

Fine-Grained Vibration Based Sensing Using a Smartphone

Kamran Ali¹ and Alex X. Liu², *Fellow, IEEE*

Abstract—Recognizing surfaces based on their vibration signatures is useful as it can enable tagging of different locations without requiring any additional hardware, such as near field communication (NFC) tags. However, previous vibration based surface recognition schemes either use custom hardware for creating and sensing vibration, which makes them difficult to adopt, or use inertial (IMU) sensors in commercial off-the-shelf (COTS) smartphones to sense movements produced due to vibrations, which makes them coarse-grained because of the low sampling rates of IMU sensors. The mainstream COTS smartphones based schemes are also susceptible to inherent hardware based irregularities in vibration mechanism of the smartphones. Moreover, the existing schemes that use microphones to sense vibration are prone to short-term and constant background noises (e.g., intermittent talking, exhaust fan, etc.) because microphones not only capture the sounds created by vibration but also other interfering sounds present in the environment. In this paper, we propose VibroTag, a robust and practical vibration based sensing scheme that works with smartphones with different hardware, can extract fine-grained vibration signatures of different surfaces, and is robust to environmental noise and hardware based irregularities. We implemented VibroTag on two different Android phones and evaluated in multiple different environments where we collected data from four individuals for 5 to 20 consecutive days. Our results show that VibroTag achieves an average accuracy of 86.55 percent while recognizing 24 different locations/surfaces, even when some of those surfaces were made of similar material. VibroTag's accuracy is 37 percent higher than the average accuracy of 49.25 percent achieved by one of the state-of-the-art IMUs based schemes, which we implemented for comparison with VibroTag.

Index Terms—Vibration sensing, acoustic sensing, symbolic localization, surface recognition, mobile sensing

1 INTRODUCTION

1.1 Motivation

VIBRATION based sensing has been shown to be a low-cost and effective approach to recognizing different surfaces [1], [2], [3], [4]. The key intuition is that different surfaces respond to the same vibration differently because the surfaces may be made of different materials, and even if they are made of the same material, they may have different shapes and sizes. Even if two surfaces are made of the same material and have the same shape and size, they may have different objects placed on them, such that those surfaces still exhibit different vibration patterns because the objects placed on them may respond to the same vibration differently. Recognizing surfaces based on their vibration signatures is useful as it can enable tagging of different locations without requiring any additional hardware such as Near Field Communication (NFC) tags. Such tagging of locations can provide us with indirect information about the user activities and intentions without any dedicated infrastructure, based on which we can enable useful services such as context aware notifications/alarms [5]. For example, a user can set their phone to automatically go into silent mode

when it is placed on their bed side table. Such vibration based tagging can also be used to reduce the search space for the user when they lose their phone. For example, the “find my phone” feature can report the GPS location of a lost phone as well as what material the phone is laying on.

A robust and practical vibration based sensing scheme should satisfy three key requirements. First, it should work with commercial off-the-shelf (COTS) smartphones with different hardware, so that it can be easily deployed and widely adopted. Second, it should be able to extract fine-grained vibration signatures, so that it can accurately differentiate different surfaces. Third, it should be robust to environmental noise and hardware based irregularities, so that its accuracy stays consistent across different environments and devices.

1.2 Limitations of Prior Art

Several vibration based sensing schemes have been proposed in the past to realize different kinds of applications [2], [3], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. However, none of them satisfies all the above three requirements. Existing vibration based sensing schemes can be divided into two categories: custom hardware based and COTS smartphones based. The custom hardware based schemes use separate hardware including a micro-controller, a vibrator motor, and some piezoelectric (e.g., a microphone) or IMU sensors [2], [3], [11], [12], [13], which gives them fine-grained low level control over different physical layer parameters of their underlying hardware. However, these schemes are incompatible with COTS smartphones and are not easily generalizable to different hardware because most COTS smartphones have limited sensing capabilities and control over the hardware installed in

• Kamran Ali is with the General Motors Research and Development, Warren, MI 48557 USA. E-mail: kamran.lumininte@gmail.com.

• Alex X. Liu is with the Qilu University of Technology, Jinan 250353, Shandong, China. E-mail: alexliu360@gmail.com.

Manuscript received 20 Oct. 2019; revised 15 Mar. 2021; accepted 17 Mar. 2021. Date of publication 22 Mar. 2021; date of current version 3 Oct. 2022.

(Corresponding author: Kamran Ali.)

Recommended for acceptance by B. Liang.

Digital Object Identifier no. 10.1109/TMC.2021.3067679

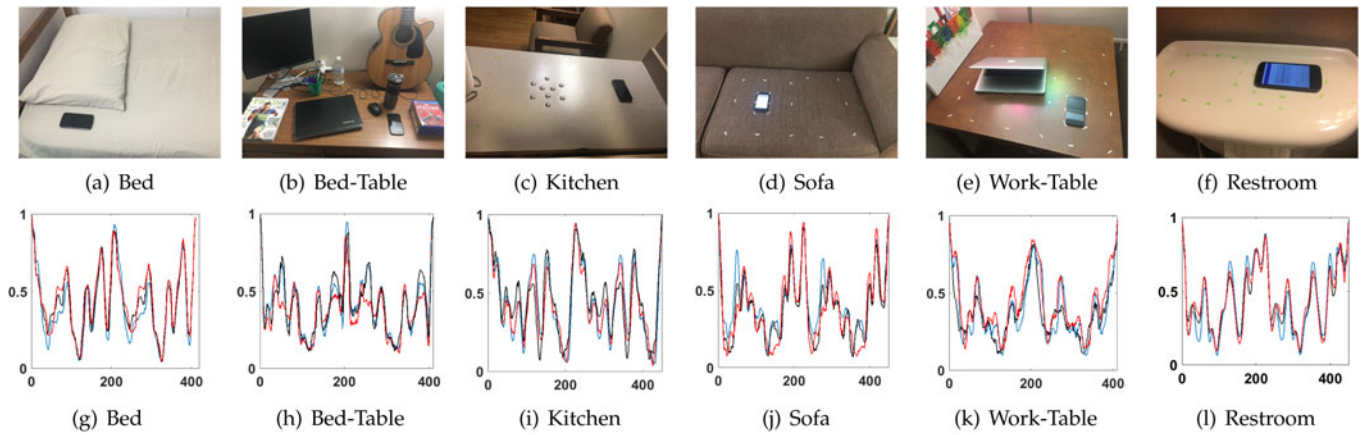


Fig. 1. Experimental scenarios and their corresponding extracted acoustic time-series based vibration signatures.

them. Most COTS smartphones based schemes rely on motion based features extracted from built-in IMU sensors [1], [4], [6], [7], [10]. However, these features are coarse-grained because the sampling frequencies of IMU sensors on COTS smartphones are much lower (usually between 200-300 Hz and less than 4 kHz even after driver modifications [11]) compared to microphones (usually above 44 kHz), which naturally leads to lower classification accuracies. Because of this reason, such schemes can only broadly differentiate between different types of surfaces (e.g., wood and plastic) and cannot differentiate between similar surfaces (e.g., two different wooden tables in Figs. 1b and 1e). Also, IMU readings get significantly affected by the smartphone's own motion in space (e.g., when a user moves his hand while holding his smartphone or when the smartphone slides during vibration when placed on a smooth surface). Moreover, the mainstream COTS smartphones based schemes (may they be IMU or microphone based) are susceptible to inherent hardware based irregularities in vibration mechanism of the smartphones. We discuss the impact of such irregularities in Section 4.2. Finally, the existing schemes that use microphones to sense vibration [3], [8], [9] are prone to short-term and constant background noises (e.g., intermittent talking, clapping, exhaust fan, etc.) because microphones not only capture the sounds created by vibration but also other interfering sounds present in the environment. We discuss the impact of such noises in Section 4.1.

1.3 Proposed Approach

In this paper, we propose VibroTag, a vibration based sensing approach that can recognize different surfaces based on their unique vibration signatures. Compared to previous work, VibroTag is robust and practical because it works with COTS smartphones, it can extract fine-grained features representative of different surfaces, and it is robust to hardware irregularities and background environmental noises. The key intuition is that as the vibrating mass inside a smartphone's vibrator motor repeatedly moves to and fro, the vibrating mass causes the whole smartphone structure and the hardware inside it to vibrate in a peculiar pattern, which depends upon the vibration response (or absorption properties) of the surface that smartphone is placed on. These vibrations produce peculiar sound waves that VibroTag detects using the smartphone's microphone. Fig. 1

shows the unique vibration signatures that VibroTag extracted for 6 different surfaces. We observe that vibration signatures of even two similar surfaces, i.e., Bed and Sofa, are quite different from each other.

To make VibroTag easily scalable and compatible with COTS smartphones, we design VibroTag's signal processing pipeline such that it relies only on built-in vibration motors and microphone for sensing, and it is applicable to different phones with different hardware. To reliably extract fine-grained vibration signatures from the sound signals recorded during vibration, we propose a novel time-series based approach, which is robust to hardware irregularities and environmental noise. The key idea behind our vibration signature extraction approach is that even if there are irregularities in vibration frequencies due to hardware imperfections, the time-series patterns created during different vibration cycles are very similar. When a phone vibrates during a specific period of time, such as 3 seconds, multiple such patterns occur and get distributed all over the time-series of recorded sound signals. VibroTag finds multiple of these patterns in randomly selected intervals of the time-series, and then combines them into single time-series features that ensures consistency even if there are irregularities in the occurrence of those patterns and/or if the environment is slightly noisy. Afterwards, it uses these features to differentiate between surfaces.

1.4 Technical Challenges and Our Solutions

The first technical challenge is to reliably extract fine-grained vibration signatures. Based on our experiments on two different phones, we observed that the frequency response of a surface to vibrations introduced by vibration motors installed in COTS smartphones exhibit repeated irregularities, which makes extraction of reliable features a challenging task. This happens because the phone, its vibrator motor, and the rest of its hardware vibrate at irregular frequencies during every experiment, which occurs due to hardware imperfections. This behavior is random and uncontrollable, and therefore, is bound to create significant variations within features extracted at the same location, which will lead to classification inaccuracy. The existing techniques that use microphone to extract sound (sampled in the order of kHz) based straightforward frequency domain features (e.g., vibration sound spectrum [3]) are not only considerably susceptible to such hardware based irregularities, but also to short-term and constant environmental

noises, where even intermittent talking or noise from a rest-room's exhaust fan can significantly affect their performance. To address this challenge, we take a time-series based vibration signature extraction approach. First, we differentiate the recorded sound signals and take their root mean square (RMS) envelope, which removes most of the unrelated constant and higher frequency background noise. Second, we develop a specialized peak-detection based algorithm to extract unique time-series patterns corresponding to vibrations from the RMS envelope, and then use them as *vibration signatures* to represent different surfaces. Our extraction algorithm is based on the observation that even if there are irregularities in vibration frequencies due to hardware imperfections, the time-series patterns created during different vibration cycle are very similar. When a phone vibrates during a specific period of time, multiple such patterns occur all over the time-series of recorded sound signals, which can be successfully extracted by our algorithm. To make the vibration signatures robust to environmental noise, VibroTag extracts numerous such vibration patterns across time during an experiment and combines them by taking their median.

The second technical challenge is to compare vibration signatures of any two surfaces. The midpoints of extracted vibration signatures of the same surface rarely align with each other because the start and end points determined by extraction algorithm are never perfectly aligned. Moreover, the lengths of different vibration signatures also differ slightly because the duration of vibration cycle can often be a little different due to hardware irregularities. Consequently, the midpoints and lengths of vibration signatures do not match either. Another issue is that the shape of different vibration signatures of the same surface are often distorted versions of each other, which occurs due to hardware based irregularities in the vibration mechanism. Therefore, two vibration signatures cannot be compared using standard measures like *correlation coefficient* or *euclidean distance*. To address this challenge, we use the Dynamic Time Warping (DTW) to quantify the distance between any two vibration signatures. DTW can find the minimum distance alignment between two waveforms of different lengths. For classification, we employ a *Nearest-Neighbor* (NN) classifier with DTW distance as the comparison metric between different vibration signatures.

1.5 Key Novelty and Advantages

The key technical novelty of this paper is on proposing the first fine-grained vibration based sensing scheme that can recognize different surfaces using the vibration mechanism and microphone of a single COTS smartphone. Furthermore, we propose a novel signal processing technique to extract fine-grained vibration signatures that are robust to hardware irregularities and background environmental noises. The key insight is that even if there are irregularities in vibration frequencies due to hardware imperfections, the time-series patterns created during different vibration cycles are very similar. VibroTag finds many such patterns in the sound signals recorded during vibration, and combines them into single consistent vibration signatures. Compared to previous schemes, VibroTag works with COTS smartphones, it can extract fine-grained features representative of different surfaces, and it is robust to hardware irregularities and background environmental noises.

1.6 Summary of Experimental Results

We implemented VibroTag on two Android based smartphones, i.e., Nexus 4 and OnePlus 2, for which we developed an application for generating vibrations and to sample sound signals simultaneously. We tested our system for 4 different individuals, from whom we collected data for 5 - 20 days. We show that VibroTag achieves an average accuracy of 86.55 percent while recognizing 24 different locations, even when some of those surfaces were made of similar material, with as few as 15 training samples per location. Moreover, VibroTag maintains an average accuracy of up to 85 percent without any re-training requirements after 3-4 days of training. We also implement one of the state-of-the-art IMUs based vibration sensing schemes for single COTS smartphones proposed in [4], and compare its surface recognition accuracy with VibroTag. We chose this scheme for comparison because this is the only existing scheme that is insusceptible to acoustic noises in the environment. We show that VibroTag achieves more than 37 percent higher accuracy when compared to the IMUs based scheme, while recognizing the 24 different locations.

2 RELATED WORK

Existing work related to our paper consists of some vibration based sensing schemes [1], [2], [3], [4], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [15], [16], [17] and sound based symbolic localization schemes [18], [19].

Vibration Based Sensing. Vibration based sensing schemes leverage the response of different surfaces to a specific vibration pattern to recognize those surfaces. Existing vibration based sensing schemes can be divided into two categories, i.e., custom hardware based, and COTS smartphones based. The custom hardware based schemes use separate customized hardware made using a set of micro-controller, vibrator motor, and piezoelectric (e.g., microphones) or IMU sensors [2], [3], [11], [12], [13], so that they have fine-grained low level control over different physical layer parameters of their hardware. ViBand uses variations introduced due to vibrations produced by different objects to identify those objects, e.g., electric tooth brush [11]. VibKeyboard [12] and VibSense [13] develop a virtual keyboard based on the idea that the impact of a touch on a surface such as a table or door causes a shockwave to be transmitted through the material that can be passively detected with accelerometers or more sensitive piezo-vibration sensors. Kunze *et al.* [3] develop customized hardware to recognize surfaces through active sampling of acceleration and sound signatures. However, the above schemes are incompatible with COTS smartphones and are not easily generalizable to different hardware because most COTS smartphones have limited sensing capabilities and control over the hardware installed in them. Most COTS smartphones schemes rely on motion based features extracted from built-in IMU sensors [1], [4]. Cho *et al.* [4] and Shafer *et al.* [2] use built-in vibrator and accelerometer of a COTS smartphone to recognize surfaces. Griffin *et al.* use vibration detected by an acceleration signal to determine if a phone is in the user's hand [1]. However, these features are coarse-grained because the sampling frequencies of IMU sensors on COTS smartphones are much lower (usually between 200-300 Hz and less than 4 kHz even after driver modifications [11]) compared to microphones (usually above 44 kHz), which naturally leads to lower classification accuracies. Because of this reason, such

schemes can only broadly differentiate between different types of surfaces (e.g., wood and plastic) and cannot differentiate between similar surfaces (e.g., two different wooden tables in Figs. 1b and 1e). Also, IMU readings get significantly affected by the smartphone's own motion in space (e.g., when a user moves his hand while holding his smartphone or when the smartphone slides during vibration when placed on a smooth surface). Moreover, the mainstream COTS smartphones based schemes (may they be IMU or microphone based) are susceptible to inherent hardware based irregularities in vibration mechanism of the smartphones. We discuss the impact of such irregularities in Section 4.2. Finally, the existing schemes that use microphones to sense vibration [3], [8], [9] are prone to short-term and constant background noises (e.g., intermittent talking, clapping, exhaust fan, etc.) because microphones not only capture the sounds created by vibration but also other interfering sounds present in the environment. We discuss the impact of such noises in Section 4.1. Compared to all the above schemes, VibroTag is robust and easily deployable because it works with COTS smartphones, it can extract fine-grained features representative of different surfaces/locations, and it is robust to hardware irregularities and background environmental noises.

Sound Based Symbolic Localization. Sound based symbolic localization systems leverage the propagation of the sound generated using speakers of a device, such as a smartphone, to determine the symbolic location of that device (e.g., whether the device is in the user's kitchen or at his bedroom table). SurroundSense uses sensor data from a microphone, a light sensor, the wireless radio, and passive accelerometer data for localization [19]. However, their technique can only be used for very coarse-grained localization (e.g., room level) and not for finer-grained localization (e.g., whether the phone is on user's study table or his bedroom table). EchoTag generates ultrasound signals and then uses the reflections from the environment to achieve centimeter level tagging [18]. However, their work requires strict millimeter level marking of the tagged locations because the ultrasound signals based signatures that they use are highly location dependent, where even small variations in the phone's position leads to significant localization errors. This makes their scheme unsuitable for symbolic localization, and also puts significant calibration effort on the user end. In contrast to above schemes, VibroTag uses vibration instead of speaker generated sound signals for such symbolic localization. Moreover, VibroTag achieves finer-grained localization, and does not require strict marking of the tagged locations.

3 UNDERSTANDING VIBRATIONS

3.1 Vibrator Motors in Smartphones

Electric vibrator motors generate vibrations by periodically moving an unbalanced mass around a center position using the principles of electromagnetic induction. The vibrator motors used in today's smartphones are often known as *coin-type vibration motors* due to their coin-like shapes and sizes. There are two types of coin-type vibrator motors that are widely adopted in smartphones: (i) *Linear Resonant Actuator* (LRA) based (e.g., used in Nexus 4) and (ii) *Eccentric Rotating Mass* (ERM) based (e.g., used in OnePlus 2). Fig. 2 shows the internals of ERM and LRA based vibration motors. ERM based

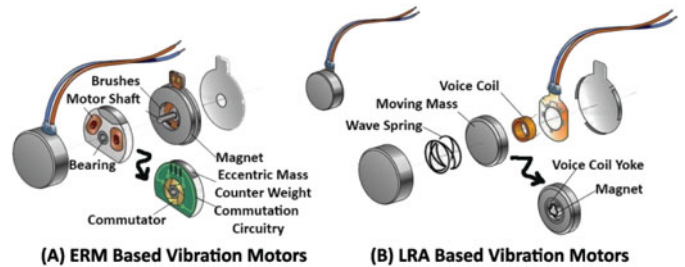


Fig. 2. ERM and LRA based vibration motors [20].

motors use a DC motor to rotate an eccentric mass around an axis. As the mass is not symmetric with respect to its axis of rotation, it causes the device to vibrate during the motion. Both the amplitude and frequency of vibration depend on the rotational speed of the motor, which can in turn be controlled through an input DC voltage. With increasing input voltages, both amplitude and frequency increase almost linearly and can be measured by an accelerometer. In LRA based motors, vibration is generated by the linear movement of a magnetic mass suspended near a coil, called the "voice coil". When an AC current is applied to the motor, the coil behaves like a magnet (due to the generated electromagnetic field) and causes the mass to be attracted or repelled, depending on the direction of the current. This generates vibration at the same frequency as the input AC signal, while the amplitude of vibration is dictated by the signal's peak-to-peak voltage. Thus, LRAs offer control on both the magnitude and frequency of vibration.

3.2 Physics of Surface Response to Vibrations

Sound is essentially pressure waves created by vibrating matter. These waves are longitudinal, i.e., they oscillate along the axis of travel, where the oscillation is composed of compression and rarefaction of molecules in the medium (e.g., air). For example, human speech is based on vibrations created inside our vocal chords, and audio speakers generate sound by translating an electrical signal into physical vibrations via mechanical excitation of a diaphragm using an electromagnet.

VibroTag is based on the intuition that different surfaces exhibit different response to vibrations introduced by smartphone. When a smartphone vibrates, it mechanically excites not only its own structure and hardware inside it, but also the surface on which it is placed. On one hand, some surfaces tend to absorb most of the vibration energy (e.g., Sofas), while on the other hand, some surfaces may exhibit a resonant response where they start vibrating in sync with the smartphone (e.g., the smartphone's surface vibrates in sync with the vibrator motor inside). Moreover, the effect of these vibrations can reach different objects placed nearby, which may get mechanically excited as well (especially the lighter objects); therefore, leading to more peculiar sounds. As different surfaces respond to the vibrations differently (in terms of their absorption/dampening effect on smartphone's movements), and as different surfaces often have different objects placed on them, which also respond to those vibrations differently, pressure waves peculiar to those surfaces are created during the vibration, which we can sense using a piezoelectric device (e.g., a built-in microphone) and then leverage to differentiate those surfaces.

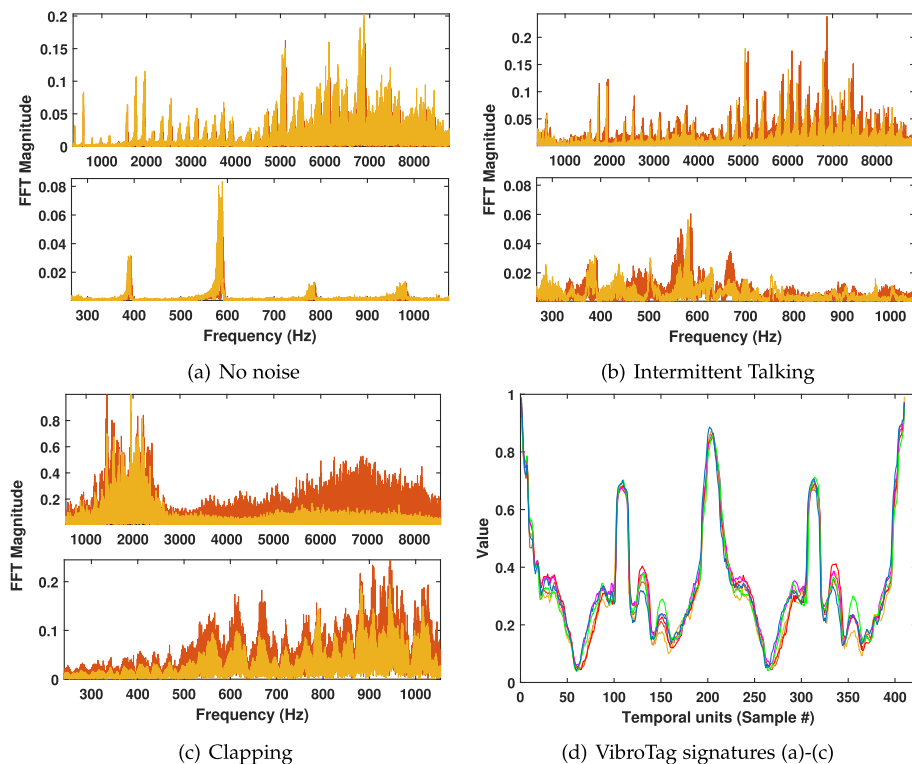


Fig. 3. Impact of background noises on features extracted by traditional techniques and VibroTag.

4 FEATURE EXTRACTION

To differentiate between different locations, we need to extract features that can uniquely and consistently represent those locations. In VibroTag, a smartphone is vibrated for about 3 seconds while the surface response to the vibration is recorded simultaneously via the phone's built-in microphone. Sounds produced during vibration are sampled at fixed $F_s = 44.1$ kHz. The recorded sound is analyzed in both frequency and time domains to extract robust surface/location specific vibration signatures. There are two key challenges in feature extraction for VibroTag to be robust. The first challenge is on reducing impact of background noises (such as those created by fans and short-term human speech). The second challenge is on accommodating smartphone hardware imperfections (i.e., microphones and vibrator motors mainly) that degrades the quality of the signals collected when a smartphone vibrates.

4.1 Robustness to Background Noise

To understand the challenge posed by background noise, we use Fast Fourier Transform (FFT) based Power Spectral Density (PSD), which is one of the mainstream frequency based feature extraction techniques for acoustic sensing. Figs. 3a, 3b, and 3c show the FFT coefficients for both lower and higher frequency ranges corresponding to our experiments conducted at the same location on a wooden chair's cushion for three scenarios: (a) no noise, (b) intermittent human speech, and (c) clapping, respectively. We can observe that the FFT features are significantly affected by the background noises because the frequencies produced by these noises directly interfere with the frequency bands for vibration based sensing. It also shows that mainstream techniques such as FFT or PSD are unsuitable for vibration based sensing on COTS

smartphones when there are background noise sources present in the environment. In this paper, we propose two schemes to reduce the impact of constant and intermittent short-term background noises, respectively.

To reduce the impact of constant background noises in VibroTag, we take the *first order difference* of the recorded sound signals and then take their *root mean squared* (RMS) envelope. We choose RMS envelope for our analysis as it gives us a measure of the power of the vibration signals, while producing a waveform that is easy to analyze. Moreover, higher frequency noisy variations are averaged out in the envelope signal, while it still keeps most of the useful vibration response related information intact. We take the RMS envelope of the signals over a sliding window of N samples, where $N = 15$ audio samples in our current implementation of VibroTag. In the rest of this paper, when we mention sound signals, we mean the first order difference of the RMS envelope of those sound signals.

To reduce the impact of intermittent short-term noises (similar to Figs. 3b and 3c), we vibrate the phone for at least 3 seconds, and extract multiple vibration patterns across time from the processed sound signals; then, we combine these vibration patterns to get a single consistent vibration signature. We will discuss how we extract such signatures in Section 4.3. Fig. 3d shows the signatures extracted by VibroTag for the experiments corresponding to Figs. 3a, 3b, and 3c, from which we observe that our signatures are consistent and almost identical even though there are intermittent short-term noises.

4.2 Robustness to Hardware Imperfections

To understand the challenge posed by smartphone hardware imperfections, we use PSD and Short Time Fourier Transform (STFT). Figs. 4a and 4b show the PSD of the recorded

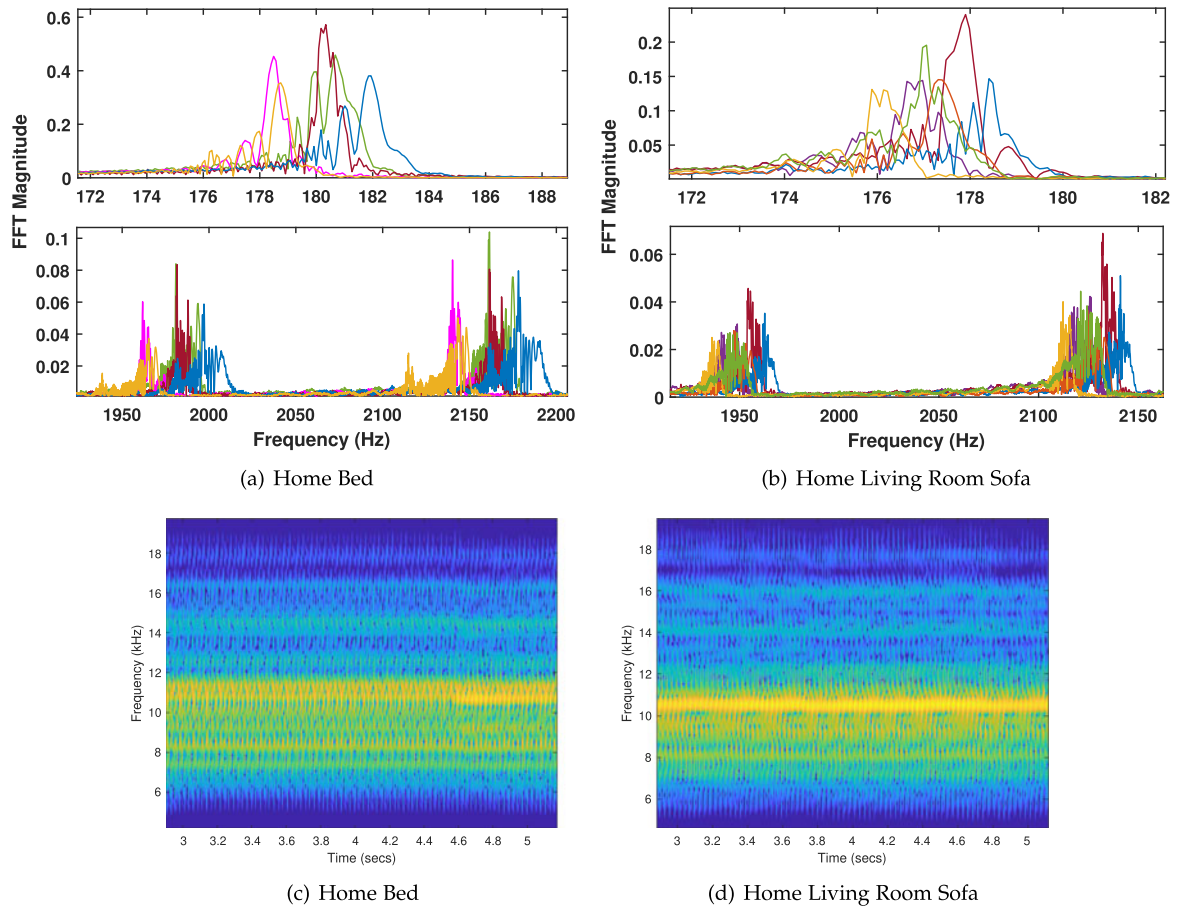


Fig. 4. Impact of hardware imperfection on features extracted by traditional techniques PSD and STFT.

unprocessed sound signals for multiple different experiments, which we performed by placing a smartphone at the same location on a sofa and a bed, respectively. Each figure shows the PSD coefficients over two different frequency ranges.

We can observe that PSDs for both the sofa and the bed are not consistent for repetitive experiments, even when the smartphone's location and the environmental scenario while performing the experiments remained unchanged. This occurs because the smartphone, its vibrator motor, and the rest of its hardware vibrate at slightly different frequencies in each different experiment. This behavior is random and uncontrollable, and therefore, is bound to cause intra-class (i.e., within samples of the same class) variations, which will lead to classification errors. Moreover, due to this inconsistency, it often happens that the variations due to vibration on two different surfaces occur in similar set of frequencies (as shown by some samples in Figs. 4a and 4b), which further makes the use of such frequency domain features difficult as they do not always represent different surfaces uniquely.

We hypothesize that even if there are irregularities in the repetition frequency of such vibration patterns, the patterns themselves must be very similar. Figs. 4c and 4d show the STFT of the unprocessed sound signals from one experiment corresponding to each of the two surfaces. Interestingly, we observe that some patterns repeating along time; however, the time period of their repetition is not consistent due to the aforementioned hardware irregularities. As we discussed in Section 3, smartphone vibrations are generated

because of the to and fro motion of a mass inside its vibration motor. The vibration motor tries to move the smartphone (and the hardware inside) at its own vibration frequency (which is often irregular due hardware imperfections). However, the smartphone's motion is constrained due to its own weight/structure and the absorption properties of the surface that it is placed on. This whole process during vibration gives rise to peculiar pressure waves, which can be sensed by the built-in microphone. Moreover, as the mass inside the motor repeatedly moves to and fro, it will give rise to similar pressure waves in every cycle of its "irregular" vibration period, which will reflect in time-series of the sound signals. These intuitions form the basis of our time-series based analysis of the surfaces' vibration response (Section 4.3).

In this paper, we propose to address the issues due to smartphone hardware imperfections by extracting time-series based vibration signatures from the processed sound signals. Figs. 5a and 5b show time-series of the sound signals (i.e., first order difference of the RMS envelope) corresponding to one experiment from each of the two different surfaces (i.e., Bed and Sofa), whose PSDs are shown in Figs. 4a and 4b and whose STFTs are shown in Figs. 4c and 4d, respectively. The two time-series correspond to a window of 4,800 sound samples (i.e., ~ 0.1088 seconds for $F_s = 44.1$ kHz). We can easily observe distinguishing patterns repeating in both time-series, which repeat approximately with the frequency of the vibrating mass in the smartphone's vibration motor. We also observe that the patterns in both scenarios are consistent across time, and that

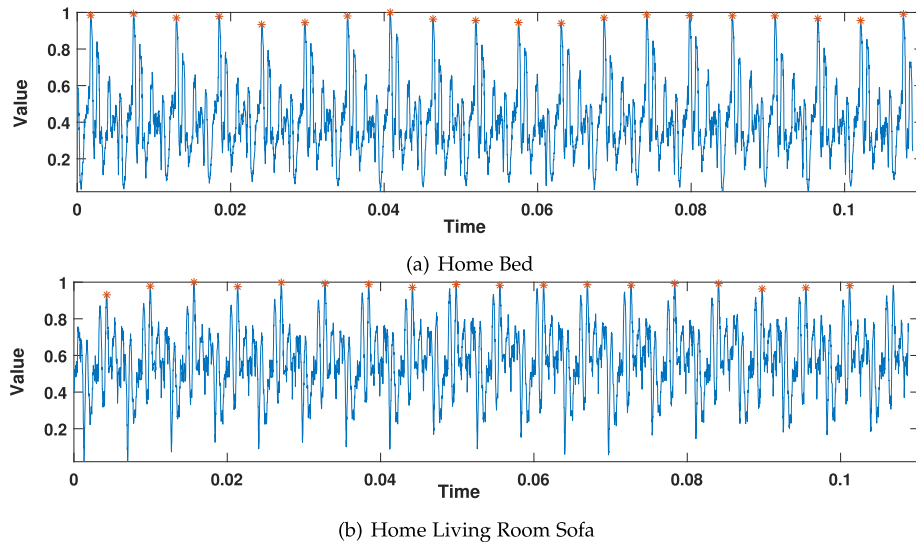


Fig. 5. Repetitive patterns appearing in the processed sound signals corresponding to vibration.

the patterns in one scenario are different from the ones in other scenario.

4.3 Extraction of Vibration Signature

To robustly differentiate surfaces, we need to extract vibration patterns from time-series of the processed sound signals and then use those patterns to obtain consistent vibration signatures. However, we face multiple challenges. The first challenge is that intermittent short-term noises in real-life scenarios are uncontrollable, and therefore, can affect any part of the time-series. VibroTag needs to extract the vibration patterns that are representative of the whole time-series, i.e., the vibration patterns extracted from one segment of the time-series (i.e., a time window) should repeat in other segments, and therefore, truly represent the surface's vibration response. A naive approach is to extract all vibration patterns in the recorded signals and then combine them (e.g., by taking their average), which is computationally expensive. To address this challenge, we take a *randomized* approach, where we first divide the whole time-series of the sound signal into equally sized time windows, and then randomly select multiple of those time windows to extract vibration patterns from. Each window is of size S , where $S = 4800$ sound samples in our current implementation of VibroTag. Moreover, the windows are selected without replacement, i.e., once selected, they are not selected again. VibroTag keeps randomly selecting new time windows until M vibration patterns are extracted ($M=100$ in our implementation). *In real-life scenarios, the number of iterations required to extract M vibration patterns of a surface can be used to tell the user whether their environment is too noisy to extract a valid vibration signature or not.* For example, the average (taken over 20 different samples) number of iterations it took for convergence when loud music (i.e., high variable noise) was played on a laptop in the background was ~ 1351 , for medium noise/volume level the number was ~ 495 , whereas for no variable noise scenarios it was ~ 136 . In the cases where our algorithm cannot find enough vibration patterns and runs out of possible time windows to search for patterns, it will not converge. However, we did not experience any such scenarios during our testing.

The second challenge is to extract the vibration patterns from different randomly selected time windows by

localizing their place of occurrence in those time windows. However, because of the inconsistencies in the vibration behavior of smartphone due to hardware imperfections, we cannot know the frequency of repetition of the vibration patterns, which makes it harder to localize the place of occurrence of such patterns. To address this challenge, we develop a *peak detection* based algorithm to extract vibration patterns. Our algorithm is based on the observation that every vibration pattern has a peak value that occurs consistently at around the same part of every vibration cycle (as evident in Figs. 5a and 5b). Based on this algorithm, VibroTag determines the locations of multiple such peaks in each of the randomly selected time windows. Afterwards, VibroTag uses the consecutive peaks detected in each window to extract the vibration patterns between those peaks.

4.3.1 Extraction of Vibration Patterns

There are two key challenges in developing our peak detection based vibration pattern extraction algorithm. First, based on our experiments, we observe that the scale of variations due to vibration in the time-series of different randomly selected windows varies, even when the same smartphone is placed at the same location of the same surface, which happens due to hardware imperfections based inconsistencies in the vibration process. Moreover, different surfaces and different smartphones exhibit different scale of variations due to vibration. This makes the parametrization of our peak detection algorithm difficult to generalize. To address this challenge, VibroTag performs *max-min normalization* on the time-series corresponding to every selected window before feeding it to the peak detection algorithm. This step ensures that the parametrization of our algorithm can be easily generalized to different time windows and to different smartphones and surfaces.

The second challenge is to robustly determine the locations of peaks corresponding to different vibration patterns present in a time window. To address this, VibroTag's peak detection algorithm determines the location of such peaks based on three key parameters, namely *minimum peak prominence* (MINPRO), *minimum peak distance* (MINDIST), and *minimum peak strength* (MINSTR). The prominence of a peak measures how much the peak stands out, due to its height

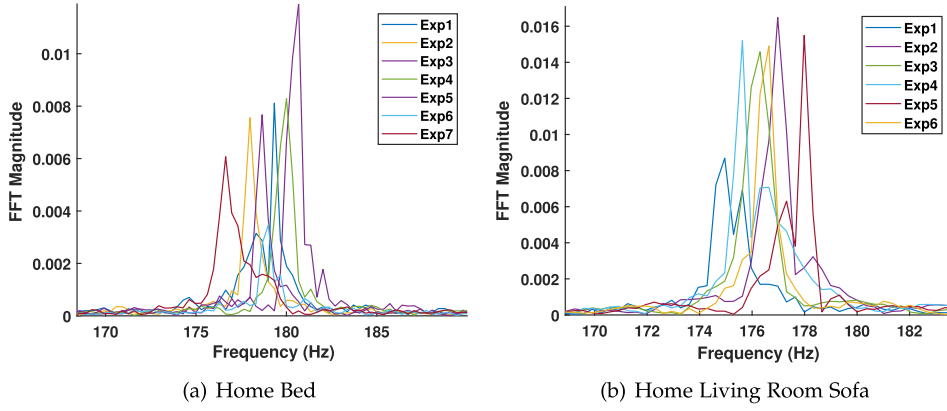


Fig. 6. PSD of the sound signals in time windows corresponding to the scenarios in Figs. 5a and 5b.

and location, relative to other peaks around it. We tune MINPRO such that we only detect those peaks which have a relative importance of at least MINPRO. We tune MINDIST according to the fact that maximum repetition rate of patterns is approximately $\hat{f}_o + \delta f$, such that the redundant peaks are discarded, where \hat{f}_o is an approximate number for the frequency of repetition (f_o) of vibration patterns. As we discussed before, the frequency of repetition of vibration patterns in the processed sound signal is irregular. In order to determine f_o , VibroTag calculates PSD of the time-series in the selected window. Figs. 6a and 6b show example PSD's corresponding to one of the randomly selected time windows from seven different experiments performed on the Bed and the Sofa, respectively. VibroTag determines the peak frequency from the PSD, which corresponds to the approximate repetition frequency (i.e., \hat{f}_o) of the vibration patterns present in the window. The term δf represents the second standard deviation of the variation in vibration frequency around \hat{f}_o , which we estimate for every smartphone based on multiple different experiments. δf is required as some vibration patterns can repeat earlier than $1/\hat{f}_o$.

To further sift out redundant peaks, we only choose peaks of value greater than MINSTR times the median value of the peaks detected in the window. In our current implementation, we chose MINPRO = 0.65, MINDIST = $\frac{1}{\hat{f}_o + \delta f}$ seconds, $\delta f = 6.5$, and MINSTR = 0.5. We perform this parametrization only during the design time, which generalizes well for multiple different surfaces and smartphones (i.e., Nexus 4 and OnePlus 2). Our algorithm does not require any end-user calibration effort. VibroTag uses the consecutive peaks detected in each randomly selected window to extract multiple vibration patterns between those peaks.

4.3.2 Construction of Vibration Signature

To construct a single consistent vibration signature, VibroTag first collects at least a total of $M = 100$ patterns extracted from the randomly selected time windows. Once an enough number of vibration patterns are extracted from different time windows, VibroTag combines all those patterns using median (i.e., 50th percentile) to get a single vibration signature. The median operation helps VibroTag remove short-term noisy variations in different vibration patterns, and therefore, allows it to extract a single robust vibration signature of the surface. Figs. 7a and 7b show the example signatures extracted for the Bed and the Sofa related experiments, where we can observe that the vibration signatures of each surface are consistent and almost

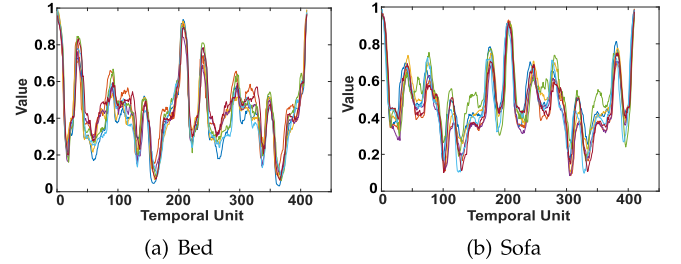


Fig. 7. Extracted time-series features (low noise).

identical. Moreover, the extracted signatures uniquely represent their respective surfaces.

Fig. 8 shows the vibration signatures extracted for two similar tables during lunch time in a cafeteria on a university campus (i.e., a highly noisy environment). We can see that although some vibration signatures that VibroTag extracted are inconsistent, yet it was able to extract several consistent vibration signatures even in such a highly noisy environment.

5 CLASSIFICATION AND RECOGNITION

We use the shapes of the extracted waveforms as features because the shapes retain both time and frequency domain information of the waveforms and are thus more suited for use in classification. After obtaining the time-series based vibration signatures, VibroTag uses them to build training models for classification. As VibroTag needs to compare vibration signatures obtained for different surfaces, we need a comparison metric that provides an effective measure of the similarity between vibration signatures of two surfaces. To achieve this, VibroTag uses the technique of Dynamic Time Warping (DTW) that calculates the distance between waveforms by performing optimal alignment between them. Using DTW distance as the comparison metric between vibration signatures, VibroTag trains a k-nearest neighbour (kNN) classifier using those signatures.

DTW is a dynamic programming based solution for obtaining the minimum distance alignment between any two waveforms. DTW can handle waveforms of different lengths and allows a non-linear mapping of one waveform to another by minimizing the distance between the two waveforms. In contrast to euclidean distance, DTW gives us the intuitive distance between two waveforms by determining the minimum distance warping path between them even if they are

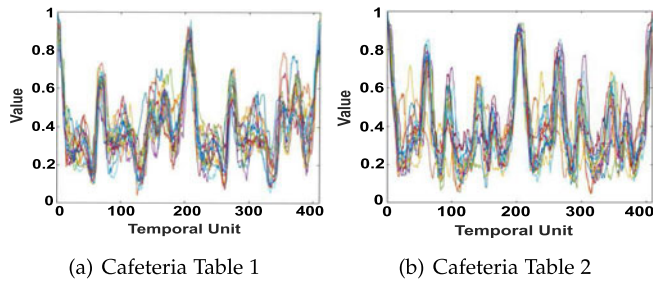


Fig. 8. Extracted time-series features (high noise).

distorted or shifted versions of each other. DTW distance is the euclidean distance of the optimal warping path between two waveforms calculated under boundary conditions and local path constraints. In our experiments, DTW distance proves to be effective for comparing two vibration signatures of different surfaces. Fig. 9a shows the colormap of DTW distance between the vibration signatures extracted by VibroTag from the experiments corresponding to the bed and the sofa (12 signatures each). The average DTW distance among signatures of the bed is ~ 2.3 and that for the sofa is ~ 3.1 . However, the average DTW distance between the vibration signatures of the two surfaces was 16.59. Fig. 9b shows the color map of euclidean distance between features obtained using the IMUs based scheme proposed in [4]. We can see that IMU based features cannot successfully differentiate between the two surfaces due to high similarity, whereas VibroTag's signatures are significantly better at differentiating the two seemingly similar surfaces.

VibroTag requires training data for the surfaces to be recognized. Afterwards, it trains a kNN classifier using the vibration signatures corresponding to those surfaces. To recognize a surface, VibroTag feeds the detected vibration signature of that surface to the trained kNN classifier. The kNN classifier searches for the majority class label among k nearest neighbors of the corresponding vibration signature using the DTW distance metric. VibroTag declares the majority class label obtained from the kNN classifier as label of the tested surface. In the current implementation of VibroTag, we chose $k = 5$ so that the classification process averages more voters in each prediction, which makes our classifier more resilient to outliers.

6 IMPLEMENTATION AND EVALUATION

6.1 Implementation Details

We developed an Android application for generating vibrations and sampling sound signals simultaneously (Fig. 10a shows VibroTag's UI). Our application can record sound in a separate high priority asynchronous thread which helps minimize sampling related irregularities. We use a 16 bit PCM encoding on Mono channel with a sampling rate of 44,100 Hz for sound recording. We also record the data from the smartphone's IMU sensors (i.e., accelerometer and gyroscope) in a separate high priority thread. We use this data to implement the state-of-the-art IMUs based vibration sensing approach for single COTS smartphones proposed in [4], and then compare its surface recognition accuracy with VibroTag. Each data instance constitutes ~ 3 seconds of sound and IMU data, during which the vibration motor keeps vibrating. Our application controls a smartphone's vibrator motor only in terms of turning it ON or OFF, and therefore, does not change the amplitude or pattern of the

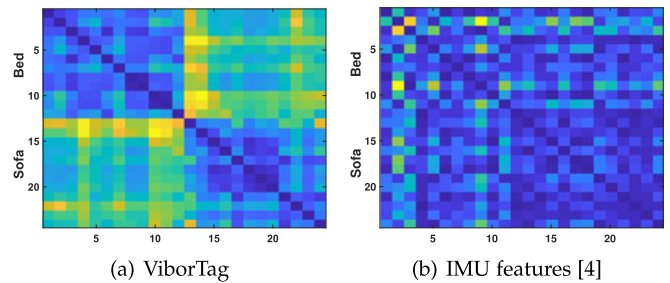


Fig. 9. Colormaps of distance between features of (a) VibroTag (DTW) & (b) IMU scheme (Euclidean).

vibrations. This allows the smartphone to vibrate at its default vibration settings, which helps keep data samples collected at the same surface/location consistent. Moreover, it makes VibroTag applicable to smartphones which do not provide any amplitude control over their vibration motors. We evaluated VibroTag using two smartphones, i.e., Google Nexus 4 and OnePlus 2.

6.2 Evaluation Setup

We evaluated VibroTag's performance by conducting extensive experiments in two different type of environments, i.e., office and apartment. We selected these environments because they represent real-world use case scenarios where a user interacts with different objects and surfaces regularly. We collected data from 4 different volunteers, three with Nexus 4 and one with OnePlus 2, whom we name User-1 (Nexus), User-2 (Nexus), User-3 (OnePlus 2), User-4 (Nexus), respectively. All volunteers were university students who lived in different apartment complexes. No restrictions were imposed on the movement or work conditions of people residing/working in the apartments/office. For example, when we collected data in the office environment, other people in the office were working and chatting as they do on a normal working day. Similarly, data collection did not cause interference in the daily activities (cooking, eating, watching TV, cleaning, etc.) in the volunteers' apartment mates. Therefore, our evaluation of VibroTag takes into account realistic environments where noise from human activities is present most of the time. We used metrics such as confusion matrices, True-Positive-Rates (TPRs) and False-Positive-Rates (FPRs) to evaluate VibroTag's classification performance. We also compare VibroTag's performance with the IMUs based approach in [4].

6.3 VibroTag's Sensitivity

We define VibroTag's sensitivity as its ability to differentiate between different positions and orientations of the smartphone placed on the same location/surface. For example, a user can place his smartphone on his office table in several possible positions and different orientations. VibroTag's sensitivity is an important metric since we claim that a user can place his smartphone on a surface with reasonable flexibility, without having to worry about centimeter level differences in its position and orientation (unlike e.g., EchoTag [18]). This claim will not be satisfied if VibroTag is too sensitive.

To understand VibroTag's sensitivity, each volunteer collected data in his restroom (on the toilet tank), on his bed, bedroom table, living room table and living room sofa. Users collected 25 to 30 samples from each surface for three different smartphone placement scenarios i.e., (1) *least*



Fig. 10. VibroTag Setup (a) VibroTag App, (b) office environment, (c) example of data collection locations in office, (d) example of surfaces used for data collection, and (e) example of data collection locations in apartment.

restricted, (2) moderately restricted and (3) highly restricted. Each scenario corresponded to three rectangular regions of different sizes. We marked the highly restricted region to be approximately within a few inches of the same dimensions as that of the smartphone, the moderately restricted region to be ~ 4 times larger than that of the highly restricted region, and the least restricted region to be about ~ 3 times larger in size compared to moderately restricted region. For example, Figs. 1d and 1e show the marked regions for a sofa and a worktable, respectively, in an apartment. For the experiments related to the least restricted region, volunteers were allowed to place their smartphone even beyond the third zone, as long as their smartphone was placed on the same surface. Fig. 11 shows confusion matrix plots for the tested 5 classes, (namely Bed, Living Room Table, Living Room Sofa, Restroom Ledge and Kitchen Counter), for both User-1 using Nexus 4 (Figs. 11a, 11b, and 11c) and User-2 using OnePlus 2 (Figs. 11d, 11e, and 11f). For each scenario, confusion matrices are plotted using results from 2-fold cross-validation. We observe that for both users, highly restricting the device placement results in highest average prediction accuracy, (i.e., 92.16 and 97.03 percent respectively) which gradually decreases as restriction on smartphone's position and orientation changes from high to least. From the confusion plots (Fig. 11), we observe that the accuracies corresponding to User-2 are higher than User-1's. This may be because either OnePlus 2 is able to extract better quality vibration signatures than Nexus 4, or because User-2's environment and the tested surfaces therein were different from User-1's (e.g., User-1's bedroom table might have some light objects (e.g., keys) placed close to the smartphone that created noise when responding to the vibration). We discuss such impact of surrounding objects on VibroTag in Section 6.4.

Our results show that average accuracy corresponding to User-1's moderately restricted scenario are higher than highly restricted scenario (Figs. 11a, 11b, and 11c), which may be attributed to more noisy samples obtained during highly restricted scenario. Figs. 12a and 12b show average accuracy of all 5 classes for all 3 restriction scenarios obtained with 2-fold, 3-fold, 4-fold and 5-fold (i.e., increasing percentage data used for training from 50 to 80

percent) cross-validation classification experiments. We observe that VibroTag performs well for all 3 restriction scenarios even when only 50 percent data is used for training and remaining for testing, and accuracies for even least restricted scenarios reach as high as 87 percent for User-1 and 95 percent for User-2 when percentage of training data is 80 percent.

6.4 VibroTag's Accuracy

6.4.1 Location Recognition Accuracy

VibroTag achieves an average 4-fold accuracy of 86.55 percent when identifying different locations, whereas the IMU based approach achieves only 49.25 percent. Table 1 shows average 2-fold and 4-fold classification accuracies obtained for 24 different locations/surfaces by User-1. For these experiment, User-1 collected 30-35 samples from each of those locations in a moderately restricted manner. Our results show that VibroTag achieves an average (4-fold) recognition accuracy of 86.55 percent, which is 37 percent higher than the average accuracy achieved by the latest IMUs based approach [4], which achieves only 49.25 percent. Moreover, the lowest accuracy achieved by VibroTag is 70.33 percent, whereas the IMUs based method's accuracy goes as low as 16.38 percent. This shows that the features extracted using VibroTag can successfully differentiate between different locations/surfaces, even when the surfaces are made of very similar material (e.g., wood chair versus wood table, or metal drawer versus metal shelf).

6.4.2 Location Recognition Accuracy Over Days

VibroTag can maintain an average accuracy of up to 85 percent using training samples obtained for 3-4 days only. VibroTag's accuracy can change over days due to several different reasons, for example, changes in environmental noise and/or changes in position of other items placed on a surface (e.g., light objects such as keys, etc.). Next, we explore how VibroTag's accuracies change over days and how much training VibroTag requires to maintain high accuracies when testing on data from a new day.

Fig. 13b shows average cross-validation accuracy over all classes for data obtained from User-1 on 5 different days.

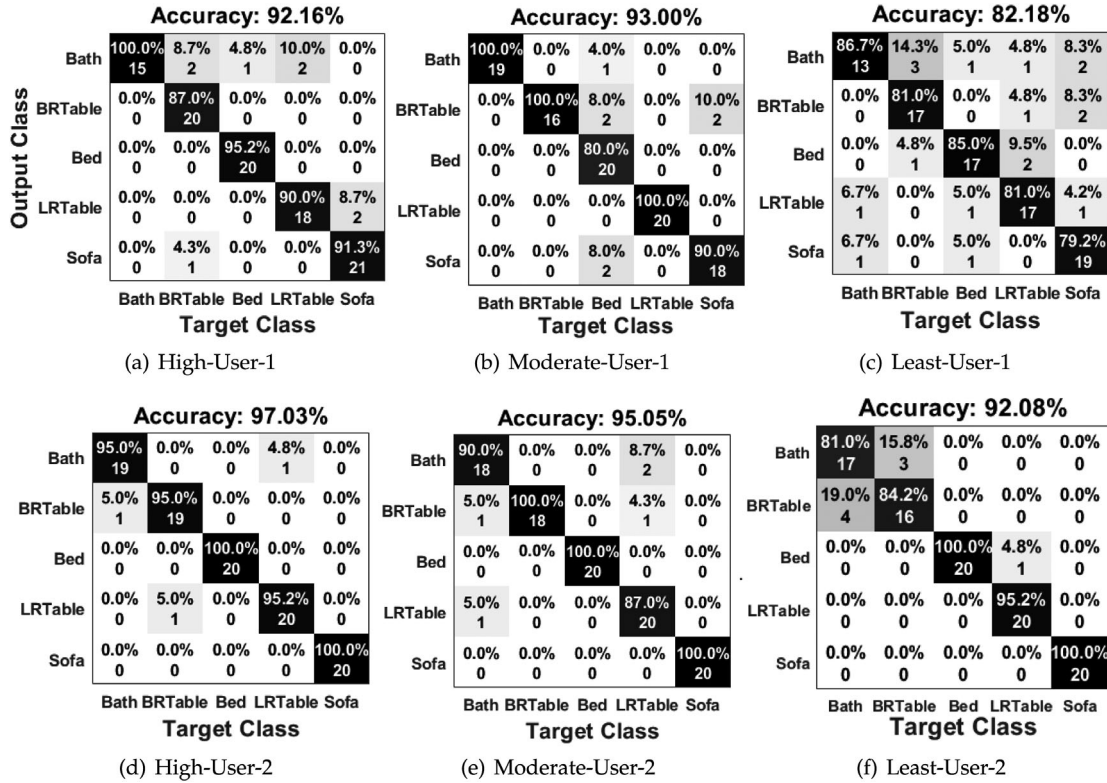


Fig. 11. Confusion matrices for experiments performed by User-1 and User-2 to determine VibroTag's sensitivity.

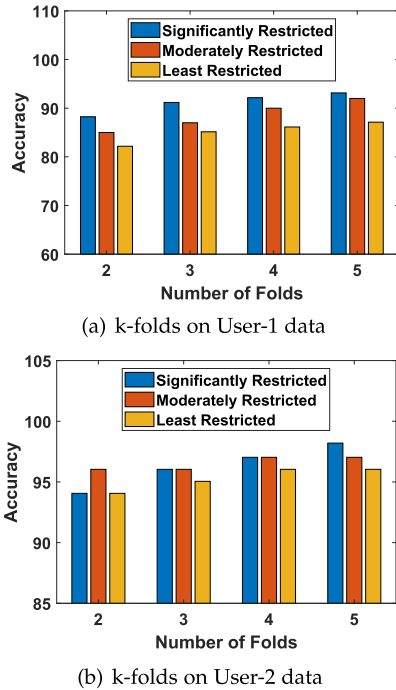


Fig. 12. Average accuracy with increasing number of training samples (VibroTag's sensitivity experiments).

The figure also shows cross-validation accuracy and confusion matrix obtained when data from all 5 days was combined. For these experiments, User-1 collected 6-20 samples from 10 different locations (i.e., Workplace (office table), Bed, Living Room Table, Kitchen (on marble counter), Car (small compartment in front of the gear stick), Hand, Living Room Sofa, Restroom (on toilet tank), Bedroom Table and

Pocket) every day. For this set of experiments, data was collected for both highly and moderately restricted smartphone placement scenarios. Our results show that VibroTag achieves at least 79 percent (and at most 87 percent) accuracy everyday when using only 50 percent of a day's data for training. Fig. 13c shows how combining data from previous days improves accuracy for data collected on the subsequent days from User-1. We observe that VibroTag can achieve accuracy of more than 80 percent on the unknown samples on day 5 for both restriction scenarios. Moreover, we observe that the accuracy of moderately restricted smartphone placement scenarios approaches highly restricted scenarios.

To understand how VibroTag's accuracy changes over days across multiple users, we collected 5 samples from Users 2, 3, and 4 for 4 different locations (i.e., Kitchen (on marble counter), Living Room Sofa, Restroom (on toilet tank), Bedroom Table and Living Room Table) for 20, 10, and 10 consecutive days, respectively. Fig. 14 shows how the classification accuracy changes for different users, where VibroTag is trained using the data from previous days and test on the subsequent days. We observe that the accuracy generally increases with days for User-2, however, there is a major dip for User-3 on day 5 and for User-4 between days 5-7, which can be attributed to major changes in the surrounding environment in terms of noise and/or addition/removal of different objects (such as keys or a pen) on the surface.

6.4.3 Impact of Surrounding Objects

To understand the impact of surrounding objects on VibroTag's accuracy, we performed two different sets of experiments. In the first set of experiments, we collected data on a participant's bedroom table before and after removing 4 different heavier objects (i.e., a guitar, an LCD, a laptop and a mug) from the table one by one. Fig. 15 shows

TABLE 1
Average Accuracy of Recognizing Different Surfaces in Office and Apartment Scenarios

Office environment												Apartment environment												
	Pages Bundle	Printer	Foam Chair	Metal Drawer	Carpet	Wooden Chair	Metal Shelf	Mouse Pad	Leather Chair	Center Desk (Wood)	Cardboard Box	XBox	Work Desk (Wood)	Window Ledge (Marble)	Bathtub Ledge	Living Table (Wood)	Glass Table	Kitchen Counter	Fridge	Wooden Floor	Bedroom Table (Wood)	TV Table (Wood)	Living Room Sofa	Microwave
VibroTag's Accuracy																								
2-Fold	77.60	97.72	68.50	83.31	86.55	88.31	89.09	78.53	91.76	79.02	95.79	73.46	80.62	92.86	93.19	81.9	79.03	87.36	87.65	84.31	82.29	76.39	72.94	91.06
4-Fold	80.87	99.28	70.33	86.77	91.15	88.69	90.84	81.25	93.19	81.53	95.52	75.96	83.49	96.16	94.06	82.84	82.43	91.17	90.09	88.26	88.26	77.61	76.99	92.56
State-of-the-Art IMUs Based Method's Accuracy [4]																								
2-Fold	68.64	87	30.35	68.3	35.13	94.94	89.96	64.05	75.45	27.07	34.29	19.55	49.41	65.08	38.78	25.18	16.89	19.42	56.60	32.98	51.10	30.07	29.31	64.98
4-Fold	75.63	90.39	31.23	70.02	37.48	95.91	91.52	66.31	78.23	29.69	40.21	21.56	52.92	66.95	40.26	25.92	17.04	20.23	63.40	33.98	52.35	32.53	32.27	65.27

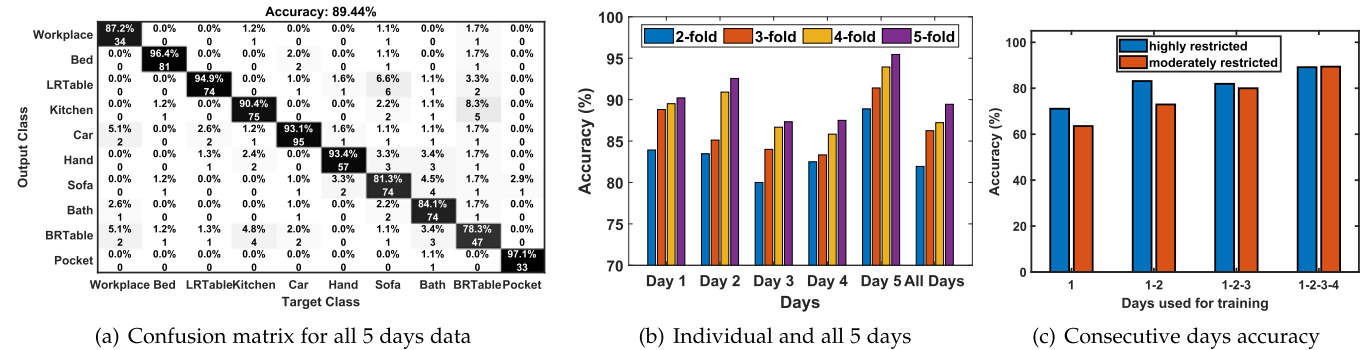


Fig. 13. (a) Average 4-fold cross-validation accuracies over all classes (User-1) (moderately restricted experiments), (b) Confusion matrix after cross-validation, and (c) Training on data from previous days, testing on subsequent days.

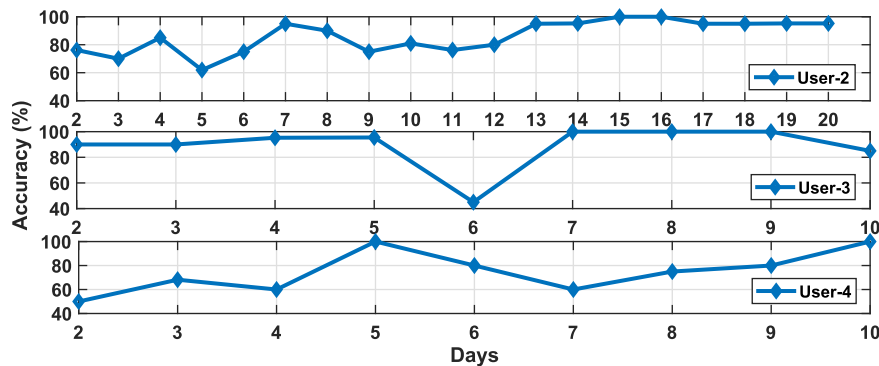


Fig. 14. Accuracies on consecutive days.

the setup for these experiments, where Fig. 15a corresponds to the scenario where all objects were on the table, and Fig. 15d corresponds to the scenario where all three objects were removed. In the second set of experiments, we collected data on the participant's living room table by bringing 4 different lighter objects (i.e., a cup, a set of keys, a pen, and a water bottle) closer to the smartphone. Fig. 16 shows the setup, where Fig. 16a shows the different objects used in these experiments, and Figs. 16b, 16c, and 16d shows a set of keys being brought closer to the smartphone. Fig. 17 shows results for the aforementioned sets of experiments.

We observe from Fig. 17a that as the lighter objects come closer to the smartphone (i.e., down from 12 inches to 3 inches closer), the classification accuracy of the table decreases significantly. For example, when the pen is within 3 inches of the phone, the accuracy goes as low as 9 percent. From Fig. 17b, we observe that the impact of heavier objects on VibroTag's accuracy is not as significant as the lighter ones, which happens because the energy transferred by smartphone's vibration is not enough to make those objects vibrate significantly. However, we observe that the classification accuracy still drops more than 10 percent as we slowly remove the objects that were previously placed on the table. This is

because each of those objects has its own vibration response which was contributing to the overall vibration signature of the bedroom table. So, when those objects are removed one by one, their response is subsequently omitted in the new vibration signature of the surface, which leads to loss in the classification performance.

6.4.4 Impact of Upper Cut-Off Frequency

VibroTag achieves best accuracy when frequencies above 5500 Hz are filtered out from the recorded sound signals. A smartphone's microphone can usually capture sounds in the frequency range of 20 Hz - 20 kHz. However, the smartphone's vibration usually causes variations in lower frequencies, and therefore, filtering out higher frequencies can reduce impact of background noise and any unwanted noisy variations. To understand which frequencies can be filtered out to achieve best accuracy in VibroTag, we employ a Butterworth band-pass filter, and determine the average accuracy for 5 different upper cut-off frequencies of the filter. Fig. 18 shows how User-3's (OnePlus 2) multi-fold cross-validation accuracies vary as upper cut-off frequency increase from 1500 to 12,000. We observe that VibroTag achieves best accuracy at cut-off frequency of 5500 Hz. As the upper cut-off frequency

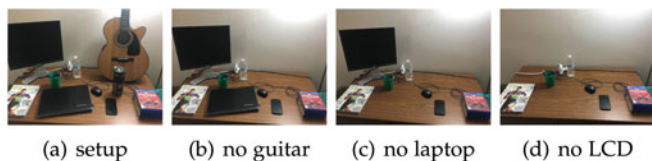


Fig. 15. Removing things from bedroom table.

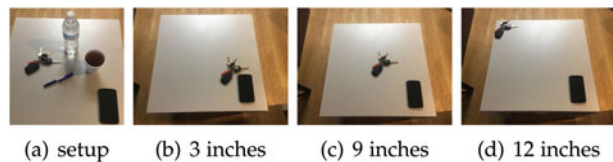
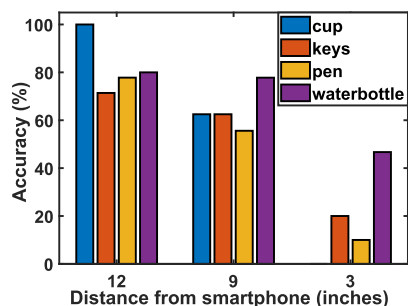
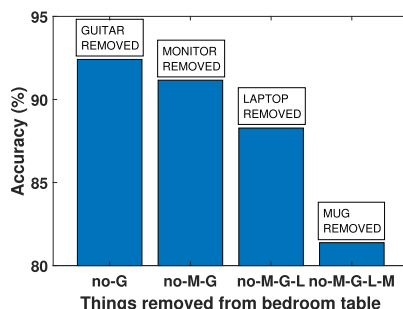


Fig. 16. Bringing light objects closer to smartphone.



(a) Living Room Table



(b) Bedroom Table

Fig. 17. Effect of (a) moving objects closer and of (b) removing objects on classification.

increases, it allows higher frequency noisy variations in the vibration signatures, which leads to lower classification accuracies. We observed similar results for other users as well. Therefore, all accuracies reported in this paper correspond to 5500 Hz cut-off frequency. Note that other smartphones may exhibit better accuracies for cut-off frequencies which are slightly different from 5500 Hz, however, this is an aspect which is out of the scope of this paper.

6.5 Usability Study

We carried out a usability study and asked 24 participants (20 male, 4 female), recruited at university, about the flexibility and usability of the VibroTag application in daily life. The participants comprised of students and university employees of ages 19 to 35. They were first briefed about the working of VibroTag and then its target applications such as symbolic localization. The volunteers were shown the VibroTag application interface as pictured in Fig. 10a and given a demo of the acoustic trace collection. They were

Authorized licensed use limited to: GM eLibrary. Downloaded on March 25, 2025 at 01:05:09 UTC from IEEE Xplore. Restrictions apply.

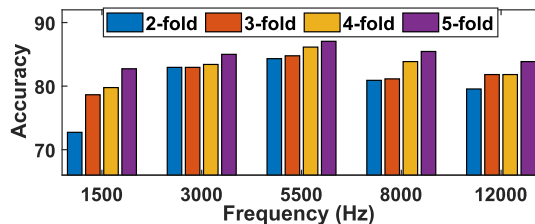


Fig. 18. Cross-validation accuracies for different band-pass filter upper cut-off frequencies (User-3).

also briefed on how the smartphone can be placed on a surface with 3 different levels of restriction flexibility. At the end, they were given a set of usability questions given below and summarized in Fig. 19: Q1. Are you comfortable using smartphone for location recognition? Q2. Can VibroTag help you save time by setting reminders? Q3. Are you comfortable with VibroTags' use of vibration? Q4. Is it easy to place smartphone on preferred locations for learning? Q5. Can VibroTag help you in setting smart notifications linked to locations? Q6. Is VibroTag useful in activating other smart applications? Q7. Do you find VibroTag application valuable and fun to use? Our study indicates reasonably high agreement on the usefulness of VibroTag based smart notifications and reminders.

7 DISCUSSIONS AND FUTURE WORK

Impact of Smartphone's Motion. During our experiments, we did not constrain the motion of smartphones during vibration. Therefore, there were several occasions (e.g., on a glass table) where smartphones moved as well while they vibrated. We assume that such motions are not significant and essentially a feature of the surface. However, if such motions are significant, VibroTag's accuracy may decrease as it can be sensitive to different positions and orientations on the same location/surface as shown in Section 6.3.

Comparison With Microphone-Based Approaches. The existing works that use microphones to sense vibration ignore hardware based irregularities in a smartphone's vibration pattern, as well as, the impact of short-term and constant background noises. The focus of our work is not to compare with existing microphone-based approaches but instead to identify their loop-holes and propose a better technique for fine-grained vibration-based sensing that is relatively more robust to hardware-based irregularities and environmental noises. Moreover, it is difficult to perform an apples-to-apples comparison of our work with prior art due to three key reasons: (1) lack of a shared dataset, (2) lack of prior art's code implementations where all system parameters are clearly defined, and (3) creating scenarios that clearly show the advantage of our technique over previous ones. Although, we have discussed some such scenarios in Figs. 3 and 4 of Section 4, yet to compare the performance of VibroTag features versus FFT features (similar to [8], [9]), we collected 35 samples each from two very similar looking surfaces (a sofa and a cushion-chair) in a noisy living room with a movie playing on TV and people talking in the background, and then evaluated the classification accuracy of both approaches. We collected data using two different smartphones: i.e., OnePlus 2 and Galaxy S8. For OnePlus 2, during feature extraction, VibroTag was not able to extract minimum required number of vibration patterns from one

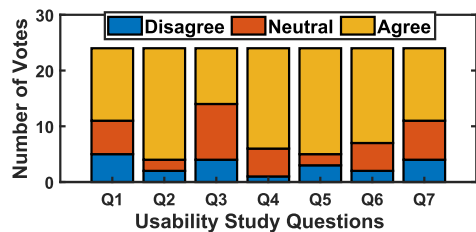


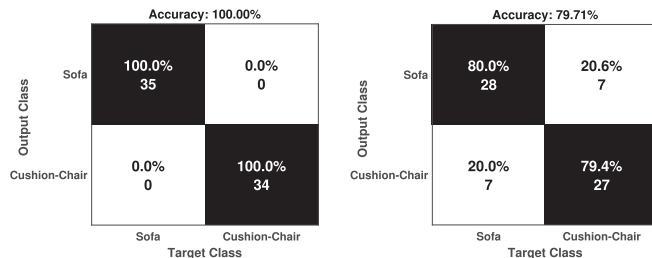
Fig. 19. Vote distribution of 7 VibroTag's usability questions asked from 24 participants.

of the cushion-chair samples. Therefore, we removed that sample from FFT based analysis as well for a fair comparison. Similar issue happened with Galaxy S8 where 4 out of the 35 cushion-chair samples were discarded. Figs. 20 and 21 show the classification results (2-fold cross-validation) showcasing the advantage of our approach. Although we acknowledge the fact that our technique needs to be further compared with FFT based techniques over longer periods of time and on various different smartphones, yet we feel that such an extended comparative study would require significantly more resources and data collection efforts, which are out of this paper's scope.

VibroTag Activation Mechanisms. VibroTag can be activated based on motion dynamics of a smartphone. For example, a smartphone can detect if it was just placed on a surface and then start vibrating after a certain threshold amount of time has passed during which it stayed stationary.

User Discomfort Due to Vibrations. Note that the smartphones we use are off-the-shelf, and therefore, already satisfy the noise constraints set by the standardization bodies. However, we still observed a significant number of 'Neutral' comments regarding the use of vibration from our usability study (as shown in Fig. 19, Section 6.5). These comments were mainly related to the use of VibroTag in specific scenarios, e.g., vibrating on the surface of the table in a meeting room, vibrating in pockets of a pant, etc. It is possible to reduce such concerns by allowing the users to adjust the level of vibration used by our VibroTag app installed on their phones. Although, the smartphones we use in our evaluation do not allow changing the vibration levels, yet some smartphones do come equipped with such capabilities and therefore might be more conducive to VibroTag's application scenarios.

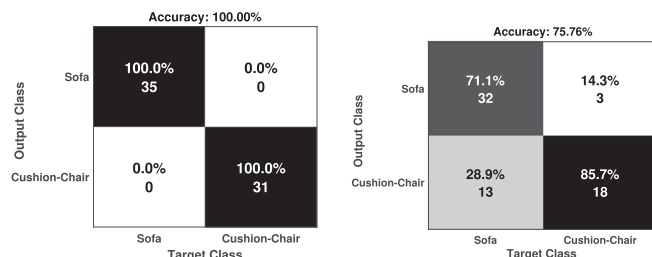
Computational and Memory Footprint. Right now, VibroTag runs offline on a remote server, where the acoustic and IMU data is automatically uploaded after every experiment by VibroTag's smartphone app. However, VibroTag can be implemented completely on a smartphone as well given that its computational and memory footprint is reasonable enough for an on-device implementation. Although kNN classifiers are exhaustive search based algorithms and there is a concern that the memory and computational complexity will increase as the number of training samples per surface increases, yet we assume that an average user will collect only a few samples from a few surfaces during the course of use. Note that we choose kNN-based classification process mainly because kNN is considered a "gold standard" for DTW based time-series classification. Our results also show that the kNN-DTW combination works reasonably well and gives high classification accuracies. We leave the



(a) VibroTag features with kNN

(b) FFT features with SVM

Fig. 20. Classification performance of VibroTag versus FFT features in terms of differentiating even two similar looking surfaces in a noisy living room (OnePlus 2).



(a) VibroTag features with kNN

(b) FFT features with SVM

Fig. 21. Classification performance of VibroTag versus FFT features in terms of differentiating even two similar looking surfaces in a noisy living room (Galaxy S8).

task of exploring computationally more efficient alternatives for our classification module as part of future work.

Impact of Different Hardware. VibroTag is an early step towards time-series based vibration signature extraction for surface recognition using off-the-shelf smartphones. There is obviously room for continued evaluation such as evaluating VibroTag on more smartphone models, determining impact of different types of vibration motors and smartphone sizes on sensing accuracy, etc. However, the goal of this paper was to dive deeper into the physical nature of vibrations and focus on extracting fine-grained vibration signatures that are robust to hardware irregularities and background environmental noises. To the best of our knowledge, this is the first work that addresses the hardware based irregularities in a smartphone's acoustic vibration pattern, as well as, the impact of short-term and constant background noises (e.g., intermittent talking, clapping, exhaust fan, TV programs, etc.) on vibration based sensing using smartphone microphones. In this first work, we have shown that the proposed techniques work well for two different smartphones with two different types of vibration motors. Since the primary focus of this work was to develop the techniques to achieve high accuracy fine-grained vibration based surface recognition using a single smartphone and show its possibility on a small testbed in a resource-limited setting, covering every aspect related to the effect of various variables on the experimental results is out of the scope of this study and left as part of future work.

8 CONCLUSION

In this work, we make the following contributions. First, we propose the first fine-grained vibration based sensing

scheme, that can recognize different surfaces using the vibration mechanism and microphone of a single COTS smartphone. The intuition is that the smartphone's vibration causes the whole smartphone structure and the hardware inside it to vibrate in a peculiar pattern, which depends upon the absorption properties of the surface that the smartphone is placed on. These vibrations produce peculiar sound waves that we detect using the smartphone's microphone. Second, we propose a time-series based signal processing technique to extract fine-grained vibration signatures that are robust to hardware irregularities and background environmental noises. We implemented VibroTag on two different Android phones and evaluated in multiple different environments. Our results show that VibroTag achieves an average surface recognition accuracy of 86.55 percent, which is 37 percent higher than the average accuracy of only 49.25 percent achieved by the state-of-the-art IMUs based schemes.

ACKNOWLEDGMENTS

The authors would like to thank Salman Ali for his significant help with experiments and conducting the usability study.

REFERENCES

- [1] J. Griffin and S. Fyke, "User hand detection for wireless devices," CA Patent 2 532 692, May 2011.
- [2] I. Shafer, "Learning location from vibration," 2013. [Online]. Available: <http://www.mrcaps.com/>
- [3] K. Kunze and P. Lukowicz, "Symbolic object localization through active sampling of acceleration and sound signatures," in *Proc. Int. Conf. Ubiquitous Comput.*, 2007, pp. 163–180.
- [4] J. Cho, I. Hwang, and S. Oh, "Vibration-based surface recognition for smartphones," in *Proc. IEEE Int. Conf. Embedded Real-Time Comput. Syst. Appl.*, 2012, pp. 459–464.
- [5] J. Ho and S. S. Intille, "Using context-aware computing to reduce the perceived burden of interruptions from mobile devices," in *Proc. ACM Conf. Hum. Factors Comput. Syst.*, 2005, pp. 909–918.
- [6] M. Goel, J. Wobbrock, and S. Patel, "Gripsense: Using built-in sensors to detect hand posture and pressure on commodity mobile phones," in *Proc. 25th Annu. ACM Symp. User Interface Softw. Technol.*, 2012, pp. 545–554.
- [7] J. Wiese, T. S. Saponas, and A. J. Brush, "Phoneprioception: Enabling mobile phones to infer where they are kept," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2013, pp. 2157–2166.
- [8] S. Hwang and K. Wahn, "Vibrotactor: Low-cost placement-aware technique using vibration echoes on mobile devices," in *Proc. 18th Int. Conf. Intell. User Interfaces*, 2013, pp. 73–74.
- [9] R. Darbar and D. Samanta, "Surfacesense: Smartphone can recognize where it is kept," in *Proc. ACM 7th Int. Conf. HCI*, 2015, pp. 39–46.
- [10] J. Cho, I. Hwang, and S. Oh, "Vibephone: Efficient surface recognition for smartphones using vibration," *Pattern Anal. Appl.*, vol. 19, no. 1, pp. 251–265, 2016.
- [11] G. Laput, R. Xiao, and C. Harrison, "Viband: High-fidelity bioacoustic sensing using commodity smartwatch accelerometers," in *Proc. 29th Annu. Symp. User Interface Softw. Technol.*, 2016, pp. 321–333.
- [12] J. Liu, Y. Chen, and M. Gruteser, "Vibkeyboard: Virtual keyboard leveraging physical vibration," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 507–508.
- [13] J. Liu, Y. Chen, M. Gruteser, and Y. Wang, "Vibsense: Sensing touches on ubiquitous surfaces through vibration," in *Proc. 14th Annu. IEEE Int. Conf. Sens., Commun., Netw.*, 2017, pp. 1–9.
- [14] M. Ono, B. Shizuki, and J. Tanaka, "Touch & activate: Adding interactivity to existing objects using active acoustic sensing," in *Proc. 26th Annu. ACM Sympo. User Interface Softw. Technol.*, 2013, pp. 31–40.
- [15] M. Fan and K. N. Truong, "Soqr: Sonically quantifying the content level inside containers" in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2015, pp. 3–14.
- [16] M. Goel et al., "Surfacelink: Using inertial and acoustic sensing to enable multi-device interaction on a surface," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2014, pp. 1387–1396.
- [17] Y.-C. Tung and K. G. Shin, "Expansion of human-phone interface by sensing structure-borne sound propagation," in *Proc. 14th Annu. Int. Conf. Mobile Syst., Appl., Serv.*, 2016, pp. 277–289.
- [18] Y.-C. Tung and K. G. Shin, "EchoTag: Accurate infrastructure-free indoor location tagging with smartphones," in *Proc. 14th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2015, pp. 525–536.
- [19] M. Azizyan, I. Constandache, and R. R. Choudhury, "Surroundsense: Mobile phone localization via ambient fingerprinting," in *Proc. 15th Annu. Int. Conf. Mobile Comput. Netw.*, 2009, pp. 261–272.
- [20] Lightninghawk, "Haptic motor driver hook-up guide," 2018. [Online]. Available: <https://learn.sparkfun.com/tutorials/haptic-motor-driver-hook-up-guide/erm-and-lra-motors>



Kamran Ali received with the PhD degree in computer science and engineering from Michigan State University, USA, in 2019. He is currently at General Motors R&D. He received the BS degree in electrical engineering and computer sciences, from the School of Science and Engineering (SSE) of Lahore University of Management Sciences (LUMS), Pakistan, in 2013. He has worked with Hewlett Packard Labs, NEC Laboratories, Nokia Bell Labs, and Microsoft's Applied Sciences Group. His research interests include the areas of communication systems, sensing systems, RF sensing, machine learning, signal processing, pervasive and mobile computing, and Internet of Things.



Alex X. Liu (Fellow, IEEE) received the PhD degree in computer science from The University of Texas at Austin, Austin, TX, USA, in 2006. He is currently an honorary professor at the Qilu University of Technology. He was a professor with the Department of Computer Science and Engineering, Michigan State University. His research interests include networking, security, and privacy. He was the recipient of the IEEE and the IFIP William C. Carter Award in 2004, the National Science Foundation CAREER Award in 2009, the MSU Withrow Distinguished Scholar (Junior) Award in 2011, and the MSU Withrow Distinguished Scholar (Senior) Award in 2019. He was the editor of the *IEEE/ACM Transactions on Networking* and an area editor for the *Computer Communications*. He is currently an associate editor for *IEEE Transactions on Dependable and Secure Computing* and *IEEE Transactions on Mobile Computing*. He was the technical program committee co-chair of the ICNP 2014 and the IFIP Networking 2019. He was the recipient of the Best Paper Awards from the SECON-2018, ICNP-2012, SRDS-2012, and LISA-2010. He is also an ACM distinguished scientist.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.