Moisil)," *Synthese*, vol. 30, pp. 407–428, 1975.

[53] A. Jain, M. Murty, and P. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[54] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, pp. 90–110, 1993.

[55] H. Sun, S. Wang, and Q. Jiang, "FCM–based model selection algorithms for determining the number of clusters," *Pattern Recognition*, vol. 37, pp. 2027–2037, 2004.

[56] R. Krishnapuram and J. M. Keller, "The possibilistic c–means algorithm: Insights and recommendation," *IEEE Transactions on Fuzzy Systems*, vol. 4, pp. 385–393, 1996.

[57] F. Masulli and S. Rovetta, "Soft transition from probabilistic to possibilistic fuzzy clustering," *IEEE Transactions on Fuzzy Systems*, vol. 14, pp. 516–527, 2006.

[58] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c–means clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 13, pp. 517–530, 2005.

[59] J. H. Bentley and H. E. Shaalan, *Electrical engineering: A referenced review*. Kaplan AEC Education, 2005.

[60] L. Wuidart, "Understanding power factor," *ST Microelectronics Application Note AN824/1003*, 2003.

[61] C. Chatfield, *The Analysis of Time Series: An Introduction*. Chapman and Hall/CRC, 2004.

[62] P. S. Addison, *The Illustrated Wavelet Transform Handbook*. Bristol Institute of Physics, 2002.

[63] A. Robbins and W. C. Miller, *Circuit analysis with devices: theory and practice*. Thomson Delmar Learning, 2004.

[64] R. C. Dugan, S. Santoso, and M. F. McGranaghan, *Electrical power systems quality*. McGraw-Hill, 2003.

[65] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley and Sons, Inc., 2001.

[66] H. K. Seifoddini, "Single linkage versus average linkage clustering in machine cells formation application," *Computers and Industrial Engineering*, vol. 16, pp. 419–426, 1989.

[67] B. King, "Step–wise clustering procedures," *Journal of the American Statistical Association*, vol. 62, pp. 86–101, 1967.

[68] J. Joe H. Ward, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, pp. 236–244, 1963.