ENERGYCON 2014 • May 13-16, 2014 • Dubrovnik, Croatia

Evaluating the impact of data sample-rate on appliance disaggregation

Anders Dalén[1], Christof Weinhardt[2]

IISM, Karlsruhe Institute of Technology
Karlsruhe, Germany

[1] dalen@kit.edu

[2] weinhardt@kit.edu

***Abstract*—This study evaluates how the disaggregation of appliances from a central location is impacted by the sample-rate used to measure a household's energy use. The sample-rate parameter is the deciding factor for choosing hardware requirements necessary for designing an appliance-level information system that can scale to general households. However, the impact of sample-rate on appliance disaggregation accuracy has so far not been evaluated.**

The results show that the harmonics feature and the real/reactive power feature have the highest disaggregation accuracy of 80-90%. These two features are, however, most affected by the sample-rate downscaling from 12kHz to 100Hz. The harmonics feature is stable down to 6kHz while the real/reactive power feature is stable 1.5kHz, in terms of disaggregation accuracy. The most stable features, at lower disaggregation accuracy, are the switching transients waveform and the instantaneous admittance waveform.

This study also makes a contribution to the ongoing research on appliance disaggregation by using a public dataset. Other research teams can thus verify, compare and extend this research in a cumulative way.

***Keywords:* Appliances, Disaggregation, Sample-rate.**

I. INTRODUCTION

Investments made in advanced meter infrastructure (AMI) had an estimated value of $3 to $4 billion worldwide in 2009 and are expected to continue to grow [1]. One of the major reasons behind the large investments is to measure and supply energy information in a timely and flexible manner [2]. In Germany it is mandatory for new buildings to install a digital energy meter that has the ability to supply near real-time information about the current household load (cf. Energiewirtschaftsgesetz §21d Messsysteme). Sweden and Italy, among other countries, have already replaced all of their analog household measurement devices.

First and foremost, the new sensors will allow for a more frequent, accurate and convenient way of billing for energy used. The energy bill is however on a household-level while the inhabitants operate on an appliance-level. Thus, even with this new technology the information is still disconnected from how individuals interact with their energy-based services. By providing appliance-level information, studies have found that users are better able to discern between the impact of different actions and save energy [3]. The potential benefit of user centric energy information has been quoted as the most likely form of feedback to transfer knowledge and motivate a more energy-saving behavior [4].

Making energy use information more transparent is therefore of top priority on many research agendas. Most common is central solutions, which has been the focus of several dissertations [5], [6]. These systems parse and use algorithms to recognize individual appliances based on whole household's aggregated energy use data. Furthermore, Belkin's recent programming competition, where the public could win $25,000 for developing ways of distinguishing individual appliances from a household-level meter reading, show that the interest in disaggregating appliance-level energy use information goes beyond a purely academic one.

Instead of measuring at a central location and dividing up the individual appliances from there, every single appliance could naturally have their own sensor. The reason such a system has not made it beyond laboratory environments is due to the prohibitive invasiveness and cost it would mean to install and maintain such a system. The predominant focus, therefore, is to develop more sophisticated software, which scale more easily than hardware [7].

Reporting values with a 15-minute resolution, which the new AMI proposes, is already a great improvement from the yearly billing cycle that has been predominant for households in the energy domain. However, measuring more often with a higher frequency provides more even detail of a given signal. This way, more subtle changes can be detected and evaluated. To separate single appliances among noise and several other sources of load, research laboratories are currently analyzing values on a resolution of up to 100 million data points per second [8]. At these sample-rates more sophisticated and costly hardware is needed. For example Nunes et.al [9] and Patel et.al [8] both use personal computers in their experiments. Granted that computer power is becoming ever more powerful and affordable as Moore once predicted [10], however to make scalable solutions for general households this power still has to be used as efficiently as possible. This requires an understanding of how data sample resolution impacts the results of appliance disaggregation, as it directly influences the necessary hardware sophistication level.

The main goal of this study is to analyze how measurement sample-rate influences the ability to disaggregate appliances with the commonly used appliance load features. The related work in the next section (2) give a brief introduction to the topic of central appliance disaggregation and then describe some of the general concepts. Based on these concepts, an experiment model will be developed in MATLAB to parse and evaluate the laboratory energy data, which will be presented in the experiment section (3). After presenting the results from the analysis in Section 4, the findings will then be discussed as well as their potential ramifications on future AMI developments in Section 5. This paper will then conclude with the main results and contributions of the study.

II. Related Work

Hart [7] first implemented a platform for distinguishing patterns of distinct loads and appliances in households in the late eighties. The patented system used embedded electronics to measure and analyze the load patterns at a central point (the electrical mains power line). Since then, several alternatives with the same goal of disaggregating single load features and individual appliances have been presented.

The frameworks used by the different proposed systems all follow a general pattern [11]. Fig 1 gives a graphical overview of the general process of sensing, extracting and recognizing appliance loads and the related sections reviewing the current literature on these topics. First, the load is measured over the cycles of alternating current and voltage, which are digitized for further processing. This step will be introduced in Section A and cover the current range of methods of sensing appliances. Second, the data is then analyzed for changes that signify a load event. This is, for example, when an appliance is turned on or off, or when the appliance changes its state automatically. The event detection will be explained further in Section B. Third, when an event has been detected, the data around this point is evaluated to extract and compare specific load features. A range of known appliance features will be presented in Section C to provide some background for the design of this study's analysis. Fourth, the calculated features are then compared to known appliance features to single out individual appliances in Section D.

Fig 1. General overview of an event-based framework for household measurements (A) and the following appliance disaggregation steps (B-D).

A. Data acquisition

The flow of alternating current (AC) electricity can be sensed either directly inline with the electricity flow or indirectly through its magnetic fields. The number of different loads that could potentially be included depends on where in the tree-like electrical circuit structure of a household the AC is measured. To monitor a whole household either a few sensors could be given a central location or multiple sensors need to be distributed throughout the house [12].

Multipoint sensing is normally located at the connection between the fixed household outlets and the appliances. University campus laboratories facilities are normally equipped with this kind of sensors to collect measurements that represent the individual appliances. The data from the distributed sensors is well suited as training data for evaluating different methods of central appliance load disaggregation. This study will use a publicly available dataset from such a facility in Pittsburg, USA [13]. The use of a public dataset reflects our goal to enable a more cumulative research development, as other research teams can more easily verify, compare and extend their analyses based on the openly available dataset.

Measuring at several points throughout a house is more invasive and more expensive than the option of using a central measuring point. The potential of applying more sophisticated software instead of more hardware is a major motivation for focusing on a central solution [14]. A single-point sensor is most often placed at the electrical mains by the fuse box or electricity meter. However, depending on the features that are measured, other placements are also possible [8]. While the central location of single point sensing is often convenient, the measurements can be influenced by other connected loads or the wire infrastructure [12], [15].

Apart from the amount of hardware necessary for the system the factor that impacts the cost, and thus the scalability of an appliance load disaggregation system, is the data sample-rate with which the electricity flow is measured. The sample-rate is generally divided into a macro and a micro level [16].

The macro category of disaggregation studies the sample-rate range from 1 data point per second (1Hz) [7], to 1 data point every 16 seconds (0.0625Hz) [17]. There is also research on disaggregating data on a 15-minute basis [18]. However, this is done "offline" by parsing the data multiple times and is therefore not suitable for real-time implementations.

At the micro frequency level, with more frequent measuring than once per second, additional information about the measured appliances' load characteristics is provided. The studies in this category range from a data sample-rate of a few kHz [19], [20] to 100MHz [8]. The publicly available dataset that this study is based on has a sample-rate of 12kHz [13]. This data thus allows for an exploration of both highly detailed features and, by downscaling the data, comparative data for lower frequencies.

B. Event detection

The event detection is an optional step depending on the subsequent method of analyzing the data. For example, an artificial neural network method analyzes the most likely combination of appliances for each evaluated dataset [21]. This method does not wait until a change has occurred and will therefore simply confirm that the same appliances are running if no change has been made. With this method no assumptions need to be made about individual appliance events. The downside of using non-event based methods, is that all combinations of potential appliances are needed for the analysis. This makes it challenging for the non-event approach to scale beyond 20 individual appliances [14].

The Building Level fUlly labeled Event based Dataset (BLUED) used in this study is based on individual appliance events. The event detection allows for focused analysis around certain points of interest, signified by the events [22]. Being able to parse the data in defined sample windows has the added benefit of limiting the number of appliance "finger prints" that need to be stored, since theoretically no combinations of appliances will be evaluated. Naturally frequently occurring events from a combination of appliances can also be stored for recognition.

This study will not use an event detection step as the primary focus is on the feature matching step and as the event detection has already been well described by the authors behind the BLUED [23]. However, a brief explanation of their design is included here for completeness. Their event detector method is based on the well-established General Likelihood Ratio (GLR), which can detect changes in signals as they pass through the filter "online" [24]. The online aspect is important for creating a responsive system. The specific event detection design follows the GLR implementation described by Luo et.al [25], with the addition of a power change threshold proposed by Berges et.al [26].

The GLR works on a sample set of incoming data. This sample set is divided into two smaller sample windows around the data point that is being analyzed for a potential event. First, a threshold change in power use of 30W is analyzed through subtracting the two sample windows. The 30W limit was considered by Berges et.al [26] to be a good balance between detecting events while removing much of the falsely labeled noise as events. The same threshold was later used by Anderson et.al [13] when parsing the BLUED. Second, a likelihood ratio is computed on the data that is above the threshold of power change, or otherwise set to zero. Third, the results from the likelihood ratio test are summed up into a test statistic. In the fourth step, the data point with the highest test statistic receives a vote, signifying the most probably event point. Finally the whole sample set is shifted to include new data points, which through the same procedure adds another vote to the most likely event point in the set. When a certain point has received enough votes, the sample set of data for that event is extracted and the specific features are calculated. The current methods for the feature extraction will be reviewed in the next section.

C. Feature extraction

The measured voltage and current has several layers of information or features that reflect certain characteristics of the appliances [27], [28]. In this section, the different features that have been used in related research will be presented. These features will be used in this study's matching step.

The most well-known feature of any appliance is its real power, which is the average of the product of multiplying the supplied instantaneous voltage and current. Purely resistive loads, for example electrical heating elements or incandescent lights, use real power. By evaluating the real power of an appliance operating modes it can be singled out if it has a stable and large power use. However, many appliances vary and overlap in this feature, which makes it necessary to provide more dimensions to the solution space [20]. For modern appliances with rectifiers and pulse-width-modulation or appliances with motors, the load creates a mismatch between the current and voltage sine waves. This results in a reactive power. By combining the real power (P) and the reactive power (Q), each appliance can be placed in two dimensions. Having two distinguishing characteristics assists in detecting unique appliances. The so-called PQ feature was, until recently, frequently the only feature used in appliance disaggregation research [7], [29], [30]. One common critique of the PQ feature is that it requires distinct power levels to recognize appliances' state changes. Dimmable lights or computers, which ramp the CPU up or down depending on current demand, are hard to detect with the PQ feature [20]. Furthermore, the variability of some appliances' normal operation overshadows a whole range of devices. For example, an office lamp, which use little power is difficult to detect with the PQ feature when it is running in parallel to a kettle, which has a variable steady-state power use [31].

One approach to distinguishing appliances that overlap within the PQ feature was suggested by Norford and Leeb [22]. These researchers found that even though the startup of different appliances reached similar power levels, the shape of how the appliances transitioned between the states was different. For example, a fluorescent light and an induction motor can both reach 400W power surge at the moment they were switched on. However, the power transient of the motor falls slowly, due to inertia, while the fluorescent light's power transient falls more abruptly. The so called "Switching Transient Waveform" (STW) [22], was used by Berges et.al to disaggregate appliances with a correct classification result of 82% [26].

While the STW feature focuses on appliance characteristics at times of state change, appliances can also be identified from their signal form at a steady-state as well. Both the Instantaneous Power Waveform (IPW) and the Instantaneous Admittance Waveform (IAW) are examples of these kinds of features. The IPW uses the instantaneous power over time to describe the shape, while the IAW uses the instantaneous admittance (current divided by voltage) over time. Both features have been used quite successfully with between 70 and 80% appliance detection accuracy [32].

The activity of an appliance can also be described without using voltage. For example, the "Current Waveform" (CW) feature is describing the altered sinusoidal curve of the current measurement. The CW feature has received mixed reception by the research community: one team claims it to be a distinct appliance feature [33], while another criticizes it for being too similar across different appliances [16]. In disaggregation tests the CW feature has been equally successful as the IPW and IAW features, with around 75% detection rate [32]. At much higher measurement frequencies and by limiting the tested appliances to seven distinct ones 97% were found using the CW feature [34]. However, in the same study comparisons with 15 appliances were made and subsequently the detection rate fell to a little more than 60%.

Current specific metrics neglect the voltage and evaluate the peak current (I_{peak}) and the root-mean-square current (I_{rms}) features [19]. These two features were reported to have a detection rate of 80% among 94 appliances. However, the experimental design and whether or not appliances had overlapping times of operation was not reported.

The final feature reviewed is increasing in popularity and is based on the harmonics (H) of the current measurements. The harmonics are multiples of the fundamental frequency (either 50Hz or 60Hz depending on the country) and are directly related to an appliance operation, particularly for semiconductors and power converters [35]. Fast Fourier analysis can analyze an appliance's harmonics in the frequency domain and with its information many overlapping appliances in the traditional PQ feature can now be separated [20]. Harmonics have also been used as a separate feature successfully. For example Srinivasan et.al [21] managed to detect 60-100% of the 8 appliances they were experimenting with.

The harmonics feature is limited by the available measurement frequency, as each higher harmonic requires the fundamental frequency multiplied by an additional integer. Furthermore, to recreate a signal, at least double the measurement bandwidth is needed according to the sampling theorem. This means that the measurement frequency must be at least double that of the harmonic level in focus. In practice this is not a problem as harmonics above the 11th are in most cases neglected [16]. The 11th harmonic level could thus theoretically be evaluated at a sample-rate beginning at 1320Hz (1100Hz) for a fundamental frequency of 60Hz (50Hz).

Features that will not be included in this study's evaluation are the electrical noise feature [8], electromagnetic interference feature [12] or features that are based solely on voltage [36]. The first two features (electrical noise and interference) are not considered in this paper since the necessary measurement frequency to detect them calls for sophisticated and costly hardware [14]. Additionally, the first method is path sensitive, which means that appliance noise is different depending on where the appliance is installed in the building [8]. Thus, if the appliance changes location, its outputted noise also changes. The second feature has a narrow focus and is specific to appliances with a switch mode power supply (SMPS). These SMPS are becoming more prevalent but still exclude several household appliances [12]. Pure voltage features in buildings have only been suggested in a study by Ruzzelli et.al [36] are not commonly studied in research. Cox et.al [37] suggested that voltage could be used as a feature in confined spaces like ships, however this environment is different to most building where voltage is actively managed in the network and kept within narrow tolerance limits.

D. *Appliance identification*

Based on the features derived from power and current the process of identifying individual appliances can begin. In this section the current theoretical concepts of appliance matching and how they relate to presented features will be reviewed. A brief overview of the identification algorithms will be given last.

Hart [7] defined three broad classes of appliances: binary state appliances, finite state appliances and infinite state appliances. The binary state appliances are, for example, normal lights that are simply turned on and off. Depending on the power levels between which these appliances switch and on what other appliances are running in parallel, binary state appliances can be detected with all the features presented in Section C. The finite state appliance can be thought of as a collection of binary state appliances that work consecutively or in parallel to produce the expected service. A washing machine is a good example that contains a binary state valve to let water into the machine, a binary state heating coil to warm the water, a motor capable at running at different binary states for the wash and spin cycles and finally a binary state pump to eject the used water. For an appliance identifier, all the different possible states that the appliance can assume must be known, which is similar to any binary state appliance. However, since many finite state appliances follow a predetermined path through a given program, the search space for the next event can be made relatively narrow [7]. Infinite state appliances lack distinct states and are therefore not detectable by the normal event filter that evaluates sudden changes. Even if an infinite state appliance would be detected, it would be challenging to build a database with which to match the recognized state.

A fourth class of single state appliances was later added by Baranski and Voss [18]. These types of appliances use a constant power and are never switched off. For example, aquarium pumps, electrical mains connected clocks and smoke detectors would fall into this category. These single state appliances do not produce any events, as they do not change states. To recognize this kind of appliances other non-event based methods that use the aggregated data for the analysis like neural networks must be used [21].

The process of identifying an appliance from a feature or a set of features follows either an optimization method or a pattern recognition method [38]. The optimization method uses a certain algorithm to minimize the difference of already-known appliance features and the measured ones. Commonly used optimization algorithms are the least square [27], [39], dynamic programming [40] and integer quadratic programming [34]. The pattern recognition, in contrast, commonly use a training set of several combined appliances that is then compared against the measured aggregated signal of active appliances according to a specific approach. Currently used approaches are the Naïve Bayes [26], [41], k-nearest neighbor [26], [31] and artificial neural networks [32], [36].

This study will focus on the optimization method as it lends itself well to event-based detection, which is scalable over several appliances. This last criterion is important for a solution that has to support several households and a large set of different appliance states [14].

III. Experiment Model

This section will detail the implemented system to evaluate the impact of data sample-rate on disaggregation accuracy.

First the dataset will be introduced followed by an explanation of the different treatments. Second, the experimental framework, which is based on the reviewed literature, will be presented.

E. The Public Dataset

So far, most published implementations of appliance recognition have been highly successful, with between 60 and 95% rate of detection [38]. However, this research has been done with specific datasets that have been exclusive to each research team. This has made it difficult to quantify the value of the additional hardware and software solutions between the studies [38]. Therefore, large datasets from detailed measurements have been made public in recent years. Most notably are the REDD and the BLUED dataset that provide open access and thus the ability to verify approaches between studies and research teams [13], [42].

The BLUED is an analogue version of the Reference Energy Disaggregation Dataset (REDD) that was made publically available by Kolter and Johnson [42]. The current and voltage measurements are gathered at the circuit panel in the single-family residential building. The current transformer sensors were attached to the two phases A and B, which supply the household with energy. These two power lines physically divide the measured appliances in two datasets that is also kept in the evaluation. The BLUED, as the name implies, is additionally labeled according to the appliances' events, which was made possible by plug-level and environmental sensors. This setup is shown in **FIG 2** and allows for a ground truth for the event based analyses. Thirty-one different appliances are recorded (8 on phase A and 23 on phase B) and span a power usage between 15W (kitchen light) to 1600W (hair dryer). The data was accumulated over a week at a sample-rate of 12kHz and take up 307 GB of computer storage space when uncompressed. The appliances were labeled through a system of environmental sensors and circuit level meters. 95% of the appliances were thus labeled, while the final 5% were labeled as "unknown" [13].

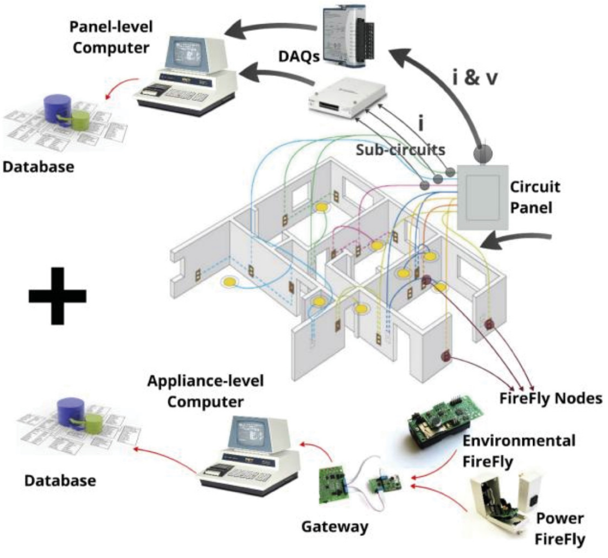

FIG 2. BLUED SETUP FOR DATA AND GROUND TRUTH COLLECTION [13]

To keep the integrity of the data, the treatments of different sample-rate levels is scaled in integer steps by removing data points from the dataset. Thus, the specific treatments that will be in focus are 12kHz, 6kHz, 4kHz, and so on down to 100Hz. This means that the frequency will be more detailed as the time in between data points becomes longer. Other potential methods of creating a more even down sampling would be to use a statistical average over a number of data points and interpolate the desired data point. Interpolation was not chosen because of the artifacts that this would introduce.

F. Experimental Design

The steps of a disaggregation framework were introduced in the related works (Section II) and are visualized above in Fig 1. Below, the specific setup for this experiment will be detailed. In general, the focus has been on creating a system that can handle a variety of sample frequencies and the algorithmic solutions have been chosen to support an online disaggregation and embeddable hardware. However, it should be noted that the simulation in this study uses a normal PC to parse the BLUED dataset offline to make the evaluation process more efficient.

As previously mentioned the step of the event detection will not be considered here. Anderson et.al's [23] results (which is also the group behind the BLUED) show 49-78% event detection accuracy when focusing on the number of appliances and 74-80% detection accuracy when focusing on the total detected power. The second step, of feature extraction, was based on the BLUED labeled event data.

The disaggregation accuracy was evaluated based on the different features presented in the related work (Section II-C) and the specific sample-rate treatment. The comparison between the measured features and the ones stored as known in a database was done using a 4-fold cross validation. This means that one part of the dataset was used as completely known labeled data used to create the load signature database (LSDB). The other three parts were treated as unknown trial data, which were matched against the LSDB. These two steps in the analysis are shown in **FIG 3**.

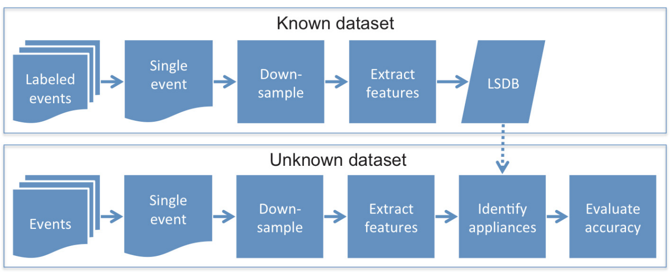

FIG 3. 4-FOLD CROSS-VALIDATION DATASETS, WITH ONE PART KNOWN TO BUILD LOAD SIGNATURE DATABASE (LSDB) AND THREE UNKNOWN PARTS FOR EVALUATING THE APPLIANCE DISAGGREGATION ACCURACY.

To match a feature against the known labeled features the least residue approach is used for features detailing brief events like the PQ, I_{max} or I_{peak}, while the features based on a shape are evaluated through cross-correlation. The least residue method is an optimization algorithm that minimizes the squared difference (residual) between the features of the found event against known appliance features. The cross-correlation on the other hand compares two time-series by running one of them over the other. This way, an exact timing of the feature between the two time-series is not necessary for matching the waveform, as every phase shift in the given range will be evaluated [43].

In the previous literature several metrics for determining disaggregation accuracy has been used. Hart [7] and Anderson et.al [13], for example, used a disaggregation definition based on the total amount of power that was correctly found. Although this metric is beneficial for developing methods for extracting the appliances with the highest power usage, it might not correlate with the appliances that the user would be willing to operate in a different and more sustainable way. This study, therefore, follows Liang et.al's [32] method, which was recommended in a review of current disaggregation methods by Zeifman and Roth [16]. This accuracy metric is calculated as the factor between the total number of disaggregated appliances and the total number of labeled events.

IV. Results

By benchmarking the results against a related study, which also has evaluated several different methods against each other, a sense of the implementation of the algorithms can be captured. It is natural that the results deviate as the underlying data of this study is based on the publicly available BLUED, while it previously has been done on exclusive research team data [32]. In Table I the results can be seen from this benchmark. This study uses a lower detection limit of 30W instead of Liang et.al's [32] 100W. A comparable higher limit is therefore provided for direct comparison and shows that the current study's implementation of the feature matching is comparable to the results from previous literature.

TABLE I
LOAD DISAGGREGATION ACCURACY COMPARISON BETWEEN THIS STUDY AND THE RESULTS OF LIANG ET.AL [32]

	Threshold / Phase	Appliance disaggregation accuracy P > 30 W A	P > 30 W B	P > 100 W A	P > 100 W B	P > 100 W Liang
Features	PQ	86.51%	60.82%	89.80%	81.52%	75-80%
	CW	72.57%	42.62%	88.16%	48.60%	80-85%
	IAW	61.67%	18.13%	80.62%	36.58%	75-80%
	IPW	69.66%	29.69%	80.62%	36.58%	75-80%
	H	81.71%	72.78%	90.03%	83.36%	65-70%

The main goal of the study is to analyze the impact of data sample-rate on disaggregation accuracy. Here below are the results from scaling the sample-rate down. The average disaggregation accuracy of the single-family household's Phase A show a quite consistent detection-rate between 12kHz and 100Hz, which is visualized in **Fig 4**. The standard error over the different features and validation runs are between 0.2 and 3.3 %.

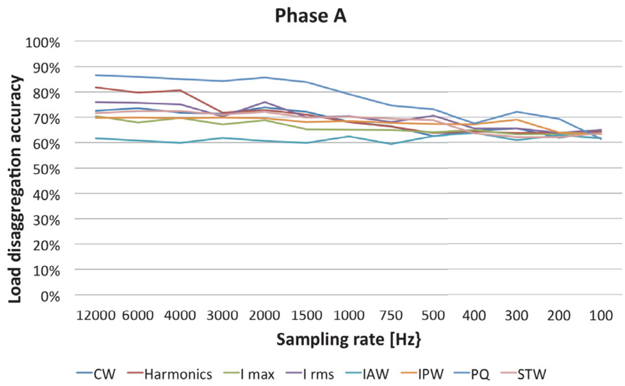

FIG 4. DISAGGREGATION ACCURACY FOR TESTED FEATURES OVER DIFFERENT SAMPLING RATES ON PHASE A.

The PQ feature provides the best results and stays above 85 % for sample-rates above 1.5 kHz. The harmonics feature show similar results for the higher sample-rates but proves to be more sensitive to a lower sample-rate as it drops from 80% to 70 % detection accuracy when the sample-rate is changed from 4 kHz to 3 kHz. The features based on current (I_{peak} and I_{max}) and waveforms (IAW, IPW, CW) all show none or a slight decline in detection accuracy between 70 % and 60 %.

Phase B is more sensitive to downscaling the data sample-rates than Phase A, which can be seen in **Fig 5**. It also has a higher standard error from 1.2 to 6.6 %.

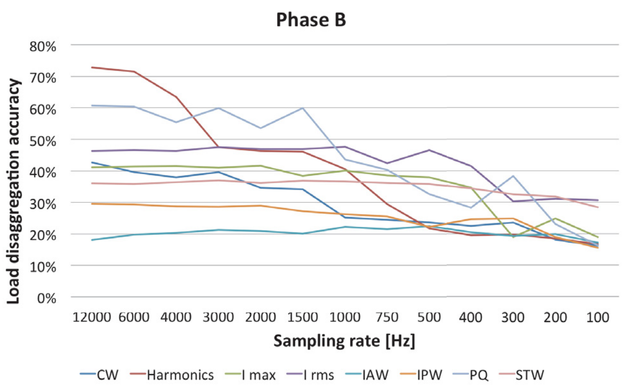

FIG 5. DISAGGREGATION ACCURACY FOR TESTED FEATURES OVER DIFFERENT SAMPLING RATES ON PHASE B.

In Phase B, the harmonics feature showed the best disaggregation result for the highest sample-rate. This feature, however, drops off rapidly and falls below the PQ feature after 4 kHz. The PQ feature is more stable at appliance detection over different sample-rates. However, below 1.5kHz it also drops substantially. Similar to Phase A, the current (I_{max} and I_{peak}) and waveform (STW, CW, IAW and IPW) features performed worse in terms of disaggregation accuracy and were also more stable over the data rate downscaling than the PQ and H features. Overall these features drop at most 20 % of detection accuracy (from 40 % to 20 % for CW) between 12 kHz and 100 Hz's sample-rate.

V. DISCUSSION

The disaggregation of individual appliances has developed much over the last few years. More precise measurement technology is used [12] and several sophisticated methods have been applied [32]. However, beyond the complexities of singling out appliances among similar ones, scaling these solutions for general use in households has not been evaluated. This study therefore focused on the influence of data sample-rate on disaggregation accuracy, which is one of the main parameters for choosing the necessary hardware.

By only focusing on the feature extraction and feature matching, this study simplifies the process of disaggregating appliances. More research comparing the different event-based detection is however warranted as Anderson et.al [13] was only able to detect the events of Phase A with high certainty, while Phase B resulted in 50% missed events or falsely detected events. Jin et.al [44] has suggested the goodness-of-fit event detection method to be superior to the GLR method used by Anderson et.al [13], which should be an interesting starting point for future research in this direction.

The detection accuracy was negatively affected by lowering the sample-rate, as expected. Phase A proved more robust than Phase B. The more stable impact over different sample-rates is due to the underlying appliances supplied by the two phases. First of all, the number of appliances are fewer, 8 compared to Phase B's 23 individual appliances. Second, one of Phase A's 8 appliances is dominant in terms of events (the refrigerator), which is also recognizable with all tested features.

Overall, the PQ and the harmonics features showed the best results, the higher the data sample-rates were, on both phases in the BLUE Dataset. These results are comparable to the ones found by Liang et.al [32], who compared different features under a given sample-rate. Contrary to our findings however, Liang found that the waveform features (CW, IAW, IPW) were superior to the harmonics feature. This could be explained by the difference in the underlying data or the parameterization of the feature extraction algorithms. The difference in used data has been an inhibiting factor for comparing results between studies and this is why this study has elected to use the publicly available dataset provided by Anderson et.al [13].

An important result from this study is that the features proved to be robust over quite a range of down sampling. The PQ feature, for example, started its decline around 1.5kHz and the harmonics feature was negligibly affected by the lower sample-rate until 4kHz on Phase A and 6kHz on Phase B. This implies that disaggregation systems can be designed in the range between 1.5-6kHz and still accurately detect appliances comparable to other similar studies [32]. The most stable over all data samples were the STW and IAW features. Improving the accuracy of these features by using different parameters or alternative analysis methods than the cross-correlation used in this study would be interesting points for further research.

VI. CONCLUSION

This study contributes with an evaluation and parameterization of the requirements for future scalable household disaggregation systems. Several different features from this data were evaluated. The real and reactive (PQ) feature and the harmonics (H) feature were superior in our results in terms of overall detection accuracy, while the Switching Transient Waveform (STW) and the Instantaneous Admittance Waveform (IAW) proved to be most robust over the whole range of evaluated data sample-rates (12 kHz to 100 Hz). The PQ feature was slightly affected by the downscaling of the sample frequency between 12 kHz and 1,5 kHz, while the H feature was more sensitive and dropped in accuracy below 4 kHz (Phase A) and 6 kHz (Phase B).

By using a public dataset (BLUED) for the analysis, this study also contributes with much needed disaggregation research that can be verified and continually developed by other research teams. We hope that these results will spur on the development of a general and scalable energy information system that can support more knowledgeable and sustainable decisions.

ACKNOWLEDGMENT

The authors would like to express their gratitude to student Johannes Beck, whose development and testing of the experiment framework has been invaluable to this study.

REFERENCES

[1] A. Booth, N. Demirdoven, and H. Tai, "The smart grid opportunity for solution providers," Atlanta, 2010.

[2] D. G. Hart, "Using AMI to realize the Smart Grid," in 2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008, pp. 1–2.

[3] K. Ehrhardt-Martinez, K. A. Donnelly, and S. Laitner, "Advanced metering initiatives and residential feedback programs: A meta-review for household electricity-saving opportunities," in American Council for an Energy-Efficient Economy, 2010.

[4] C. Fischer, "Feedback on household electricity consumption: a tool for saving energy?," Energy Effic., vol. 1, no. 1, pp. 79–104, May 2008.

[5] H. Pihala, Non-intrusive appliance load monitoring system based on a modern kWh-meter, 356th ed. Espoo, Finland: VTT Publications, 1998.

[6] O. Parson, "Unsupervised Training Methods for Non-intrusive Appliance Load Monitoring from Smart Meter Data," University of Southampton, 2014.

[7] G. W. Hart, "Nonintrusive appliance load monitoring," Proc. IEEE, vol. 80, no. 12, pp. 1870–1891, 1992.

[8] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, and G. D. Abowd, "At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line," in UbiComp 2007: Ubiquitous Computing, 2007, vol. 4717, pp. 271–288.

[9] N. J. Nunes, L. Pereira, F. Quintal, and M. Berges, "Deploying and evaluating the effectiveness of energy eco-feedback through a low-cost NILM solution," in Persuasive'2011, 2011.

[10] R. R. Schaller, "Moore's law: past, present and future," IEEE Spectr., vol. 34, no. 6, pp. 52–59, Jun. 1997.

[11] Y. Du, L. Du, B. Lu, R. Harley, and T. Habetler, "A review of identification and monitoring methods for electric loads in commercial and residential buildings," in 2010 IEEE Energy Conversion Congress and Exposition, 2010, pp. 4527–4533.

[12] S. Gupta, M. S. Reynolds, and 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 - Ubicomp '10, 2010, pp. 139–148.

[13] K. Anderson, A. Ocneanu, and D. Benitez, "BLUED: a fully labeled public dataset for Event-Based Non-Intrusive load monitoring research," in Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability, 2012.

[14] M. Zeifman, "Disaggregation of home energy display data using probabilistic approach," IEEE Trans. Consum. Electron., vol. 58, no. 1, pp. 23–31, Feb. 2012.

[15] A. Dalen and C. Weinhardt, "Fusebox dialects: Central appliance recognition for a smarter grid," in 2012 IEEE PES Innovative Smart Grid Technologies (ISGT), 2012, pp. 1–7.

[16] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," IEEE Trans. Consum. Electron., vol. 57, no. 1, pp. 76–84, Feb. 2011.

[17] L. Farinaccio and R. Zmeureanu, "Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses," Energy Build., vol. 30, no. 3, pp. 245–259, 1999.

[18] M. Baranski and J. Voss, "Non-intrusive appliance load monitoring based on an optical sensor," in 2003 IEEE Bologna Power Tech Conference Proceedings,, 2003, vol. 4, pp. 267–274.

[19] T. Saitoh, Y. Aota, T. Osaki, R. Konishi, and K. Sugahara, "Current sensor based non-intrusive appliance recognition for intelligent outlet," in The 23rd International Technical Conference on Circuits/Systems, Computers and Communications, 2008, pp. 349–352.

[20] C. Laughman, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong, "Power signature analysis," IEEE Power Energy Mag., vol. 1, no. 2, pp. 56–63, Mar. 2003.

[21] D. Srinivasan, W. S. Ng, and A. C. Liew, "Neural-Network-Based Signature Recognition for Harmonic Source Identification," IEEE Trans. Power Deliv., vol. 21, no. 1, pp. 398–405, Jan. 2006.

[22] 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 Build., vol. 24, no. 1, pp. 51–64, 1996.

[23] K. D. Anderson, M. E. Berges, A. Ocneanu, D. Benitez, and J. M. F. Moura, "Event detection for Non Intrusive load monitoring," in IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, 2012, pp. 3312–3317.

[24] M. Basseville and I. V. Nikiforov, Detection of abrupt changes: theory and application. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[25] D. Luo, L. Norford, S. Leeb, and S. Shaw, "Monitoring HVAC equipment electrical loads from a centralized location- methods and field test results.," ASHRAE Trans., vol. 108, no. 1, pp. 841–857, 2002.

[26] [26] M. Berges, E. Goldman, H. S. Matthews, L. Soibelman, and K. Anderson, "User-Centered Nonintrusive Electricity Load Monitoring for Residential Buildings," J. Comput. Civ. Eng., vol. 25, no. 6, pp. 471–480, Nov. 2011.

[27] [27] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng, "Load Signature Study—Part I: Basic Concept, Structure, and Methodology," IEEE Trans. Power Deliv., vol. 25, no. 2, pp. 551–560, Apr. 2010.

[28] [28] M. Ito, R. Uda, S. Ichimura, K. Tago, T. Hoshi, and Y. Matsushita, "A method of appliance detection based on features of power waveform," in CCECE 2003 - Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology (Cat. No.03CH37436), 2004, pp. 291–294.

[29] [29] M. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, "Learning Systems for Electric Consumption of Buildings," in Computing in Civil Engineering (2009), 2009, pp. 1–10.

[30] [30] A. I. Cole and A. Albicki, "Algorithm for nonintrusive identification of residential appliances," in ISCAS '98. Proceedings of the 1998 IEEE International Symposium on Circuits and Systems (Cat. No.98CH36187), 1998, vol. 3, pp. 338–341.

[31] [31] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake, "Leveraging smart meter data to recognize home appliances," in 2012 IEEE International Conference on Pervasive Computing and Communications, 2012, pp. 190–197.

[32] [32] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng, "Load Signature Study—Part II: Disaggregation Framework, Simulation, and Applications," IEEE Trans. Power Deliv., vol. 25, no. 2, pp. 561–569, Apr. 2010.

[33] W. Lee, G. Fung, H. Lam, F. Chan, and M. Lucente, "Exploration on load signatures," in International Conference on Electrical Engineering (ICEE), 2004, pp. 1–5.

[34] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, and K. Ito, "Nonintrusive appliance load monitoring based on integer programming," in 2008 SICE Annual Conference, 2008, pp. 2742–2747.

[35] J. Arrillaga and N. R. Watson, Power System Harmonics. Chichester, UK: John Wiley & Sons, Ltd, 2003.

[36] A. G. Ruzzelli, C. Nicolas, A. Schoofs, and G. M. P. O'Hare, "Real-Time Recognition and Profiling of Appliances through a Single Electricity Sensor," in 2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2010, pp. 1–9.

[37] R. W. Cox, P. L. Bennett, T. D. McKay, J. Paris, and S. B. Leeb, "Using the Non-Intrusive Load Monitor for Shipboard Supervisory Control," in 2007 IEEE Electric Ship Technologies Symposium, 2007, pp. 523–530.

[38] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey.," Sensors (Basel)., vol. 12, no. 12, pp. 16838–16866, Jan. 2012.

[39] C. Beckel, W. Kleiminger, T. Staake, and S. Santini, "Improving device-level electricity consumption breakdowns in private households using ON/OFF events," ACM SIGBED Rev., vol. 9, no. 3, pp. 32–38, Jul. 2012.

[40] M. Baranski and J. Voss, "Detecting Patterns of Appliances from Total Load Data Using a Dynamic Programming Approach," in Fourth IEEE International Conference on Data Mining (ICDM'04), 2004, pp. 327–330.

[41] K. Chahine, K. E. K. Drissi, C. Pasquier, K. Kerroum, C. Faure, T. Jouannet, and M. Michou, "Electric Load Disaggregation in Smart Metering Using a Novel Feature Extraction Method and Supervised Classification," Energy Procedia, vol. 6, pp. 627–632, Jan. 2011.

[42] J. Kolter and M. Johnson, "REDD: A public data set for energy disaggregation research," in proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, 2011, pp. 1–6.

[43] M. B. Figueiredo, A. de Almeida, and B. Ribeiro, "Wavelet Decomposition and Singular Spectrum Analysis for electrical signal denoising," in 2011 IEEE International Conference on Systems, Man, and Cybernetics, 2011, pp. 3329–3334.

[44] Y. Jin, E. Tebekaemi, M. Berges, and L. Soibelman, "Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities," in 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 4340–4343.