Contactless Material Identification with Millimeter Wave Vibrometry

Hailan Shanbhag*
École Polytechnique Fédérale de Lausanne

Sohrab Madani*
University of Illinois Urbana-Champaign

Akhil Isanaka
University of Illinois Urbana-Champaign

Deepak Nair
University of Illinois Urbana-Champaign

Saurabh Gupta
University of Illinois Urbana-Champaign

Haitham Hassanieh
École Polytechnique Fédérale de Lausanne

ABSTRACT

This paper introduces RFVibe, a system that enables contactless material and object identification through the fusion of millimeter wave wireless signals with acoustic signals. In particular, RFVibe plays an audio sound next to the object that generates micro-vibrations in the object. These micro-vibrations can be captured by shining a millimeter wave radar signal on the object and analyzing the phase of the reflected wireless signal. RFVibe can then extract several features including resonance frequencies and vibration modes, damping time of vibrations, and wireless reflection coefficients. These features are then used to enable more accurate identification, with a step towards generalizing towards different setups and locations. We implement RFVibe using an off-the-shelf millimeter-wave radar and an acoustic speaker. We evaluate it on 23 objects of 7 material types (Metal, Wood, Ceramic, Glass, Plastic, Cardboard, and Foam), obtaining 81.3% accuracy for material classification, a 30% improvement over prior work. RFVibe is able to classify with reasonable accuracy in scenarios that it has not encountered before, including different locations, angles, boundary conditions, and objects.

CCS CONCEPTS

• **Hardware** → **Wireless devices**; *Digital signal processing*; • **Computing methodologies** → *Neural networks*; • **Networks** → **Sensor networks**.

KEYWORDS

Millimeter-wave Sensing, Material Classification, Object Classification, Wireless Vibrometry

ACM Reference Format:
Hailan Shanbhag*, Sohrab Madani*, Akhil Isanaka, Deepak Nair, Saurabh Gupta, and Haitham Hassanieh. 2023. **Contactless Material Identification with Millimeter Wave Vibrometry** . In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys '23), June 18–22, 2023, Helsinki, Finland.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3581791.3596850

*indicates equal contribution.

MobiSys '23, June 18–22, 2023, Helsinki, Finland

ACM ISBN 979-8-4007-0110-8/23/06…$15.00
https://doi.org/10.1145/3581791.3596850

1 INTRODUCTION

Sensing the world around us has become increasingly important in the age of automation and smart devices. Non-invasive contactless material sensing and identification is a key primitive that can enable or enhance a plethora of applications like robotic grasping and mapping [38], liquid and food quality monitoring [12, 19], soil and plant sensing in agriculture [13], quality control in warehouses and sorting facilities, simpler security scanning as well as structural health monitoring of building, bridges, planes, trains, etc. Unfortunately, today material sensing requires bulky and specialized equipment like optical spectroscopy, X-Ray, and ultrasonography which typically cost tens of thousands of dollars and are not practical for everyday use [12, 22, 24].

Recently, there has been significant interest, in the wireless and mobile research community, to work towards enabling a scalable and practical solution [12, 19, 40, 47, 48]. In particular, recent work leverages wireless signals from IoT radios to capture material properties which can enable a cheap, non-invasive, and ubiquitous alternative that can be used for everyday applications. Most of the recent techniques either require physically touching the material [16, 51] or attaching an RFID tag to the object a priori [19, 40, 48] which is sometimes not feasible if the object being sensed is unreachable or does not have a rigid surface. Other techniques analyze the wireless signal properties after it has penetrated through the material to extract the material's permittivity [12, 15]. However, such techniques have only been demonstrated for liquids, require careful placement of the liquid container between two radios, and depend on the container the liquid is in. mSense [47] addresses these limitations by using millimeter wave (mmWave) radios and measuring the reflection coefficient of the wireless signal from the object to identify the material. mSense enables a portable and non-invasive solution but requires the mmWave radio to have a large phased array with 32 TX by 32 RX antennas and its performance can degrade with smaller arrays like 2 TX by 4 RX antennas as we show in Sec. 6. It is also fairly limited in how many scenarios it can accurately classify materials, as the reflection coefficient depends not only on the material of the object but the shape and size, the smoothness of its surface and the surrounding environment.

In this paper, we ask: *How can we push the performance of wireless sensing towards a contactless, cheap, and more generalizable solution?* To this end, we introduce RFVibe, a mmWave wireless system that can accurately identify different objects and their materials. RFVibe does not require careful placement of the object and can work for various locations and objects. The core idea behind RFVibe is to combine mmWave radar signals with acoustic waves to analyze the properties of the material. In particular, RFVibe shines mmWave

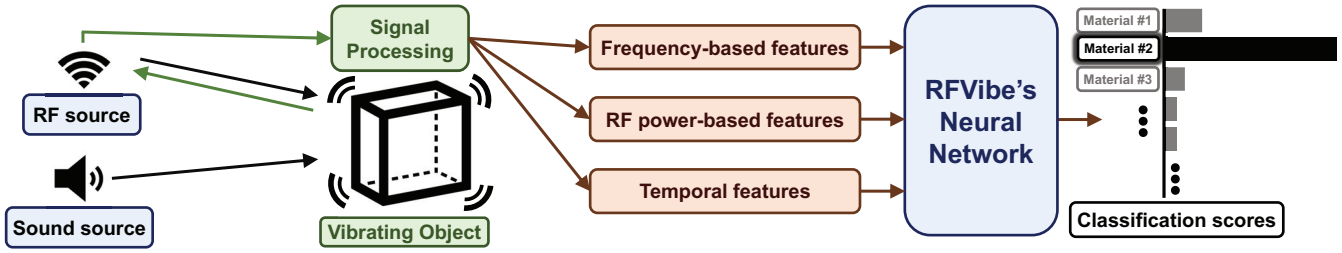


Figure 1: RFVibe's overall design.

radar signals at an object while playing a sound source near the object without any contact. The sound creates micro-vibrations in the object which gets picked up by the mmWave signals that carry rich information about the object's material and shape. As compared to simply using the reflection coefficient, vibration properties of material add another dimension, enabling us to identify material and objects more uniquely even for objects with non-uniform surfaces.

Translating the above idea into a practical system requires addressing several challenges. First, we must design an acoustic sound that allows us to extract resonance frequencies and vibration modes spanning various objects and frequencies. In other words, we need an acoustic sound which excites objects consistently and equally across a range of frequencies, ensuring that objects are excited in a reproducible manner, allowing for their unique frequencies to be distinguishable. Moreover, the excited frequencies must also be discoverable by the mmWave radar hardware as high vibration frequencies tend to be filtered out as we explain in detail in Sec. 4. Finally, as each object's vibration damps at a different speed, the acoustic sound must also allow us to extract the damping properties of the object to help with its classification. To address this, we design a weighted impulse train sound source that effectively stimulates vibrations, can be detected by the mmWave radar, and allows the vibrations to damp during the silence intervals between impulses (see Sec. 4.1 for more detail).

Second, feeding the raw data directly to a learning scheme results in poor performance as we show in Sec. 6.5. Hence, we need to extract the maximum amount of information useful for material classification from the radar data before feeding it to machine learning approaches. We note that the raw signal contains time-frequency information relating to the micro-vibrations of the object. In Sec. 4.2, we show how to extract visible vibration frequencies and how these frequencies change over time as our acoustic impulse train plays. We also show how to extract the speed at which the object excitation dampens by processing the time-domain signal. As these properties heavily depend on the object's material type, we introduce two new features, namely Frequency, and Damping, that play a key role in RFVibe's material and object classification. Finally, the raw signal also contains reflection power information, which depends both on the material type and traveled distance. Therefore, after compensating for the traveled distance, it can be used as a third feature, which we call the Power Feature. RFVibe then combines all three features to infer the object and its material.

Third, the extracted features need to be combined to perform classification. RFVibe's dataset includes different setups, objects, and environments, and since the time-frequency features (Frequency and Damp) are large vectors, performing classification using these features becomes a complex task. As such, using ready-to-use classification methods such as SVMs and random forest does not fully exploit the information in these features as we show in Sec. 6.5. Instead, RFVibe uses a deep neural network that can achieve accurate classification based on these features. To do so, RFVibe's neural network architecture uses three parallel branches which convert the vibration frequency, reflection power, and damping time features in one compressed latent space. To make sure each branch is trained to saturation, we leverage auxiliary supervision [35, 41] by combining the intermediate outputs and feeding them to a final combined classification head.

We implement and extensively evaluate RFVibe using commercial speakers and 77-GHz millimeter-wave radars equipped with 2 Tx and 4 Rx antennas in different environments and conditions. We tested RFVibe on 23 different objects consisting of 7 different material types (metal, cardboard, wood, ceramic, glass, plastic, and foam). Our results show that RFVibe achieves 81.3% overall material classification accuracy among all objects as well as 74.1% accuracy in object identification; a 30.0% improvement over prior work. In unseen settings and environments, RFVibe can still achieve 73.1% accuracy in material classification over all objects. This performance goes up to above 83% to 95% for materials with stronger vibrations like metals, cardboard, and foam. In Sec. 6.5 we show how RFVibe performs with new, unseen objects, different speakers, and unseen distances between the speaker, the radar, and the object, as well as different boundary conditions. Moreover, when trained only on forward-facing objects, RFVibe continues to perform accurately for unseen, rotated objects (up to 80°). Finally, an added benefit of RFVibe is its ability to operate in scenarios where vision-based material identification fails, for example when an object is occluded, inside a bag, painted over, or covered. We show that RFVibe can perform reasonably well in these unseen scenarios even if they are not in the training set. Because of that, we envision practical applications for material sensing abilities in distribution facilities, to verify that contents of a package match the label or to detect packages with broken glass.

It is worth noting, that while RFVibe takes a significant step towards non-invasive, contact-less material identification, there is still a long way to go in terms of its robustness, reliability, and generalizability to enable practical applications. Further discussion on RFVibe's limitations can be found in Sec. 7.

Contributions: The paper makes the following contributions:
• We introduce RFVibe, the first system that uses acoustic vibrations together with RF signals to identify objects and materials.

• We introduce novel vibration and damping features based on carefully designed acoustic sources that together characterize objects and their material types.
• We build a prototype of RFVibe using commercial radars and speakers and evaluate its performance against state-of-the-art baselines that relies on RF signals to identify materials.

2 RELATED WORK

A. Non-RF Vibrometry. Visual Vibrometry [11] captures video with high speed cameras to estimate material properties of objects based on the vibrational modes of an object. Other works [5] have since improved over [8, 11] or introduced new applications such as estimating tree structure [49]. Other camera based methods for material identification include [33, 36] that use time-of-flight (ToF) cameras to identify materials. Vibrosight [54] senses activities across rooms using long-range laser vibrometry. Laser vibrometers have been previously used in [4, 6, 7, 9, 10, 14, 20] spanning many applications such as locating mines, fruit quality testing, material defect identification, and antique quality assessment. Finally, [16] uses a contact-based method to detect potentially damaged products along a supply chain.

B. RF Based Material Sensing. Authors of [37] use Channel State Information (CSI) of commodity WiFi to identify suspicious objects inside a baggage. IntuWition [52] uses WiFi signals and methods in radar polarimetry to identify the location and types of materials. In another work, [45] uses one-shot learning to classify between objects based on the reflection properties of mmWave signals. Strobe [13] senses soil moisture using RF propagation in existing Wi-Fi bands. RadarCat [51] takes the reflected signal from different points on and within the target for classification using a random forest classifier. However, for the system to work, the target needs to be placed on the device, touching the sensor. Most relevant to our work is mSense [47]. It proposes a reflection-based material sensing method, using a material's reflection coefficient to classify materials. However, mSense needs controlled environments to work well. Another work [21] also relies on the power of reflected signals and a machine learning model to identify between five materials. However, their system is only tested on a small number of flat objects of similar size and thickness, and is not tested on less physically constrained settings. We implement and test systems from [21, 47] against RFVibe described in Sec. 6.4.

C. RF Based Liquid Sensing LiquID [12] and WiEps [32] identify the liquid type within a container using UWB and WiFi signals respectively, that pass through the container and liquid inside it. FG-LiquID [26] uses mmWave sensors and improves over [12] by leveraging neural networks. WiMi [15] classifies liquids that are placed between a router and a WiFi connected device using CSI. Authors of [31] further generalize liquid identification by using RF signals to do fine-grained liquid recognition regardless of container material, shape, and liquid height. HearLiquid [50] attaches a commodity speaker and microphone to a container to detect liquid fraud based on the acoustic absorption and transmission curve of the liquid. Akte-Liquid [34] uses a smartphone speaker and microphone to sense different amplitude-frequency features of various liquids. Vi-Liquid [23] identifies liquids based on their viscosity using phone vibrations attached to the liquid container.

D. RFID Based Sensing. Several works use RFID tags [1, 3, 17] to sense the presence of nearby objects, without identifying the type of object. TwinLeak [18] uses RFID tags to detect liquid leakages in industrial environments. Authors in [29] design a sensor antenna transmitter system whose frequency shifts depend on the permittivity of the material under test. TagScan [40] uses RFIDs and a feature which uses the phase and attenuation constant of the air to generalize the phase and attenuation constant of the target. Similarly, Tagtag [48] attaches overlapping RFID tags to targets and identifies them by looking at the high-resolution phase change caused by the tag antenna's impedance difference. Finally, [39] uses RFID tags to verify the internal status of packages and detect abnormal changes of the internal items. RF-EATS [19] senses food and liquids inside a container using the properties of signals reflected off RFID tags attached to their containers. While RFID based methods are effective, they rely on RFID tags being in direct contact with the object.

E. Recovering sound and vibrations using wireless signals. mmVib [25] measures object vibrations in industrial settings. Work in [44] and subsequently in [27, 42, 43, 53] demonstrate the possibility of picking up acoustic vibrations using RF signals. However, they do not use it to identify materials but rather to extract voice and speech from very thin materials.

3 BACKGROUND

3.1 Millimeter Wave Sensing

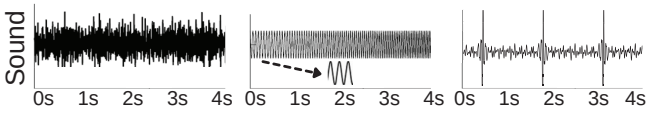
RFVibe uses a mmWave device to capture the vibrations of the objects in the scene. Specifically, it uses FMCW chirps that are continuously transmitted towards the vibrating object. The reflected signal is captured by the mmWave receiver. The phase of the received signal contains high-resolution information of the object vibrations. Specifically, each FMCW chirp gives a phase value that corresponds to the object distance:

$$\phi = 2\pi \frac{d}{\lambda} \tag{1}$$

where d is the distance from the object, and λ is the wavelength of the mmWave signal. Each FMCW sweep results in one phase measurement, so the frequency of chirps constrains the maximum detected vibration frequency without aliasing.

3.2 Object Vibration and Non-Destructive Testing

Based on the material, geometry and boundary conditions, systems will have different frequency response [30]. When the frequency of the applied force is close to the natural frequency of a system, the amplitude of the vibrations will increase and we can denote these as the resonant frequency of the system. These are frequencies with which objects will naturally vibrate at, and also vibrates at its peak amplitude when an external force is applied. Other frequencies can also incur a response, however at lower amplitudes. The natural



(a) White Noise (b) Linear Chirp (c) Impulse Train

Figure 2: Potential sound sources.

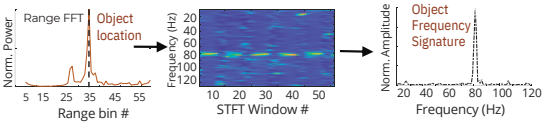


Figure 3: Extracting the frequency feature.

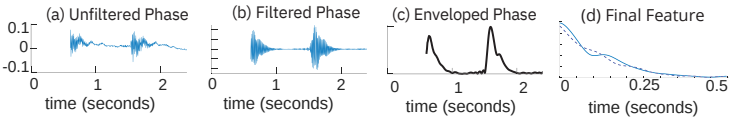


Figure 4: Extracting the damp feature.

frequency of an object is a fundamental property of objects based on the specific material and geometry.

There are multiple ways to perform vibration or stress testing. One way to monitor stress testing is by using a non-destructive testing technique called acoustic emission testing. This form on non-destructive testing measures the ultrasonic waves that are emitted from structures upon stress testing. However, the three most common are touch based techniques: sinusoidal, random and shock testing. 1) Sinusoidal vibration testing consists of vibrating objects at single sinusoidal tones or a frequency sweep. The resonant frequencies can be determined by the vibration measurements that give the largest amplitude. 2) Random vibration testing uses random noise within the frequency range to vibrate the object. 3) Shock testing measures the reaction of an object to a sudden impulse that excites a broadband range of frequencies. In vibration testing, all these methods are produced with contact based devices. We use equivalents of these methods in Sec. 4.1 to determine the best way to excite the objects.

4 RFVIBE'S DESIGN

RFVibe identifies materials by combining RF, acoustics, and vibrometry. It works by transmitting a sound signal generated using a speaker and measuring the vibrations of the object using a mmWave device. RFVibe analyzes these vibrations to identify the object. Fig. 1 shows an overview design, which can be divided into three main steps, each discussed in more detail in the rest of this section:

(1) **Capturing the object vibration.** First, RFVibe captures the mmWave reflections off the target object. It estimates the objects location in 3D space while eliminating the effect of other dynamic and static reflectors in the environment. The output of this stage is a time-domain RF signal that contains vibration and reflection characteristics of the object.
(2) **Extracting features**. RFVibe then extracts different features from the received mmWave signal that help identify the object and its material. Specifically, RFVibe introduces a Frequency, Power, and Damping feature that carry useful information for object classification.
(3) **Object and material identification**. In the final step, RFVibe feeds the features acquired from the received RF signal into a feed-forward neural network that classifies the material of the object and further identifies the object itself.

4.1 Capturing the Object Vibration

To capture the object vibration, RFVibe excites the object using sounds originating from a speaker. Different objects and materials exhibit different vibration frequencies, we need an acoustic source which excites objects consistently and equally across a range of frequencies, ensuring that the objects are excited in a reproducible manner that allow for their unique frequencies to be distinguishable. These excited frequencies must also be discernible by the mmWave transceiver. The first step is to determine the working frequency range of RFVibe.

Working frequency range. This choice depends on three factors; one relating to acoustic signal processing, one from a vibrometry perspective, and one concerning the RF signals.

(1) First, the speaker must be able to generate the frequencies that are present in the sound signal. This limits the sound source frequencies to the range of around 20 Hz to 20 KHz, the typical range for commercial speakers.
(2) Second, the working frequency range must have an adequate overlap with the objects natural frequency to make them distinguishable. Our experiments showed that working with the range above 20 Hz has sufficient overlap with objects' frequencies. Additionally, we observed that most objects have natural frequencies that fall in the range of below 150 Hz.
(3) Finally, the frequency range must be discoverable by the mmWave device. It is therefore infeasible to measure vibrations that are in the KHz range, as commercial mmWave devices cannot transmit chirps at such high frequencies. In RFVibe we transmit chirps at 250 Hz, which can identify vibration frequencies of up to 125 Hz.

Based on these insights, RFVibe operates in the frequency range of 20 to 125 Hz. The speaker must generates a sound that contains frequencies in that range. As a result, RFVibe uses a low-frequency speaker. Sec. 6.5 reviews the effect of using a smaller speaker with an non-ideal frequency range.

Sound source signal. The next step is to find a source signal that contains frequencies in the 20 to 125 Hz range. We considered several options shown in Fig. 2:
(1) *Filtered white noise.* One obvious option is to use white noise filtered to the desired range of 20 to 125 Hz.
(2) *Linear chirp.* Another option is to sweep frequencies starting from 20 Hz and move up to maximum of 125 Hz.
(3) *Pseudo-impulse train.* A third option is to sum phase-aligned sinusoidal signals ranging from 20 to 125 Hz.

While white noise and chirp sources both contain frequencies from 20 to 125 Hz, they suffer in two ways. First, they both have relatively lower peaks compared to the pseudo-impulse train, as their power is more evenly distributed over time. Second, they both lack intervals during which the signal amplitude is low, which allows the objects' vibrations to damp. We therefore use the pseudo-impulse train as the source that the speaker uses to generate the sound.

Audio Output. The corresponding power for every frequency a speaker emits is tied to the frequency response of that speaker, which is generally tuned to emphasize certain frequencies for humans. Thus, when using a commercial speaker, some frequencies will produce more pronounced movement. This may result in the speaker's most powerful frequencies overpowering the natural frequencies of the object. To correct this issue, we measure single sinusoidal tones

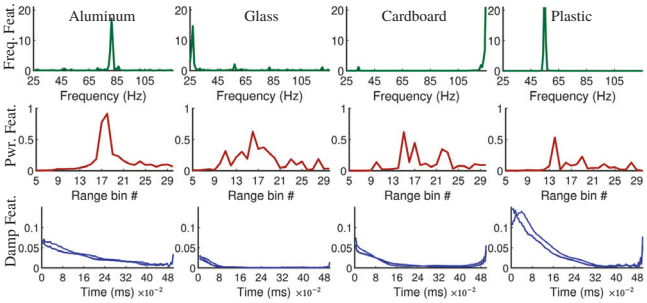


Figure 5: Examples of Frequency, Power and Damping Feature.

ranging from 20 to 125 Hz using a measurement microphone. The measured amplitude per frequency is used for a one time calibration of the speaker. We created an audio source by summing sine tones from 20 to 125 Hz and dividing each sine tone by the speakers measured amplitude for each frequency.

Experiment Setup. Described in more detail in Sec. 5, the speaker and radar are placed opposite each other with the object in the middle. This is to mitigate any vibrations seen at the radar and amplify the acoustic signals hitting the object. Since acoustic signals attenuate extremely quickly, placing the object within 0.15m ensured that the vibrations hit the objects. In addition, keeping the speaker more than 0.25m away from the mmWave device ensured that the radar picked up minimal vibrations from the speaker itself. This was validated by looking at range bins close to the radar and checking for any vibrations that were correlated to the speakers vibrations.

Object localization. As the sound wave hits the object and stimulates vibrations, the mmWave device transmits FMCW chirps that reflect back. However, the target is not the only object that reflects the mmWave signal. Therefore, we need to isolate the object in 3D to capture its vibrations. To achieve that, we first build on past work that takes the FFT of the received FMCW signal (*range FFT*) and uses range bins to isolate objects in terms of distance from the radar [2]. As mentioned in Sec. 5, each range bin is an interval of around 5cm. Within the range bin where the object is, we leverage the 2D antenna array of our mmWave device to beamform and zoom into the angles with the strongest reflection. Now, we can isolate the object of interest and analyze its vibrations.

4.2 Extracting Features

Do we need features? RFVibe extracts features from the reflected signal that help classify objects. A natural question is whether we need to extract features from the raw vibration signal, as using the raw signal and feeding it to a neural network might suffice. However, the useful information of this raw signal is hidden in several aspects of the signal, mixed with irrelevant information which might be difficult for a neural network to learn. As shown in Sec. 6.5, using the raw phase signal will perform worse than any of the individual features extracted in this section, and much worse than the combined system that RFVibe uses for object classification.

RFVibe uses three such features: Frequency Feature, Power Feature, and a Damping Feature, examples of which are shown in Fig. 5. The Frequency and Damping features are both extracted from the unwrapped phase of the mmWave signal (Sec. 3.1). The Power Feature is computed using the amplitude of the received mmWave signal. Below we explain each feature in detail.

1) Frequency Feature. We take the Short Time Fourier Transform (STFT) over the received phase samples. The STFT is more expressive over simply the FFT since the vibrations of the object are transient, as the impulse train momentarily excites the object to vibrate, and then waits for it to damp. Hence using the STFT better captures object vibrations as it is a sliding windowed FFT which gives both time domain and frequency domain information.

a) Removing ineffective frequencies. Frequencies of 0-20 Hz contain more noise and represent large movements rather than micrometer-level vibrations. They do not correspond to the range of frequencies generated by the sound source (Sec. 4.1). Therefore, they are removed from the STFT as they do not contain any information about the material itself. The resulting STFT heatmap is shown in Fig. 3, where we can see the natural frequency of around 80 Hz of an aluminum sheet that occurs with each impulse of sound source.

b) Removing time and keeping important frequencies. While the STFT performs well in capturing the resonance frequencies, we are not interested in the *time* when these occur. For example, if the STFT in Fig. 3 is shifted by a few windows, it does not affect the 80 Hz resonance frequency, while the STFT heatmap changes. We therefore need to remove the time dimension from the STFT and only keep the frequency information. To do so, we begin by normalizing each window of the resulting STFT to a standard normal distribution by subtracting the mean and dividing by the standard deviation:

$$\text{Frequency Feature}(j) = \sum_{i=1}^{\#\text{ windows}} \underset{j}{\text{normalize}}(\text{STFT}_{ij}). \quad (2)$$

The signal is summed along the time window axis, giving a frequency signature invariant to shifts in time. As seen in Fig. 3, the 80 Hz frequency of the aluminum sheet is more clear in the object frequency signature compared to the STFT heatmap. We also compare RFVibe's Frequency Feature to using the STFT in Sec. 6 and show that RFVibe outperforms using the STFT in material classification.

2) Power Feature. This feature is generated by taking the Range FFT of the raw data and multiplying the amplitude by the square distance each range bin corresponds to. To see why, note that the free space loss equation for RF signals is as follows $A = A_0 G_T G_R/(\lambda 4\pi d)$ where A_0 is the transmitters signal amplitude, G_T and G_R are the transmitter and receiver's gains, λ is the wavelength of the carrier frequency, and d is the distance from the transmitter to the receiver. Since RFVibe uses the reflected signal, and the transmitter and receiver are co-located, we account for the loss in amplitude by a factor of d^{-1} once on the outgoing and once again on the incoming path resulting in an amplitude loss proportional to d^{-2}. Once we compensate for the loss, we select the amplitude of a few bins surrounding the object location and use them as the Power Feature.

3) Damping Feature. This feature shows the damping of an object's resonance frequencies. The process of obtaining this feature is depicted in Fig. 4.

a) Extracting the unwrapped phase. The first step is to compute the unwrapped phase of the received mmWave phase signal. This step was explained in detail in Sec. 4.1.

b) Applying bandpass filter. To better capture the damping of the object and throw out irrelevant information, a bandpass filter is applied to the signal in the previous step. The passband cutoffs are determined by selecting the maximum frequency after applying Welch's method [46] to the vibration signal. Once the peak is selected, the vibration data is filtered with a butter bandpass filter around that peak frequency ± 10 Hz.

c) Computing the envelope of the signal. We next compute the envelope of the filtered phase by taking the absolute value of the Hilbert Transform.

d) Aligning the peaks. Finally, we need to make sure our feature is consistent and resistant to time shifts. Note that for each impulse that the speaker emits, there is a corresponding rise and fall for the object vibration as shown in Fig. 4(c). To create a consistent feature, we search for the peaks corresponding to each rise and fall in the object's enveloped, filtered phase. As we have approximately 250 samples between consecutive impulses, we search for the maximum within each 250 points of the enveloped phase. For each maximum, we take the next 125 samples which approximately includes the entire time the object loses energy before the next impulse starts. RFVibe uses these curves as its final Damping Feature.

4.3 Object and Material Identification

The final component of RFVibe uses a neural network that takes in the extracted features from the previous section and outputs one of the N possible classes. In general, the network can classify between any set of categories; for instance:

- $N = 2$, can classify between two classes of metal and non-metal objects as we see in Sec. 6.
- $N = 7$, can classify between seven classes of different materials (metal, wood, plastic, etc.) as depicted in Fig. 9(a,c).
- $N = 23$, classifies between objects, depicted in Fig. 9(b,d).

At a high level, all of RFVibe's neural network architectures are composed of three stages as shown in Fig. 6. First, a set of feature heads that convert an input feature to an intermediate feature map. Second, the feature maps are separately used to classify between N classes. Finally, an aggregation head combines all intermediate feature maps to perform classification. The loss of RFVibe's network is computed based on both separate and aggregate classifications. We now explain RFVibe's architecture in more detail.

1) Feature heads. RFVibe uses feature heads to bring the input features from different sources into a common latent space. Each feature head comprises a feed forwards neural network that stacks convolutional or fully connected layers. The goal of having feature heads is to bring features from different sources into a common space where intermediate features can be summed or concatenated. For example, in RFVibe, the output of all feature heads are of size $n \times 128$ for some n depending on the input size. Fig. 6 shows the three features heads for the features described in Sec. 4.2:

(1) The Frequency Feature head takes a 1×286 input feature as described in Sec. 4.2 and outputs a 8×128 intermediate feature map through 8 convolutional layers.

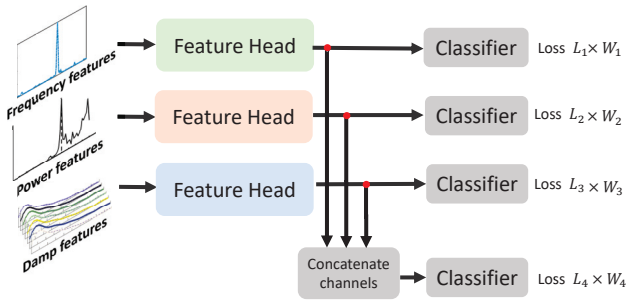


Figure 6: RFVibe's Neural Network.

(2) The Power Feature head takes in a vector of length 25 and outputs a 1×128 feature map. It consists of two stacked fully connected layers with maxpool and dropout.

(3) The Damping Feature follows the Frequency Feature head architecture but takes inputs of size 8×125 and outputs 8×128-sized intermediate feature maps.

2) Classification heads. The second component of RFVibe's network are the classification heads, that take in intermediate feature maps and output one of the N possible classes. Shown in Fig. 6, all final classification heads follow the same stack of fully connected, dropout and maxpooling architecture.

3) Aggregation head. The final component combines the intermediate feature maps from all sources by summing each set of intermediate features across its channels, and then concatenating the resulting feature maps. The result is passed through a classification head to give the aggregate classification results.

RFVibe uses the prediction from the aggregate classification head as its final output. In particular, we do not use the results of individual classification heads as final predictions since they would not express the intersection of all the features. Instead they are used to modify the loss function in order to improve the classification accuracy of the final model, as shown in Sec. 6. The final loss function is a weighted linear combination of the losses of different classification heads ($W_1, \cdots W_4$ in Fig. 6). This enforces the network to incorporate each of the feature heads and ensure that the final performance of the network is trained for. As shown in Sec. 6.5, the weighted combination outperforms a single aggregate classification head.

5 IMPLEMENTATION

Hardware Setup. RFVibe uses TI's IWR1843BOOST evaluation module with the DCA1000EVM. Two transmitters and four receivers with a two-dimensional virtual array were used to capture data at 77 GHz. The FMCW chirp was designed to have a frame rate of 250Hz. The chirp parameters created a range resolution of 5.63 centimeters, a maximum range of 3.24 meters, and had a sweep bandwidth of 2.6 GHz. Received data was sent to a host PC using mmWave Studio. A Klipsch SUB100NA subwoofer was used to produce acoustic signals in the lower frequency range with a frequency response of 32 Hz to 120 Hz ± 3dB.

Software Setup. We modified the OpenRadar GitHub [28] to connect the mmWave Studio and scripting for our experiment setup. We

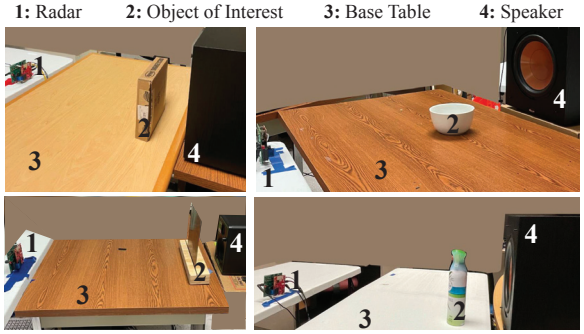


Figure 7: Experiment setups.

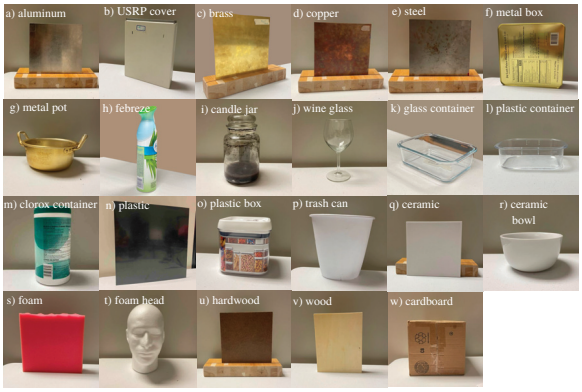


Figure 8: Experiment Objects.

calibrated the raw data in MATLAB, and used Python to extract the features used.

Speaker and Sound Specifications. The sound used for experiments was a summation of sinusoidal tones from 20 Hz to 125 Hz, with a step size of 1 Hz. The amplitude of each sinusoid was compensated by the amplitude of the speakers output measured with a microphone for a one time calibration process. The sound had an average power of 58 dB, an intensity similar to normal conversation. The speaker was selected based on its ability to clearly produce low frequency tones within the 20-125 Hz frequency range, which is set by the mmWave radar's sampling frequency. These low frequency speakers, or subwoofers, tend to be larger in size because the more surface area allows for a less distorted output sound. Within the selection of commercial speakers, it is rare to find speakers with a frequency range below 30dB. However, when measuring the power of 20-125 Hz single tones produced by the Klipsch speaker, the power for lower frequencies was adequate for our evaluation.

Neural Network Parameters. RFVibe is designed with PyTorch. It uses a Adam optimizer with learning rate 0.001. The loss is calculated using cross-entropy loss. The batch size is 32, and the dropout probability is 0.25. After being fed to RFVibe all three features are resized to size 128 for each of the intermediate layers. The final loss weights are set to $(W_1, W_2, W_3, W4) = (0.9, 0.3, 0.3, 1)$.

Experiment Environment. The data collection setup consisted of three rooms (various parts of each room were used), shown in Fig. 7. This way, the dataset was diversified with different multipath profiles for the acoustic and the RF signals. Objects were placed on different surfaces: cardboard box, and tables made of plastic, metal, and wood. Each experiment ran for 13 seconds, during 10 seconds of which the speaker was playing the sound source. The object was positioned 0.5-2m from the radar, and placed between the radar and speaker, close to the line connecting the speaker with the radar.

Experiments were performed in an imprecise manner to mimic everyday use. This meant objects were not placed in an exact location or orientation to the audio source or the radar. The surroundings of the experiments had various amounts of background noise and movement. In addition, the volume of the sound source varied greatly, which affects the visibility of certain frequencies, which made sure that the network did not overfit to a specific volume.

Dataset. The dataset includes 23 different objects of various shapes and sizes consisting of the 7 materials, shown in Fig. 8. The dataset was split into training, validation and test sets. The validation set was approximately 12% of the training set size, and the test set was approximately 15% of the training set size. Experiments were split into three datasets such that they had no overlap in experiment runs. Between train, validation and test sets, there are no experiments done in the exact same environment. In other words, for any two experiments were run in the same batch, they were used both for training, or both for testing. An experiment run includes a set of 10-20 experiments per object where, the physical setup did not change. While the power feature for these experiments remains relatively similar between the 10-20 experiments, the objects vibration features still fluctuates. After each run, the radar was turned off and the physical set up was adjusted. For the results in Sec. 6.2, 6.3 and 6.4 the dataset we used has 12154, 1437, and 1795 experiments for the training, validation, and test split respectively. In addition, we limit the samples of each material in all three datasets so that all materials have equal representation in the datasets.

6 RESULTS

6.1 Evaluation Metrics

We evaluate our system based on accuracy which signifies the overall percentage of how often RFVibe correctly classifies between candidate classes.

Train and test repetitions. For more confidence in our metrics, we train and test the network 10 times with different random seeds to get an average accuracy over different initializations and subsets of the train, validation and test dataset.

In the rest of this section, we will first present the main classification results and compare our results to the current baseline for material classification [21, 47]. Then, to better understand the role of each component of RFVibe, we perform microbenchmarks on different aspects of our system.

6.2 Material Classification

Experiment. We first present the overall performance of RFVibe in classifying between different materials. We evaluated RFVibe's performance among 7 classes of materials: Cardboard, Metal, Glass, Ceramic, Plastic, Foam and Wood. A total of 23 objects are tested that belong to one of the 7 classes such that each class includes multiple objects. We use the network described in Sec. 4.3 with $N = 7$. As outlined in Sec. 5, no setup was used both in train and test datasets throughout the experiments. We further report results

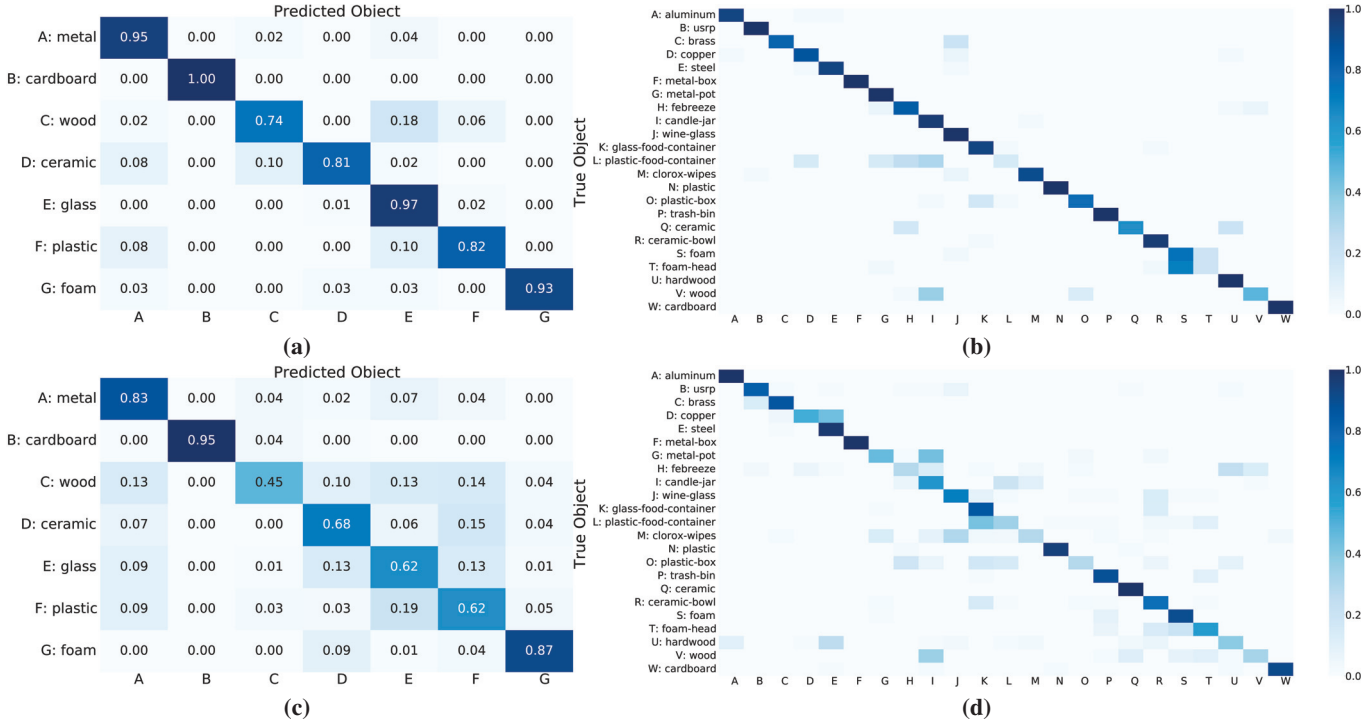


Figure 9: Confusion matrices. **(a)** material, similar setups. **(b)** object, similar setups. **(c)** material, different setups. **(d)** object, different setups.

for another train and test split, where the setups in the train and test datasets were drastically different.

Results. The overall accuracy of RFVibe in classifying between the seven classes was 81.3%, with the detailed accuracy breakdown shown in Fig. 9(a). As seen from the matrix, RFVibe has a high accuracy in classifying metal, cardboard, ceramic, glass, plastic and foam (> 80%). In addition, when generalizing to new, unseen setups, RFVibe achieves 73.2% accuracy (see Fig. 9(c)) showing its ability to generalize to new environments. It shows high accuracy for metal, cardboard, and foam, while maintaining a moderate accuracy for ceramic, glass, and plastic.

Analysis. Overall, RFVibe accurately classifies all materials. By relying on features that depend heavily on the material type as described in Sec. 4.2, RFVibe can generalize to new environments and unseen setups.

We attribute the higher performance of RFVibe in classifying metals to them having distinct resonance frequencies. Objects made of cardboard and foam also demonstrate clear vibrational signals as they are usually lightweight, resulting in high classification accuracy. The drop in accuracy around wood is primarily due to the fact that strong vibrations were not visible with the acoustic source. Even with contact-based methods such as stimulation by means of tapping, wooden objects did not demonstrate strong vibrations, resulting in poor frequency-based and temporal features as discussed in Sec. 4.2. As a result, the classification network has to rely mostly on the power-based feature, bringing the classification accuracy of wood down to 45%. Finally, RFVibe tends to confuse glass and ceramic. This could be because ceramic and glass have similar properties, being rigid structures.

6.3 Object Classification

Experiment. We evaluated RFVibe's classification performance among 23 classes of different objects, where we ran our network in Sec. 4.3 with $N = 23$. Other conditions of the experiment were similar to those described in Sec. 6.2.

Results and Analysis. RFVibe demonstrates an overall accuracy of 74.4% for object classification. The confusion matrix among all 23 objects is shown in Fig. 9(b). In our results, 10 of the objects have more than 90% average accuracy and 6 objects have over 70% average accuracy. Shown in Fig. 9(d), RFVibe generalizes to new, unseen setups and achieves 63.4% accuracy where 11 of the objects have more than 80% average accuracy and 16 objects have over 50% average accuracy While we do lose accuracy when specifying objects compared to materials, it gives a good stepping stone towards more specific objects recognition and potentially geometry classification. Similarly to material classification, higher accuracy is seen within metals, which not only have distinctive frequencies when compared to other materials, but also exhibit unique frequencies within themselves. In fact, when RFVibe is trained to classify between four metals, RFVibe attains 83% accuracy. Another visible trend is that objects with larger surface areas produce vibrations that are much easier for the radar to detect. While RFVibe accounts for this in the design of features, there are some limitations in the radar's angular resolution which makes this task much harder.

6.4 Comparison to Baselines

Baseline Implementation. We compare our results to mSense [47] and He et al [21] that introduce systems for material classification based on power of reflected signals. We implemented both systems

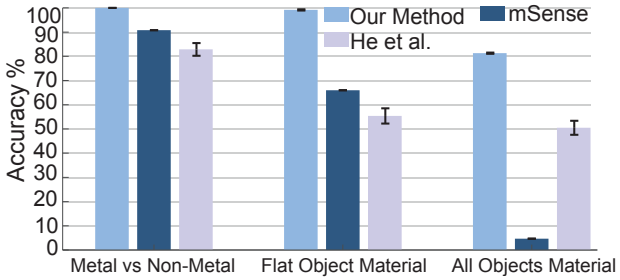


Figure 10: RFVibe vs Baselines mSense [47] and He et al [21].

using the TI IWR1843BOOST evaluation module connected to a DCA1000EVM. For the mSense implementation, while we replicate their method and algorithms, our hardware implementation differs in the following two ways:
1) Antenna diversity: While mSense uses a virtual antenna array of 32×32, giving a total number of 1024 antenna combinations, our implementation only uses 2 Tx and 4 Rx, resulting in only 8 combinations. This means our implementation has much lower antenna diversity yielding poorer performance.
2) Object sizes used in the experiments: In mSense, the experiment objects have surfaces facing the mmWave device that are larger than those used in our implementation. This means that our implementation has a lower reflection power due to smaller reflection surface which results in features being less distinguishable in our implementation.

Experiment. To better compare with the baselines, we test all three methods on three separate classification tasks: 1) Classifying between only two classes: metal vs non-metal. The purpose of this is to show how well each method works when there are only two classes, and when there is no need for generalizability. This scenario is ideal for the power-based methods, as metals generally demonstrate much stronger reflections than non-metals. 2) Two baselines are designed for objects which have a flat surface, thus we evaluate all three systems on 7 materials consisting of flat objects only. This task was designed to show how the methods perform when the number of classes increases, making it more difficult to classify solely based on signal power. 3) Finally, we compare the main results for RFVibe with the two baselines. This includes all objects for material classification, including curved objects. For all three classification tasks, we provide results for train and test datasets that contain experiments from similar setups, such as those in manufacturing settings where the environment and conditions remain relatively constant.

Results and Analysis. As seen in Fig. 10, RFVibe performs at 100% accuracy for binary classification between metal vs non-metal and 99.2% for objects which have a flat surface area for material classification. In comparison, while mSense performs relatively well for binary classification at 90.8%, its performance drops to 66.0% when classifying between flat materials. This is expected since mSense relies on a scalar feature (namely normalized power) to classify materials. Finally, the network of [21] achieves an average accuracy of 82.9% for metal vs non-metal classification and drops significantly to 55.3% for the material-wise classification.

In the last classification task, RFVibe achieves 81.3% average accuracy. mSense's accuracy drastically drops to 5.1% with the introduction of curved objects, since their feature does not take into account the size and curvature of an object, all of which affect the reflected signal power. The system from [21] achieves 50.5% accuracy for material classification with similar train and test environments and 45% accuracy for different train and test environments. The system from [21] does better due to use of all the channels from the transmitter-receiver pairs and using a range of 7 bins surrounding the object's range bin location for the features.

Features Used	Acc (%)	STD
Freq., Power, Damp	73.1	1.2
STFT, Power, Damp	54.8	2.4
Raw	24.7	3.0

Features Used	Acc (%)	STD
Frequency	47.3	1.6
Power	34.5	2.4
Damp	32.5	2.2

(a) Using different features

Loss Weights	Acc (%)	STD
0.9,0.3,0.3,1	73.1	1.2
0,0,0,1	68.0	1.8
1,1,1,1	67.1	2.7

(b) Using different loss weights.

Classifer Used	Acc (%)	STD
RFVibe Network	73.1	1.2
SVM	58.8	-
Random Forest	32.95	0.8

(c) Results of various classifiers.

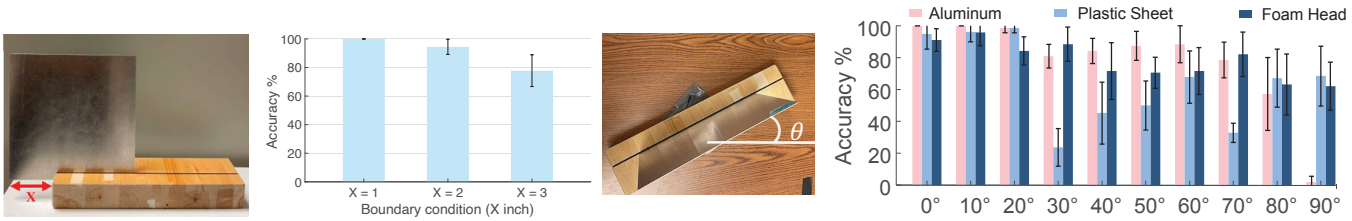
Table 1: Feature and network microbenchmark results.

6.5 Microbenchmarks

1) Impact of Individual Features on the Final Performance. It is important to see how useful each of the three features are to the performance of RFVibe. We train and test each of the features separately to perform classification the 23 objects. In each experiment, to remove the other features, we trained the single branch for that feature and removed the rest of the network as well as the aggregate head. The results are shown in Tab. 1(a). RFVibe outperforms each individual feature by a significant margin, by around 26-40%. The frequency feature performs the best among the features at 47.3% accuracy, showing the effectiveness of this feature. The power feature achieves 34.5% accuracy, primarily due to object's size and its orientation with respect to radar making a significant difference in the reflected power. In addition, when the network was trained with all three features, but sound absent from the experiments, we achieve 35.9% accuracy, showings it's similarity to training the network with only the power feature. The temporal feature also provides relevant information achieving 32.0% accuracy. This experiment shows that while all three features contribute to the accuracy of RFVibe, the frequency feature plays a bigger role in material classification, as reflected by the fine-tuned weights of the network branches reported in Sec. 5.

We also compare these experiments with an experiment where we use the raw phase signal (Sec. 4.1) as the only feature in training. The intuition is, since the frequency and temporal features are based on the raw phase signal, it might suffice to use the raw phase alone to classify materials. However, as the results in the table show, this leads to the worst overall performance of only 25%. This implies that the raw phase signal on its own is not a very useful feature.

2) Frequency Feature vs STFT. One question raised in Sec. 4.2 was whether the frequency feature more clearly represents a material, or whether and STFT would adequately characterize the vibration frequencies. We compare by substituting the STFT rather than using the frequency feature and adjusted the network architecture accordingly, displayed in Tab. 1(a). This change resulted in a drop of close to 8%



(a) X:unconstrained edge. **(b)** Generalizing unseen X. **(c)** θ: angle rotation. **(d)** Generalizing to unseen rotation angles θ.

Figure 11: Microbenchmarks for generalizing to boundary condition, rotation and affect of using different amount of time domain samples.

in overall accuracy, which demonstrates that while the STFT shows both temporal and time domain information, the frequency feature better emphasizes the important frequencies, leaving the temporal information for the damping feature to capture.

3) Length of Vibration Time Used. RFVibe uses 8 seconds of data to classify the materials, however, here we show that using just 3 seconds of data is enough. We train and test the the network using 650, 1150, 1650, 2150 time domain samples for the features. The accuracy across all 4 time lengths was 69-73% with standard deviation of 2%, showing that even with 3 seconds of data, RFVibe does not lose much in accuracy.

4) Comparison with Basic Classifiers. To test whether training simpler classifiers would suffice to train with RFVibe's features, we train an SVM and a Random Forest Classifier. Scikit-learn was used for both the SVM and Random Forest Classifier and the parameters were tuned to the optimal numbers. Tab. 1(c) shows that both the SVM and Random Forest Classifiers have much worse classification accuracies compared to RFVibe. While the SVM is closer in accuracy at 61.4%, the Random Forest Classifier is drastically worse at 29.7%. This result shows that the features of RFVibe require a more complicated network to characterize the signal.

5) Binary Classification: Metal vs Non-Metal, Flat vs Curved. Binary classification gives us insight into how useful RFVibe is when simply separating between two classes for the object properties. Even the simple task of classifying between two classes, such as metal vs. non-metal can be extremely important when determining material properties. We trained our network with simple binary labels for curved versus flat objects and again for metal versus non-metal objects. Fig. 10 showed that RFVibe can separate metal and non-metal objects with 92.8% accuracy irrespective of shape. This shows that our features contain useful information regarding the material of an object that are independent of its shape. RFVibe can also RFVibe can separate flat and curved objects with 96.9% accuracy irrespective of material. These results imply that vibration properties of different objects retain a great deal of information regarding the geometry of the object.

6) Impact of Loss Weights. The loss weights, described in Section 4.3, greatly impact the performance of RFVibe. Varying the loss weights for each branch of the network indicates which features to put more weight towards when calculating the loss. Loss weights of [0.9, 0.3, 0.3, 1] correspond to a weight of 0.9 for the Frequency Feature, 0.3 for the Power Feature, 0.3 for the Damping Feature and 1 for the overall network. Shown in Tab. 1(b), using these loss weights, RFVibe has an accuracy of 73%. As compared to other selections of weights: [0, 0, 0, 1] gives an accuracy of 68% and [1, 1, 1, 1] gives an accuracy of 67.1%. This shows that weighing the loss more towards the better performing features and the overall network tends to give better accuracy.

7) Same Shape, Different Material. Material classification between different 8×8 inch square sheets of metal allow us to determine RFVibe's performance when the same shape is used but the material is varied. The metals were placed in a wooden block with a slit so the metals could face the radar. Material classification for these four metals gave an overall average accuracy of 83.2% shown in Fig. 13(d). It should be noted that brass is an alloy, usually made of around 60% copper which explains the confusion between brass and copper when classifying.

8) Effect of Weight. Given that the vibration properties of an object are influenced by its weight, understanding the resulting impact on RFVibe's classification accuracy is of interest. To this end, we look at the trend of accuracy vs weight of objects within the range of 80g to 600g for object wise classification of similar environment. The results are reported in Fig. 13(a) showing in fact, that lower-weight items have a slight drop in accuracy as the movement on the object is heavily correlated with the speaker's sound profile. However, within the range of 200g to 600g we retain 77% accuracy for the objects classification.

9) Background Noise. Another question that arises is whether audible background noise will affect RFVibe's accuracy. Thus, we study the impact of how background noise affects the classification of 4 metal sheets. In these experiments, we run experiments as described before, however, we are producing a large background noise using speakers throughout the room. The results are tested against a network that has not seen any experiments with large background noise. In fact, we see that there is almost no difference when testing these experiments and see 82.3% accuracy when classifying between these objects as compared to the prior 83.2%. This indicates that the excitation of the object requires a considerable intensity of acoustic

(a) Speakers used.

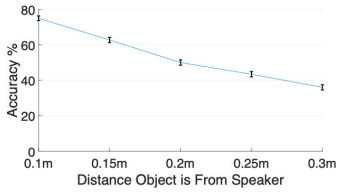
Speaker	Acc (%)	STD
Klipsch	83.2	3.3
CR-X	79.6	6.0

(b) Acc vs speakers.

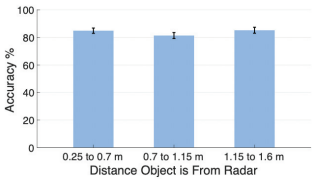
Figure 12: Different speakers used in experiments.

Object Weights	Acc (%)	STD
10g to 200g	67.0	5.1
200g to 400g	77.6	6.64
400g +	77.5	6.4

(a) Different weighted objects.



(b) Unseen distance from speaker.



(c) Unseen distance from radar.

Metal Name	Acc (%)	STD
Aluminum	98.7	5.6
Brass	68.4	7.1
Copper	58.6	10.1
Steel	98.0	0.8

(d) Same shape metals.

Figure 13: Different weighted object results, generalizing to unseen distances and same shape, different material classification.

waves. In addition, the distance between the object and the background audio is much larger, thus resulting in a rapid attenuation in the background noise's resulting power.

6.6 Generalizability of RFVibe

Here we study aspects of RFVibe's generalizability.

1) Generalizing to unseen boundary conditions. The boundary conditions of an object also affect its visible frequencies. Here we demonstrate the impact of training the metal sheets on a fully constrained wooden base and testing RFVibe on experiments where we incrementally move the metal sheet out of the wooden base. This incremental movement results in changes in the boundary condition of the object, as less of the base is constrained to the wooden base. Fig. 11(b) demonstrates the results for $X = 1$, $X = 2$, $X = 3$ when 1 inch, 2 inches, and 3 inches of the base of the metal sheet are unconstrained. Once 4 inches of the metal sheet, or more than half of the metal is unconstrained, the accuracy drops to almost 0%. This shows that up to a certain degree, RFVibe is resilient against boundary condition variations.

2) Generalizing to unseen rotation. It is possible the object cannot be exactly perpendicular to the radar in practice. Here, we trained on the sheet metals that are within ±1° perpendicular to the RF signals and tested on an aluminum sheet, thick plastic sheet and foam head with angles varying from 10° to 90° rotation, with 0° denoting perpendicular to the radar. For the aluminum sheet, Fig. 11(d) shows accuracy is above 90% for up to 20° rotation. After 30° rotation, the accuracy becomes more unreliable which is shown by the standard deviations increasing. This is because the reflection from the metal sheet is weak caused by the rotation away from the radar and thus the vibration signals measured are inconsistent. A similar trend appeared for plastic where the initial rotation of up to 20° retained above 90% accuracy and after 30° rotation the accuracy drops to around 60% with a larger standard deviation. The lower accuracy for plastic as compared to the aluminum sheet is attributed to the much weaker power of the reflected signal which also degrades as the object rotated. However, with the plastic, the accuracy remained around 40% even as the object was rotated to 90°. This was because the plastic is not as flat as the aluminum sheet allowing for some minor reflections even at 90°. A similar trend is seen with the foam head, however, since this object is not flat, the average accuracy remains above 70% till 80° rotation where it dips to around 60%. The more stable accuracy is due in part to the foam head having enough surface area for the RF signal to reflect back to the radar. However, the standard deviation of the accuracy still increases as the object is rotated because of the difference in vibration characteristics due to the rotation.

3) Generalizing to unseen distance of object to speaker. We analyze the performance of RFVibe against different distances of the object to the speaker. There is a trade-off between the proximity of the object to the speaker because a closer distance allows for stronger vibrations induced on the object, however, there is more leakage from the speakers vibrations when the range bins are so close together. Here we train on a set of experiments done on a sheet of metal, that is 10cm from the speaker, and then test RFVibe on experiments done farther from the speaker. Fig. 13(b) shows the trend of the accuracy decreasing as the object is moved further from the speaker. This is because the acoustic waves can no longer excite the object as effectively as the distance is increased.

4) Generalizing to unseen distance of object to radar. Another variable that is important to material sensing systems is a distance agnostic system. Classifying a material incorrectly because of the variability in regards to distance is a common issue with power based methods because of the loss in power as objects move further from the radar. In this result, we train RFVibe on the 7 materials (23 objects) that are 0.5m away and test on experiments done 0.6m, 1m, 1.5m away. Fig. 13(c) shows the results. RFVibe has relatively the same accuracy for all distances, showing that the vibration characteristics of the materials do not change as the distance is varied.

5) Generalizing to obstructed objects. There is the question as to how different methods of disguising the objects will affect RFVibe's performance. For vision-based material classification methods, not having sight of the material itself would completely trick the classifier. We show that RFVibe performs well under three circumstances of disguising the objects. Experiments were done on the metals sheets. First, different textures were printed on paper and taped to the metals. These textures included images of various other materials that could be used to trick a material classifier, such as putting an image of wood on metal, or putting an image of brass on a sheet of steel. Second, we placed the objects into a paper bag which obscured the object from the radar. Finally, the metals were painted, thus hiding their material. All three of these methods can cause the vibration frequencies to shift, however, we show that RFVibe is resilient to these minor impacts. Examples of the occlusion method is shown in Fig. 14(c). These three different experiments were all tested on the same network which was only trained with metals that had nothing on or in front of them. As shown in Fig. 14(b), we only see a slight drop in accuracy of RFVibe which is due to the change in vibrations because of the paper, bag, or paint.

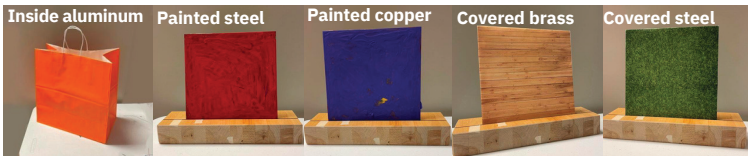
6) Generalizing to unseen objects. It is important to be able to classify the materials of objects we have not seen before, since it is not possible to train RFVibe on every possible object it might encounter. We take a few objects out of the train set when doing

Unseen Object	Acc (%)	STD
Steel	99.9	0.01
Plastic Sheet	90.5	13.3
Metal Box	80.7	10.3
Plastic Box	53.5	6.0
Foam Head	47.0	9.5

(a) Unseen objects.

Hidden Method	Acc (%)	STD
Uncovered	83.2	3.3
Paper Cover	70.0	4.4
In Bag	73.0	4.8
Painted	69.8	3.6

(b) Occluded objects.



(c) Occluded metal sheets (inside bag, painted, and covered).

Figure 14: Occluded and new objects.

material classification and test on these objects. Results in Fig. 14(a) show that large flat objects such as a steel sheet, sheet of plastic or a metal box are classified, they perform extremely well achieving 80-100% accuracy. However, for objects with smaller reflections such as a foam head and plastic box, they only achieve 47% and 53.5% respectively. While this is not as accurate as the other unseen objects classification, it still shows that RFVibe can generalize to unseen objects for material classification.

7) Generalizing to different speakers. A smaller speaker is another consideration when designing RFVibe. We evaluate RFVibe's system replacing the Klipsch Subwoofer with the CR-X series speaker shown in Fig. 12(a), which has a 3dB frequency range of 80-20 KHz. A different sound was used since the impulse train resulted in much power lower for the CR-X speaker, including frequencies in the 3dB range. Instead, a logarithmic chirp was used in order to get the most power for each frequency. As shown from in the table of Fig. 12(b) there is a small drop in accuracy, around 3.5%. However, this comes at the cost of a much louder sound because of the smaller speakers weaker frequency response for lower frequencies.

7 DISCUSSION LIMITATIONS

Several limitations of RFVibe are worth noting:

• **Multiple Objects**. Currently, RFVibe assumes that only one object is in view of the radar and it is directly in front of the radar. However, future work can look into differentiating between multiple objects by beamforming and differentiating the vibrations seen at different angles to separate multiple objects or objects made of multiple materials.

• **Thick Occlusions**. While RFVibe generalizes well for thin occlusions and painted-over objects, thick occlusions between the radar and object as well as occlusions between the object and the speaker pose another challenge. The attenuation resulting from the signal passing through the occlusion weakens the vibration signals from the object and RFVibe might instead pick up the vibration from the occlusion.

• **Location of the Speaker**. In our experiments, the radar and speaker were on opposite sides of the object since the radar's vibrations would be more pronounced because of the proximity of the two. In addition, the acoustic signals drastically lose power due to being farther away from the object as shown in Sec. 6.5. In order to create a more mobile friendly setup, this calls for future work in a robust algorithm to decouple the radars vibrations from the speaker. Another possible solution would be to use an array of speakers in order to focus the signal towards the object.

• **Object Weight**. The weight of the object is another factor that affects vibrations. In this paper, we primarily focus on objects that can be handheld, and thus tend to be lighter and easier to vibrate. However, it is import to note that there are scenarios where purely using an acoustic source to vibration the object would be unsuccessful due to the extreme weight and rigidity of the objects. In this case, the network would be forced to classify purely based on power which seen in Sec. 6.5 is not as robust as using vibration-based features. In addition, in certain applications, it would make sense to use a touch based vibration transducer which would more easily vibrate larger and heavier objects.

• **Environmental Factors**. RFVibe shows a level of robustness against boundary condition variations, however, drastic changes would completely alter the vibration frequencies. This challenge is not unique to RFVibe's wireless method would also hold for contact-based techniques. Other environmental factors, such as the surfaces that the radar and object are placed on can affect the vibration characteristics. A variety of different surfaces were included for both the object and radar to be placed on, however, adding more experiments with objects on different surfaces would help with the bias towards the environment.

• **Liquids**. While RFVibe is shown with solid objects, the system could be extended to liquids. As liquids have also been studied to exhibit vibration characteristics [23].

8 CONCLUSION

In this paper, we show the feasibility of leveraging micrometer level vibrations for an additional layer of information when determining material and object properties. By placing an object in the presence of an audio source, we can extract the vibrational modes and in addition to the other reflected signals, accurately classify materials and even further objects. We demonstrate using a 77-GHz mmWave radar how we classify 7 different materials and investigate multiple variables that affect reflection-based material sensing.

ACKNOWLEDGEMENTS

We would like to thank the reviewers and shepherd for their valuable feedback. We would also like to thank Ali Kanj for the useful background knowledge on vibration testing and vibrational modes. This research is funded in part by NSF award number 1750725.

APPENDIX

The research artifact accompanying this paper is available via https://doi.org/10.5281/zenodo.7932020 and https://doi.org/10.5281/zenodo.7913318.

REFERENCES

[1] N. Adair. Radio frequency identification (rfid) power budgets for packaging applications. *PGK-491*, pages 2–11, 2005.

[2] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)*, 34(6):1–13, 2015.

[3] E. M. Amin, R. Bhattacharyya, S. Kumar, S. Sarma, and N. C. Karmakar. Towards low-cost resolution optimized passive uhf rfid light sensing. In *WAMICON 2014*, pages 1–6. IEEE, 2014.

[4] V. Aranchuk, A. K. Lal, C. F. Hess, and J. M. Sabatier. Multi-beam laser doppler vibrometer for landmine detection. *Optical Engineering*, 45(10):104302, 2006.

[5] O. Buyukozturk, J. G. Chen, N. Wadhwa, A. Davis, F. Durand, and W. T. Freeman. Smaller than the eye can see: Vibration analysis with video cameras. In *World Conference on Non-Destructive Testing 2016*, 2016.

[6] O. Büyüköztürk, R. Haupt, C. Tuakta, and J. Chen. Remote detection of debonding in frp-strengthened concrete structures using acoustic-laser technique. In *Nondestructive Testing of Materials and Structures*, pages 19–24. Springer, 2013.

[7] P. Castellini, N. Paone, and E. P. Tomasini. The laser doppler vibrometer as an instrument for nonintrusive diagnostic of works of art: application to fresco paintings. *Optics and Lasers in Engineering*, 25(4-5):227–246, 1996.

[8] J. G. Chen, A. Davis, N. Wadhwa, F. Durand, W. T. Freeman, and O. Büyüköztürk. Video camera–based vibration measurement for civil infrastructure applications. *Journal of Infrastructure Systems*, 23(3):B4016013, 2017.

[9] J. G. Chen, R. W. Haupt, and O. Buyukozturk. The acoustic-laser vibrometry technique for the noncontact detection of discontinuities in fiber reinforced polymer-retrofitted concrete. *Materials evaluation*, 72(10), 2014.

[10] L. Collini, R. Garziera, and F. Mangiavacca. Development, experimental validation and tuning of a contact-less technique for the health monitoring of antique frescoes. *NDT & E International*, 44(2):152–157, 2011.

[11] A. Davis, M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, and W. T. Freeman. The visual microphone: Passive recovery of sound from video. 2014.

[12] A. Dhekne, M. Gowda, Y. Zhao, H. Hassanieh, and R. R. Choudhury. Liquid: A wireless liquid identifier. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '18, page 442–454, New York, NY, USA, 2018. Association for Computing Machinery.

[13] J. Ding and R. Chandra. Towards low cost soil sensing using wi-fi. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–16, 2019.

[14] T. Emge and O. Buyukozturk. Remote nondestructive testing of composite-steel interface by acoustic laser vibrometry. *Materials evaluation*, 70(12), 2012.

[15] C. Feng, J. Xiong, L. Chang, J. Wang, X. Chen, D. Fang, and Z. Tang. Wimi: Target material identification with commodity wi-fi devices. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pages 700–710, 2019.

[16] A. Gadre, D. Vasisht, N. Raghuvanshi, B. Priyantha, M. Kotaru, S. Kumar, and R. Chandra. Milton: Sensing product integrity without opening the box using non-invasive acoustic vibrometry. pages 390–402, 2022.

[17] J. D. Griffin, G. D. Durgin, A. Haldi, and B. Kippelen. Radio link budgets for 915 mhz rfid antennas placed on various objects. In *Texas Wireless Symposium*, volume 44, 2005.

[18] J. Guo, T. Wang, Y. He, M. Jin, C. Jiang, and Y. Liu. Twinleak: Rfid-based liquid leakage detection in industrial environments. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 883–891. IEEE, 2019.

[19] U. Ha, J. Leng, A. Khaddaj, and F. Adib. Food and liquid sensing in practical environments using rfids. In *Proceedings of the 17th Usenix Conference on Networked Systems Design and Implementation*, NSDI'20, page 1083–1100, USA, 2020. USENIX Association.

[20] R. W. Haupt and K. D. Rolt. Standoff acoustic laser technique to locate buried land mines. *Lincoln laboratory journal*, 15(1):3–22, 2005.

[21] S. He, Y. Qian, H. Zhang, G. Zhang, M. Xu, L. Fu, X. Cheng, H. Wang, and P. Hu. Accurate contact-free material recognition with millimeter wave and machine learning. In L. Wang, M. Segal, J. Chen, and T. Qiu, editors, *Wireless Algorithms, Systems, and Applications*, pages 609–620, Cham, 2022. Springer Nature Switzerland.

[22] A. Hind. Agilent 101: An introduction to optical spectroscopy, 2013.

[23] Y. Huang, K. Chen, Y. Huang, L. Wang, and K. Wu. Vi-liquid: unknown liquid identification with your smartphone vibration. In *MobiCom*, pages 174–187, 2021.

[24] J. F. James, R. S. Sternberg, and S. A. Rice. The design of optical spectrometers. *Physics Today*, 23(12):55, 1970.

[25] C. Jiang, J. Guo, Y. He, M. Jin, S. Li, and Y. Liu. Mmvib: Micrometer-level vibration measurement with mmwave radar. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, MobiCom '20, New York, NY, USA, 2020. Association for Computing Machinery.

[26] Y. Liang, A. Zhou, H. Zhang, X. Wen, and H. Ma. Fg-liquid: A contact-less fine-grained liquid identifier by pushing the limits of millimeter-wave sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3):1–27, 2021.

[27] M. Z. Ozturk, C. Wu, B. Wang, and K. J. R. Liu. Radiomic: Sound sensing via mmwave signals, 2021.

[28] E. Pan, J. Tang, D. Kosaka, R. Yao, and A. Gupta. Openradar. https://github.com/presenseradar/openradar, 2019.

[29] H. Saghlatoon, R. Mirzavand, M. M. Honari, and P. Mousavi. Sensor antenna transmitter system for material detection in wireless-sensor-node applications. *IEEE Sensors Journal*, 18(21):8812–8819, 2018.

[30] A. A. Shabana. *Theory of Vibration*, volume 2. Springer, 1999.

[31] F. Shang, P. Yang, Y. Yan, and X.-Y. Li. Liqray: Non-invasive and fine-grained liquid recognition system. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, MobiCom '22, page 296–309, New York, NY, USA, 2022. Association for Computing Machinery.

[32] H. Song, B. Wei, Q. Yu, X. Xiao, and T. Kikkawa. Wieps: Measurement of dielectric property with commodity wifi device—an application to ethanol/water mixture. *IEEE Internet of Things Journal*, 7(12):11667–11677, 2020.

[33] S. Su, F. Heide, R. Swanson, J. Klein, C. Callenberg, M. Hullin, and W. Heidrich. Material classification using raw time-of-flight measurements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3503–3511, 2016.

[34] X. Sun, W. Deng, X. Wei, D. Fang, B. Li, and X. Chen. Akte-liquid: Acoustic-based liquid identification with smartphones. *ACM Trans. Sen. Netw.*, aug 2022. Just Accepted.

[35] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[36] K. Tanaka, Y. Mukaigawa, T. Funatomi, H. Kubo, Y. Matsushita, and Y. Yagi. Material classification using frequency-and depth-dependent time-of-flight distortion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 79–88, 2017.

[37] C. Wang, J. Liu, Y. Chen, H. Liu, and Y. Wang. Towards in-baggage suspicious object detection using commodity wifi. In *2018 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9. IEEE, 2018.

[38] C. Wang, X. Zhang, X. Zang, Y. Liu, G. Ding, W. Yin, and J. Zhao. Feature sensing and robotic grasping of objects with uncertain information: A review. *Sensors*, 20(13):3707, 2020.

[39] G. Wang, J. Han, C. Qian, W. Xi, H. Ding, Z. Jiang, and J. Zhao. Verifiable smart packaging with passive rfid. *IEEE Transactions on Mobile Computing*, 18(5):1217–1230, 2018.

[40] J. Wang, J. Xiong, X. Chen, H. Jiang, R. K. Balan, and D. Fang. Tagscan: Simultaneous target imaging and material identification with commodity rfid devices. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, MobiCom '17, page 288–300, New York, NY, USA, 2017. Association for Computing Machinery.

[41] L. Wang, C.-Y. Lee, Z. Tu, and S. Lazebnik. Training deeper convolutional networks with deep supervision. *arXiv preprint arXiv:1505.02496*, 2015.

[42] Z. Wang. *Towards Robust and Secure Audio Sensing Using Wireless Vibrometry and Deep Learning*. University of California, Los Angeles, 2020.

[43] Z. Wang, Z. Chen, A. D. Singh, L. Garcia, J. Luo, and M. B. Srivastava. Uwhear: through-wall extraction and separation of audio vibrations using wireless signals. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 1–14, 2020.

[44] T. Wei, S. Wang, A. Zhou, and X. Zhang. Acoustic eavesdropping through wireless vibrometry. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 130–141, 2015.

[45] J. Weiß and A. Santra. One-shot learning for robust material classification using millimeter-wave radar system. *IEEE Sensors Letters*, 2(4):1–4, 2018.

[46] P. Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73, 1967.

[47] C. Wu, F. Zhang, B. Wang, and K. R. Liu. msense: Towards mobile material sensing with a single millimeter-wave radio. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3):1–20, 2020.

[48] B. Xie, J. Xiong, X. Chen, E. Chai, L. Li, Z. Tang, and D. Fang. Tagtag: Material sensing with commodity rfid. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, SenSys '19, page 338–350, New York, NY, USA, 2019. Association for Computing Machinery.

[49] T. Xue, J. Wu, Z. Zhang, C. Zhang, J. B. Tenenbaum, and W. T. Freeman. Seeing tree structure from vibration. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 748–764, 2018.

[50] Y. Yang, Y. Wang, J. Cao, and J. Chen. Hearliquid: Non-intrusive liquid fraud detection using commodity acoustic devices. *IEEE Internet of Things Journal*, pages 1–1, 2022.

[51] H.-S. Yeo, G. Flamich, P. Schrempf, D. Harris-Birtill, and A. Quigley. Radarcat: Radar categorization for input & interaction. pages 833–841, 10 2016.

[52] D. Zhang, J. Wang, J. Jang, J. Zhang, and S. Kumar. On the feasibility of wi-fi based material sensing. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, New York, NY, USA, 2019.

Association for Computing Machinery.

[53] J. Zhang, Y. Zhou, R. Xi, S. Li, J. Guo, and Y. He. Ambiear: Mmwave based voice recognition in nlos scenarios. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(3), sep 2022.

[54] Y. Zhang, G. Laput, and C. Harrison. Vibrosight: Long-range vibrometry for smart environment sensing. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pages 225–236, 2018.