Building and Environment 100 (2016) 1–9

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

# A novel method of determining events in combination gas boilers: Assessing the feasibility of a passive acoustic sensor

Thomas Neeld [a, *], James Eaton [b], Patrick A. Naylor [b], David Shipworth [a]

[a] UCL Energy Institute, 14 Upper Woburn Place, London, WC1H 0NN, UK
[b] Dept. of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ, UK

ARTICLE INFO

*Article history:*
Received 3 December 2015
Received in revised form
28 January 2016
Accepted 29 January 2016
Available online 3 February 2016

*Keywords:*
Energy monitoring
Acoustic classification

ABSTRACT

To assess the impact of interventions designed to reduce residential space heating demand, investigators must be armed with field-trial applicable techniques that accurately measure space heating energy use. This study assesses the feasibility of using a passive acoustic sensor to detect gas consumption events in domestic combination gas-fired boilers (C-GFBs). The investigation has shown, for the C-GFB investigated, the following events are discernible using a passive acoustic sensor: demand type (hot water or central heating); boiler ignition time; and pre-mix fan motor speed. A detection algorithm was developed to automatically identify demand type and burner ignition time with accuracies of 100% and 97% respectfully. Demand type was determined by training a naive Bayes classifier on 20 features of the acoustic profile at the start of a demand event. Burner ignition was determined by detecting low frequency (5–10 Hz) pressure pulsations produced during ignition. The acoustic signatures of the pre-mix fan and circulation-pump were identified manually. Additional work is required to detect burner duration, deal with detection in the presence of increased noise and expand the range of boilers investigated. There are considerable implications resulting from the widespread use of such techniques on improving understanding of space heating demand.



## 1. Introduction

### 1.1. Motivation

Approximately 72% of energy consumption in the domestic environment is done so in gas-boilers for the purposes of domestic central heating (DCH) and domestic hot water (DHW) production [4]. Being such a large contributor to domestic energy consumption, gas-boilers have received significant attention from policy makers and researchers alike. In accordance to the UK Government's $CO_2$ target of an 80% reduction in emissions by 2050 [28], amongst other actions, policy makers have modified building regulations [11] so that boilers installed after 2006 are required to have a SEDBUK [2] efficiency of 88% or above. In addition to policy changes, many researchers look to evaluate environmental or behavioural interventions designed to reduce DCH and/or DHW consumption. To evaluate such interventions it is necessary to accurately monitor changes in disaggregated operational energy consumption, often in field trials.

Measuring energy consumed for DCH in field trials has however proven a challenge, as Hong et al. [12] note: due to the "… high cost associated with sophisticated fuel-use monitoring equipment required for different types of heating system and the complexity of its installation". Estimates of DCH energy use have been made in a number of studies by subtracting the summer fuel load from the winter fuel load [12,36]. These methods however require more than one years' worth of data and assume no variation occurs for non-space heating energy use between seasons. Other studies, such as Martin and Watson [22] and Love [23], used temperature sensors on flow pipes or radiators and relate temperature increases to boiler firing. These methods however miss short burner cycles and burner modulation. If methods were available that could accurately detect burner duration and modulation, space heating energy consumption estimates would be more accurate.

In addition to the above, it is well established that unnecessary cycling of boilers is undesirable [24]. Frequent on/off cycles, for example, cause boilers to operate at efficiencies "well below (their)

* Corresponding author.
*E-mail addresses:* tom.neeld.13@ucl.ac.uk (T. Neeld), j.eaton11@imperial.ac.uk (J. Eaton), p.naylor@imperial.ac.uk (P.A. Naylor), d.shipworth@ucl.ac.uk (D. Shipworth).

http://dx.doi.org/10.1016/j.buildenv.2016.01.029
0360-1323/

full-load bench value" [38]. This issue has been observed in practice, for example Ren et al. [32] note "Space heating systems cycled more frequently than anticipated due to a tight range of room thermostat settings and potentially oversized heating capacities." Likewise a survey of 35 non-residential buildings revealed that "all of the installed heating capacity was oversized by a minimum of 30% …" [21]. Furthermore unnecessary cycling causes additional wear and tear on components. These problems can be due to oversizing [10], incorrect return flow temperature sensor settings, inadequate boiler modulation ranges etc. Data on cycling rates and boiler modulation could thus help identify potential issues causing efficiency losses.

Thus to accurately evaluate interventions designed to reduce demand, new field-trial applicable methods are required that can accurately measure operational burner duration for space heating purposes. Additionally such methods could be used to provide data on operational boiler cycling behaviour and help identify potential efficiency losses due to setup, design and/or sizing issues.

### 1.2. Aim

The aim of this study is to assess the feasibility of a single-point acoustic sensor and associated detection device that can be retrofitted to Combination Gas Fired Boilers (C-GFB) to provide data on energy consumption, demand type, boiler operation and boiler failure modes. Primary events for detection include demand type and the burner duration. Secondary events include the activity of the pre-mix fan and circulation-pump. Note, the circulation-pump circulates heating water around the closed heating system circuit (boiler, radiators etc.) and the pre-mix fan draws both natural gas and air into the burner chamber and helps to expel the exhaust gases. C-GFBs were selected for this study because, as of 2011, 60% of DCH and DHW systems in the UK were of this type. Additionally the number of C-GFBs has risen steadily since 1990, with no indication of it decreasing in popularity [4].

For this investigation, event detection will use the acoustic signal produced by the physical process of the event itself. This is to avoid making the assumption that the expected process flow of the system is being followed. For example, changes in the pre-mix fan motor speed are directly related to the period of firing, however using the pre-mix fan motor speed as a basis for detecting firing would assume the boiler is behaving as expected. Thus, the investigation bases burner identification on the acoustic signal produced by the burner and not the acoustic signal produced by any other component of the system. By doing so, events occurring outside of the expected process flow of the boiler can potentially be detected.

Alternative methods of event detection exist. One method is to monitor the boiler's Central Processing Unit (CPU). Junkers (the German brand of Bosch), for example, have developed a smartphone application which is linked via a wireless network with the CPU of some Junkers boilers [18]; however the application doesn't give detailed information on boiler events such as ignition time (only failure codes are reported back to the user). In general, accessing a boiler's CPU requires manufacturers' consent and substantial proprietary software, and if performed by research fieldwork teams, could invalidate boiler warranties. The advantage of using a non-invasive retrofitted method, such as an acoustic sensor, is that warranties are not invalidated and theoretically all boilers are accessible irrespective of age, software or hardware. Another option would be to use multiple sensors: The sensors however would need to be situated in various locations depending on the boiler. This increases both cost, complexity and the probability of sensor failure. The advantage of a single acoustic sensor is that most events of interest could, in principle, still be detected, and that the exact positioning of such an acoustic sensor would be of less importance and could sit externally to the boiler.

## 2. Related work

This study is focused on identifying techniques to determine events of interest from the acoustic signals produced by domestic C-GFBs. No studies can be found in the literature of this nature, however related studies exist that apply signal analysis techniques to determine resource usage in the domestic environment and they include: the determination of electrical component usage from the electrical mains signals [9,29]; the localisation of water valve usage from water pipe pressure fluctuations [7]; and derivation of gas component usage from the acoustics of gas relief valves [3]. The general technique applied in all these studies involved the use of supervised-machine-learning (SML) algorithms. SML algorithms comprise of a set of procedures which automatically create models to determine the events of interest when the events are unknown, from a set of training data when the events are known; Refs. [20,41] provide overviews. The training data used in these studies comprise of a specifically selected set of features within the signal called feature vectors.

In the determination of electrical devices (switched-mode power supply only) used in the home, Gupta et al. [9] analysed the electromagnetic interference signal created in the domestic mains voltage supply during device operation. They selected the amplitude, mean and standard deviation of any voltage peak in the frequency domain as feature vectors. As the number of dimensions were low the k-Nearest Neighbour (kNN) classification algorithm was applied. The cross validation accuracy (Refer to Ref. [19]; p.2–3) of the classifier was 94%. After calibration this detection method was then tested across seven individual homes and found to work with accuracy greater than 90%. Patel et al. [29] develop a similar classification algorithm for detection of resistive and inductive electrical devices. The research of Patel et al. [29] and Gupta et al. [9] is complimentary; together they cover the detection of all electrical component types.

Sensing and classification algorithms have also been applied to the analysis of domestic water consumption. Froehlich et al. [7] built on the work done by Fogarty et al. [6] in the development of a sensor to detect faucet flow-rate and location. Froehlich et al. [7] used a customized pressure sensor attached to a fixture within the household to detect shockwaves created when faucets open or close. The signature of the shockwaves were found to be unique depending on the valve type and location. Thus Froehlich et al. [7] were able to create classification algorithms to determine the fixture, its location and estimate the flow-rate. Two layers of hierarchical classifier were applied to determine if the valve had been opened or closed and its location. Once a valve event was determined features were extracted to identify the fixture location, these were: (1) the matched filter; (2) the matched derivative; (3) the real part of the Mel Frequency Cepstral Coefficients (MFCCs); and (4) the mean squared error: mathematical definitions of which can be found in Ref. [37]. This method resulted in an average cross-validation accuracy of 98% when identifying fixtures across multiple homes (10 homes were tested). Estimates of the flow rates on three houses showed average errors below 8% (comparable to utility supplied water meters). Note that MFCCs are widely used as features for speech recognition technologies [25] as they correspond more closely with the human auditory perception [14].

In the application to gas usage, Cohn et al. [3] present a method for single-point acoustic sensing of individual domestic gas appliance usage. They analysed acoustic signals emanating from a Government regulated gas relief valve; a valve that is installed in most US homes. Signals from the gas relief valve were isolated from unwanted background noise using a high-pass filter. A linear

relationship was observed between the gas flow rates and the acoustic intensity of the gas relief valve. Thus the flow rate was estimated via linear regression analysis. Additionally they found the appliance type could be accurately estimated using a kNN classifier, with a feature vector consisting of three dimensions at the time of a gas usage event: (1) the relative magnitude of signal changes, (2) the rate of change in signal and (3) the period of the signal increase due to the unknown event. The classifier developed was, on average, 95% successful in determining the correct appliance across the nine homes investigated. These techniques may not be applicable outside of the US as they rely heavily on the gas valve type. Additionally, it should be noted that in the context of this report, the study did not attempt to distinguish between different boiler demand types.

In summary, in order to determine events within complex signals, a process of signal analysis in combination with the application of SML algorithms is generally required. Signal analysis is performed to extract feature vectors: The features selected can range from one dimension [7] to 2048 dimensions [29]. Feature vectors are selected using a process of logical deduction; the features should provide sufficient information to make the events of interest discernible without including unnecessary data. Selecting the appropriate SML algorithm depends on the application. Refs. [20,41] amongst others provide general guidance.

In addition to the above, SML algorithms have recently been applied extensively in the area of occupant thermal comfort. Refs. [13,15] applied SML algorithms to predict occupant thermal comfort, in HVAC controlled environments, using very similar data sets: Both studies used four environmental (temperature, humidity etc.) and two occupant factors (metabolic rate and clothing) as feature inputs; and thermal-sensation-index questionnaire responses as classification inputs. Jiang and Yao [15] found Support Vector Machines (SVMs) to be applicable in accurately predicting (over 89%) the correct thermal-sensation-index of the occupant. Hu and Li [13] restricted their investigation to SML algorithms based on logic statements; they did this to limit computational cost and because logic statements can be interpreted easily for use in HVAC control systems.

## 3. Methods

Analysis was made of the acoustic signals emanating from C-GFBs. Relevant techniques were applied to attempt to automatically detect demand and ignition events. The acoustic signatures corresponding to the operation of the pre-mix fan and circulation-pump were investigated.

### 3.1. Equipment

Two C-GFBs were selected from which to collect acoustic data: 'C-GFB A' was a Worcester Greenstar 28i Junior [42] and was investigated for the majority of this study. 'C-GFB B' was a Vaillant ecoTEC pro 28 [39] and was used as a reference C-GFB to confirm ignition event signals (section 4.2). These C-GFBs were selected because: a. 'C-GFB A' is quite representative of C-GFBs as it is the most common boiler sold by Worcester and b. these boilers were readily available to the research team. To capture acoustic signals, a Samson Go Mic [35] portable USB microphone, labelled 'Microphone A' was selected since it could be easily positioned in various positions on or within the boiler. Recordings were made using Audacity [1] at a sampling frequency of 16 kHz.

In order to provide the ground truth time at which the burner fired, $t_1$, a separate event detector was used. It consisted of a photoresistor positioned to face the ignition indicator on the C-GFB front panel, combined with a Raspberry Pi data logger (courtesy of Stephen Hailes - see Acknowledgements).

The instant at which an initial noise was produced by the boiler, in response to demand, was labelled $t_0$. This instant was determined manually by listening to audio recordings. The event detector recorded the instant when the burner turned on and turned off, $t_1$ and $t_2$ respectively. Note, for $t_1$ and $t_2$ an estimated error of $\pm 1$ s was induced due to the process of manually synchronising the clock on the event detector to the clock of the recording device.

### 3.2. Procedure

To facilitate time alignment of the recorded acoustic signal and the event detector, the real time clock on the Raspberry Pi data logger was manually synchronized with the clock on the laptop used to make audio recordings. Microphone A was then placed on the outer casing of the C-GFB under investigation: The base of the microphone was in contact with the outer casing of the C-GFB but the microphone head was not. The position of the microphone was approximate but always on the lower right side of the C-GFB under investigation (as shown in Fig. 1). The event detector was positioned to detect the green light indicator (light indicating firing) for C-GFB A (as shown in Fig. 2), or the flame, via the flame viewing hole, for C-GFB B. Note, the event detector was purely used as an independent validation of the firing event for classifier training. Both the event detector and Microphone A were set to record. Single, continuous, demand events were created by either opening a hot water tap valve or initiating DCH. The demand events occurred between 30 and 120 s to ensure some irregularity in the data gathered. Noise in the vicinity of the boiler was kept to a minimum during recordings. The recording was stopped after the acoustic signal emanating from the boiler ceased.

## 4. Results and discussion

### 4.1. Demand type

The C-GFB under investigation (C-GFB A) meets the household demands for DHW and DCH. In response to a particular demand type, the initial acoustic signal (first 5 s) produced by the C-GFB consistently followed a repeatable pattern. From the data collected (40 instances in total) DHW demand produced a single acoustic configuration whereas DCH demand produced two acoustic configurations. DCH demand was thus split into two classes; DCH1 and DCH2. Fig. 3 shows the acoustic patterns for the three demand classes identified during the first 5 s of C-GFB activity. By comparing this signal with the expected process flow of C-GFB A [42], the expected components operating during this period were deduced as indicated on Fig. 3a, b, and c.

Each demand class had a different component operating before the pre-mix fan initiated i.e. before the relay switch operated: For DCH2 there was no initial component whereas for DCH1, the diverter valve activated before the electronic relay and pre-mix fan. It is likely the case that for DCH2 the diverter valve was already in the correct position to meet DCH demand. However, unlike DCH1, in about half of the DCH2 instances recorded there existed an unexpected process. The unexpected process was the activation of the burner and a complete demand process flow cycle after the initial demand period had finished, i.e. the C-GFB was responding to a demand which was not present. The C-GFB specifications [42] indicate that this behaviour must be that of the 'Preheat Demand' function: A demand cycle designed to ensure the temperature of water sitting in the DHW heat exchanger does not drop below a certain point (10 K below the DHW set point). Why the Preheat Demand function seemed to operate after DCH2 demand cycles instead of DCH1 is unknown. Additionally why this occurred

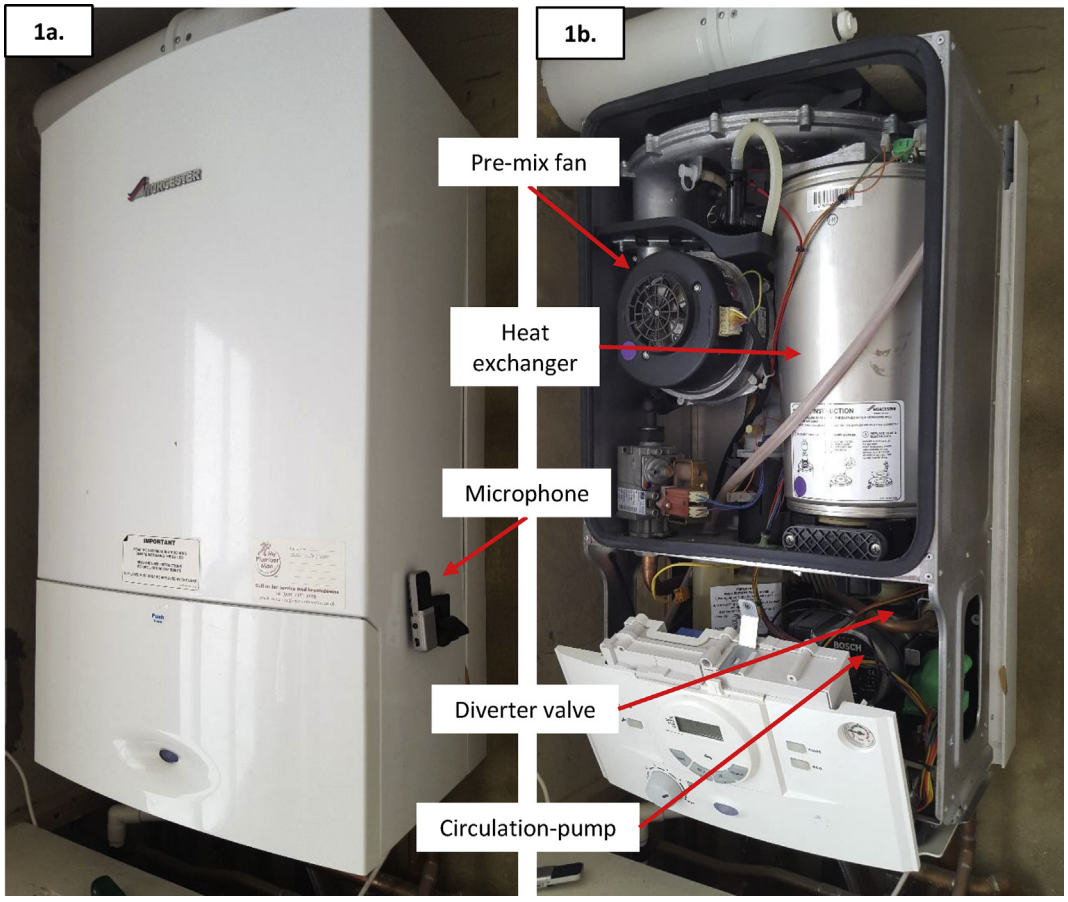


**Fig. 1.** (a) Picture showing the position of Microphone A on the side of C-GFB A. (b) Picture of C-GFB A with the outer casing removed showing the internal components. Highlighted in (b) are the Pre-mix fan; heat exchanger (housing the burner assembly); diverter valve and circulation-pump.

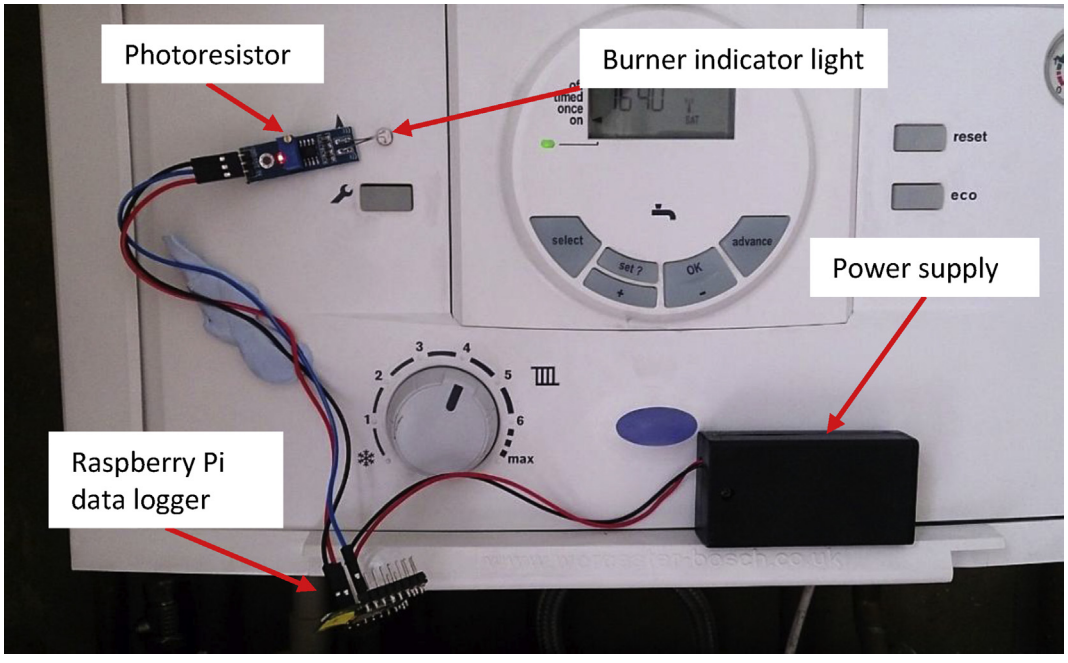


**Fig. 2.** Picture of the control panel on C-GFB A. Attached to the control panel is the event detector setup consisting of a photoresistor, Orisen Prime data logger and power supply. As observed in the figure the photoresistor is placed directly over the burner light indicator.

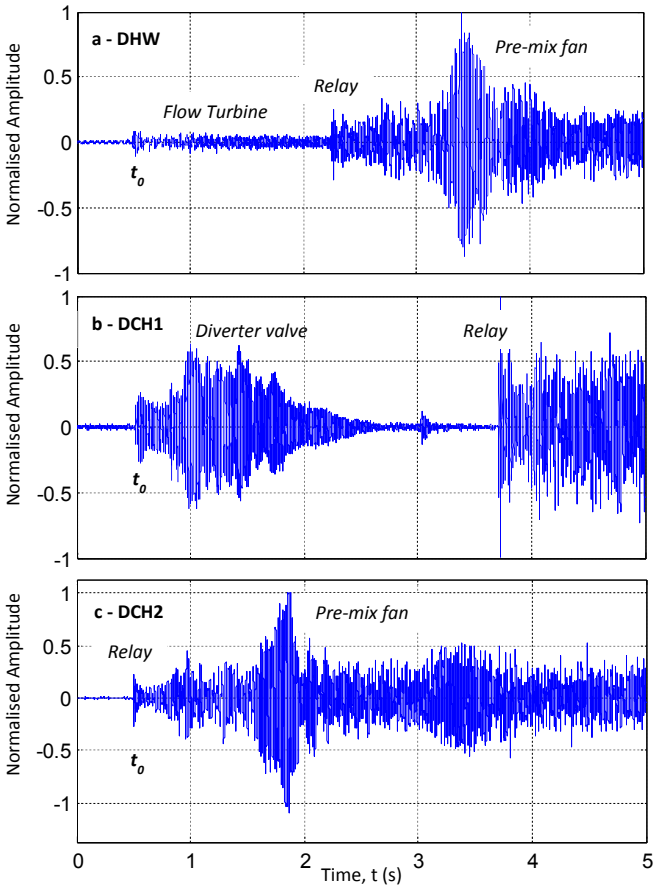


**Fig. 3.** Plots of the initial acoustic signal emanating from C-GFB A for the three demand classes (a) DHW, (b) DCH1 and (c) DCH2. Amplitude was normalised for each recording. For each demand class the likely C-GFB components causing the acoustic signal is noted above the signal.

directly following a DCH2 demand cycle is also unknown. However with respect to this study, such analysis highlights how a detection device could be used to assist in determining if the C-GFB is following the designed process flow.

To automatically determine the demand response instant, $t_0$, the detection algorithm was designed to trace through the acoustic signal until an extended period of increased acoustic signal energy was observed. The demand response instant was then identified by working backwards along the period of increased signal energy until it dropped below a trained threshold. This method for detecting $t_0$ was found to be sufficient for the data recorded (40 instances), however introducing additional background noise into the recording may cause inconsistency in accurately identifying $t_0$, in which case other techniques may be required. Once $t_0$ was determined, efforts were made to identify the demand class.

Fig. 3 shows that each demand class had a visually distinct initial signal profile. Accordingly the initial acoustic signal from the C-GFB was used to automatically determine the demand class. To do this SML techniques were tested because no exact rule differentiating the demand classes could be manually derived. For SML algorithm analysis the following feature vectors were used: the normalised energy profile of the signal divided into 200 ms frames for the first 4 s of boiler activity (20 dimensions in total). These feature vectors were selected because they were found to provide all the information required to accurately determine the demand class while minimising the total number of features required. A total of 40 instances were used for classification training and testing. After testing numerous classifiers using Weka [40] the naive Bayes (nB) classifier [33], with the assumption of a kernel density distribution (refer to Appendix A. for a description), was found to be the most accurate - giving an average 10-fold cross-validation accuracy of 100%. An overview of the main classes of SML algorithms tested and their performance has been provided in Table 1. The kernel nB classifier was incorporated into the overall detection algorithm coded in Matlab. Note, Weka [40], is a purpose built open source program containing most machine learning algorithms for application in data mining.

Analysing the feature data, it was found that the assumption of normality (comparing each feature across all instances) was generally a poor one - indicating in most cases a distribution with high kurtosis. Therefore, as suggested by John and Langley [16], it is expected that a kernel density estimation would be better suited. This can be seen in Table 1 by comparing the performance of the nB classifier using a kernel compared with a Gaussian density estimate; the nB classifier performs 1.5% more accurately with a kernel density estimate.

The application of the nB classifier assumes conditional independence between the features. Conditional independence requires that, for a given class $C$, the probability of obtaining a particular signal energy for the 1st feature, $F_1$, is independent the signal energy of the 2nd feature, $F_2$ (this applies across all n features):

$$P(F_1|C,F_2) = P(F_1|C).$$

The conditional independence equation above infers, in this instance, that knowledge of the signal energy for one 200 ms interval tells us nothing of the signal energy for the next, or any other 200 ms interval. However the signal energy is dictated by the operational mechanics of the system and the pre-programmed process flow. Thus the variation of the signal energy from one time to the next is inextricably linked. The strength of the link is dependent on the time between signal energy data points; the longer the time the less likely a prediction can be made. Nevertheless for the case of 200 ms variations the assumption of conditional independence appears to be a poor one. Refs. [5,34] however find nB performs unexpectedly well for some cases of functionally dependent features, as observed in this instance.

### 4.2. Burner firing

Once the demand type and $t_0$ are known, the instant at which the burner fires, $t_1$, can be estimated by assuming that the C-GFB behaves as expected and the demand event is continuous. With these assumptions it was found that $t_1$ could be estimated to an accuracy of ±2 s. However, if the assumptions are invalid, a method to detect the acoustic signal of the ignition event itself needs to be derived as described below.

Low frequency analysis of all recorded instances (40 in total) revealed that within 1 s of ignition, there existed a relatively small energy spike in the frequency band 5–15 Hz. The energy of the frequency spike was approximately between 60 and 80 dB lower than the highest energy signal produced by the C-GFB. Fig. 4 shows an example of this frequency spike for a DCH2 demand event. As observed in Fig. 4, the 5–15 Hz range contains very little energy beside for the spike produced during firing at $t_1 \pm 1$ s. Note, Fig. 4 has been constrained to show a maximum of −50 dB in order to emphasise events with a relatively low level of energy.

By summing the 5–15 Hz components of the signal, the spike seen at the point of ignition was emphasised. This summation can be expressed by the equation:

**Table 1**
A range of the main classes of machine learning algorithms which were tested for their ability to estimate the correct demand type class - not all algorithms tested are shown above. The measure of the algorithm's accuracy is given by repeating a 10-fold cross validation accuracy calculation 10 times with a different (random) data partition in each cross validation calculation; this has been labelled the 'Average 10-fold accuracy'. The standard deviation is calculated based on the variation of the cross validation result for each of the 10 repetitions.

| Demand type SML testing | | |
|---|---|---|
| Classification algorithm | Average 10-fold accuracy | Standard deviation |
| 1R Learning Algorithm: Minimum-error attribute | 94.9% | 0.0 |
| NB Algorithm with Gaussian density estimate | 98.5% | 1.3 |
| NB Algorithm with kernel density estimate | ***100%*** | ***0.0*** |
| k-Nearest Neighbours Algorithm | 95.1% | 0.8 |
| Sequential Minimal Optimization [30] | 97.4% | 0.0 |
| C4.5 Decision Tree: Top-down recursive divide-and-conquer algorithm | 94.6% | 0.8 |

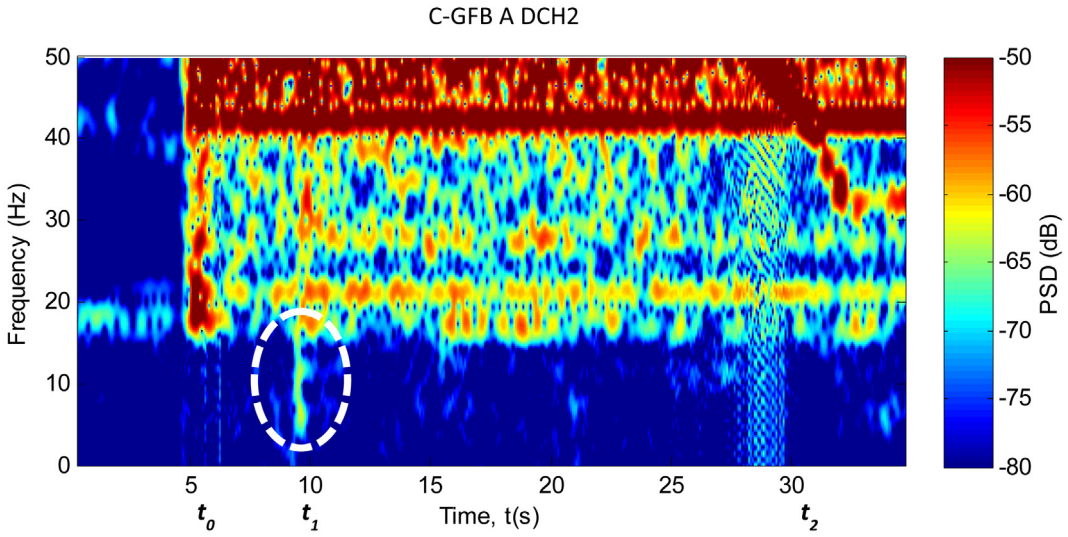


**Fig. 4.** Plot of the spectrogram produced by C-GFB A for a DCH2 demand event. Frequency is constrained to maximum of 50 Hz and Power Spectral Density (PSD) to −50 dB. The low frequency spike produced at $t_1$ has been highlighted by the white ellipse.

$$S(t) = \sum_{i=5}^{15} \mathrm{PSD}(f_i, t)$$

Where PSD($f_i$,$t$) is the Power Spectral Density of the acoustic signal for frequency $f_i$ and time $t$. The point in time when $S(t)$ reached a maximum was found to align with the moment of ignition in 97.1% of recorded instances. There was only one instance when the maximum of $S(t)$ did not align with the point of ignition, however it was found the second maximum of $S(t)$ did align with the point of ignition for this data point. Non-ignition spikes in $S(t)$ were likely caused by other noises in the system; generally they had very different frequency distributions compared with ignition spikes.

To test whether the signal produced during ignition was only a feature of C-GFB A or common amongst other C-GFBs, the same frequencies were analysed for C-GFB B. C-GFB B also showed activation of frequencies ranging from 5 to 15 Hz at the point of ignition in all recorded instances (five in total). This indicated that the activation of low frequency components at the point of firing is a common feature of C-GFBs. Consequently indicating that the same detection methods could be applied to other C-GFBs. Additionally the literature suggests that ignition events should produce low frequency acoustic signals indicating further that the detection algorithm is detecting the acoustic signal of an ignition event: Jones [17] notes that pressure pulsations from pulverised coal burners occur in the range of 1–10 Hz; Ottemöller and Evers [26] found the most prominent acoustic wave produced by a chemical explosion was infrasound (around 0.4 Hz).

Note, the amplitude of the sub-20 Hz signals detected were about 20 times lower than other C-GFB signals. However the frequency response of Microphone A starts to roll-off below 50 Hz and is only quoted as low as 20 Hz [35]. It could be assumed therefore that below 20 Hz the response of Microphone A continues to roll-off. Thus the acoustic signals produced during ignition may contain more relative energy than detected.

### 4.3. Burner ending

For C-GFB A, within the period of the burner ending instant, $t_2$, it was observed that a strong acoustic signal in the range of 40 Hz seemed to consistently drop to 30–35 Hz. After testing various SML algorithms with a range of frequency features, it was found the algorithms heavily weighted the 30–35 Hz components of the signal. This produced, at best, algorithms that estimated $t_2$ to within 4 s. As discussed in section 4.4 the 30–35 Hz signal was found to be caused by the pre-mix fan decreasing in motor frequency. It could be conjectured that a direct physical relationship existed between burner ending and the pre-mix fan motor speed: A drop in pressure in the burner chamber will occur when the burner ends, this could have an impact on the resistance experienced by the blades on the pre-mix fan and thus on pre-mix fan motor speed. Additionally changes in pre-mix fan motor speed may be due to restriction of the gas valve closing or change in Heat Exchanger resistance when

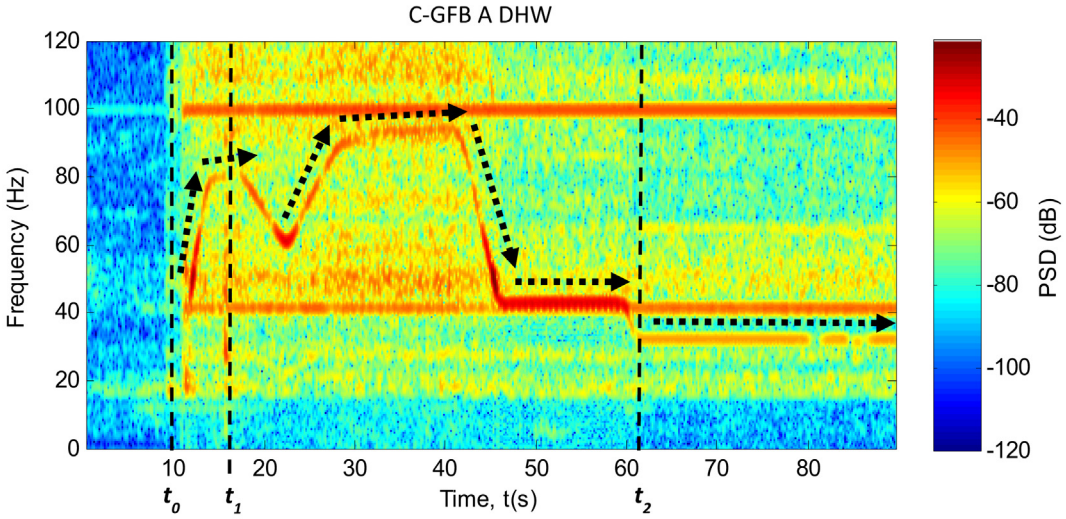


**Fig. 5.** Plot of the spectrogram of the acoustic signal for the entire period of C-GFB activity for DHW demand event. Frequency is constrained to maximum of 120 Hz. The Power Spectral Density (PSD) is not constrained. The black arrows highlight the profile of the pre-mix fan.

there is no flame. Furthermore such a change in pre-mix fan motor speed may just be a consequence of the C-GFB's process flow and not an indication of a direct physical relationship. To improve the detection of $t_2$ additional analysis needs to be performed to either find the acoustic features of the burner, or determine if a direct physical relationship exists between burner ending and the pre-mix fan motor speed.

### 4.4. Pre-mix fan and circulation-pump operation

For C-GFB A, the pre-mix fan was the only component expected to modulate during operation. Additionally the pre-mix fan motor was expected to start at a frequency between 45 and 55 Hz and reach maximum frequency between 90 and 100 Hz [42]. As seen in Fig. 5, alongside the black arrows, there is a continuous signal which changes in frequency. It starts at 45–55 Hz and reaches a maximum of 95 Hz during demand. This signal is expected to be due to the pre-mix fan motor. As observed the pre-mix fan activates 30–35 Hz components just after $t_2$, accounting for the result seen in section 4.3. All recordings of C-GFB A showed a similar pre-mix fan modulation pattern.

Circulation-pumps typically operate at shaft speeds between 12 and 33 Hz [31]. Upon analysis of C-GFB A, frequencies were found within this range for the period of time the pump was expected to be in operation. Fig. 6 below highlights these frequency components during DCH1 demand for C-GFB A. From $t = 150$ s onwards, according to C-GFB A's process flow [42], the only component operating until $t = 250$ s is that of the pump. Thus it is expected that the 21–22 Hz signal highlighted by the black arrow is caused by the circulation-pump. For C-GFB B the same analysis was performed, in this case the signal produced by the pump appeared to be in the region of 35 Hz. For C-GFB B the pump was a Grundfos Super Selectric, a pump that has three speed settings up to 50 Hz [8], thus the 35 Hz signal identified was likely to be caused by the pump.

Figs. 3 and 4 show that the motor frequencies of the pre-mix fan and circulation-pump may be detectible by analysing the acoustic

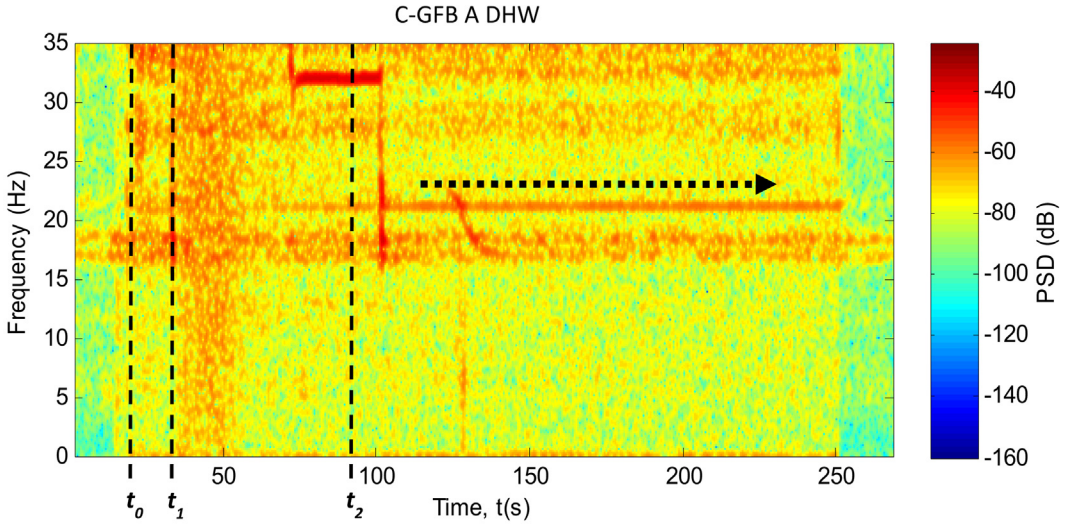


**Fig. 6.** Plot of the spectrogram of the acoustic signal for the entire period of C-GFB activity for a DCH1 demand event. Frequency is constrained to maximum of 35 Hz. The Power Spectral Density (PSD) is not constrained. The black arrow highlights the profile of the circulation-pump.

signal emanating from the C-GFB. This was done by manually analysing the signals produced by the C-GFB. To automate this detection however requires application of additional signal processing techniques. Note, the signal produced by the pump in C-GFB A (Fig. 6) is approaching the frequency limits of Microphone A, thus the actual acoustic signal produced by the pump may contain more relative energy than detected in comparison with the acoustic signals produced by other C-GFB components.

## 5. Conclusions

Our investigation has shown that acoustic sensing methods, using a single-point sensor, can be used to detect specific events of interest in domestic C-GFBs. The methods developed allow for the accurate automatic detection of demand type (100 ± 0.0% accuracy) and ignition time (97.1% accuracy). Algorithms to determine demand type were trained on features based on the normalised signal energy for the first 4 s of C-GFB activity. Infrasound pressure pulsations produced during ignition were analysed to identify the time burner ignition occurred. Furthermore our investigation has shown that it is also feasible to determine the motor frequencies of the pre-mix fan and possibly the circulation-pump from the acoustic data gathered.

In order for the techniques developed to be applicable in energy demand field-trials, the methods need to be modified to deal with noise sources external to the C-GFB (application of methods such as spectral subtraction), to incorporate the detection of burner duration and intensity to determine gas consumption, to deal with overlapping demands (DHW interrupting DCH), and to apply to a wide range of C-GFB makes and models.

With respect to energy efficiency improvements, analysis of the data from field trials could track the extent to which operational boiler cycling occurs and help identify potential issues causing unnecessary cycling i.e. efficiency losses. Such analysis could affect control settings, future design and policy.

The project has therefore started a process for the development of a widely applicable set of tools for accessing energy systems, which, for example, might help in specific energy system challenges or to diagnose the states of future energy systems such as heat pumps.

## Acknowledgements

Stephen Hailes and Sarah Chisholm of the Computer Science department at UCL for providing assistance in the application and gathering of equipment for detecting the burner firing event, and for initial discussions regarding application of machine learning algorithms to the analysis of acoustic signals. George Bennett of Bosch and UCL for reviewing the work completed and advising on combination boiler operation. Lastly I acknowledge the London-Loughborough Centre for Doctoral Research in Energy Demand for funding my research.

## Appendix

### A. Kernel Naive Bayes Classifier

Naive Bayes is an algorithm based on Bayes' rule. Bayes' rule states that the probability of class (or hypothesis) $c$ occurring given $n$ feature variables (or evidence) $\boldsymbol{F}$ is given by the following equation [27]. Note, $\boldsymbol{F}_i$ is a vector containing all the instances for feature $i$:

$$P(c|\boldsymbol{F}_1,\ldots\boldsymbol{F}_n)=\frac{P(c)P(\boldsymbol{F}_1,\ldots\boldsymbol{F}_n|c)}{P(\boldsymbol{F}_1,\ldots\boldsymbol{F}_n)} \qquad \mathbf{A}.1$$

Above $P(c)$ represents the prior probability of the class independent of the features; $P(F_1, \ldots F_n|c)$ represents the probability of the features given the class; and $P(F_1, \ldots F_n)$ represents the probability of the features independent of the class. $P(F_1, \ldots F_n)$ is simply the addition of the feature probabilities for all the classes and thus is a normalising coefficient. Naive Bayes applies the "naive" assumption that each feature of the system is independent of every other feature. With this assumption and by applying the chain rule for conditional probability the term $P(F_1, \ldots F_n|c)$ can be written as the product of the individual probabilities. Thus:

$$P(c|\boldsymbol{F}_1,\ldots\boldsymbol{F}_n)_{NB}\propto P(\boldsymbol{c})\prod_{i=1}^{n}P(\boldsymbol{F}_i|c) \qquad \mathbf{A}.2$$

If a set of instances is available for training i.e. the possible classes are known for a set of features $\boldsymbol{F}$, one can calculate the individual probabilities $P(\boldsymbol{F}_i|c)$ and consequently calculate the probability of a particular class given a new set of features $\boldsymbol{f}$ – the new features are aligned with the training set and probabilities extracted. The selected class, $C$, is the class which results in the highest probability for $P(C|\, f_1 \ldots f_n)$. Accordingly naive Bayes will estimate class $C$ using the following function. Note, the factor $P(\boldsymbol{F_1}, \ldots \boldsymbol{F_n})$ does not depend on the class and therefore can be removed:

$$classify(f_1,\ldots f_n)\leftarrow \underset{C}{\operatorname{argmax}}\, P(c=C)\prod_{i=1}^{n}P(\boldsymbol{F}_i=f_i|c=C) \qquad \mathbf{A}.3$$

In the application of naive Bayes where the features are numeric, exactly aligning the new instances with the probabilities of the training set instances is not possible. In order to handle numeric features, a probability distribution can be assumed for each feature within a known class. For example, a kernel distribution can be assumed for each feature. In this case the probability of a new (numeric) feature value $f_i$ is given by the probability density function [16]:

$$P(\boldsymbol{F_i}=f_i|c)=\frac{1}{T}\sum_{j=1}^{T}K\left(\frac{f_i-F_{ij}}{h}\right) \qquad \mathbf{A}.4$$

Where $T$ is the total number of instances for the $i$th feature of class $c$. $K$ is the kernel function - a non-negative function that integrates to one and has a mean of zero. $h$ is the smoothing parameter and is approximated to the standard deviation of the instances for the $i$th feature of class $c$. This particular formulation of the naive Bayes algorithm is called kernel naive Bayes.

## References

[1] Audacity, Recording Software, Available at:, 2016 http://audacityteam.org (Accessed: January 2016).
[2] BRE, Building Energy Performance Assessment: DECC, Available at:, 2015 http://www.ncm-pcdb.org.uk/sap (Accessed: November 2015).
[3] G. Cohn, S. Gupta, J. Froehlich, E. Larson, S. Patel, GasSense: Appliance-Level, Single-Point Sensing of Gas Activity in the Home, in: P. Floréen, A. Krüger, M. Spasojevic (Eds.), Pervasive Computing Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2010, pp. 265–282.
[4] DECC, United Kingdom housing Energy Fact File: 2013, 2014.
[5] P. Domingos, M. Pazzani, On the optimality of the simple Bayesian classifier under zero-one loss, Mach. Learn. 29 (2–3) (1997) 103–130.
[6] J. Fogarty, C. Au, S.E. Hudson, Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition, in: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, ACM, Montreux, Switzerland, 2006, pp. 91–100, 1166269.
[7] J.E. Froehlich, E. Larson, T. Campbell, C. Haggerty, J. Fogarty, S.N. Patel,

HydroSense: infrastructure-mediated single-point sensing of whole-home water activity, in: Proceedings of the 11th International Conference on Ubiquitous Computing, ACM, Orlando, Florida, USA, 2009, pp. 235–244, 1620581.
[8] Grundfos, Grundfos Selectric/Super Selectric Circulators, Available at:, http://net.grundfos.com/Appl/ccmsservices/public/literature/filedata/Grundfosliterature-1229.pdf (Accessed: 10 January 2015).
[9] S. Gupta, M.S. Reynolds, S.N. Patel, ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home, in: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, ACM, Copenhagen, Denmark, 2010, pp. 139–148, 1864375.
[10] K.E. Heselton, Cycling efficiency: A basis for replacing outsized boilers, Energy Eng. 95 (4) (1998) 32–44.
[11] HM Government, Approved Document L1A: Conservation of Fuel and Power in New Dwellings, 2013 edition, 2014.
[12] S.H. Hong, T. Oreszczyn, I. Ridley, The impact of energy efficient refurbishment on the space heating fuel consumption in English dwellings, Energy Build. 38 (10) (2006) 1171–1181.
[13] C.-C. Hu, H.-L. Li, Deducing the classification rules for thermal comfort controls using optimal method, Build. Environ. 98 (2016) 107–120.
[14] C. Ittichaichareon, S. Suksri, T. Yingthawornsuk, Speech recognition using MFCC, 2012, pp. 28–29.
[15] L. Jiang, R. Yao, Modelling Personal Thermal Sensations Using C-Support Vector Classification (C-SVC) algorithm, Build. Environ. 99 (2016) 98–106.
[16] G.H. John, P. Langley, Estimating continuous distributions in Bayesian classifiers, in: Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers Inc, 1995, pp. 338–345.
[17] A.R. Jones, Flame failure detection and modern boilers, J. Phys. E Sci. Instrum. 21 (10) (1988) 921.
[18] Junkers, Junkers Multi-Home Phone Application, Available at:, 2015 http://www.junkers.com/fachkunde/interaktiv/apps/multihome/multihome (Accessed: 10 August 2015).
[19] R. Kohavi, A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection, 1995, pp. 1137–1145.
[20] S.B. Kotsiantis, Supervised Machine Learning: a Review of Classification Techniques, 2007.
[21] Z. Liao, M. Swainson, A.L. Dexter, On the control of heating systems in the UK, Build. Environ. 40 (3) (2005) 343–351.
[22] J.A. Love, Understanding the Interactions between Occupants, Heating Systems and Building Fabric in the Context of Energy Efficient Building Fabric Retrofit in Social Housing, University College London, UCL, 2014 [Online] Available at: http://www.lolo.ac.uk/wp-content/uploads/2015/10/1418647283_LOVEthesisfinalreducedfilesize.pdf.
[23] C. Martin, M. Watson, Measurement of Energy Savings and Comfort Levels in Houses Receiving Insulation Upgrades, Energy Saving Trust, 2006.
[24] J. McKeegan, Short-cycling can hamper boiler efficiency, Air Cond. Heat. Refrig. News 220 (11) (2003) 18.
[25] L. Muda, M. Begam, I. Elamvazuthi, Voice Recognition Algorithms Using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques, arXiv preprint arXiv:1003.4083, 2010.
[26] L. Ottemöller, L.G. Evers, Seismo-acoustic analysis of the Buncefield oil depot explosion in the UK, 2005 December 11, Geophys. J. Int. 172 (3) (2008) 1123–1134.
[27] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: Sentiment Classification Using Machine Learning Techniques, Association for Computational Linguistics, 2002, pp. 79–86.
[28] U.K. Parliament, Climate Change Act 2008, London, UK, 2008.
[29] S. Patel, T. Robertson, J. Kientz, M. Reynolds, G. Abowd, At the flick of a switch: detecting and classifying unique electrical events on the residential power line (nominated for the best paper award), in: J. Krumm, G. Abowd, A. Seneviratne, T. Strang (Eds.), UbiComp 2007: Ubiquitous Computing Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2007, pp. 271–288.
[30] J. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, Adv. large margin Classif. 10 (3) (1999) 61–74.
[31] J. Reginald, Central Heating Combination Boilers: Fault Finding and Repair, Copperjob Publishing, 2004.
[32] X. Ren, D. Yan, T. Hong, Data mining of space heating system performance in affordable housing, Build. Environ. 89 (2015) 1–13.
[33] J.D. Rennie, L. Shih, J. Teevan, D.R. Karger, Tackling the poor assumptions of naive bayes text classifiers, in: International Conference on Machine Learning Washington DC, 2003, pp. 616–623.
[34] I. Rish, An empirical study of the naive Bayes classifier, in: *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*: IBM New York, 2001, pp. 41–46.
[35] Samson, Go Mic User Manual, Available at:, 2014 http://www.samsontech.com/samson/products/microphones/usb-microphones/gomic/ (Accessed: 10 July 2014).
[36] A. Schuler, C. Weber, U. Fahl, Energy consumption for space heating of West-German households: empirical evidence, scenario projections and policy implications, Energy Policy 28 (12) (2000) 877–894.
[37] S.W. Smith, The Scientist and Engineer's Guide to Digital Signal Processing, 1997.
[38] D.M. Tanton, R.R. Cohen, S.D. Probert, Improving the effectiveness of a domestic central-heating boiler by the use of heat storage, Appl. energy 27 (1) (1987) 53–82.
[39] Vaillant, ecoTEC Pro Installation and Servicing Manual, Available at:, 2014 http://www.vaillant.co.uk/products/domestic-boilers/combi-boilers/ecotec-pro-24-28 (Accessed: 10 July 2014).
[40] Weka, Weka 3: Data Mining Software in Java, The University of Waikato., 2014. Available at: http://www.cs.waikato.ac.nz/ml/weka/ (Accessed: 10 July 2014).
[41] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2005.
[42] Worcester Bosch Group, Installation Manual for Greenstar 28i Junior (Manufactured from Jul 13), Available at:, 2014 http://www.worcester-bosch.co.uk/installer/literature/gas-boiler-literature/greenstar-28i-junior-literature (Accessed: 10 July 2014).