

AUDIOSENSE: Leveraging Current to Acoustic Channel to Detect Appliances at Single-Point

Yijie Li[†], Xiatong Tong[†], Qianfei Ren[†], Qingyang Li[†], Lanqing Yang[†],

Yi-Chao Chen[†], Guangtao Xue^{†*}, Xiaoyu Ji[‡], Jiadi Yu[†],

[†] Department of Computer Science and Engineering, Shanghai Jiao Tong University, China

[‡] College of Electrical Engineering, Zhejiang University, China

Email: {yijieli,txt0313, cordial.Daniel, yanglanqing, disward2017, yichao_gt_xue}@sjtu.edu.cn, xji@zju.edu.cn, jiadiyu@sjtu.edu.cn

Abstract—Over the past years, smart ecology has attracted much attention, especially for smart home applications. As a key component, monitoring appliances performs significant impact. However, appliances under monitoring usually contain smart modules such as WiFi or Bluetooth, which are limited to traditional appliances. Existing approaches such as distributed sensing, energy disaggregation, and infrastructure-mediated sensing, require the installation of external hardware or have a limited sensing range. In this study, we developed AUDIOSENSE to leverage the acoustic signal generated by the power supply to monitor electrical appliances throughout the house remotely from a single point. In realizing AUDIOSENSE, we proposed an optimized Variation Mode Decomposition scheme to extract the frequency components, as well as a data augmentation scheme to improve generalizability and enable multi-label classification. In experiments, AUDIOSENSE achieved mAP values of 99.3% in multi-label classification.

Index Terms—Appliance detection; Acoustic; Side-channel

I. INTRODUCTION

Technology visionaries have long sought to create a smart home ecosystems to enable monitoring and communication among appliances. As the smart speakers have proliferated, and provided voice interactive capabilities using a microphone, which played a critical role to monitor smart appliances. However, commercial off-the-shelf (COTS) services require communication modules (Bluetooth, WiFi, etc.) be pre-installed in appliances, thereby necessitating the replacement of old appliances or the installation of smart modules. This has greatly curtailed the dissemination of centralized smart home ecosystems. Therefore, enabling the ability of a smart speaker to monitor traditional appliances **remotely** without installing **additional hardware** is of extraordinary significance.

Existing mechanisms used to monitor the operating status of electrical appliances can be divided into three categories. First involves the installation of *distributed sensors* on each appliance, including but not limited to, RFID tags [1], magnetometers [2], microphones [3], or smart plugs [4], which imposes additional costs. Second approach is referred to as *non-intrusive load monitoring* (NILM) [5], which involves tracking appliances by monitoring energy consumption based on total energy. Unfortunately, it is unable to deal with complex time-varying appliances (e.g., computers) or differentiate

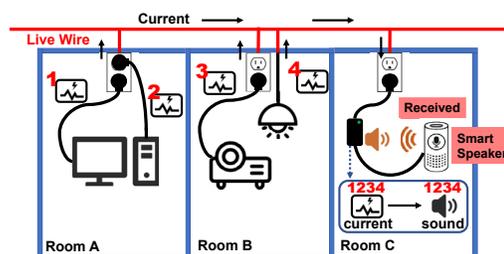


Fig. 1: Illustration of AUDIOSENSE using a smart speaker to identify electrical appliances from a single point.

appliances of the same model [6]. Third approach is referred to as *infrastructure-mediated sensing* (IMS), which utilizes side-channels that affect the infrastructure of the house (including gas [7], water [8], and electricity [9]–[13]). All of these methods require the installation of additional hardware or are ill-suited to monitor multiple appliances at the same time due to limited range.

In the current study, we sought to propose AUDIOSENSE, a novel software-based solution to appliance monitoring, wherein a smart speaker monitors the status of appliances **remotely** (e.g., appliances in different rooms) without needing to install additional hardware. As shown in Fig. 1, the working principle of AUDIOSENSE is that when an appliance is working, its power factor correction (PFC) module¹ in the power supply generates high-frequency **current** (i.e., PFC signal) (Sec. III-A), which thereby interferes with the power network and propagates to the branch **current** of a smart speaker's power supply (Sec. III-B). Current passes through the power supply and activates internal electric components, which emit high-frequency **sounds** (Sec. III-C), due to Magnetostriction and Piezoelectric effect. Specific components contain specific PFC components, which can be detected by the built-in microphone in smart speakers. Analysis of these sounds allows AUDIOSENSE to infer the working states of the appliance.

The detection and identification from longer distances (e.g., $> 50m$), imposes several challenges. First, the frequency bands of PFC signals from different appliances tend to overlap, resulting in a complex mix of signals. In situations involving

¹Not all appliances have a PFC module but it's becoming more and more common. We will discuss it in Sec. III-A

* Corresponding author.

multiple appliances operating simultaneously, it is necessary to differentiate among individual appliances in the overlapping spectrum. Second, PFC signals from appliances located farther from the smart speaker are subject to attenuation, resulting in weaker sound. Noise from the environment, other appliances, and the smart speaker itself also degrade signal quality. In addition, different types of PFC modules can generate different signal patterns. Thus, we developed a denoising scheme to extract PFC features from noisy data, while dynamically optimizing its configuration to accommodate various sets of working appliances.

To deal with these issues, we implemented a spectral subtraction scheme to mitigate the noise emitted by the smart speaker and external environment (Sec. VI-A1). Optimal variational mode decomposition (OVMD) is then used to adaptively optimize the bandwidth of each intrinsic mode function in decomposing the PFC signal (Sec. VI-A2). To address the problem of overlap and signal attenuation over long distances, we propose a data augmentation scheme aimed at increasing generalizability (Sec. VI-B). Finally, the obtained data is used to build a multi-label classification model by which to identify and differentiate among multiple appliances (Sec. VI-C).

We implemented AUDIOSENSE on 28 different electrical appliances (13 of appliances were the same model.) The proposed system was also tested in a real-world family household. The contributions are summarized as follows:

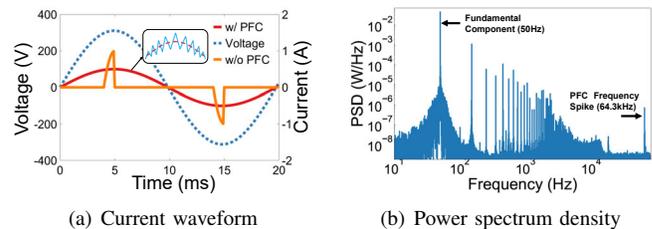
- We proposed a novel study to use the **sound** generated by power supplies to remotely identify electrical appliances at a **single point** based on a **current-acoustic model**.
- We developed an Optimized Variation Mode Decomposition scheme to adaptively optimize the bandwidth of each mode in decomposing PFC features in audio signals.
- We proposed a data augmentation scheme to deal with the issue of distance attenuation and signal overlap. We also developed a multi-label classification model using augmented training data from single-label data to improve usability by avoiding multi-label collection.
- The system was evaluated by conducting exhaustive experiments in real-world scenarios. The overall performance of AUDIOSENSE achieved mean average precision of 99.3%.

II. RELATED WORK

A. Electrical Appliance Detection

Close-range sensing: A common idea for detecting electrical appliance is to install distributed sensors on each appliance [14], [15]. Device-level detection performs straightforward sensing results but requires costly installation and maintenance. Another indirectly sensing technologies such as visual-based [16], audio-based [3], radiated-emission based approaches [2] must be used with limited distance and number of devices. Our approach on contrary can detect multiple electrical events with a single-point detection.

Non-intrusive load monitoring: Non-intrusive load monitoring (NILM) [5] detect electrical appliances based on their



(a) Current waveform (b) Power spectrum density
Fig. 2: Illustration of power factor correction. (a) Current waveform generated by PFC module. (b) Power spectrum density of a Dell desktop current.

energy usage patterns. State-of-the-art solutions disaggregate power consumption data and extract the consumption characteristics from the total energy usage [6], [17], [18]. However, NILM techniques are unsuccessful in identifying complex time-varying consumption devices (e.g. desktops, projectors) and low-power equipments (e.g. CFLs, LEDs). Moreover, they also fail in distinguishing appliances of the same model. Our approach utilizes time-invariant feature which makes up the shortcomings of conventional NILM techniques.

Infrastructure-mediated sensing: Infrastructure-mediated sensing (IMS) [19] is another single point technology which detects events that affect house infrastructure including gas [7], water [8], electricity [9]–[13], and infrastructure vibration [20]. Patel, et al. [9], [10] and Gulati [12], [13] measure EMI using expensive systems (e.g., USRP and spectrum analyzers.). NoDE [21] and OutletSpy [22] sense powerline voltage using an oscilloscope. Our approach also belongs to IMS technologies but it does not need any additional device unlike the approaches mentioned above.

B. Acoustic-based Application

Acoustic-based sensing [23] technologies have been widely employed with the explosive growth of speakers and microphones on commodity devices, ranging from indoor localization [24], tracking [25], health sensing [26], and so on. For the electrical events, there exists techniques by hearing audible motor sound [3], [27]. However, these approaches may endanger users' privacy and they require a close-range detection. Our approach mainly focuses on inaudible sound ($> 20kHz$) without violating privacy. Moreover, we innovatively propose a scheme to monitor appliances using sound generated by power supplies based on a current-acoustic model.

III. BACKGROUND

A. Power Factor Correction Circuit

Power factor correction (PFC) circuits are meant to improve energy utilization by generating high frequency ripple using pulse width modulator (PWM) to reduce harmonics as shown in Fig. 2(a). The PWM switches periodically to fit the sine wave, generating a current of specific switching frequency. Fig. 2(b) presents an example of a frequency spike generated by a Dell desktop. According to IEC61000-3-2 standard [28], PFC modules must be installed in lighting equipment exceeding 5W and Class D equipment (e.g., monitors, TVs, computers, etc.). PFC modules are also used in chargers to

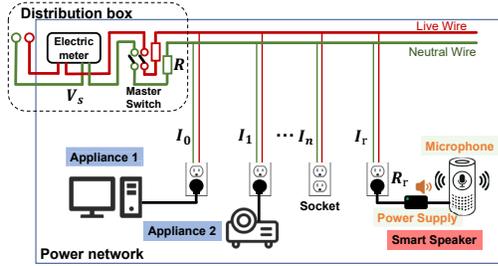


Fig. 3: Illustration of an electrical network in a house with the distribution box delivering electricity to sockets.

reduce energy consumption [29]. Therefore, the appliances containing PFC modules are common in daily life.

B. Powerline Interference

We then describe how PFC circuits in appliances interfere with power networks. A typical indoor electrical supply network (shown as Fig. 3) contains a distribution box who delivers electricity in parallel to outlets in different rooms. The smart speaker and other appliances can be plugged into any outlet. The relationship between the current in each branch I_i and the source supply voltage V_s satisfies the following equation: $I_r R_r + (I_0 + I_1 + \dots + I_n + I_r)R = V_s$, where I_i denotes the current of the i^{th} appliance, R denotes the common resistance of the distribution box, R_r denotes the self-resistance, and I_r denotes the current of the smart speaker. Thus, the current of the smart speaker (I_r) can be derived:

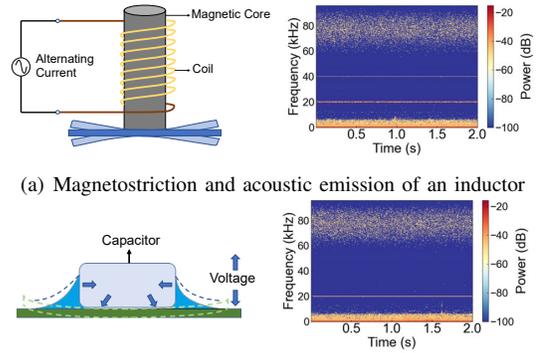
$$I_r = \frac{V_s}{R + R_r} - \frac{R}{R + R_r} \sum_{i=1}^n I_i \quad (1)$$

From Eq. 1, the current I_r of the smart speaker contains the current components of other appliances (Sec. IV-B). As mentioned in Sec. III-A, the PFC circuit generates output current ripples with special frequency spikes, which makes it possible to detect appliances by the smart speaker.

C. Acoustic Emission

Then we show how the current I_r causes the power supply to emit sound. A typical power supply structure contains inductors and capacitors, who will generate acoustic signal caused by Magnetostriction and Piezoelectric effects. The *magnetostriction*, illustrated in Fig. 4(a), refers to the phenomenon that a magnetic material elongates or contracts in a magnetic field, most occurs in coils, resulting in **the same** and double vibration frequency to the frequency of alternating current passing through the coil [30], accordingly the sound will be heard. The *Piezoelectric effect* [31] is a phenomenon that a high frequency electrical signal is applied to a capacitor, producing high frequency mechanical vibrations. The vibrating capacitor generate acoustic noise as **the same** frequency of current (as shown in Fig. 4(b)). This noise mainly occurs in capacitors and crystal oscillator.

Remarks. To summarize, inductors and capacitors both emit sounds and a frequency **corresponding to** the frequency of the current passing through the circuit. Fig. 4 presents a spectrogram of the sound emitted the same frequency as the



(b) Piezoelectric effect and acoustic emission of a capacitor

Fig. 4: Illustration of Magnetostriction and Piezoelectric effect; and the spectrogram of the acoustic emission.

passing current of 20kHz . This makes it possible to convert specific components of current channel to acoustic channel, which is the main principle of AUDIOSENSE.

IV. FEASIBILITY STUDY

In this part, we focused on answering two questions: i) Can PFC signals be used as a fingerprint to identify electrical appliances? ii) Does the acoustic signal contain corresponding PFC current components, and can they be detected remotely? In this section, we used an ACS712 current sensor of 160kHz to measure the current of 25 appliances of 15 models connected in series over 10 days.

A. PFC Signal as Fingerprint

Uniqueness. We first assessed whether the PFC signals of the appliances were unique. The dataset was divided into a subset of different models and another subset of the same model. We utilized t-SNE to visualize the clustering relationships among data points. Fig. 5(a) plots the t-SNE between 15 models. The figure reveals obvious variations across different models. Fig. 5(b) presents the t-SNE of the same model, including 5 lights, 4 desktop computers, and 4 monitors. Note that inter-class variation far exceeded intra-class variation. Nonetheless, we still observed differences in PFC, due to variations in the manufacturing processes. These results reveal that the PFC frequencies of both different model and the same model of appliances are distinguishable.

Stability. We then assessed whether the PFC signals were stable over time. Fig. 6 presents variations in the CDF of PFC center frequency over various durations (*1day, 2days, 3days*). Overall, the variation in PFC signal is within 40Hz which confirms that the PFC signal is stable. To summarize, the PFC signal was unique to each appliance and stable over time, thereby confirming that it could be used as a fingerprint in identifying electrical appliances.

B. Current-Acoustic Model

The PFC signal can be attenuated through the powerline due to line resistance. In this section, we verify whether the sound emitted corresponds to the branch current and whether the effects can be detected remotely.

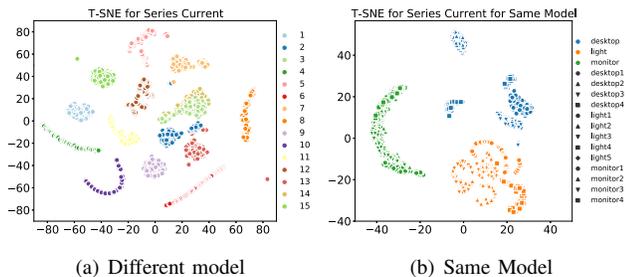


Fig. 5: The t-SNE for series current of different models and same models, respectively.

Distance attenuation. Line resistance increases with distance with a corresponding decrease in the amplitude of the PFC signal. We first assessed whether the PFC signal is able to activate the audio signals from a remote location. We conducted experiments by adjusting line length on the transmitter (from $1.8m$ to $18m$). Fig. 7 presents the SNR of the audio signal from 3 appliances. Interestingly, the SNR of the projector presented a slight improvement due to compensation for voltage loss [32]. The SNR of the other two different lights decreased, whose dropping slope was roughly $-0.13dB/m$ by a linear regression. It means that the maximum distances approximately $53m$, which can meet the requirements for covering a common family.

Consistency. We then assessed whether PFC signals can be transmitted to a smart speaker and whether doing so would activate the supplies to emit sound. A Dell desktop was used as a transmitter and a Tmall smart speaker power supply was used as the receiver in another room. Note that an electret condenser microphone² was used to sample the sound at $192kHz$. Fig. 8 presents the same frequency spike at $64.385kHz$ of the spectrum of the current passing through the smart speaker and audio signal. These results demonstrate the consistency between the audio signal and current.

Remarks. AUDIOSENSE leverages the PFC feature in emitted audio signal of power supply for appliance identification. It is meant to convert the current/voltage channel into a sound channel, which can be implemented on any device with a built-in microphone without additional equipment. Moreover, AUDIOSENSE extends the function to be applied to a wider range of equipment for different types of PFC modules (Sec. V).

V. PRELIMINARY ANALYSIS OF AUDIO SIGNAL

The audio signal is strongly correlated to the PFC component. In this section, we highlight the analysis of the received audio signal with various PFC components.

Impact of Type of Devices. Typically, PFC modules take different conduction modes in different devices according to different power levels [33]. As for high-power appliances ($> 300W$), it usually performs a constant frequency. In comparison, PFC modules in low-power appliances perform changing frequency, resulting in a wide bandwidth. Fig. 9 presents the audio signal emitted by the power supply of

²Both electret condenser and MEMS microphones can effectively reduce electromagnetic interference to ensure the picked up signal is sound, while dynamic microphones will suffer EMI.

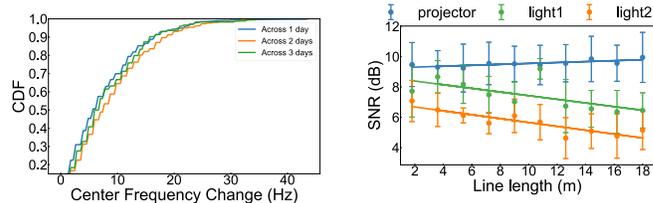


Fig. 6: The CDF of PFC center frequency variation.

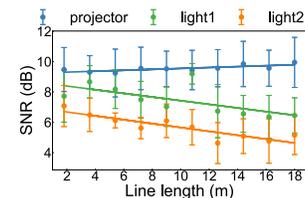


Fig. 7: SNR of the sound of three models under distance attenuation.

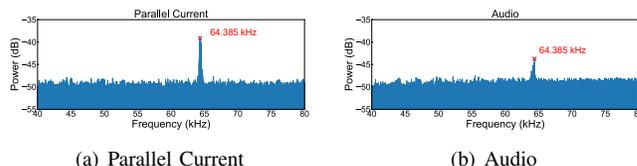


Fig. 8: Spectrum of the parallel current and the audio signal

a projector ($330W$) and phone charger ($65W$) while under working. The results revealed the projector produced a narrow-band frequency whereas the charger produced a fluctuating frequency over a wide band. Note that even in cases where the frequency changes often, it remained a periodic signal with a constant center frequency.

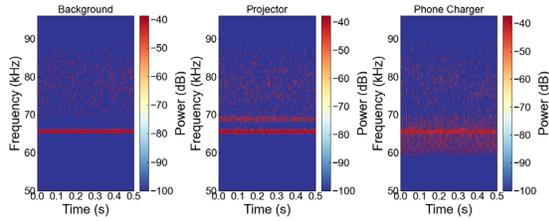
Impact of Working States. An increase in load prompts an increase in current to provide sufficient power. The role of a PFC module is to fit a sinusoidal current wave in accordance with the current. Fig. 10 presents the PFC component in the acoustic signal from an induction cooker operating under various power levels. Clearly, the frequency of the PFC component varied as a function of power, thereby demonstrating that the PFC component can be used to identify the working states.

Impact of Charging States. Charging is a common issue with many devices, and monitoring the charging state is important. The processing involved in charging has three modes according to battery level: pre-charge mode (battery level $< 20\%$), constant current charging mode ($20\% - 80\%$), and constant voltage charging mode ($> 80\%$) [34]. Fig. 11 presents the acoustic signal received from a phone charger operating under various battery levels. We found that the battery level influenced the PFC signal, thereby confirming that AUDIOSENSE could potentially be used to detect the charging state of devices.

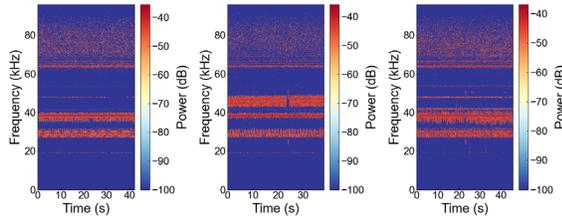
Impact of Aliased Sampling. In reality, the sampling rate of the microphone can be lower than the Nyquist rate ($f_s > 2f$). When sampling a high-frequency signal at the sub-Nyquist rate, the frequency component is aliased or folded back, as the following:

$$f_a = \begin{cases} (N+1)f_s - f & f_s/2 < f - Nf_s < f_s \\ f - Nf_s & 0 \leq f - Nf_s \leq f_s/2 \end{cases}$$

where f_a refers to the aliased frequency, f is the original frequency, f_s is the sampling frequency and $N = 0, 1, 2, \dots$. Fig. 12 shows the audio signals obtained by sampling at different sampling rates. The original frequency was $69kHz$ sampled at $192kHz$ (shown as Fig. 12(a)). When the sampling rate was reduced to $96kHz$ and $48kHz$, the correspond-



(a) Background (b) Projector (c) Phone charger

 Fig. 9: Audio signal of power supply that the projector performs constant frequency ($69kHz$) and the phone charger performs changing frequency ($59kHz-69kHz$).


(a) Battery < 20% (b) Battery 20%-80% (c) Battery > 80%

Fig. 11: Audio signal of power supply from a phone charger as a function of battery level.

ing frequencies dropped to $27kHz$ (Fig. 12(b)) and $21kHz$ (Fig. 12(c)), respectively. This means that existing COTS smart speakers are able to sample PFC features in audio signals.

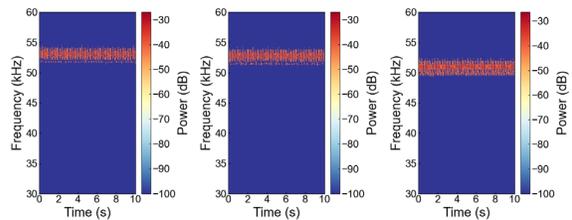
VI. SYSTEM DESIGN

In this work We proposed AUDIOSENSE, which leverages the acoustic signals generated by power supply to detect electrical appliances. Fig. 14 illustrates the architecture of the proposed AUDIOSENSE system, which involves a *single-label registration phase* and a *multi-label classification phase*. In *single-label registration phase*, AUDIOSENSE begins collecting 2-min data when an appliance is firstly accessed. The collected signal then undergoes spectral subtraction (Sec. VI-A1) to boost the SNR, Optimized Variational Mode Decomposition (OVMD) and periodicity detection (Sec. VI-A2) to extract PFC feature. The signal then is segmented into multiple parts and applied STFT to obtain spectrograms. A data augmentation scheme (Sec. VI-B) is applied to improve generalizability. Finally, a CNN model is trained a classifier for multi-label classification (Sec. VI-C). *Multi-label classification* involves classifying multi-label data using augmented data from single-label data. It undergoes the same process as the registration phase. AUDIOSENSE captures the sound of the power supply and predict the result of which appliance is working.

A. Preprocessing

The collected audio signals comprise the background noise and the sound produced by external PFC signals. The clean extraction of the band is crucial to subsequent feature extraction and classification. The preprocessing is presented below.

1) *Spectral subtraction*: Our aim is to boost SNR by weakening the background noise and extracting the band includes



(a) Low power (b) Medium power (c) High power

Fig. 10: Audio signal of power supply when an induction cooker works in different power.

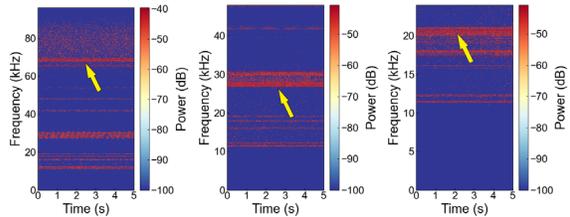

 (a) $f_s = 192k$ (b) $f_s = 96k$ (c) $f_s = 48k$

Fig. 12: Audio signal of power supply when sampled at different sampling rates.

the PFC component. The signal received by the microphone Y_r (Fig. 13(b)) can be denoted as follows:

$$Y_r = S_{pfc}(\omega_0) + Y_{bg}$$

where $S_{pfc}(\omega_0)$ refers to the external PFC signal and ω_0 refers to the central frequency of PFC component. Y_{bg} denotes the background noise. Such background noise is time invariant which obeys a specific distribution. This allows using spectral subtraction to obtain an enhanced set of PFC component $\hat{S}_{pfc}(\omega_0)$ as follows:

$$\hat{S}_{pfc}(\omega_0) = F^{-1}[\{|Y_r| - E[|Y_{bg}|]]e^{j\theta}] = S_{pfc}(\omega_0) + \epsilon$$

where $E[|Y_{bg}|]$ refer to an estimate of the background noise spectra, F^{-1} is the IFFT, θ is the original phase, and ϵ is estimation error. Fig. 13(c) presents the signal under SNR boosting. Note that spectral subtraction can only weaken the noise but cannot completely remove it. Thus, we need to further extract the band of the PFC component without interference from extraneous noise.

2) *Optimized Variational Mode Decomposition*: Even after the spectral subtraction phase, weakened background noise still exists. While the center frequency of the PFC signal remains unchanged, Variational Mode Decomposition (VMD) [35] is used to decompose the audio signal into several band components, referred to as Intrinsic Mode Functions (IMFs), from which we can extract the PFC component.

Conventional VMD. When dealing with a 1D signal $\hat{S}_{pfc}(\omega_0)$, VMD separates it into k narrow band signals, which are denoted by u_k , with different values for the estimated central frequency ω_k . The analytic signal of each u_k is shifted from the pass band to the base band whereupon the penalty term is computed. The objective of optimization in the frequency domain can be represented as follows:

$$\min_{u_k(\omega_0)} \left\{ \left\| \sum_k u_k(\omega_0) - \hat{S}_{pfc}(\omega_0) \right\|_2^2 + \alpha \sum_k \left\| j\omega u_k(\omega_0 - \omega_k) \right\|_2^2 \right\}$$

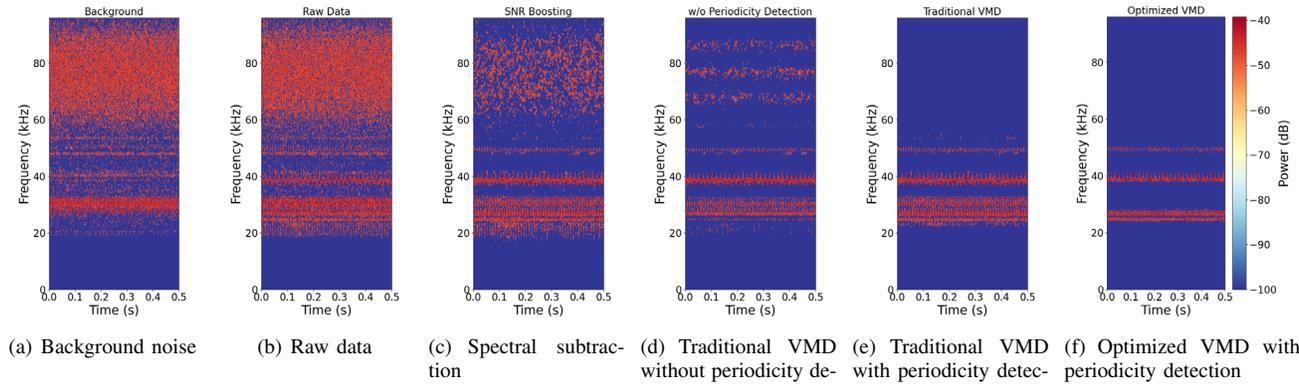


Fig. 13: Spectrograms of the acoustic emission from a power supply during the preprocessing pipeline. (a) Background noise containing noise from the sound card ($60\text{kHz}-90\text{kHz}$); (b) Raw data obtained from two lights and a Lenovo laptop computer during working; (c) Background noise weakened via spectral subtraction (d) Conventional VMD without periodicity detection, containing a random noise band at $60\text{kHz}-90\text{kHz}$ (e) Conventional VMD with periodicity detection, appearing broadband with noise at $25\text{kHz}-32\text{kHz}$. (f) Optimized VMD with periodicity detection, appearing the PFC frequency are correctly distinguished.

Note that in the frequency domain, $\|j\omega u_k(\omega_0 - \omega_k)\|_2^2$ is equal to $\|j(\omega_0 - \omega_k)u_k(\omega_0)\|_2^2$. To guarantee the fidelity of decomposition, the constrained variational model corresponding to the decomposition process of the response signal $\hat{S}_{pfc}(\omega)$ is as follows:

$$\begin{aligned} & \min_{u_k(\omega_0)} \left\{ \left\| \sum_k u_k(\omega_0) - \hat{S}_{pfc}(\omega_0) \right\|_2^2 + \right. \\ & \left. \alpha \sum_k \|j(\omega_0 - \omega_k)u_k(\omega_0)\|_2^2 \right\} \\ & s.t. \sum_k u_k(\omega_0) = \hat{S}_{pfc}(\omega_0) \end{aligned}$$

The optimal solution can be obtained using a Lagrange multiplier, after which the ADMM algorithm is used to find the saddle point. Please refer to [35] for more details.

Parameter optimization. The VMD algorithm decomposes the original signal based on pre-defined parameters (i.e. k and α). For parameter k , a high value can result in spurious modes (consisting of noise content) or mode splitting (the same component shared by several IMFs). A small value for parameter k can lead to mode loss or mode mixing. The choice of parameter α is equally important. In our experiments, the bandwidth varied from device to device, such that a set of universal parameters would not restore the original PFC waveform well. Thus, we developed a bandwidth-based optimization scheme to facilitate component extraction.

Take as an example a decomposed narrow band signal $u_k(\omega)$, in which the bandwidth is equivalent to the width of a rectangle of the power spectrum of $u_k(\omega)$, and whose height is the amplitude at the center of power spectrum. Thus, the bandwidth (bw) can be derived as follows

$$\begin{aligned} \Delta_f^2 &= \int_0^\infty (\omega - \omega_k)^2 \frac{|\hat{u}_k(\omega)|^2}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} d\omega \\ bw_k &= 2\Delta_f \end{aligned}$$

where $\hat{u}_k(\omega)$ denotes the amplitude spectrum of $u_k(\omega)$ and bw_k denotes the bandwidth of decomposed signal $u_k(\omega)$. The total bandwidth of the decomposed signal is indicated as

$BW = \sum_{k=1}^k bw_k$. Theoretically, the degree to which the decomposed mode fits the real signal is inversely proportional to the total BW . Inappropriate values for k or α would skew the value of BW . We therefore initialized the parameters using two sets: k within the range of $[10, 40]$ and α within the range of $[10000, 100000]$, and adjust the parameters. We chose the combination of k and α that can achieve the smallest BW as the optimal solution.

Periodicity detection. The signals were decomposed into several IMFs; however, we focused exclusively on the sound produced by the PFC feature. Note that the PFC component is periodic and noise is non-periodic. Here, we managed to extract the external PFC signal from the IMFs. For each IMF, we first computed the upper envelope, and then derived the corresponding auto-correlation coefficients. We segmented the IMF into multiple parts in which the start point is where the peak of the auto-correlation coefficients. The duration of each part should exceed the fundamental period, which we empirically set at 0.05s . We computed the Pearson Correlation Coefficient for pairs of segments and utilized a threshold to determine which IMF is periodic. Fig. 13(f) compares our best results obtained using OVMD (w/ periodicity detection) against conventional VMD (w/o periodicity detection). Overall, the characteristic frequency band was effectively extracted.

B. Data Augmentation

Our objective was to enable the identification of multiple appliance combinations when the user registers only once. To ensure the extraction of robust features and expand the generalizability, the data augmentation is implemented:

Augmentation against distance attenuation. Moving an appliance to another location in the house could alter the signal amplitude due to the effects of signal attenuation as a function of distance. Thus, we augmented the spectrum of each signal part by randomly amplifying or reducing the amplitude by a given multiple. The SNR can drop 6.5dB at a distance of 50m , means the range of amplitude range is within a range of $[0.47, 2.11]$. Data augmentation within this range can extract

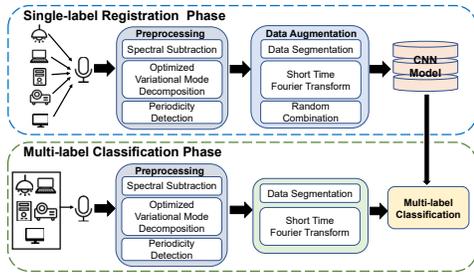


Fig. 14: The system overview of AUDIOSENSE.

features irrelevant with the amplitude, so as to combat the distance attenuation and appliance placement.

Augmentation against overlap. When multiple electrical appliances are operating at the same time, the frequency bands tend to overlap. Thus, we augment the data by taking addition with random combinations of different classes to enable the extraction of features from overlapping bands. The label of an n -categories data is an n -dimensional binary vector. Note that each appliance operates independently; therefore, data augmentation assumes that the probability of any appliance appearing is equal. The augmented data is then fed into the CNN model for feature extraction.

C. Multi-label Classification

After preprocessing and augmentation, we obtain a large multi-label dataset augmented with single-label data. Each data sample is processed by STFT into a 2D-spectrogram. All spectrograms $X = \{\vec{x}_1, \dots, \vec{x}_n\}$, $\vec{x}_i \in \mathcal{X}$ are associated with a ground truth label \vec{y}_i , and we seek the classification function $\vec{f}: \mathcal{X} \rightarrow \mathcal{Y}$ that minimizes loss function using N training sample-label pairs (\vec{x}_i, \vec{y}_i) , $i = 1, 2, \dots, N$. In the label layer, we encode labels as binary vectors $\vec{y} \in \{0, 1\}^L = \mathcal{Y}$ (with L labels). Since each appliance is independent, the classifier treats each label independently as an L binary classification. Moreover, the probability of any given appliance appearing in the signal is equal; therefore, we employed class-averaged binary cross entropy (BCE) loss as our loss function:

$$(\vec{y}, \vec{f}) = \frac{1}{L} \sum_{i=1}^L -y_i \log f_i - (1 - y_i) \log(1 - f_i)$$

In our feature extraction model, we assessed several convolutional neural networks (CNNs) to determine their ability to account for local similarities in the spectrogram during a given time period. We selected the ResNet-18 model [36] as the CNN model. Our ultimate objective was multi-label classification; therefore, the outputs of the network were the activations of L independent Sigmoid functions, where L indicates the size of the class vocabulary. We employed the Adam optimizer with a scheduled learning rate to train the model. The testing process involved passing the multi-label data through the preprocessing block employed in the registration phase before feeding it into the trained model. The output is a binary vector \vec{y} with L labels, indicating which appliance is working at that time.

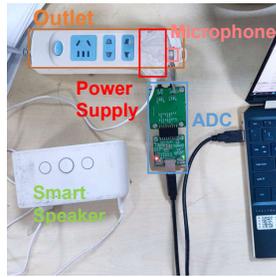


Fig. 15: The prototype of AUDIOSENSE system.

TABLE I: 28 Appliances with 34 states are employed in the experiments. The corresponding information is listed.

Appliance	Company	#States	#Same	#Different	#Total num	#Total states
Light	Xiaomi, PHILIPS, etc.	1	5	2	7	7
Monitor	Dell, PHILIPS, etc.	1	4	1	5	5
Desktop	Dell, HP, etc.	1	4	2	6	6
Laptop	Hp, Lenovo	1	0	2	2	2
Projector	Canon, Sony	1	0	2	2	2
Phone charger	Huawei, Xiaomi	3	0	2	2	6
Induction cooker	Supor	3	0	1	1	3
TV	Skyworth	1	0	1	1	1
UPS	SANTAK	1	0	1	1	1
Accumulator	Chilwee	1	0	1	1	1

VII. EVALUATION

A. Experiment Setup

Prototype: Due to intellectual property issues, we were unable to obtain microphone sampling information by COTS smart speaker devices. We therefore constructed a prototype to verify the efficacy of AUDIOSENSE. A schematic illustration is shown as Fig. 15. We plugged a smart speaker into an outlet to receive the PFC signals from remote outlets, while continuously sampling the acoustic signals received by a microphone. The acoustic signals were analyzed further.

Dataset: We collected data in real-world from 28 appliances with 34 states over a period of 10 days, sampled at 192kHz, 96kHz, 48kHz, and 44.1kHz, respectively. The dataset was divided into a single-label dataset (for training and testing) and a multi-label dataset (only for testing). Table. I lists the appliances used in the experiments. For the multi-label dataset, we representatively select 10 appliances (12 states) and randomly combined them.

Performance metrics. Evaluating performance involving multi-label learning is more complicated. To this end, we employed two common metrics to access multi-label learning [37], including average precision (AP) to measure the performance on each class, and mean average precision (mAP) to measure all classes. Specifically, AP can be computed as follows:

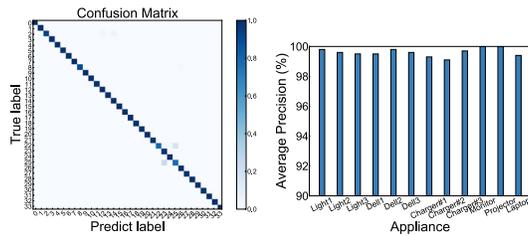
$$AP(h) = \frac{1}{L} \sum_{i=1}^L \frac{1}{|\mathcal{Y}|} \cdot \frac{|\mathcal{P}_i|}{\text{rank}^h(\vec{x}_i, y)}, \text{ where}$$

$$\mathcal{P}_i = \{y' | \text{rank}^h(\vec{x}_i, y') \leq \text{rank}^h(\vec{x}_i, y), y' \in \mathcal{Y}\}$$

Where, h returns the predicted labels of \vec{x}_i ; $\text{rank}^h(\vec{x}_i, y)$ returns the rank of y derived from the confidence for y to occur in a predicted label of \vec{x}_i . AP is used to determine the average fraction of predicted labels ranked above a particular label $y \in \mathcal{Y}$. mAP is the mean of average score of all labels. Essentially, the values of AP and mAP, are proportional to the performance of the classifier.

B. Overall performance

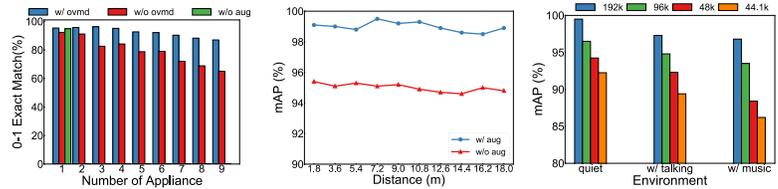
Single-label classification We evaluated the overall performance of the proposed system using 28 electrical appliances with 34 states. Fig. 16(a) presents the confusion matrix obtained using the single-label dataset, resulting in overall accuracy of 99.5%. The results demonstrate that AUDIOSENSE can work well on single-label classification.



(a) Single label

(b) Multi label

Fig. 16: Overall classification performance on single label dataset and multi-label dataset.



(a) Number of appliance

(b) Distance attenuation

(c) Environment interference

Fig. 17: System performance as a function of number of appliances, distance and environmental interference.

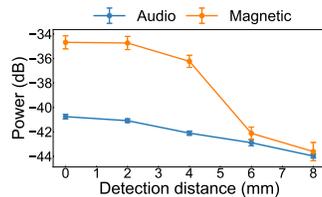


Fig. 18: Comparison of power attenuation between audio and magnetic signal.

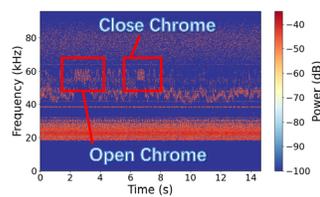


Fig. 19: The spectrogram of a laptop when a browser is opening and closing.

Multi-label classification. We evaluated the overall performance on multi-label classification. 10 appliances with 12 states are selected at random to form combinations. Among the appliances, 3 lights and 3 dell desktops are of the same model, and a phone charger has 3 states. As shown in Fig. 16(b), the AP of all classes exceeded 99% which demonstrated that AUDIOSENSE achieves good performance.

C. Robustness

Impact of the number of appliances. We first examined the impact of multiple electrical appliances superimposed on the multi-label classification. We employed a 0-1 exact match rate as the metric to eliminate interference of the number of categories. As shown in Fig. 17(a) the complete scheme (w/ ovmd) produced the best results (blue bar), resulting in a match rate of more than 87%. When OVMD was not applied, the exact match rate decreased sharply as the number increased for the reason that OVMD can effectively extract the PFC feature. The match rate dropped to just 65% when 9 appliances were included. When data augmentation was not applied, the exact match rate when using one appliance reached 94.3%; however, matching rate perform worse when multiple devices worked simultaneously because it cannot learn the multi-label classifier from only single-label data. Therefore, the data augmentation is significant to our system.

Impact of distance. We also examined the impact of transmission distance on mAP values. As shown in Fig. 17(b) shows that all of the mAP values obtained using data augmentation exceeded 98%, regardless of distance (from 1.8m to 18m). When data augmentation was not applied, mAP dropped to around 95%. These results indicate that AUDIOSENSE is robust to transmission distance.

Impact of environmental interference. Finally, we examined the impact of interference in different environments

including quiet, talking, and playing music. Data was collected under various sampling rates. As shown in Fig. 17(c), in a quiet environment, mAP decreased with the sampling rate; however, all mAP values exceeded 93%. In situations where someone was talking or playing music, the mAP values dropped slightly in terms of sampling rate: 192kHz (2.7%), 96kHz (3%), 48kHz (5.8%), and 44.1kHz (6%). Lower sampling rates had a more profound effect in noisy environments. Higher sampling rates resulted in higher performance; All the mAPs values exceeded 85% regardless of the sampling rates and surrounding environments. Note that the sampling rate is increasing to support Hi-Res recording and playback, indicating the performance on smart speaker should improve in the future.

VIII. DISCUSSION AND CONCLUSION

In this work, we proposed AUDIOSENSE, a novel approach to leverage current-acoustic channel to achieve single point appliances detection with a microphone. It is potentially to be updated on COTS smart speakers for more applications.

A. Acoustic vs. Magnetic Sensing

Moreover, acoustic solutions have certain advantages over traditional electromagnetic solutions:

Propagation distance. As signals propagate outward, the attenuation rate of acoustic signals is lower than that of magnetic field. Acoustic signals decays linearly [38] and the magnetic strength decays cubically [39]. Thus, acoustic signals can be detected at a greater distance. Fig. 18 presents the power attenuation of two signals against detection distance (0 – 8mm), the magnetic field strength decays much faster.

Magnetic shielding. In many situations, magnetic signals are regarded as a form of pollution. Electromagnetic interference (EMI) disrupts the operation of electronic equipment, often leading to malfunctions and faulty reading. Thus, most electronic equipment is equipped with magnetic shielding, such as a Faraday cage. High-frequency inaudible acoustic noise is largely disregarded, with the result that the transmission of audio signals is unimpeded.

No additional hardware. The collection of magnetic field data depends on specialty items, such as Hall sensors. By contrast, every smart device comes equipped with a microphone. Moreover, the sampling rate of built-in magnetometer in mobile devices is too low (e.g., 50Hz). The rate of microphone is far higher (e.g., $\geq 48kHz$). A shift to devices of Hi-Fi is pushing sampling rates even higher. The microphone on the

LG V60 has ability to record audio at $192kHz$ [40], which indicates AUDIOSENSE could be widely used in the future.

B. Extended Applications

AUDIOSENSE also has other potential applications. First, PFC signals are a form of EMI occurring at frequencies of $40\text{--}150kHz$. AUDIOSENSE can detect EMI in current, as long as the bandwidth is below $96kHz$. Second, we observed variations in acoustic signals when the programs are launched (e.g., Chrome), as shown in Fig. 19. Specifically, [22] used this phenomenon to identify which application was launched when using a desktop. It means AUDIOSENSE can also perform app-level monitoring in future work using audio signal.

ACKNOWLEDGEMENT

This work is supported in part by the NSFC (62072306, 61936015) and Program of Shanghai Academic Research Leader (20XD1402100).

REFERENCES

- [1] Michael Buettner, Richa Prasad, Matthai Philipose, and David Wetherall. Recognizing daily activities with rfid-based sensors. In *Proceedings of the 11th Ubicomp*, pages 51–60, 2009.
- [2] Hua Huang and Shan Lin. Met: a magneto-inductive sensing based electric toothbrushing monitoring system. In *Mobicom*, 2020.
- [3] Nilavra Pathak and et al. Acoustic based appliance state identifications for fine-grained energy analytics. In *2015 PerCom*. IEEE, 2015.
- [4] Shih-Hsiung Lee and Chu-Sing Yang. An intelligent power monitoring and analysis system for distributed smart plugs sensor networks. *International Journal of Distributed Sensor Networks*, 2017.
- [5] George William Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, pages 1870–1891, 1992.
- [6] Wang Dan, Huang Xiao Li, and Ye Shu Ce. Review of non-intrusive load appliance monitoring. In *2018 IEEE 3rd IAEAC*, pages 18–23. IEEE, 2018.
- [7] Gabe Cohn, Sidhant Gupta, and et al. Froehlich. Gassense: Appliance-level, single-point sensing of gas activity in the home. In *International Conference on Pervasive Computing*, pages 265–282. Springer, 2010.
- [8] Edison Thomaz and et al. Bettadapura, Vinay. Recognizing water-based activities in the home through infrastructure-mediated sensing. In *UbiComp*, pages 85–94, 2012.
- [9] Sidhant Gupta, Matthew S Reynolds, and Shwetak N Patel. Electriscense: 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*, pages 139–148, 2010.
- [10] Shwetak N Patel and Thomas et al. Robertson. At the flick of a switch: Detecting and classifying unique electrical events on the residential power line (nominated for the best paper award). In *International Conference on Ubiquitous Computing*, pages 271–288. Springer, 2007.
- [11] Ke-Yu Chen, Sidhant Gupta, and et al. Dose: Detecting user-driven operating states of electronic devices from a single sensing point. In *2015 IEEE PerCom*, pages 46–54. IEEE, 2015.
- [12] Manoj Gulati and Shobha Sundar et al. Ram. Single point conducted emi sensor with intelligent inference for detecting it appliances. *IEEE Transactions on Smart Grid*, pages 3716–3726, 2016.
- [13] Manoj Gulati, S Sundar Ram, and et al. Detecting it and lighting loads using common-mode conducted emi signals. In *Proc. NILM2016 3rd International Workshop on Non-Intrusive Load Monitoring*, 5p, 2016.
- [14] Matthai Philipose, Kenneth P Fishkin, and et al. Fox. Guide: Towards understanding daily life via auto-identification and statistical analysis. In *Ubihealth*, 2003.
- [15] Emmanuel Munguia Tapia and et al. Intille. Activity recognition in the home using simple and ubiquitous sensors. In *Percom*, pages 158–175, 2004.
- [16] Chenyang Zhang and Yingli Tian. Rgb-d camera-based daily living activity recognition. *Journal of computer vision and image processing*, page 12, 2012.
- [17] M Zeifinan, Craig Akers, and Kurt Roth. Non intrusive appliance load monitoring (nialm) for energy control in residential buildings. *Energy Efficiency in Domestic Appliances and Lighting 20 II*, 2011.
- [18] Jian Liang and Simon KK et al. Ng. Load signature study—part i: Basic concept, structure, and methodology. *IEEE transactions on power Delivery*, pages 551–560, 2009.
- [19] Shwetak N Patel. *Infrastructure mediated sensing*. Georgia Institute of Technology, 2008.
- [20] Wei Sun and Tuochao et al. Chen. Vibrosense: Recognizing home activities by deep learning subtle vibrations on an interior surface of a house from a single point using laser doppler vibrometry. *IMWUT*, pages 1–28, 2020.
- [21] Zhihui Shao, Mohammad A Islam, and Shaolei Ren. Your noise, my signal: exploiting switching noise for stealthy data exfiltration from desktop computers. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, pages 1–39, 2020.
- [22] Juchuan Zhang, Xiaoyu Ji, Yuehan Chi, Yi-chao Chen, Bin Wang, and Wenyuan Xu. Outletsy: cross-outlet application inference via power factor correction signal. In *Proceedings of the 14th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2021.
- [23] Chao Cai, Rong Zheng, and Menglan Hu. A survey on acoustic sensing. *arXiv preprint arXiv:1901.03450*, 2019.
- [24] Anran Wang and Shyamnath Gollakota. Millisonic: Pushing the limits of acoustic motion tracking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2019.
- [25] Yongzhao Zhang and Wei-Hsiang et al. Huang. Endophasia: Utilizing acoustic-based imaging for issuing contact-free silent speech commands. *IMWUT*, pages 1–26, 2020.
- [26] Fusang Zhang, Zhi Wang, and Beihong et al. Jin. Your smart speaker can” hear” your heartbeat! *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, pages 1–24, 2020.
- [27] Ahmed Zoha and Alexander et al. Gluhak. Acoustic and device feature fusion for load recognition. In *2012 6th IEEE International Conference Intelligent Systems*, pages 386–392. IEEE, 2012.
- [28] IECCE. Iec61000-3-2:electromagnetic compatibility (emc)- part 3-2, 2018.
- [29] Pui Yee Kong, JA Aziz, MR Sahid, and Low Wen Yao. A bridgeless pfc converter for on-board battery charger. In *2014 IEEE Conference on Energy Conversion (CENCON)*, pages 383–388, 2014.
- [30] Rohan Dayal, Suman Dwari, and Leila Parsa. A new design for vibration-based electromagnetic energy harvesting systems using coil inductance of microgenerator. *IEEE Transactions on Industry Applications*, pages 820–830, 2010.
- [31] Henno Allik and Thomas JR Hughes. Finite element method for piezoelectric vibration. *International journal for numerical methods in engineering*, pages 151–157, 1970.
- [32] SI MALAFEEV and SS MALAFEEV. Compensation of voltage loss in the power line. *DAAAM International Scientific Book*, 2018.
- [33] ON Semiconductor. Power factor correction (pfc) handbook. *Choosing the right power factor controller solution.*, 2011.
- [34] Texas Instruments. Flash battery charging pushes the boundary of charging current, 2017.
- [35] Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. *IEEE transactions on signal processing*, pages 531–544, 2013.
- [36] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on CVPR*, pages 770–778, 2016.
- [37] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, et al. *Modern information retrieval*, volume 463. ACM press New York, 1999.
- [38] Xiaofeng Sun and Xiaoyu Wang. *Fundamentals of Aeroacoustics with Applications to Aeropropulsion Systems: Elsevier and Shanghai Jiao Tong University Press Aerospace Series*. 2020.
- [39] Salvatore Celozzi and Marcello D’Amore. Magnetic field attenuation of nonlinear shields. *IEEE transactions on electromagnetic compatibility*, pages 318–326, 1996.
- [40] SoundGuys. Best smartphones for audio. <https://www.soundguys.com/best-smartphones-for-audio-16373/>, 2020.