SingSprite: Non-Tactile Appliance Interaction for the Blind through Power-Supply Acoustic Signatures

Lanqing Yang
Shanghai Jiao Tong University
Shanghai, China
yanglanqing@sjtu.edu.cn

Weiyi Wang
Shanghai Jiao Tong University
Shanghai, China
weiyiiwang@gmail.com

Yongzhao Zhang
UESTC
Chengdu, Sichuan, China
zhangyongzhao@uestc.edu.cn

Yixin Liu
Shanghai Jiao Tong University
Shanghai, China
louise_l@sjtu.edu.cn

Yi-Chao Chen*
Shanghai Jiao Tong University
Shanghai, China
yichao@sjtu.edu.cn

Guangtao Xue*
Shanghai Jiao Tong University
Shanghai, China
gt_xue@sjtu.edu.cn

Ahmad Ali
Shenzhen University
Shenzhen, Guangdong, China
ahmadali@szu.edu.cn

Abstract

Touchscreen interfaces have become ubiquitous in modern household appliances, yet they remain largely inaccessible to the blind due to their lack of tactile feedback. Existing accessibility solutions, such as tactile overlays, remote control applications, and proximity-based interfaces, suffer from failing to capture real-time appliance states, limited generalizability, indirect interactions, or strong environmental dependencies. In this work, we introduce *SingSprite* a novel appliance interaction system that allows blind users to seamlessly identify and control appliances by recognizing their unique acoustic signatures. The key insight is that many appliances emit distinct, device-specific humming sounds during operation, primarily generated by internal power supply components. *SingSprite* captures these acoustic fingerprints using commodity smartphone microphones, applying a tailored background noise cancellation scheme to enhance signal clarity. To address the issue of low mobile sampling rates, we integrate a Variational Mode Decomposition (VMD) approach with an autoencoder framework for effective feature extraction. The *HumNet* classifier, which concurrently classifies appliance types and operational states, demonstrates robust performance with an F1 score of 0.95 across a diverse set of 100 appliances. Additionally, user studies conducted with blind participants confirm the system's usability and its practical effectiveness in real-world scenarios. Additionally, user studies conducted with blind participants confirm the system's usability and its practical effectiveness in real-world scenarios. *SingSprite* offers a hardware-free and scalable accessibility solution, providing proximity-aware interaction in smart homes without the need for additional infrastructure, paving the way for more inclusive and intuitive environments for blind users. Using natural device acoustics, *SingSprite* pioneers a universal, proximity-sensitive interaction paradigm that significantly advances accessibility in smart home environments, without requiring any modifications to existing infrastructure.

CCS Concepts

• **Human-centered computing** → **Accessibility systems and tools**.

Keywords

Smart Interaction, Assistive Technologies, Smart Living, Acoustic Signatures from Power Supply

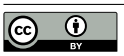
ACM Reference Format:
Lanqing Yang, Weiyi Wang, Yongzhao Zhang, Yixin Liu, Yi-Chao Chen, Guangtao Xue, and Ahmad Ali. 2026. SingSprite: Non-Tactile Appliance Interaction for the Blind through Power-Supply Acoustic Signatures. In *ACM/IEEE International Conference on Embedded Artificial Intelligence and Sensing Systems (SenSys '26), May 11–14, 2026, Saint Malo, France.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3774906.3800505

1 Introduction

The home appliance industry has experienced substantial growth, with both sales volume and per capita ownership steadily increasing. According to Statista, the global appliance market has expanded consistently, fueled by the rising demand for smart devices. Data from March 2025 shows that total retail revenue reached USD 670 billion in 2024 [52], with projections forecasting 5 billion household appliance shipments by 2029 [53]. By 2028, retail unit sales are expected to reach 850 million for major appliances and 4.03 billion for small appliances [17].

However, the growing prevalence of modern appliances has not been equally beneficial to the blind community. By 2050, the global blind population is projected to reach 61 million, while those with moderate to severe visual impairment are expected to number

*Guangtao Xue and Yi-Chao Chen are corresponding authors



This work is licensed under a Creative Commons Attribution 4.0 International License.
SenSys '26, Saint Malo, France
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2309-4/26/05
https://doi.org/10.1145/3774906.3800505

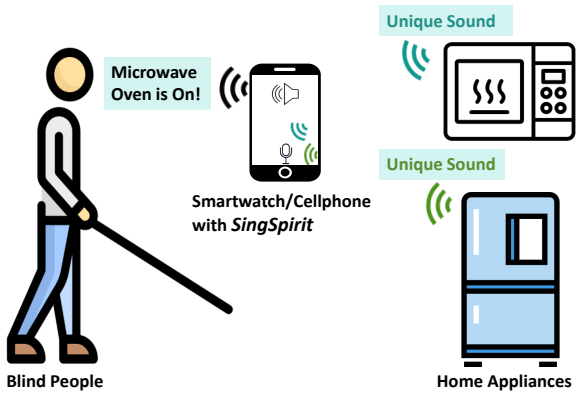


Figure 1: System overview of *SingSprite*. The wearable device (smartwatch/cellphone) detects unique sound signatures emitted by home appliances, converts them into verbal status notifications (e.g., "Microwave Oven is On!"), and provides real-time auditory feedback to the user.

474 million [15]. A Vision Atlas survey found that 35% of blind individuals need smart appliances, while 79% use smartphones—a remarkable 365% increase in usage over the past five years [36].

Unfortunately, the widespread use of touchscreens exacerbates the challenges faced by the blind. Many individuals struggle with operating appliances such as stoves, ovens, washing machines, dryers, and microwaves due to the absence of tactile feedback [7]. Over 50% of blind individuals report difficulties operating these appliances, with some expressing frustration that their community is accustomed to using older devices with physical buttons, knobs, and audio cues [1][30]. Similarly, 54% of respondents face issues with televisions, citing the lack of tactile markers and audio feedback[44]. To complicate matters, European Union regulations exclude household appliances from the European Accessibility Act, hindering the inclusion of accessibility features [43]. Additionally, the U.S. lacks laws requiring accessibility in electrical appliances [1]. From a health and safety perspective, the blind are unable to visually monitor appliance operations, which can lead to hazardous situations such as undetected sparks during short circuits, posing a significant fire risk [24]. In kitchens, blind users may unintentionally activate appliances by touching screens, unaware of which appliances they have turned on, further increasing the potential for safety hazards [1].

Several methods have been proposed to address these interaction challenges. One approach involves the use of tactile augmentation through adhesive buttons [53], but this solution lacks universality and is limited by manufacturer compatibility. Another option suggests modifying existing appliances so that touchscreen-controlled devices are converted into button-operated models [24]. The Beijing Association of Disabled Persons has even urged manufacturers to incorporate physical buttons into touch control solutions [2]. However, this method requires specialized technical support and may face issues such as discontinued warranties from manufacturers. Another approach uses remote control systems, such as Xiaomi Smart Home [59] and Apple Home [4], which enable users to control IoT-enabled appliances via smartphones. These systems offer two interaction modes: screen readers like Google TalkBack, which provide auditory feedback to navigate appliance controls, and voice assistants such as Siri [5] and XiaoAI [60], allowing users to issue commands verbally. However, a key limitation of these systems is that they are not tied to physical proximity, making them inconvenient for users who need to operate a specific nearby appliance. Furthermore, this approach also fails to provide real-time feedback on the status of appliances, which can be a significant safety concern. For instance, a blind user may struggle to ensure that a stove or oven is turned off after use, leading to a potential fire hazard. Other proximity-based solutions, such as NFC tag scanning and camera-based recognition, encounter similar limitations. These limitations highlight the need for a more reliable and proximity-based interaction method.

To solve this problem, in this paper, we propose a universal interaction method based on power supply acoustic fingerprint recognition. When blind users approach an appliance, a mobile device automatically detects its unique acoustic signature and activates the corresponding control interface, facilitating natural interactions (e.g., voice commands) for operation. This approach is characterized by three key features: (1) broad applicability, as nearly all appliances with power supplies emit distinct acoustic signals; (2) intuitive interaction, where proximity to a device triggers automatic identification and voice announcements, a method accessible to 90% of blind individuals who use voice services; and (3) implementation using existing smartphones or smartwatches, requiring no additional accessories, thereby lowering the barrier to adoption.

In this paper, we introduce *SingSprite* (Fig.1, a novel system that leverages these acoustic signatures to directly address the usability and safety challenges faced by blind users in modern smart homes, offering a seamless, non-tactile interaction paradigm. A mobile microphone array captures ambient electrical acoustic signals, employing a background noise cancellation technique to produce high-quality training samples. A deep auto-encoder module addresses signal distortion caused by limited mobile sampling rates, ensuring robust feature extraction. An end-to-end classification model identifies both the device type and its operational status by learning discriminative acoustic patterns, enabling precise and context-aware control. The main contributions of this paper are highlighted as follows:

- We pioneer the use of device-specific humming sounds to enable interaction with appliances for blind users.
- We address low sampling rate constraints using variational mode decomposition and autoencoders for robust signal feature extraction.
- We develop *HumNet*, an end-to-end classification model, for the synchronized recognition of appliance identity and operational state.
- We validate *SingSprite* through comprehensive experiments, demonstrating a significant end-to-end F1 score improvement of 0.20 (achieving 0.95) over baseline methods across a large-scale testbed of 100 appliances. A user study involving 16 blind participants indicated that *SingSprite* is more user-friendly than existing solutions like RFID tags.

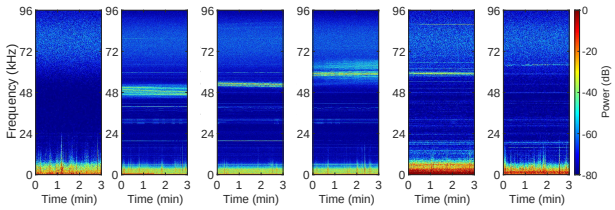


Figure 2: Spectrogram of humming sounds from a lamp and projector under various states: (From left to right) environment noise, lamp states (three levels of brightness), and projector states (stand-by and working).

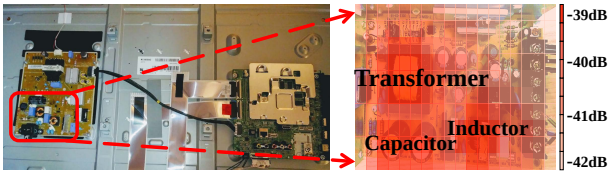


Figure 3: Signal strength over the SMPS of a disassembled SHARP LCD TV.

2 Background

In this section, we explore the cause of the high-frequency humming sounds and their feasibility for identifying the appliances from which they originate.

2.1 Cause of High-frequency Acoustic

We placed the microphone, connected to an external sound card (sampling rate of 192kHz), close to a lamp and a projector under various operating states. As Fig. 2 shows, sounds emitted by the devices ranged from 10kHz to 87kHz when the appliances worked. We observed notable differences between the signals from the two appliances, variations in the signals from individual appliances under different operating states. Moving the microphone around revealed that the strongest signal was adjacent to the power supply.

We disassembled a SHARP TV to scan its power supply and measured the sound level of each component. As shown in Fig. 3, there are multiple sounds sources (i.e., a capacitor, a transformer, and an inductor in this example). The humming sounds can be attributed to the working principle of switching-mode power supplies (SMPSs) found in most modern appliances [18].

SMPSs use a switching transistor to convert the input power into pulsed voltage, which is then smoothed by capacitors, inductors, and other components. Fig. 4 shows a typical buck-boost regulator. By switching the transistor (TR) rapidly between "ON" and "OFF," the regulator maintains the desired output voltage. As shown in Fig. 5 and Fig. 6, this high-frequency switching generates alternating current and a strong alternating magnetic field, causing the magnetic core and transformer coil to vibrate and produce high-frequency sounds. Different power loads further lead to different switching and sound frequencies through negative feedback. Similar effects also occur in capacitors and inductors. According to basic acoustic theory, the loudness of the humming sound depends on vibration amplitude and is therefore closely related to magnetic force.

To verify this principle, we employed a DRV425 magnetic sensor and a 192kHz sound card to scan different components of the disassembled TV (Fig. 3) simultaneously, including areas both inside and outside its SMPS. Fig. 7 shows that the strength of the acoustic signal is proportional to the strength of the EMI. This confirms that the high-frequency sounds are caused by the alternating magnetic field. This also demonstrates that other inner transformer modules of an appliance operate with relatively low input voltage, so the current and resulting magnetic field are also very small, which explains why this inner transformer module does not generate noticeable humming sounds.

Our measurements suggest an explanation for the difference in humming sounds produced by different electrical appliances or different states. For example, variations in the structure of connections between components (e.g., capacitors, transformers, and inductors) can lead to variations in the magnitude of the vibration and deformation, leading to variations in the magnitude of the corresponding sound. Besides, appliances operating under different working states also tend to vary in terms of acoustic signature, due to differences between the corresponding power loads (Fig. 2).

2.2 Preliminary Study

Preliminary experiments were designed to solve three questions: i) Do power supplies generate unique acoustic patterns? ii) Are the acoustics patterns temporally consistent? iii) Are the microphones in mobile devices capable of capturing sound features representative of individual appliances? iv) Are appliances' sounds audible (detectable) to humans?

2.2.1 Uniqueness: We compared the emissions from 10 different appliances (TV, refrigerator, washing machine, water heater, microwave, air cleaner, rice cooker, printer, monitor, and router) and 10 of the same model (lamps). Audio traces were obtained from each of the devices during one-minute sessions conducted over 30 days. We performed 2D correlation analysis [40] on the spectrograms from the audio traces to estimate the degree of . Fig. 8 presents the CDF of the correlations. *Dif. Models* and *Same Model* refer to different and same appliance model, respectively. The median correlations were: different appliances (0.39) and same model (0.68). These results suggest that humming sounds of both different and same appliances are distinguishable.

Another issue is that, in our daily life, it's common to have other appliances working concurrently, which may break the unique acoustic patterns of the target appliances. To study how the appliances interfere with each other, we connected three appliances (a lamp, a monitor, and a router) to the same power strip and placed them 25cm from each other. Fig. 10 shows the spectrogram of sounds when we turned on and off these appliances with various combinations. The 5 spectrograms from left to right represent *lamp*, *monitor*, *router*, *lamp+monitor*, *lamp+router*, and *lamp+router+monitor*. We can see when multiple appliances worked concurrently, their sound patterns of appliances linearly superposed, which suggests that they are distinguishable even when mixed together.

2.2.2 Stability: We collected sounds from 20 appliances every day for a period of 30 days. We then used 2D correlation analysis to determine the degree of similarity using the values collected on the

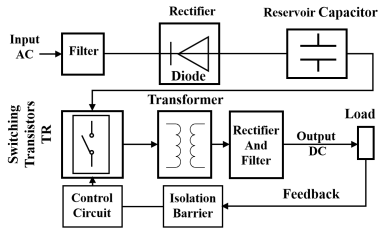


Figure 4: Circuit of typical buck-boost SMPS.

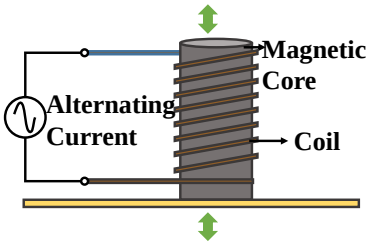


Figure 5: Magnetic cores deform in changing current field.

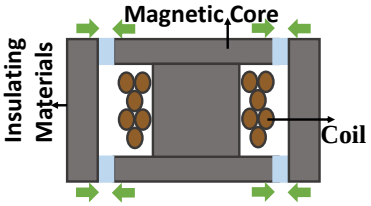


Figure 6: Different magnetic cores attract each other.

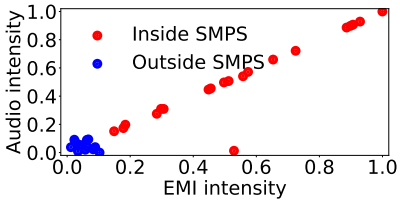


Figure 7: Relationship between EMI and sound strength.

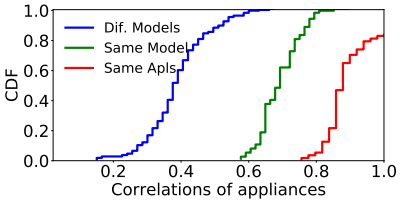


Figure 8: Similarity between any pair of the appliances.

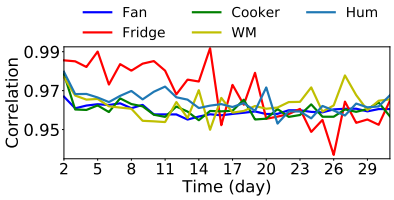


Figure 9: CDF of 2D correlations among 20 appliances over 30 days.

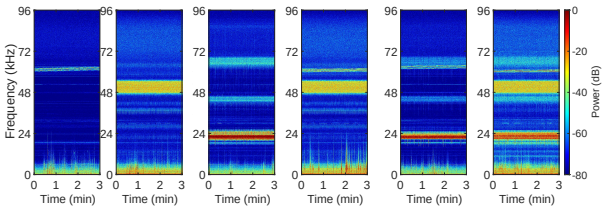


Figure 10: Sounds generated by the power supplies of 3 appliances and the combination of them.

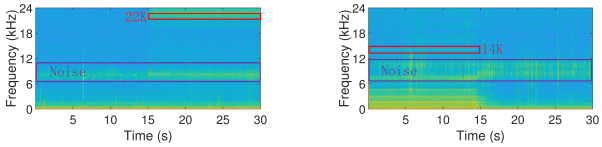


Figure 11: Spectrogram of humming sounds collected by XiaoMi K20 at a sampling rate of 48kHz.

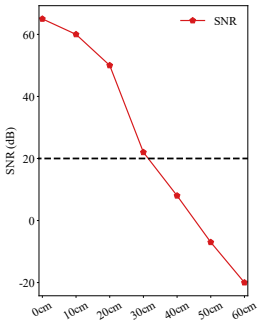


Figure 12: SNR in different distances. The black line represents human audible ranges.

first day as a benchmark. Note that for the sake of readability, we included the results from only five appliances in Fig. 9, including a *fan*, an *electric cooker*, a *humidifier* (Hum), a *refrigerator*, and a *washing machine* (WM) and recorded the 2D correlation over a period of 30 days. As shown in Fig. 9, the degree of correlation throughout the month exceeded 0.9, which suggests that the patterns remained stable over time.

2.2.3 Mobile devices: Nyquist theorem states that to prevent aliasing, a signal must be sampled at more than twice the highest frequency component of the signal. However, the sampling rate of most mobile devices (e.g., mobile phones) was 48kHz (a few of the them support 96kHz). We therefore had to determine whether features in the original high-frequency sound remained distinguishable under the effects of aliasing in a lower frequency band. As shown in Fig. 11, we used a Xiaomi K20 mobile phone to collect humming sounds at a sampling rate of 48kHz when holding the phone to get close to a TV (left), and moving it away from a laptop (right). Sound features appeared at close to 22kHz (left) and 14kHz (right). From this, we inferred that the original frequency was in the region of 70kHz and 62kHz respectively.

Despite that it is possible to implement on mobile devices, however, the aliasing [49] [48] and anti-aliasing filtering (AAF) [47] [37] increases the difficulty for reconstructing these high-frequency humming sounds using low sample-rate mobile devices. Due to the aliasing problem, some humming sound are aliased to a frequency band (0-10kHz) that can be easily disturbed by human and other noises; the AAF significantly reduces the intensity of humming sounds, which further increases the difficulty to distinguish these sounds from noises; moreover, due to the differences in sound sampling rate and frequency response of AAF on different mobile devices, the frequency range, signal-to-noise ratio, etc. of the signals of the same appliance collected using different mobile devices differ greatly. All these remain problems to interacting with appliances using COTS mobile devices.

3 System Design

Based on our preliminary findings that SMPS signatures are unique and stable, we developed *SingSprite*, a complete, end-to-end system designed for practical deployment on commodity COTS mobile and wearable devices for blind users. The system architecture, which we call HumNet, encompasses a full pipeline to address real-world challenges, as illustrated in Fig. 13. This includes: (1) a pre-processing scheme to remove environmental noise, (2) a deep auto-encoder to extract robust features from low-quality microphone data, and (3) a multi-stage classifier to identify both appliance type and operational state. This section details the design of this integrated system, which was implemented and tested on both smartphones and smartwatches.

The high-frequency humming sounds from appliance power supplies (SMPS) offer unique acoustic fingerprints, but capturing them reliably on COTS mobile devices is challenging due to environmental noise and hardware diversity. Our key insight is that a high-sampling-rate (192 kHz) sound card reveals clean, stable signatures that can serve as a ground truth. Based on this, we developed *SingSprite*, a system built on a novel, decoupled data collection strategy. Instead of requiring synchronized recordings, we use the high-fidelity sound card to build a library of clean reference signals. Our feature extraction scheme then learns a robust mapping from any noisy, low-fidelity mobile phone signal to its corresponding clean signature, using only a shared appliance-state label. This approach simplifies data collection and inherently improves robustness.

The system architecture, HumNet, is a complete end-to-end pipeline designed for practical deployment on smartphones and smartwatches. First, a pre-processing scheme removes environmental noise like fans and human voices. Second, a deep auto-encoder extracts time-invariant features from the low-quality microphone data, making the system resilient to real-world variations in distance and angle. Finally, a multi-stage classifier identifies both the appliance type and its operational state from these robust features.

To deploy on a user's device, a one-time registration phase and a tailored model transfer process adapt the system to their specific hardware. The user interaction is designed to be intuitive for blind users: a long-press on the device's power button initiates detection. *SingSprite* then provides clear voice prompts identifying the nearby appliance and its state, seamlessly integrating tactile and auditory feedback.

Based on our preliminary findings that SMPS signatures are unique and stable, we developed *SingSprite*, a complete, end-to-end system designed for practical deployment on commodity COTS mobile and wearable devices for blind users. The system architecture, which we call HumNet, encompasses a full pipeline to address real-world challenges. This includes: (1) a pre-processing scheme to remove environmental noise, (2) a deep auto-encoder to extract robust features from low-quality microphone data, and (3) a multi-stage classifier to identify both appliance type and operational state. This section details the design of this integrated system, which was implemented and tested on both smartphones and smartwatches.

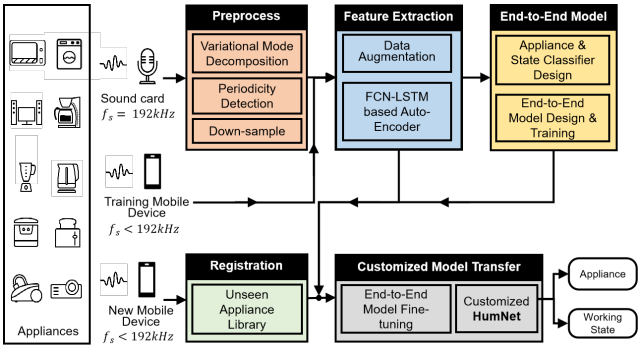


Figure 13: *SingSprite* framework.

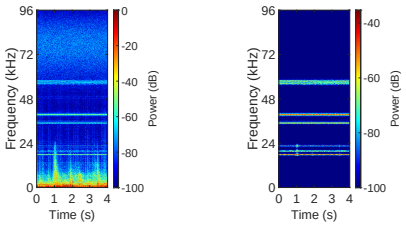


Figure 14: Acoustic signal from lamp before (left) and after denoising (right).

4 Methodology

Having established the fundamental feasibility of using SMPS signatures, the core challenge shifts to designing a system that is robust, reliable, and practical for blind users in their everyday, uncontrolled environments. Our methodology is therefore designed to systematically address these real-world challenges, starting with the critical step of background noise removal.

4.1 Background Noise Removal

The noise in the audio signals came from two main sources. As shown in Fig. 14(a), the first source was the hardware circuitry of the sound card, ranging from 60kHz to 90kHz. The second source was environmental noise below 10kHz. Noise from hardware circuitry is unvarying and can therefore be removed easily using spectral subtraction [33] [25]. Environmental noise is time-varying and unpredictable, which may overlap heavily with humming sounds in the range from 10kHz to 20kHz in some cases. As shown in Fig. 14(a), the maximum power of environmental noise was roughly 50dB greater than that of the humming sounds. Thus, removing audible noise in the lower frequency band was crucial to the performance of our classifier.

4.1.1 Variational Mode Decomposition. Choosing a universal cut-off frequency for a high-pass filter is complicated by the unpredictability of environmental noise. An interesting observation from the spectrograms above was that the central frequencies of the narrow band components remained unchanged for short periods, e.g., 0.5 seconds. Thus, we employed Variational Mode Decomposition (VMD) [57] to decompose the received signals into several narrow-band components, referred to as Intrinsic Mode Functions (IMFs), before extracting the humming sounds from the decomposed IMFs.

The signal is divided into k narrow-band signals (denoted by u_k), each of which has a different central frequency ω_{ck}. We then

Table 1: Boosted SNR (dB) vs. de-noising algorithms. WF for Wiener filter [12] [58], SS for spectral subtraction [21], MSS for multi-band spectral subtraction [31], and LMS for Least mean squares (LMS) adaptive filter [42].

Algorithms	WF	SS	LMS	MSS	Ours
Boosted SNR (dB)	10.4	15.6	20.3	24.6	35.8

compute the penalty term after converting the narrow-band signals from passband to baseband. Thus, the optimization objective in the frequency domain is denoted as:

$$\min_{u_k(\omega)} \{\| \sum_k u_k(\omega) - f_0(\omega)\|_2^2 + \alpha \sum_k \| j(\omega - \omega_{ck}) u_k(\omega)\|_2^2\}$$
$$\text{s.t.} \sum_k u_k(\omega) = f_0(\omega)$$

Note that to improve the fidelity of the reconstructed signal, we apply a restriction to the reconstructed signal. The optimal solution can be achieved using the Lagrange multiplier, followed by the ADMM algorithm to find the saddle point. Please refer to [57] for more details.

From the above optimization model, we derive two conclusions: i) k and α should be tuned together to achieve the best decomposition performance; ii) the size of α should be inversely proportional to the bandwidth of each IMF. In our case, we set $k = 40$ and ω_{ck} to be uniformly initialized, with the result that the difference in central frequency between each two initial modes was 2.4, thereby ensuring that weak periodic signals could be extracted. Moreover, we set α to a large value (e.g., 100,000) to impose a strict restriction on the bandwidth of each mode.

Note that the duration of the segments should exceed the fundamental period of each IMF, which we set (empirically) at $0.025s$. We computed the Pearson Correlation Coefficient (PCC) [9] [67] for pairs of segments, the average of which was used to indicate the periodicity of each IMF. Finally, we selected periodic IMFs by setting a threshold.

The acoustic spectrograms before and after denoising are demonstrated in Fig. 14(b), which suggests that most of the ambient and time-varying noise was eliminated. We then compared the proposed de-noising scheme with 4 other noise reduction algorithms such as the multi-band spectral reduction algorithm, and the results are shown in Table 1, and it shows that the proposed scheme boosted the SNR by 35.8dB, outperforming others by a boosted SNR of about 10dB.

4.2 Feature Extraction Scheme

4.2.1 Data Augmentation. The aliased humming sounds collected on mobile devices can be affected by various low-frequency noise sources, such as people talking, electrical appliances in operation (e.g., a TV in use), the mobile device itself (e.g., playing music or video), and other nearby appliances (e.g., a running fan).

In order to extract time-independent features and eliminate noise, we propose a data augmentation scheme [8] [41]. The idea is to add these noises into the training data, which will therefore enhance the performance of the learned feature extractor.

Let $X \in R_+^{d \times t}$ be the collected spectrogram where t is the duration time of the spectrogram, d indicates the frequency bins in each frame, λ_l as the weight factor, and X, Xh, Xa, Xp denotes appliance signals, human noises, appliance playing signals, other appliance's signals correspondingly, the process can be expressed as follows:

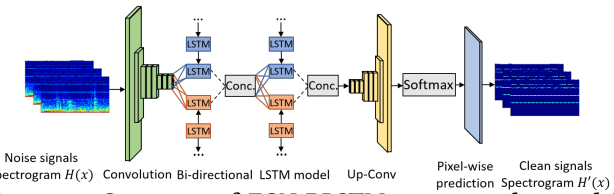


Figure 15: Structure of FCN-BLSTM auto-encoder model. *Conc.* denotes a concentration operation.

$$X_{overlap} = \sum_{l=1}^{K} \lambda_{l_0} X_l + \lambda_{l1} Xh_l + \lambda_{l_2} Xa_l + \lambda_{l_3} Xp_l,$$
$$\text{where} \qquad 1 < K < N, \sum_{i=0}^{3} \lambda_{li} = 1$$

For appliance l , we keep $\lambda_{l_0} > \lambda_{\neq l_0}$ to ensure that *SingSprite* can recognize which appliance the user is actually interacting with, even when dealing with noisy signals.

4.2.2 Feature Extraction Model Design. We developed an auto-encoder model illustrated in Fig. 15, namely FCN-BLSTM, as the feature extraction scheme to learn the transform from *Noisy Signals* to the *Clean Signals*.

As for the acoustic signature of interest, let $x \in R_+^{d \times t}$ be the raw spectrogram of *Noisy Signals*, and $y \in R_+^{d \times t}$ is its corresponding target spectrogram of *Clean Signals*, where t is the time duration of the spectrogram, and d refers to the frequency bins in each frame. When there are N pair of raw and target spectrograms, the problem can be formulated as:

$$\min_{\theta} \frac{1}{2} \sum_{i=1}^{N} ||f_\theta(x_i - y_i)||_F^2 \tag{1}$$

where f_θ refers to the mapping function from raw to target, and θ is the parameters to train.

We employed FCN [29] [32] as the feature extraction block, to account for local similarity in the spectrograms in adjacent frequency bins. After each convolution operation, a stride of half the kernel size is used to enlarge the receptive field and reduce computational overhead. All feature maps learned using the FCN block are vertically concatenated along the feature dimension to form a 2D $M \times t$ feature map ($H(x)$), where M indicates the total number of learned features. To capture the trend among contiguous spectrograms, we introduced in a bidirectional long-short term memory network (BLSTM) [68]. The BLSTM block treats $H(x)$ as an input, then splits $H(x)$ into several time steps, and then learns the long-term dependence among them. Using a many-to-many training strategy, the BLSTM block outputs $H'(x)$. We selected the mean squared error between $H'(x)$ and target y as the regression loss. The Adam optimizer with scheduled learning rate [3] [69] was used to train the model.

4.3 End-to-End Classification Scheme

Having extracted robust features, the final task is to classify the appliance and its state in near-real-time. The concern about using our MS-LSTM model for short-time recognition is addressed by our

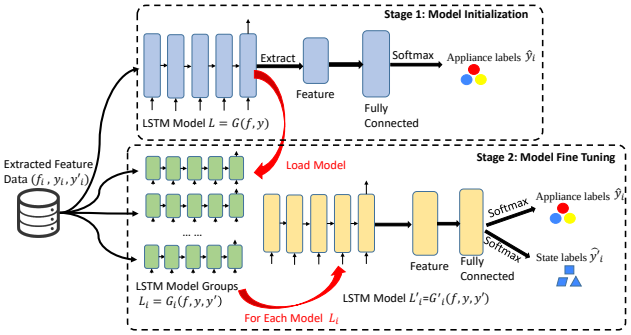


Figure 16: Structure of MS-LSTM classification model, where f denotes the extracted feature, y denotes the appliance label, y' denotes the working state label, and $\hat{y}$ for predicted labels.

data processing pipeline. Our system analyzes a continuous stream of data using a 0.5-second sliding window, not an instantaneous snapshot. The raw audio within this window is converted into a spectrogram using an STFT with a hop length of 512 samples. At a typical 48 kHz sampling rate, this 0.5-second snippet yields a rich sequence of approximately 47 feature vectors ("frames"). This sequence is then processed by our FCN-BLSTM feature extractor and fed into the MS-LSTM classifier.

This approach is physically grounded. SMPS acoustic signatures are continuous, quasi-periodic signals with stable fundamental frequencies and harmonic structures. By analyzing 47 consecutive frames, MS-LSTM can capture these stable patterns and distinguish them from transient noise, which is difficult with a single frame. This use of recurrent networks to model short-term temporal dependencies is well established in audio signal processing [39, 50, 68].

By implementing the feature extraction scheme, FCN-BLSTM model can generate feature maps of the spectrograms from *Noisy Signals* and *Clean Signals*. The feature maps are then fed into the classification scheme to identify appliances and their working states, as shown in Fig. 16. The proposed end-to-end scheme is divided into 3 steps: 1) Data augmentation to facilitate the decomposition of overlapping signals; 2) Implementation of the multi-stage LSTM (MS-LSTM) model to identify appliances and initialize the network; 3) and a fine-tuned MS-LSTM with the same structure to identify working states.

4.3.1 Classification Model Design. To begin with, the model becomes excessively complex $O(M*N)$, where N denotes *#appliances*, M denotes *#working states*. There is also a risk of mis-classification due to the similarity in acoustics's frequency changes generated in both switching appliance and working states. For example, this type of model tends to output a variety of appliances and states within one minute, although users do not typically switch between electrical appliances so frequently.

Note that we observed a correlation between the changes in the frequency of sounds produced by the power supply as a result of switching among appliances or changing their working states. The acoustic frequency of a single appliance can change considerably due to a change in working state. By contrast, different appliances differ in terms of long-term patterns and range. Note also that the switching frequency of appliances when used in a practical setting (e.g., no changes for 5 seconds) far exceeds the window siz used for analysis (e.g., 0.5 second).

Based on the above observations, we developed an LSTM model [20] to capture frequency changes over a longer period of time. This context-aware model can reduce appliance classification error caused by sound frequency fluctuations in a shorter time window and avoid gradient explosion caused by long-term dependence. We then built a multi-stage LSTM (MS-LSTM) model [62] for electrical appliances and working status classification. The MS-LSTM model structure is shown in Fig. 16. This model takes augmented feature data (f_i, y_i, y'_i) by FCN-LSTM as input. In the model initialization stage, an LSTM model $L = G(f, y)$ is trained to recognize appliances. In the fine-turning stage, a group of classifiers sharing the same structure with L, namely $L_i = G_i(f, y, y')$ is trained to recognize appliances and their working states. Finally learned features are fed into soft-max classifiers after a fully-connected layer to output final classification results for appliances and their working states.

The MS-LSTM model can be formulated as follows. In initialization stage, the model target can be defined as:

$$argmin_{\theta} \frac{1}{n_t} \sum_{f_i \in D_t, y_i \in R_t} \sum_{k=1}^{|C_y|} G_\theta(f_i, y_i), y_i)^k log G_\theta(f_i, y_i), y_i)^k$$

where G denotes the LSTM model, D_t denotes the feature space and R_t denotes label space, $|C_y|$ is the total categories of y, θ is the model parameters of G, and $\hat{y}$ denotes the predicted appliance label.

Similarly, in the fine-tuning stage, the model target is:

$$\arg\min_{\theta} \frac{1}{C_{y'_i}} \frac{1}{n_t} \sum_{j=1}^{C_{y'_i}} \sum_{f_i \in D_t, y_i \in R_t, y'_i \in R'_t} \sum_{k=1}^{|C_y|} G_{j,\theta}(f_i, y_i)^k log G_{j,\theta}(f_i, y_i)^k + \lambda \sum_{k=1}^{|C_{y'}|} G_{j,\theta}(f_i, y'_i)^k log G_{j,\theta}(f_i, y'_i)^k$$

where θ denotes the learned parameters, λ denotes the hyper-parameter that controls the performance balance between appliance and state classification.

MS-LSTM utilizes the similarity of frequency patterns caused by switching appliances and their operational states (in the initialization stage) to train a common model structure, and also explores differences to avoid classification errors (in the fine-tuning stage) to fine-tune the parameters using different classifiers for each classification task.

4.3.2 End-to-End Model Design. In order to transfer the learned model above to a COTS mobile phone, we integrate the feature extraction and classification models into an end-to-end scheme, the so called *HumNet*.

As shown in Fig. 17, the proposed *HumNet* is used for end-to-end classification of appliances and their working states. The denoised signals from the feature extraction scheme (the FCN-LSTM model) are then fed into the classification module to obtain classification results. Note that only the encoder model (for data extraction) part is used for mobile devices usage to reduce complexity. Below, we introduce the means by which the end-to-end scheme was trained and loss was back-propagated.

4.3.3 Customized Model Transfer. Using the end-to-end *HumNet* model, we can transfer learned models to non-trained mobile devices. The transfer procedure is divided into two steps: 1) Users

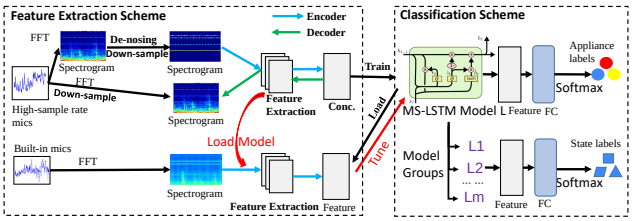


Figure 17: End-to-end structure of proposed *HumNet* divided into feature extraction scheme and classification. *Conc.* denotes a concentration operation; and model L_i denotes a group of different LSTM models sharing same model structure.

first collect data on their own appliances using a new untrained mobile device, in a process referred to as registration. 2) The trained *HumNet* model is then loaded and fine-tuned using newly collected data, in a process referred to as fine-tuning.

1) Registration process. When using a new mobile devices, the users first collects sample data on these devices to register their electrical appliances. Collecting sufficient data to infer the operating state of each appliance requires monitoring for 30 seconds. The high-sampling rate dataset pertaining to large-scale electrical appliances is denoted as (x, y, y'), whereas the dataset obtained using the mobile device is denoted as $(x_1, y_1, y_{1'})$. Note that $C(y_1)$ and $C(y_{1'})$ may or may not be a subset of $C(y)$ and $C(y')$.

2) Model fine-tuning. The acoustic data of registered appliances is fed into the end-to-end model for fine-tuning. a) Pre-processing. The trained FCN-BLSTM model is loaded for preprocessing and feature extraction, and then outputs f_1 for given input x_1. Note that under the learned auto-encoder structure, only the encoder structure is held on the mobile device for use in on-board denoising. b) Activation. After the new appliance is registered, the corresponding parts ($C(y_1)$ and $C(y_{1'})$) in the pre-trained classification model are activated, while the other parts are kept silent. Note that if the user records a new electrical appliance in a new location, the remaining model can be used to complete activation.

3) Fine-tuning. For given data $(f_1, y_1, y_{1'})$, pre-trained MS-LSTM models are used to construct a new loop and fine-tune model G_i. Training errors are also back-propagated to enable fine-tuning of the FCN-BLSTM feature extractor F. Finally, optimal learned G and F are used on the device for final classification.

5 Evaluation

5.1 Experiment Setup

Having detailed the design of *SingSprite*, we now present a comprehensive evaluation to validate its effectiveness as a practical solution to the accessibility challenges faced by blind users. Our evaluation strategy is multi-faceted, designed to answer three key questions: (1) How accurate is the core technology? (2) How robust is the system in the face of real-world interference? and (3) How usable and effective is the system for its intended users? To this end, we first detail our experimental setup, then present a series of micro-benchmarks and interference tests, and finally, report the results of a user study with 16 participants.

In the experiments, a proprietary sound card with a sample rate of 192 kHz served as the high-sample-rate device. To rigorously evaluate the system's robustness and generalizability across diverse hardware, our experiments utilized 8 different mobile phones. It is important to note that this was strictly for our cross-device validation protocol; a regular user only needs their own single device to operate SingSprite. Training devices included 8 mobile phones: iPhone 8, iPad Pro 2018, OPPO R11, Xiaomi 10, Huawei P40, XiaoMi K20, Samsung S7, and iPhone X.

5.2 Dataset

Two datasets were used in the experiments.

I. Electronic Store Dataset: This experiment was performed using a laptop equipped with a high sample rate sound card as well as the 8 mobile phones, which collected data concurrently from roughly 100 types of electrical appliances in an electronic store. We affixed stickers to the 100 appliances to identify them each time. The collection was carried out 3 times, each involving collecting 2-minute acoustic traces sampled at 192 kHz. Data was collected when the store was open, so the data may contain noise from customers. Store employees were required not to move the electrical appliances or stickers.

II. Laboratory Dataset: Within a laboratory environment, we simultaneously collected data from 10 of the 100 electrical appliances using the mentioned sound card as well as the 8 mobile phones. Similarly, we collected traces in 2-minute sessions at 3 different times over 3 weeks, with appliances also affixed with stickers for identification. The two datasets are available on GitHub.[1]

To ensure consistency and reproducibility, all data was collected following a standardized protocol. For each 2-minute acoustic trace collection, a single mobile device was placed at a fixed distance of approximately 15 cm from the primary SMPS location on the target appliance (typically near the power cord entry). This distance was chosen to ensure a high-quality signal while representing a realistic close-proximity interaction distance. The 8 mobile phones mentioned in our setup were used sequentially as part of our evaluation, not simultaneously.

5.3 Performance Metrics

The performance of the classifier was indicated by the average of 8 F1 scores, calculated as:

$$F1_i = \frac{2 \times p_i \times r_i}{p_i + r_i}$$

where p_i and r_i are the precision and recall of type i of appliances. The macro F1 score can then be computed as:

$$F1 = \frac{1}{N} \sum_{i=1}^{N} F1_i$$

where N denotes the number of different appliances or working states. We employed the same method to evaluate working state classification performance, using the macro-averaged F1 score as a metric to compare the performance of the various models.

5.4 Micro Benchmark

5.4.1 Warm-up Time. We first clarify our definition of a "working state." In this work, a state is defined by two criteria: (1) it

[1] https://github.com/SingSpirit/SingSpirit

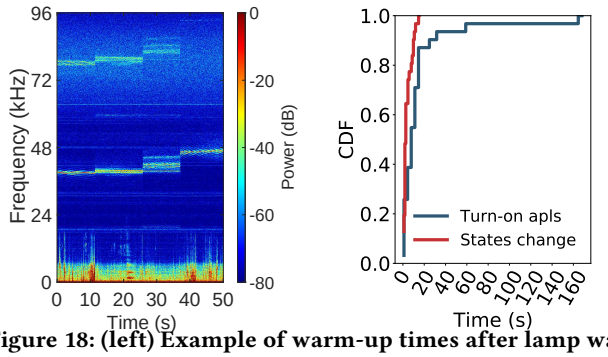


Figure 18: (left) Example of warm-up times after lamp was switched among four states; (right) CDF of warm-up times for 40 appliances.

must be a distinct, user-relevant operational mode of an appliance (e.g., different power levels of a light correspond to distinct working states), and (2) it must produce a unique and statistically stable acoustic signature that is distinguishable from other states of the same appliance. Consequently, working states are inherently appliance-specific. For example, the 'low-power' state of a fan and the 'standby' state of a television are treated as entirely separate classes by our model.

Warm-up time indicates the amount of time required for the acoustic signature to stabilize after an appliances are switched on/off or after a state change. The length of the warm-up time is proportional to the latency of *SingSprite* in correctly identifying the appliance and its working state. We evaluated this problem by considering the spectrogram of a lamp warming up. The lamp on the left side of Fig. 18 has four states. The working state is defined primarily based on the appliance's function. For example, different power levels of a light correspond to distinct working states. After turning it on or switching different operating states, the frequency of the acoustic signature was unstable for a brief time. We then assessed the warm-up times of 40 appliances, the CDF of which is shown on the right side of Fig. 18. We can see that some appliances required more than one minute (at most 180s) to stabilize after being turned on, whereas 90% of them required less than 5 seconds.

5.4.2 Training Size. Training size is the data collected on single appliance for training. To determine the optimal training size, we trained the classification model using training data of various sizes. In experiments on Electronic Store Dataset, we determined that the F1 score initially increased with the size of the training data, but then plateaued when the scale of the training data scale exceeded 120 seconds. Based on Fig. 19(a) , we can conclude that 120 seconds of training data was optimal, as a smaller training size would lead to under-fitting and a larger training size would bring no benefit. With a training size of 120 seconds, the classifier achieved an F1 score of 0.95.

5.5 Method Evaluation

5.5.1 Ablation Study of Feature Extraction Scheme. We applied an auto-encoder model, FCN-BLSTM, to extract features unrelated to environmental noise, as discussed in the methodology Section 4.2. To evaluate its ability to remove noise, we fixed the classifier as the proposed MS-LSTM model, and compared the classification performance under various settings on the Electronic Store Dataset, which includes directly perform classifying without feature extraction using data from mobile phones (No extract), using CNN, LSTM, FCN, and proposed feature extraction model. As shown in Fig. 19(b), results show that, compared to directly classifying data collected on mobile devices, the method of using a high sampling rate sound card to collect signals as ground truth achieves an F1 score improvement of at least 0.15. This is because it is difficult to collect signals that are invariant over time by using mobile devices directly due to anti-aliasing filtering and aliasing, etc. Besides, the proposed FCN-BLSTM also outperforms other feature extraction methods by an average of 0.08, achieving best appliance classification F1-score of 0.95 with a variance of 0.03.

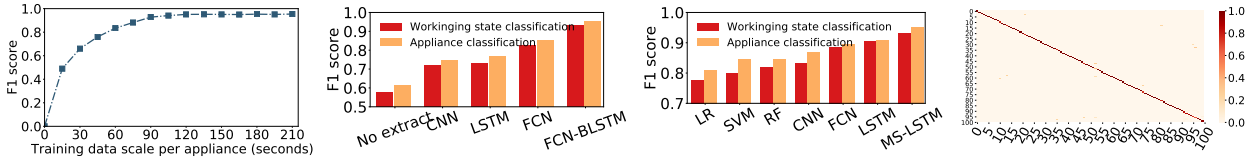
5.5.2 Ablation Study of Classification Scheme. To examine the performance of multi-stage strategy used by MS-LSTM, we employed the feature extractor as FCN-BLSTM, and compared it to other classifiers such as SVM, FCN, etc. on the Electronic Store Dataset. As shown in Fig. 19(c), the MS-LSTM model outperforms other strategies by over 0.08 (appliance classification) and 0.09 (state classification) correspondingly. This is because that the proposed MS-LSTM can both utilize the similarity of frequency patterns caused by switching appliances and working status and meanwhile explores the difference.

5.5.3 Overall Performance. We used the sound data collected from 100 appliances for classification in the validation process. As discussed in the previous section, the window size was set at 0.5 second, and the model was trained using data obtained over a period of 120 seconds. The classification results in Fig. 19(d) show an average classification F1 score (100 appliances) of 0.95. Specifically, 90 of the appliances (different models) achieved F1 scores of 0.96.

5.6 Interference Factors

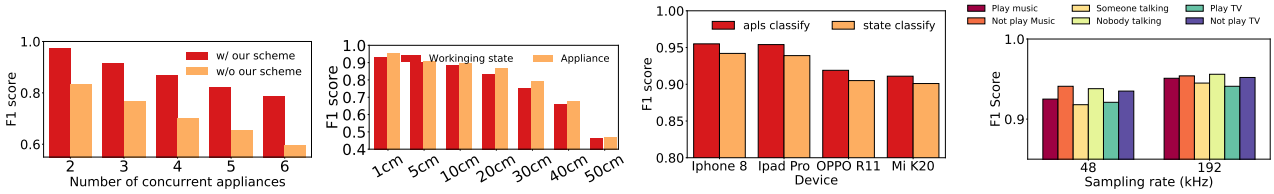
5.6.1 Interaction Distance. We first evaluated the effect of distance on the *SingSprite* system. In the experiment, we fixed the appliance in one position and continuously adjusted the distance between the phone and the appliance (1-50cm). We recorded the accuracy of recognizing the appliance and its working status at that position. We repeated the process on 10 common appliances and reported the average performance, as shown in Fig. 20(b). The experiments show that *SingSprite* can allow users to achieve better interaction with appliances within 30cm (accuracy rates are above 80%); at the same time, some appliances can still work at distances above 50cm (e.g., desktop computers, TV shows, and other high-powered appliances); since most appliances are recognized within 30cm, *SingSprite* can be used to recognize appliances in close proximity and reduce the user's choice space. For blind users who rely on the voice broadcasting service to read out the options, the reduction of options can significantly reduce the time spent on making a choice.

5.6.2 Robustness to Concurrent Appliance Interference. To evaluate the system's performance in a multi-appliance environment, we designed a protocol to measure its robustness against acoustic interference, simulating a common real-world scenario where a user



(a) Influence of training size. (b) Feature extraction scheme. (c) Classification scheme. (d) Confusion Matrix.

Figure 19: Efficiency of proposed schemes when fingerprinting appliances and working states.



(a) N concurrent appliances. (b) Detection distance. (c) Mobile device diversity. (d) Audible interference.

Figure 20: Appliance and working state classification performance of *SingSprite* under different interference.



Figure 21: Experimental scenario collecting data of concurrently working appliances.

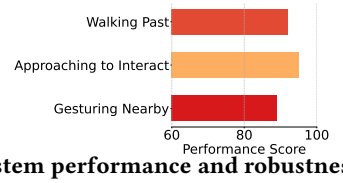


Figure 22: System performance and robustness under three distinct user motion scenarios.

interacts with one device while others operate in the background. In this setup, one target appliance was placed at a close interaction distance of approximately 15 cm from the user's smartwatch. Several other interfering appliances were positioned at varying distances of 20-50 cm to create a complex and noisy acoustic environment. The primary task was to correctly classify the target appliance amidst the mixed signals from all concurrently operating devices. We systematically evaluated the target's classification accuracy while increasing the number of active interferers from one to five. To ensure the robustness of our findings, each test condition was repeated. For instance, the performance metric reported for five concurrent appliances is the macro-average of three separate trials, each using a different combination of interfering devices. A photograph illustrating this experimental setup is provided in Figure 21).

Fig. 20(a) shows the performance when the N appliances are of different models. The existence of 2 appliances seldom disturbs the recognition. However, when there were 6 working appliances at the same time, and F1 score declines to 0.63. The proposed multi-label classification scheme can decompose humming sounds and improve the F1 score by 6.4% to 13.1%, achieving an accuracy of 0.82 when 5 appliances working concurrently.

5.6.3 Mobile Devices Diversity. The microphones in different mobile devices differ considerably in terms of frequency response, which manifests as differences in sensitivity to humming sounds. We verified this hypothesis using the four of the smartphones shown in Fig. 20(c). When using the same pre-learned model for appliance recognition, the F1 score of the four devices varied between 0.95 and 0.91, while F1 score of state recognition varied from 0.94 to 0.90, indicating that phone type can have a significant influence on recognition performance. We also observed that the iPhone outperformed Android devices by an average of 0.04, due to the fact that the iPhone system does not use a low-pass filter module, resulting in superior frequency response [38]. Our overall F1 score of 0.92 demonstrates the efficacy of the proposed *HumNet* when applied to different mobile devices.

5.6.4 Environmental Noises. We sought to determine whether audible interference would obscure the signals of interest by performing experiments in the presence of ambient music, conversation, and television. As shown in Fig. 20(d), sampling at 96kHz enabled *SingSprite* to utilize high frequency features, such that detection was unimpaired by audible interference. Also the data augmentation strategy makes it more robust to noises. When sampling at 48kHz, the environmental noise overlapped the aliased acoustic signature, resulting in a 1.5% to 2.1% drop in F1 scores. However, the overall F1 score of 0.91 demonstrates the efficacy of the proposed *HumNet* when applied to environment inference.

5.6.5 Robustness to User Motion. To validate performance in dynamic conditions, we evaluated the system's robustness to user motion by having a participant perform three realistic scenarios (Figure 22). The system proved highly resilient. Walking Past: When a user walked past the appliance at approximately 50 cm, the system correctly withheld classification, achieving a low false-positive rate of 0.08. Approaching to Interact: In the critical use case of a user

Table 2: Usability questionnaire adapted from [22].

Q1.	I found the system intuitive and easy to use.
Q2.	I feel I need more technical knowledge to use the system.
Q3.	I can interact directly with target appliances.
Q4.	I can't establish connection with target appliances.
Q5.	I feel the system can respond quickly to my requests.
Q6.	I spent a lot of time waiting for the system to response.
Q7.	I think it's easy to initiate interaction with an appliance.
Q8.	I think it's too much effort costs to initiate interaction with an appliance.
Q9.	I would use this system to interact with electrical appliances in my daily life.
Q10.	I find the system unnecessarily complex.

stopping within 30 cm of the appliance, it achieved a successful identification with an F1 score of 0.95 and an average detection latency of just 0.8 seconds. Gesturing Nearby: While simulating work on a countertop at 30 cm with continuous arm and hand motions, the system maintained a stable F1 score of 0.89.

This resilience stems from our system's core design. The FCN-BLSTM auto-encoder is specifically trained to extract time-invariant features, effectively filtering out signal fluctuations caused by movement. It distinguishes the stable, quasi-periodic nature of the SMPS signature from the aperiodic, transient noise created by user motion, a process aided by our pre-processing pipeline.

5.7 Results and Analysis

5.7.1 Quantitative Findings. Table 3 summarizes the quantitative performance results. The data reveals a critical trade-off between the two systems. While the RFID method achieved a slightly higher raw accuracy (98% vs. 95%), SingSprite was significantly faster (avg. 1.5s vs. 2.5s) and, most importantly, operated at a vastly greater distance. SingSprite offered an average recognition distance of 35cm, creating a large, forgiving "interaction volume," whereas RFID required precise positioning within a tiny 5cm radius.

5.7.2 Qualitative Analysis and Discussion. The strong user preference for *SingSprite* stems directly from its larger interaction distance, as confirmed by the questionnaire (Table 4). Participants found the RFID system frustrating, describing it as a high-effort "search-and-find" task to precisely locate a small tag without vision. This difficulty is reflected in its low score for ease of initiation (Q7: 3.4) and high score for required effort (Q8: 3.1).

In contrast, users perceived *SingSprite*'s interaction as "natural" and seamless, as they could simply point their device in the appliance's general direction. This is supported by its superior scores for ease of use (Q1: 4.8, Q7: 4.9) and low effort (Q2: 1.4, Q8: 1.1).

In conclusion, RFID's marginal accuracy advantage was rendered moot by its poor real-world usability. For blind users, the ability to interact conveniently from a distance without precise physical searching is far more valuable than a minor gain in accuracy. The overwhelming preference for the system that minimized physical and cognitive load confirms the success of *SingSprite*'s non-tactile, proximity-based interaction model.

6 Applications

To demonstrate how *SingSprite* facilitates interactions between blind individuals and household appliances across various contexts and environments, we describe four potential applications (Fig. 23).

The *SingSprite* system is triggered when the user long presses on the power button. Note that we built the *SingSprite* system for

Table 3: Experimental results vs. different solutions. Dist. stands for distance, Avrg for average among all volunteers and appliances and Accu for average accuracy.

Solution	Avrg Dist.	Min	Max	Accu	Avrg Time	Min	Max
SingSprite	35cm	20cm	80cm	95%	1.5s	1s	3s
RFID	5cm	4cm	7cm	98%	2.5s	1s	4s

Table 4: Questionnaire results vs. different solutions, listing the average scores of 16 volunteers.

Solution	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
SingSprite	4.8	1.4	4.9	1.2	4.6	2.5	4.9	1.1	4.8	1.5
RFID	4.5	1.6	4.7	2.1	4.7	1.3	3.4	3.1	4.1	2.6

different mobile devices including a **smartphone (XiaoMi 10)** and a **smartwatch (HuaWei Watch)**.

One-click appliance connection - *SingSprite* can automatically recognize the presence of nearby appliances when the blind user approaches appliances he or she wants to use. In conjunction with the smartphone's accessibility mode designed for blind users, the system audibly announces the names of all detected appliances. The user can select the desired appliance by double-tapping the screen upon hearing the corresponding prompt, as illustrated in Fig. 23(a).

Non-smart appliance interaction - *SingSprite* enables blind users to interact with non-smart appliances conveniently. For instance, when a user places food in a refrigerator, *SingSprite* detects the opening of the refrigerator door and provides an audible prompt to record the expiration date of the stored food. Additionally, the system issues auditory reminders to retrieve food approaching its expiration date, as shown in Fig. 23(b). Similarly, when using an oven, *SingSprite* can suggest appropriate cooking times and temperatures through voice guidance.

Abnormal status recognition - *SingSprite* is capable of monitoring appliance statuses and issuing alerts in case of anomalies, thereby enhancing the safety of blind users. For example, when the smartwatch worn by the user detects that the kettle is boiling dry, it does so by identifying a distinct shift in the ultrasonic signature from the kettle's power supply, caused by the drastic change in electrical load as the water evaporates. This approach is significantly more robust than using audible boiling sounds, which are unreliable, easily masked by ambient noise, and offer a delayed reaction for a critical safety alert. *SingSprite* will alert the user to add water by vibrating, as shown in Fig. 23(c).

Danger avoidance - *SingSprite* can recognize potential dangers. When integrated into a smartwatch worn by a blind user, *SingSprite* detects proximity to dangerous appliances in operation, such as an oven in use. The system immediately issues voice and a vibrational alert to warn the user of the danger ahead, as illustrated in Fig. 23(d).

7 Related Work

Previous research on appliance identification and interaction can be categorized into several approaches:

Power Signal-Based Fingerprinting: Electrical appliances induce unique power fluctuations in the electrical grid, which can be detected by power sensors. Researchers have successfully identified appliances using commercial power meters such as Plogg Kill-a-Watt and Watts Up [54] [6, 28, 51, 55, 56, 61]. However, these systems

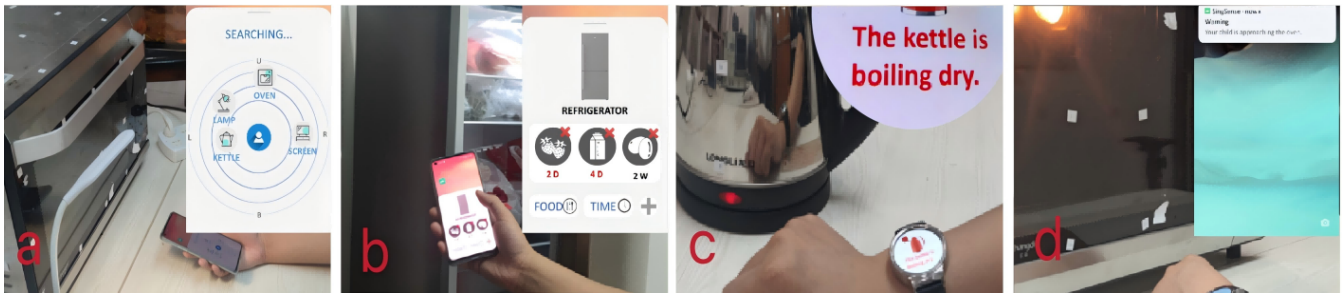


Figure 23: Typical application scenarios of *SingSprite*.

face significant limitations: (1) Deployment complexity—requiring hardwired sensors at electrical panels or outlets; (2) Scalability issues—retrofitting existing infrastructure is costly and inflexible; (3) Dynamic interaction barriers—fixed installations preclude spontaneous interactions with appliances in mobile scenarios (e.g., public spaces). Wang et al. [56] attempted to mitigate privacy concerns via edge computing but could not overcome the fundamental constraints of form factor and installation rigidity.

Electromagnetic Signal-Based Fingerprinting: Appliances emit high-frequency electromagnetic signals during operation, typically from components like switching power supplies. Magnetic sensors can capture these signals to differentiate devices. For example, Liu et al. [35] developed CamRadar, which detects hidden cameras by analyzing electromagnetic radiation leakage. Similarly, Zhang et al. [65] identified wireless devices using specific radiation patterns. Yang et al. [45, 46, 63] identify users, Apps or mics by fingerprinting the unique electromagnetic (EM) patterns generated by their physical interactions with a device. However, the low sampling rates of consumer electromagnetic sensors, such as those in smartphones and smartwatches (typically below 200 Hz), limit their identification accuracy in practical applications [16].

Optical Signal-Based Fingerprinting: Optical methods utilize visual data for appliance identification. Kim et al. [27] created a smart ring with vision capabilities to recognize appliance categories through image analysis. Systems like Snaplink [10, 34] employ smartphone cameras to capture and upload appliance images for recognition. Additionally, LiSee [11] introduced a neck-mounted system with dual cameras for assisting visually impaired users in interacting with their environment. Despite their potential, these camera-based methods raise privacy concerns and pose challenges for blind users who must capture images from specific angles. Nevertheless, these approaches inherit intrinsic challenges: (1) Viewpoint dependency—requiring line-of-sight and optimal angles for capture; (2) Ambient light sensitivity—performance degrades in low-light or high-glare environments; (3) Privacy risks—continuous imaging raises ethical concerns. For blind users, aligning cameras poses additional usability hurdles [34].

Acoustic Signal-Based Fingerprinting: Acoustic methods identify appliances based on sound emissions. Zhu et al. [70] and Yang et al. [64] monitored residential energy usage by analyzing appliance sounds recorded by smart speakers. Knocker [13, 14, 19] uses microphones to identify objects by analyzing sounds produced during tapping. Zhang et al. [66] achieve dynamic acoustic beam steering by jointly optimizing a metasurface and the signals from a small number of speakers. However, these techniques are limited to appliances that generate distinct operational sounds [26], and tapping-based interactions are impractical for routine use. In contrast, our proposed method utilizes ultrasonic emissions from switching power supplies—a nearly universal feature of appliances—offering broader applicability than traditional acoustic approaches.

Tag-Based Fingerprinting: RFID/NFC tags provide reliable identification but demand physical attachment. Commercial ecosystems (Apple Find My [5], Huawei HiLink [23]) enable item tracking via UWB/BLE tags, yet face three core constraints: (1) Deployment overhead manual tagging is infeasible at scale; (2) Durability—tags detach or degrade under wear; (3) Proximity requirements users must be within 5–10 cm for reliable reads. Wang et al. [57] improved read range using directional antennas but could not eliminate line-of-sight dependencies. For blind users, locating tiny tags on appliances remains a significant barrier, motivating our focus on contactless, proximity-tolerant (30–50 cm) interaction.

8 Conclusion

The proliferation of touchscreen interfaces on modern appliances has created significant accessibility barriers and safety hazards for blind individuals. In this paper, we addressed this challenge by introducing *SingSprite*, an innovative solution that uses acoustic signatures from electrical appliances to improve accessibility for blind users. This system operates on COST mobile devices without requiring external sensors. Our evaluations showed that the feature extraction module improved the F1 score by 0.15, the classification module by 0.08, and the overall end-to-end scheme by 0.20, achieving an impressive F1 score of 0.95 across 100 appliances. A user study with 16 blind participants indicated that *SingSprite* is more user-friendly than existing solutions like RFID tags. In general, *SingSprite* significantly improves the accessibility of the appliance for blind people in scenarios such as one-click connections and status recognition.

9 Discussion and Limitations

SingSprite has several limitations. Its functionality depends on acoustics generated by switched-mode power supplies (SMPS), and appliances without SMPS components cannot be reliably detected. Nevertheless, many modern kitchen appliances, such as kettles and rice cookers, still contain internal SMPS and are thus detectable. Practical deployment is also less convenient for devices with external power adapters (e.g., laptops and monitors), since detection

requires placing the phone near the adapter rather than the main device. Although this issue is mitigated by the market trend toward increasingly integrated SMPS designs, it still affects broader applicability.

Interaction distance is limited by mobile phone anti-aliasing filters, which attenuate acoustic signatures above 10 kHz and restrict reliable detection to about 30 cm. This range is appropriate for *SingSprite*'s close-proximity use case, where users naturally approach appliances to operate or assess them. It also matches typical safety margins, ensuring hazardous appliances are detected only at a relevant distance. Although long-range detection is not our goal, techniques such as beamforming or microphone array processing could extend the range when needed.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62502303), Natural Science Foundation of Shanghai (No. 25ZR1401171 and No. 24ZR1439100), and the Shanghai Key Laboratory of Trusted Data Circulation and Governance, and Web3.

References

[1] 2022. On flat-button kitchen appliances, blind people, and the danger of inaccessibility : r/legaladviceofftopic. https://www.reddit.com/r/legaladviceofftopic/comments/vrp0ga/on_flatbutton_kitchen_appliances_blind_people_and/?show=original [Online; accessed 2025-07-01].

[2] 2025. Beijing Disabled Persons' Federation-Building a three-dimensional barrier-free space for the blind. https://www.bdpf.org.cn/n1544/n1689/n1767/n1799/c56075/content.html [Online; accessed 2025-07-01].

[3] Kwangjun Ahn, Zhiyu Zhang, Yunbum Kook, and Yan Dai. 2024. Understanding Adam optimizer via online learning of updates: Adam is FTRL in disguise. *arXiv preprint arXiv:2402.01567* (2024).

[4] Apple. 2025. Applehome. https://www.apple.com.cn/home-app/.

[5] Apple. 2025. Applesiri. https://www.apple.com.cn/siri/.

[6] Christos Athanasiadis, Dimitrios Doukas, Theofilos Papadopoulos, and Antonios Chrysopoulos. 2021. A Scalable Real-Time Non-Intrusive Load Monitoring System for the Estimation of Household Appliance Power Consumption. *Energies* 14, 3 (2021), 767.

[7] NCHR Author. 2009. Blind Adults in America: Their Lives and Challenges - National Center for Health Research. https://www.center4research.org/blind-adults-america-lives-challenges/ [Online; accessed 2025-07-01].

[8] Brais Bosquet, Daniel Cores, Lorenzo Seidenari, Víctor M. Brea, Manuel Mucientes, and Alberto Del Bimbo. 2023. A full data augmentation pipeline for small object detection based on generative adversarial networks. *Pattern Recognition* 133 (2023), 108998. doi:10.1016/j.patcog.2022.108998

[9] Weihan Cao, Yifan Zhang, Jianfei Gao, Anda Cheng, Ke Cheng, and Jian Cheng. 2022. Pkd: General distillation framework for object detectors via pearson correlation coefficient. *Advances in Neural Information Processing Systems* 35 (2022), 15394–15406.

[10] Kaifei Chen, Jonathan Fürst, John Kolb, Hyung-Sin Kim, Xin Jin, David E Culler, and Randy H Katz. 2018. Snaplink: Fast and accurate vision-based appliance control in large commercial buildings. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–27.

[11] Kaixin Chen, Yongzhi Huang, Yicong Chen, Haobin Zhong, Lihua Lin, Lu Wang, and Kaishun Wu. 2022. Lisee: A headphone that provides all-day assistance for blind and low-vision users to reach surrounding objects. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–30.

[12] Pranav U. Damale, Edwin K. P. Chong, and Louis L. Scharf. 2023. Wiener Filtering Without Covariance Matrix Inversion. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1–5. doi:10.1109/ICASSP49357.2023.10094899

[13] Dian Ding, Lanqing Yang, Yi-Chao Chen, and Guangtao Xue. 2021. Vibwriter: Handwriting recognition system based on vibration signal. In *2021 18th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 1–9.

[14] Dian Ding, Lanqing Yang, Yi-Chao Chen, and Guangtao Xue. 2022. Handwriting recognition system leveraging vibration signal on smartphones. *IEEE Transactions on Mobile Computing* 22, 7 (2022), 3940–3951.

[15] International Agency for the Prevention of Blindness (IAPB). 2020. Projected Change in Vision Loss 2020 to 2050. https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/projected-change/.

[16] Yongjian Fu, Lanqing Yang, Hao Pan, Yi-Chao Chen, Guangtao Xue, and Ju Ren. 2024. Magspy: Revealing user privacy leakage via magnetometer on mobile devices. *IEEE Transactions on Mobile Computing* (2024).

[17] Coolest Gadgets. 2025. Home and Household Appliances Statistics By Category, Manufacturers and Usage. https://www.coolest-gadgets.com/appliances-statistics/ Accessed: 2025-04-22.

[18] geeksforgeeks. 2024. Working Principle of SMPS. https://www.geeksforgeeks.org/working-principle-of-smps/.

[19] Taesik Gong, Hyunsung Cho, Bowon Lee, and Sung-Ju Lee. 2019. Knocker: Vibroacoustic-based object recognition with smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–21.

[20] Saroj Gopali, Faranak Abri, Sima Siami-Namini, and Akbar Siami Namin. 2021. A comparison of TCN and LSTM models in detecting anomalies in time series data. In *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2415–2420.

[21] Nishi Shahnaj Haider. 2021. Respiratory sound denoising using Empirical Mode Decomposition, Hurst analysis and Spectral Subtraction. *Biomedical Signal Processing and Control* 64 (2021), 102313. doi:10.1016/j.bspc.2020.102313

[22] Thamer Horbylon Nascimento and Fabrizzio Soares. 2021. Home Appliance Control Using Smartwatches with Continuous Gesture Recognition. In *Distributed, Ambient and Pervasive Interactions*, Norbert Streitz and Shin'ichi Konomi (Eds.). Springer International Publishing, Cham, 122–134.

[23] HuaWei. 2021. Phone Meets Notebook. https://consumer.huawei.com/cn/emui10-1/tips/mobile-office-list/article1/.

[24] Hyo-won. [n. d.]. Visiting the grassroots during the Spring Festival | Busy electricity workers guarding the blind compound. http://ln.people.com.cn/BIG5/n2/2023/0124/c378489-40277667.html [Online; accessed 2025-07-01].

[25] Yasir Iqbal, Tao Zhang, Teddy Surya Gunawan, Agus Pratondo, Xin Zhao, Yanzhang Geng, Mira Kartiwi, Nasir Saleem, and Sami Bourious. 2025. A Hybrid Speech Enhancement Technique Based on Discrete Wavelet Transform and Spectral Subtraction. *IEEE Access* (2025).

[26] Liheng Jiang, Yongzhao Zhang, Ting Chen, Yi-Chao Chen, Dian Ding, Yijie Li, Fenghua Xu, Liwei Guo, Jingwei Li, Xiong Li, et al. 2025. A Survey on Acoustic Sensing in the Metasurface Era: Challenges, Advances, and Applications. *Fundamental Research* (2025).

[27] Maruchi Kim, Antonio Glenn, Bandhav Veluri, Yunseo Lee, Eyoel Gebre, Aditya Bagaria, Shwetak Patel, and Shyamnath Gollakota. 2024. IRIS: Wireless ring for vision-based smart home interaction. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 114, 16 pages. doi:10.1145/3654777.3676327

[28] Chin-Feng Lai, Wei-Che Chien, Laurence T Yang, and Weizhong Qiang. 2019. LSTM and edge computing for big data feature recognition of industrial electrical equipment. *IEEE Transactions on Industrial Informatics* 15, 4 (2019), 2469–2477.

[29] Meng Lei, Jiaqi Zhang, Shiqi Wang, Shanshe Wang, and Siwei Ma. 2023. Deep Intra Prediction by Jointly Exploiting Local and Non-Local Similarities. *IEEE Transactions on Circuits and Systems for Video Technology* 33, 5 (2023), 2396–2409. doi:10.1109/TCSVT.2022.3220434

[30] Jingyi Li. [n. d.]. State Grid Liaoning Power Shenyang Guo Family Love Team. https://qnzs.youth.cn/tsxq/202404/t20240423_15212810.htm [Online; accessed 2025-07-01].

[31] Jieye Liang and Qingning Zeng. 2023. Improved Spectral Subtraction Based on Second-Order Differential Array and Phase Spectrum Compensation. In *2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS)*. 658–661. doi:10.1109/ISCTIS58954.2023.10212993

[32] Linfeng Liu, Sitan Chen, Kejia Chen, Jiajun Xu, and Xiacan Chen. 2025. Catching the Blackdog Easily: A Convenient Depression Diagnosis Method based on Audio-Visual Deep Learning. *IEEE Transactions on Affective Computing* (2025).

[33] Tiantian Liu, Chao Wang, Zhengxiong Li, Ming-Chun Huang, Wenyao Xu, and Feng Lin. 2024. Wavoice: An mmWave-Assisted Noise-Resistant Speech Recognition System. *ACM Trans. Sen. Netw.* 20, 4, Article 86 (May 2024), 29 pages. doi:10.1145/3597457

[34] Yiming Liu, Honggui Han, Fangyu Li, and Yongping Du. 2024. Adaptive Sparse Semantic Representation for Waste Electrical Appliance Recognition. In *2024 14th Asian Control Conference (ASCC)*. IEEE, 1–6.

[35] Ziwei Liu, Feng Lin, Chao Wang, Yijie Shen, Zhongjie Ba, Li Lu, Wenyao Xu, and Kui Ren. 2023. Camradar: Hidden camera detection leveraging amplitude-modulated sensor images embedded in electromagnetic emanations. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–25.

[36] Kathryn Locke, Katie Ellis, Mike Kent, Leanne McRae, and Gwyneth Peaty. 2020. Smartphones and equal access for people who are blind or have low vision. (2020).

[37] Anton Lorenz, Christian D Matthus, and Frank Ellinger. 2025. Microstrip based Anti-Aliasing Filters for Beyond-50 GHz-Bandwidth Mix-Filter-Sample Analog-to-Digital-Converters. *IEEE Access* (2025).

[38] Haitong Luo, Xuying Meng, Suhang Wang, Hanyun Cao, Weiyao Zhang, Yequan Wang, and Yujun Zhang. 2024. Spectral-Based Graph Neural Networks for Complementary Item Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 8868–8876.

[39] Erik Marchi, Giacomo Ferroni, Florian Eyben, Leonardo Gabrielli, Stefano Squartini, and Björn Schuller. 2014. Multi-resolution linear prediction based features for audio onset detection with bidirectional LSTM neural networks. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2164–2168.

[40] MathWorks. 2020. 2-D correlation. https://ww2.mathworks.cn/help/vision/ref/2dcorrelation.html.

[41] Muhamad Muzammul and Xi Li. 2025. Comprehensive review of deep learning-based tiny object detection: Challenges, strategies, and future directions. *Knowledge and Information Systems* (2025), 1–89.

[42] M Nagabushanam, Sadashiva Chakrasali, SL Gangadharaiah, Sampath H Patel, Gurumurthy Ramaiah, and Raju Basak. 2024. An Optimized VLSI Implementation of the Least Mean Square (LMS) Adaptive Filter Architecture on the Basis of Distributed Arithmetic Approach. *Journal of The Institution of Engineers (India): Series B* (2024), 1–10.

[43] Maria OHISALO. [n. d.]. Parliamentary question | Accessibility of household appliances for visually impaired people | E-000884/2025 | European Parliament. https://www.europarl.europa.eu/doceo/document/E-10-2025-000884_EN.html?utm_source=chatgpt.com [Online; accessed 2025-04-18].

[44] Jaclyn Packer. [n. d.]. Left to Your Own Devices: Results of a Study on the Usability of Everyday Household and Electronics Products for People with Vision Loss | Accessworld | American Foundation for the Blind. https://www.afb.org/aw/12/11/15975 [Online; accessed 2025-04-18].

[45] Hao Pan, Feitong Tan, Wenhao Li, Yi-Chao Chen, Lanqing Yang, Guangtao Xue, and Xiaoyu Ji. 2022. MagDefender: Detecting eavesdropping on mobile devices using the built-in magnetometer. In *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 28–36.

[46] Hao Pan, Lanqing Yang, Honglu Li, Chuang-Wen You, Xiaoyu Ji, Yi-Chao Chen, Zhenxian Hu, and Guangtao Xue. 2021. Magthief: Stealing private app usage data on mobile devices via built-in magnetometer. In *2021 18th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 1–9.

[47] Shengju Qian, Hao Shao, Yi Zhu, Mu Li, and Jiaya Jia. 2021. Blending anti-aliasing into vision transformer. *Advances in Neural Information Processing Systems* 34 (2021), 5416–5429.

[48] Y Sri Rama, E Vijaya Babu, KV Balaramakrishna, Sanjana Mohithe, Siri Chandana Prodduturi, and Bhoomika Amgoth. 2025. VLSI based Cascaded Integrator Comb (CIC) Filter for RADAR. In *2025 5th International Conference on Trends in Material Science and Inventive Materials (ICTMIM)*. IEEE, 1841–1846.

[49] Ganesan Ramalingam. 1994. The undecidability of aliasing. *ACM Transactions on Programming Languages and Systems (TOPLAS)* 16, 5 (1994), 1467–1471.

[50] Jia-Liang Ren, Ya-Hui Chien, En-Yu Chia, Li-Chen Fu, and Jin-Shin Lai. 2019. Deep learning based motion prediction for exoskeleton robot control in upper limb rehabilitation. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 5076–5082.

[51] Hajer Salem, Moamar Sayed-Mouchaweh, and Moncef Tagina. 2020. A Review on Non-intrusive Load Monitoring Approaches Based on Machine Learning. In *Artificial Intelligence Techniques for a Scalable Energy Transition*. Springer, 109–131.

[52] statista. 2024. Home household appliances - statistics & facts. https://www.statista.com/topics/1068/home-appliances#topicOverview.

[53] statista. 2024. Household Appliances - Worldwide. https://www.statista.com/outlook/cmo/household-appliances/worldwide.

[54] Z Cihan Taysi, M Amac Guvensan, and Tommaso Melodia. 2010. Tinyears: spying on house appliances with audio sensor nodes. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*. 31–36.

[55] Sagar Verma, Shikha Singh, and Angshul Majumdar. 2021. Multi-label LSTM autoencoder for non-intrusive appliance load monitoring. *Electric Power Systems Research* 199 (2021), 107414.

[56] Haoxiang Wang, Jiasheng Zhang, Chenbei Lu, and Chenye Wu. 2020. Privacy Preserving in Non-Intrusive Load Monitoring: A Differential Privacy Perspective. *IEEE Transactions on Smart Grid* 12, 3 (2020), 2529–2543.

[57] Zhoufeng Wang, Xiangqi He, Huiling Shen, Shangjie Fan, and Yilong Zeng. 2022. Multi-source information fusion to identify water supply pipe leakage based on SVM and VMD. *Information Processing & Management* 59, 2 (2022), 102819. doi:10.1016/j.ipm.2021.102819

[58] Yuxuan Wei, Zehan Wang, Tian Guo, Hao Liu, Liquan Shen, and Hui Yuan. 2025. High Efficiency Wiener Filter-based Point Cloud Quality Enhancement for MPEG G-PCC. *IEEE Transactions on Circuits and Systems for Video Technology* (2025).

[59] Xiaomi. 2025. home.mi. https://home.mi.com/.

[60] Xiaomi. 2025. Xiaoai. https://xiaoai.mi.com/?from=aigc.izzi.cn

[61] Lei Yan, Jiayu Han, Hong Wang, Zhiyi Li, and Zuyi Li. 2020. An Online Transient-Based Electrical Appliance State Tracking Method Via Markov Chain Monte Carlo Sampling. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 1–5.

[62] Guanzhi Yang, Yundong Cao, Huaikang Jin, Shouyuan Fan, Bin Wang, and Huaiping Jin. 2024. Multi-stage Attention Mechanism based Deep Learning for Wind Power Forecasting. In *2024 43rd Chinese Control Conference (CCC)*. IEEE, 6287–6294.

[63] Lanqing Yang, Xinqi Chen, Hao Pan, Yi-Chao Chen, Guangtao Xue, Zechen Li, Yiheng Bian, Dian Ding, Linghe Kong, Jiadi Yu, et al. 2025. MagPrint++: Continuous User Fingerprinting On Mobile Devices Using Electromagnetic Signals. *IEEE Transactions on Mobile Computing* (2025).

[64] Lanqing Yang, Honglu Li, Zhaoxi Chen, Xiaoyu Ji, Yi-Chao Chen, Guangtao Xue, and Chuang-Wen You. 2020. Appliance fingerprinting using sound from power supply. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*. 160–163.

[65] Junning Zhang, Yicen Liu, Guoru Ding, Bo Tang, and Yanlong Chen. 2024. Adaptive decomposition and extraction network of individual fingerprint features for specific emitter identification. *IEEE Transactions on Information Forensics and Security* (2024).

[66] Yongzhao Zhang, Yezhou Wang, Lanqing Yang, Mei Wang, Yi-Chao Chen, Lili Qiu, Yihong Liu, Guangtao Xue, and Jiadi Yu. 2023. Acoustic sensing and communication using metasurface. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. 1359–1374.

[67] Yu Zhang, Zheng Zhang, Yanlin Zhang, Changming Zhang, Jintao Chen, Yuanyuan Luo, and Guotao Duan. 2024. A high-accurate H 2 S recognition method based on Pearson correlation coefficient feature extraction of dynamic response. *IEEE Sensors Journal* (2024).

[68] Han Zhao, Shuayb Zarar, Ivan Tashev, and Chin-Hui Lee. 2018. Convolutional-recurrent neural networks for speech enhancement. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2401–2405.

[69] Chuheng Zheng, Mondher Bouazizi, Tomoaki Ohtsuki, Momoko Kitazawa, Toshiro Horigome, and Taishiro Kishimoto. 2025. Enhancing Dementia Detection: Augmenting the Face Mesh. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*. IEEE, 1–4.

[70] Guanzhou Zhu, Dong Zhao, Kuo Tian, Zhengyuan Zhang, Rui Yuan, and Huadong Ma. 2023. Combining Smart Speaker and Smart Meter to Infer Your Residential Power Usage by Self-supervised Cross-modal Learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–26.