Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.DOI

Fingerprinting Smartphones Based on Microphone Characteristics from Environment Affected Recordings

ADRIANA BERDICH[1], BOGDAN GROZA[1], EFRAT LEVY[2], ASAF SHABTAI[2], YUVAL ELOVICI[2], and RENÉ MAYRHOFER[3]

[1]Faculty of Automatics and Computers, Politehnica University of Timisoara, Blvd. V. Pârvan Nr. 2, Timisoara, 300223, Timis, Romania (e-mail: adriana.berdich@aut.upt.ro and bogdan.groza@aut.upt.ro)

[2]Faculty of Information Systems Engineering, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva, 8410501, Israel (e-mail: elevy@post.bgu.ac.il, shabtaia@bgu.ac.il and elovici@bgu.ac.il)

[3]Institute of Networks and Security and LIT Secure and Correct Systems Lab, Johannes Kepler University Linz, Altenbergerstraße 69, Linz, 4040, Austria (e-mail: rm@ins.jku.at)

Corresponding author: Bogdan Groza (e-mail: bogdan.groza@aut.upt.ro).

ABSTRACT Fingerprinting devices based on unique characteristics of their sensors is an important research direction nowadays due to its immediate impact on non-interactive authentications and no less due to privacy implications. In this work, we investigate smartphone fingerprints obtained from microphone data based on recordings containing human speech, environmental sounds and several live recordings performed outdoors. We record a total of 19,200 samples using distinct devices as well as identical microphones placed on the same device in order to check the limits of the approach. To comply with real-world circumstances, we also consider the presence of several types of noise that is specific to the scenarios which we address, e.g., traffic and market noise at distinct volumes, and may reduce the reliability of the data. We analyze several classification techniques based on traditional machine learning algorithms and more advanced deep learning architectures that are put to test in recognizing devices from the recordings they made. The results indicate that the classical Linear Discriminant classifier and a deep-learning Convolutional Neural Network have comparable success rates while outperforming all the rest of the classifiers.

INDEX TERMS machine learning, microphone, smartphone fingerprinting

I. INTRODUCTION AND MOTIVATION

In the recent years, due to the fast evolution of the IoT (Internet of Things) and the stringent need for fast authentication mechanisms that do not call for user interaction, device fingerprinting within the scope of authentication evolved into an important research area that asked for urgent exploration. Nonetheless, privacy related topics and forensic investigations provide complementary use cases of significant interest for inimitable device characteristics.

Contemporary smartphones are equipped with numerous sensors, i.e., microphones, accelerometers, gyroscopes, magnetometers, light sensors, cameras, etc., all of which can be fingerprinted since each sensor has unique characteristics due to chemical and physical imperfections resulting from the fabrication process. The idea of circuit identification based on physical properties was explored since the early 2000s [1]. Later, Physically Unclonable Functions (PUFs) were introduced for security applications such as device authentication based on unique and unpredictable characteristics [2]. However, extracting unique sensor characteristics is challenging because sensor characteristics are also influenced by the environment, regardless of the sensor type, e.g., accelerometer [3], microphone [4], camera [5], etc. In this work, we analyze smartphone fingerprints provided by microphone characteristics using the frequency domain representation of the recorded sounds and machine learning classifiers. Concretely, we use several traditional machine learning algorithms, i.e., Linear Discriminant (LD), Ensemble-Subspace Discriminant (ENS), Decision Tree (DT), Fine K-Nearest Neighbor (KNN) and Linear Support Vector Machines (SVM), to which a deep-learning Convolutional Neural Network (CNN) is added as a comparison.

We are focusing on three distinct scenarios, as depicted in Figure 1, as a result of various types of sounds and

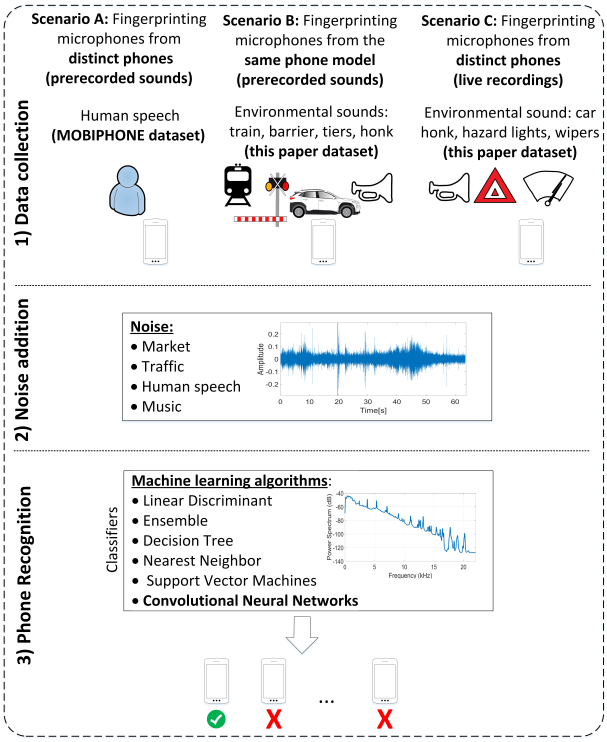


FIGURE 1. Overview of the scenarios and methodology steps in our work

environments. A specific area of application which concerns us are the vehicular environments which recently become of much interest due to the evolution of the automotive domain towards interactions with smart devices that are carried by users. The scenarios from which we collect and analyze data are the following:

Scenario A. Fingerprinting smartphones from different manufacturers and different models based on human speech: for this scenario we use the existing MOBIPHONE dataset [6] which is a public speech database containing 21 smartphones from distinct brands and models. For each smartphone the dataset contains 24 audio files from 12 female and 12 male speakers. The speakers were chosen from the TIMIT database [7]. Each recording file contains 10 spoken sentences, the first two are identical for each speaker while the rest are different.

Scenario B. Fingerprinting identical smartphones on environmental sound using prerecorded sounds: for this scenario we built our own recordings with 16 microphones from the same smartphone model (a Samsung Galaxy S6) which are used to record in-vehicle and traffic noise replayed by a high-end audio system. These experiments were performed indoors since it is much easier to work with a batch of identical microphones which are connected to the same phone in order to determine if the microphone alone (or the rest of the circuits in the smartphone) influences the fingerprinting. To generate environmental sounds, we use the SoundArchive[1] database from which we use sounds corresponding to some events that are commonly encountered in vehicular environments: (i) locomotive signaling departure, (ii) closing barriers with bells jingling, (iii) car screeching tiers and (iv) the horn sound of a car. In Figure 2 we depict the sounds which we use from SoundArchive, played in the indoor experiments (with identical microphones) in the time domain (left) as well as their power spectrum, i.e., the frequency domain representation (right). On each plot there are two signals which correspond to the two channels of a stereo recording.

Scenario C. Fingerprinting smartphones from distinct manufacturers and models based on live recordings: for this scenario we built our own recordings outdoors and inside a vehicle by using 16 distinct smartphones that record the sound at the same time. Each smartphone records sounds in three distinct sub-scenarios:

1) A car honking in an open space to avoid reflections from the nearby obstacles (for these measurements we took the car outside the city on an open area). In this scenario we performed 400 measurements with each smartphone, totaling 6400 measurements. The smartphones were placed outside the car as would be expected in case of bystanders' incidental recordings.
2) Vehicle hazard lights since these are commonly triggered inside cars in various circumstances related to traffic conditions. For this scenario we did 300 measurements for each smartphone, totaling 4800 measurements.
3) Wipers noise as this is also commonly heard inside cars (such a scenario generally occurs due to circumstances caused by the environment). For this scenario we did 300 measurements for each smartphone, totaling 4800 measurements. In the last two settings, the smartphones were placed inside the car.

In real-life circumstances, additional noise may be present in the environment. For this reason, we also analyze the influence of four types of noise on our fingerprinting procedure. For outdoor recordings we consider overlaps with music, for which we used several songs from the top 10 of the Spotify list for 2021. For indoor recordings, we used two environmental noises from the SoundArchive: (i) heavy traffic and (ii) outdoor market sounds. In Figure 3 we graphically depict the representation of these sounds from the SoundArchive in the time domain (left) and frequency domain respectively (right). Each plot contains two signals as the files from the SoundArchive are two-channel, stereo recordings.

The rest of the work is organized as follows. In Section II we analyze some related works. Section III depicts the experimental setup, devices and tools. In Section IV we attempt to fingerprint microphones using the LD, ENS, DT, KNN and SVM classifiers based on prerecorded sounds from indoor experiments. In Section V we fingerprint microphones based on live outdoor recordings using the LD classifier, which was selected as the top performer based on the experiments from the previous scenario, and we also add a more demanding

[1] https://www.soundarchive.online/?s=police

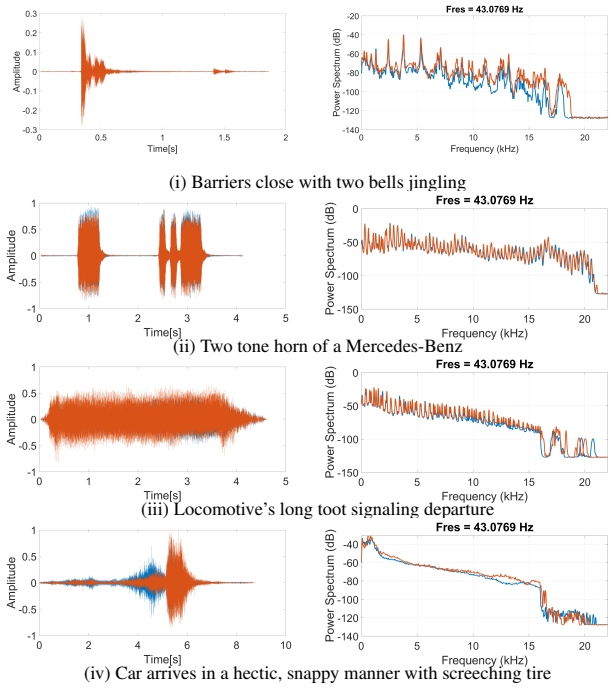


FIGURE 2. Sounds played in our experiments in time domain (left) and power spectrum (right)

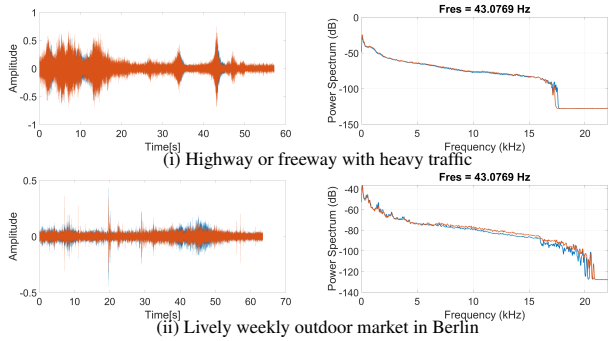


FIGURE 3. Noises used in our experiments in time domain (left) and power spectrum (right)

deep learning CNN architecture. Section VI concludes our work.

II. RELATED WORK

Several lines of work have focused on fingerprinting smartphones based on their microphones and various types of sounds and classification mechanisms were employed. We survey these in what follows, Table 1 provides an overview. Indeed, previous approaches differ not only in the algorithms that they use, i.e., traditional machine learning or deep learning, but also in the features employed for classifying the samples. For example, while most works are using the frequency spectrum extracted via the FFT transform, some works that employed human speech have also been using MFCC coefficients (which are commonly used for speech recognition). We also note that for synthetic recordings, SVM, KNN and CNN were the most used classifiers. Several details on these works follow.

In [8] smartphone microphones are identified based on the recordings of a periodic tone at 1kHz with KNN, SVM and CNNs and [9] uses a similar methodology. Noise produced by a pneumatic hammer and a gun are used in [10]. The microphone classification was realized based on the frequency representation of the recorded sound with KNN, SVM and CNN. Artificial neural networks are used in [12] for microphone identification based on the frequency response for 80 tones ranging between 100Hz and 8kHz. In [13] the microphones are identified using the inter-class cross correlation of the phase spectrum. The microphones were used to record the ambient noise generated with a fan cooler which is positioned at 0.7m from the microphones and runs at the maximum speed.

Audio signal characteristics, e.g., mean, standard deviation, dynamic range D, the crest-factor Q and autocorrelation time are analyzed in [24] within the scope of forensics applications. In [14], one-class classification is used based on noise collected from different locations, i.e., indoors or outdoors, inside a park or on a busy street. Characteristics extracted from FFT coefficients are used in [15] along with machine learning algorithms, i.e., Naive Bayes, multi class SVM, decision trees and KNN. In [11] mobile devices are identified based on two approaches. In one approach, the authors use the frequency response of the speaker and microphone based on the minimum likelihood classification of 13 tones with frequencies between 100Hz and 1300Hz. The other approach is based on calibration errors of the accelerometer sensors.

Also, human speech has been used by several papers for smartphone microphone fingerprinting. In [16] speech recordings from 25 speakers are used for microphone identification with SVM, Gaussian Supervector (GSV) and Sparse Representation-based Classifier (SRC). Speech recordings are used in [18] for microphone identification based on the band energy difference descriptor. CNN classification based on frequency domain representation of human speech is done in [22]. In [19] the smartphone is identified using CNNs based on the spectrogram extracted from speech recordings. SVM-Recursive Feature Elimination (SVM-RFE) and variance threshold are used in [20] for smartphone microphone identification based on speech recordings. Mel-frequency cepstral coefficients (MFCCs) of speech recordings are used in [17] for microphone identification. Audio source identification in the scope of anti-forensics using SVM and MFFC is proposed in [21]. In [23] the smartphone is identified based on MFCC of audio recordings.

An open-set classification algorithm is proposed in [25] for microphone identification. Microphone and environment classification using Naive Bayes is done in [26]. Distinct audio signals were used, i.e., distinct music styles, noises, speech and instrumental. Electrical network frequency (ENF) and SVM are used in [27]. Mobile device identification using deep learning algorithms, i.e., softmax regression and multilayer perceptron (MLP) based on audio recording data,

TABLE 1. Overview of various works which is proposed fingerprinting smartphones based on their microphone

Paper	Type of sound	Classifiers	Devices	Sound
[8]	1kHz and 2kHz tone	SVM, KNN, CNN	32 smartphones	synthetic sound
[9]	1kHz tone	SVM, KNN, CNN	34 smartphones	
[10]	1kHz tone, pneumatic hammer, gunshot	SVM, KNN, CNN	34 smartphones	
[11]	13 tones in the range of 100Hz-1300Hz	maximum-likelihood classification	16 smartphones	
[12]	80 tones in the range of 100Hz-8kHz	artificial neural networks	6 commercial microphones	
[13]	ambient noise generated with a fan cooler	inter-class cross correlation	8 commercial microphones	
[14]	noise: indoor, park, street	one-class classification	5 commercial microphones	
[15]	music, (metal, pop, techno), MLS noise, sine, white noise, digital silence, SQAM instrumental	naive bayes, multi class SVM, decision trees and KNN	7 commercial microphones	
[16]	25 speakers	SVM, GSV and SRC	4 commercial microphones	human speech
[17]	40 speakers	MFCC + GMM	16 commercial microphones	
[18]	4 speakers	band energy difference descriptor	31 + 141 smartphones	
[19]	number not mentioned	CNN	20 smartphones	
[20]	12 + 160 speakers	(SVM-RFE) and variance threshold	24 smartphones	
[21]	160 + 12	MFCC + SVM + CNN	16 smartphones	
[22]	24 speakers (mobiphone [6]) + 3 speakers (own))	CNN	20(mobiphone [6]) + 19 smartphones	
[23]	24 speakers (mobiphone [6])	MFCC + GMM	21 smartphones (mobiphone [6])	
this paper	human speech (mobiphone [6]) + environment sound from horns, tyres, barrier (new dataset) + real sound from horn (new dataset)	LD, ENS, SVM, DT, KNN	16 + 16 smartphones (in addition to the mobiphone dataset)	human speech + synthetic sound (affected by environment noise)

is proposed in [28].

Other papers have used the loudspeaker instead of the microphone for smartphone identification. In [29] and [30] the smartphone loudspeaker is identified based on natural sounds, i.e., instrumental, song and human speech using distinct audio features, i.e., RMS (root-mean-square), ZCR (zero crossings), Low-Energy-Rate, Spectral Centroid, Spectral Entropy, etc. In [31], the Euclidean distance is used for the smartphones loudspeaker identification based on cosine tones between 14kHz and 21kHz with 100Hz increment. SVM, Random Forest (RF), CNN and Recurrent Neural Network-Long Short-Term Memory Neural Network (RNN-BLSTM) based on MFCC and SSF sketches of spectral features extracted from human speech are used in [32] for smartphone loudspeaker identification. In a previous work, we have used a convolutional neural network (CNN) and a Bidirectional Long Short-Term Memory network (BiLSTM) to fingerprint smartphones based on the loudspeaker response to a sweep signal [33]. Interestingly, the BiLSTM network from our previous work [33] performed very poor on microphone data and the CNN required significant modifications for this task. This suggests microphone data to be more challenging for fingerprinting.

Mobile devices identification based on 20 features in time and frequency domain, extracted from accelerometer data is proposed in [34]. A more rarely employed sensor for fingerprinting is the magnetometer. In [35] the mobile devices are identified based on magnetometer fingerprints extracted from 18 features in the time and frequency domains. Multiple features extracted from distinct sensors, i.e., microphone, accelerometer, gyroscope and magnetometer are used in [36] for smartphone identification.

Other fingerprinting attempts have used camera sensors. In [37] a method for fast camera identification and verification in forensics investigations based on Photo-Response Non-Uniformity (PRNU) is proposed. Smartphone identification using camera fingerprints extracted based on hybrid green channel PRNU is proposed in [38].

In addition to smartphone fingerprinting, authentication and secure communication protocols are proposed by other works based on fingerprints extracted from speakers, microphones or other sensors. Wireless device authentication based on fingerprints extracted in the frequency domain from speakers and microphones is proposed in [39]. In [40] and [41] a secure communication system based on ambient audio is proposed. Also, [42] proposes a system for secure mobile devices pairing based on audio fingerprints extracted from the recorded audio data. An acoustic communication mechanism for smartphones based on jamming signals is proposed in [43]. In [44] a two factor authentication system is proposed which works at high frequencies, i.e., between 18kHz and 20kHz. Also, SVM is used to analyze the similarity between the recorded audio data in the time and frequency domains.

III. SETUP AND METHODOLOGY

In this section we give an overview of the devices used in the experiments, the environment configuration and software platforms.

Devices. Our experiments focus on the classification of both distinct and identical smartphones based on their microphones. To make the experiments convincing and account for differences between identical microphones, we dissembled a Samsung Galaxy S6 smartphone and bought 16 identical (original) flex cables with microphones. The Samsung Galaxy S6 microphone is placed on the same board, also referred as the flex cable, with the micro USB charging port,

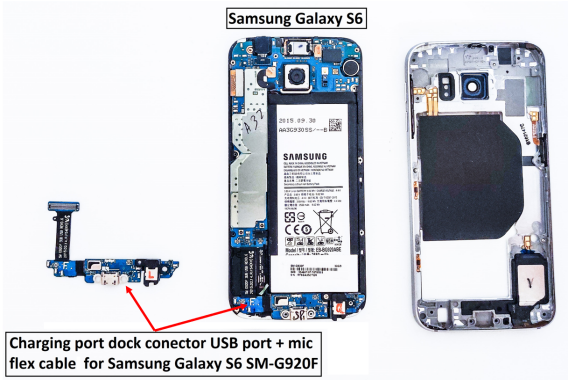


FIGURE 4. Samsung Galaxy S6 dissembled with two flex cables with microphone and charging USB port dock connector

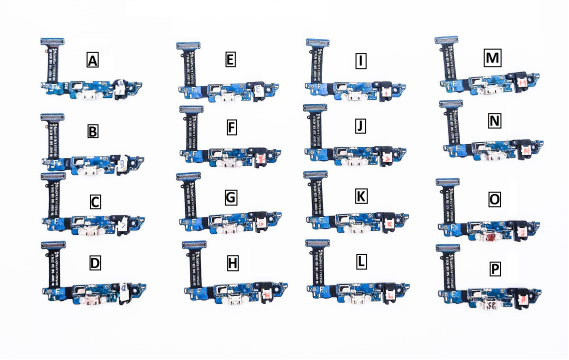


FIGURE 5. 16 flex cables with microphone and charging USB port dock connector for Samsung Galaxy S6

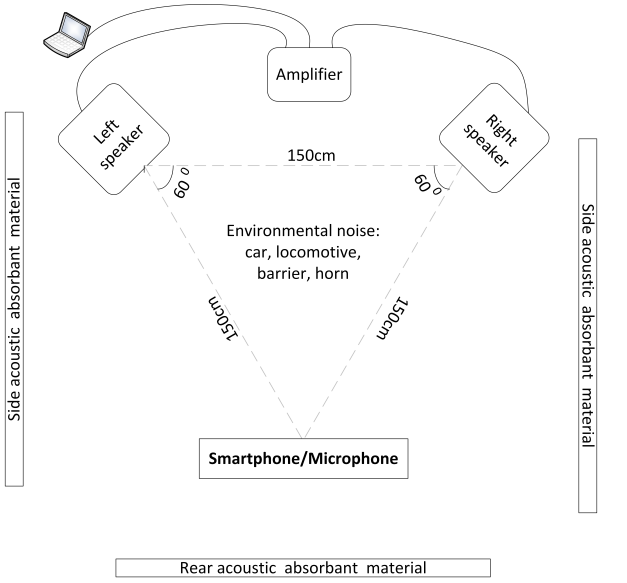


FIGURE 6. Suggestive depiction of our indoor experimental setup

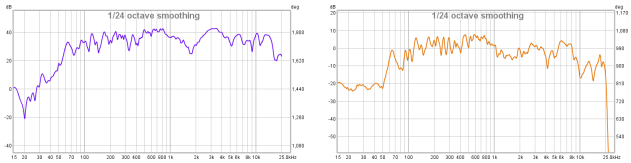


FIGURE 7. Frequency response of the audio system used in our experiments tested with miniDSP UMIK-1 microphone (left) and with a smartphone (right)

jack connector, navigation key and capacitive keys. For an easier replacement of the flex cables we cut out the capacitive keys from the board (as the capacitive keys are glued to the display they are difficult to insert in another case and are of no interest for our experiments). In Figure 4 we present both sides of a dissembled Samsung Galaxy S6 smartphone along with a flex cable nearby. In Figure 5 we present the 16 identical microphones from the Samsung Galaxy S6 on their flex cables. In Table 2 we give a summary of the devices and measurements[2]. We have fingerprinted a total of 32 devices out of which 16 are identical microphones from Samsung Galaxy S6 smartphone which were placed in the same case. The remaining 16 devices are distinct smartphones from various brands as shown in the table.

Tools and environments. For the analysis of the recorded data we used Matlab[3], which is a numerical computation environment commonly used for data analysis, algorithms and model development. In the initial analysis of the recorded data we used the Signal Analyzer application from Matlab 2021a. For the initial analysis of the classification algorithms we used the Classification Learner application from Matlab. Also, for the initial setup calibration we have used Room EQ Wizard[4] (REW), which is a free room acoustic software.

Experimental settings overview. For scenario A we used existing measurements from MOBIPHONE dataset [6].

For the experiments in scenario B, where we fingerprint microphones for the same Samsung Galaxy S6 smartphone based on environmental noise, we carried indoor measurements with already recorded noises since rewiring is necessary, i.e., to replace the microphone of the smartphone, which is difficult to perform outdoors. Moreover, recordings with distinct microphones on the same phone cannot be done at the same time, since each of them has to be separately plugged to the phone, and thus environmental conditions will be dissimilar. Figure 6 gives a graphic depiction of our indoor experimental setup. Since we want to reproduce a large frequency spectrum and low cost speakers cannot cope with this and introduce higher distortions, in our experiments we used a high-quality audio system which was able to reproduce sounds with a more linear response. The audio system used in our experiments contains two professional loudspeakers which can produce a more accurate low-frequency response. Each loudspeaker contains two drivers, one for mid-bass response and another one for high-frequency response. The

[2]the performed measurements are publicly available to serve for future investigations at https://github.com/ABerdich/Microphone-Fingerprint

[3]https://nl.mathworks.com/products/matlab.html

[4]http://roomeqwizard.com/

TABLE 2. Summary of devices and associated measurements

	Phones	Label	No.	Mic.	Meas.	Total
1.	Samsung Galaxy S6	A to P	16	1	200	3200
2.	Samsung Galaxy S6 (other)	S6 and S6′	2	1	1000	2000
3.	Allview V1 Viper I	AV	1	1	1000	1000
4.	Samsung Galaxy J5	J5, J5′ and J5′′	3	1	1000	3000
5.	One Plus 7 Pro	OP	1	2	1000	1000
6.	Samsung Galaxy Tab S7	S7t	1	2	1000	1000
7.	Leagoo Z10	LE and LE'	2	1	1000	2000
8.	Samsung Galaxy A21s	A21s	1	2	1000	1000
9.	Samsung Galaxy S7	S7	1	1	1000	1000
10.	Samsung Galaxy A3	A3	1	2	1000	1000
11.	Motorola E6 plus	MT	1	1	1000	1000
12.	Google Nexus 7	N7	1	2	1000	1000
13.	LG Optimus P700	LG	1	1	1000	1000
	Total		32			19200

amplifier that we used is a class A amplifier which is known to have low distortions. Also, the speakers placement and the audio room/environment is very important to obtain a good quality reproduction. The speakers were placed at 150cm distance one from another, with an interior angle of 60°, forming an equilateral triangle with the recorder (as recommended for stereo reproductions)[5]. The distance between the speakers and the back wall was 50cm. To avoid sound reflections and reverberations, we added an acoustic absorbing material on the side walls at mirror points, on the front and back wall and on the floor. Also, we isolated the corners of the room[6]. In Figure 7 we depict the frequency response of the audio system used in our experiments as tested in REW with a linear sweep signal generated between 0Hz and 20kHZ and recorded with the calibrated microphone UMIK-1 omnidirectional USB from miniDSP. Note that the response is sufficiently linear in the order of +/-5db. For scenario B, we played in a loop each MP3 file with in-vehicle and traffic noises from SoundArchive: (i) a locomotive's long toot signaling departure, (ii) barriers closing with two bells jingling, (iii) a car arriving in a hectic, snappy manner with screeching tiers and (iv) the two tone horn of a Mercedes-Benz. On each of the smartphones, we run an Android application which records and saves the sounds as a PCM and WAV file for analysis. The recordings were done at a sampling rate of 48kHz and 16-bit resolution.

For the experiments in scenario C, we performed outdoor experiments with 16 smartphones from distinct brands which recorded the following:

1) A car honking live for 400 times. This experiment was done in an open space. The car engine was stopped, the smartphones were placed on a board located on the front-right of the car at a distance of 3 meters from the car as we depict in Figure 8. In the recorded files as the car honks some background noise could be also heard. This scenario is more challenging because the honks are not identical, some being shorter and others longer since the honk was triggered by hand for 400 times.

[5]https://theproaudiofiles.com/better-acoustics-in-your-home-studio/
[6]http://nzacoustics.com/PolyesterPanelsColoured.htm

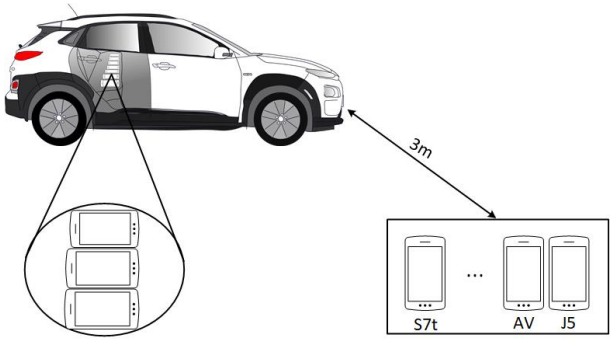


FIGURE 8. Suggestive depiction of our outdoor and in-vehicle experimental setup

2) In-vehicle hazard lights blinking for 300 times. This experiment was done inside the vehicle, the car was parked in the front of the house near a street without traffic, the engine was running at idle speed. The smartphones were placed on the rear seat next to each other as we depict in Figure 8.
3) Vehicle wipers were running at low speed 300 times. This experiment was done inside the vehicle, the car was parked in the front of the house near a street without traffic, the engine was running at idle speed and the windshield was artificially watered with the help of a garden hose. Again, the smartphones were placed on the rear seat next to each other as we depict in Figure 8.

Again, on the smartphones we run an Android application which records and saves the sounds as a PCM and WAV file for subsequent analysis.

IV. FINGERPRINTING MICROPHONES BASED ON PRERECORDED SOUNDS

In this section we analyze microphone characteristics extracted from the power spectrum of the recorded signal using traditional machine learning algorithms: LD, ENS, DT, KNN and SVM classifiers to identify the microphones. Regarding these classifiers, the following settings have been used. The linear discriminant LD was used with no regularization, i.e.,

gamma set to 0. For KNN we used the fine KNN version which uses a single neighbor with the Euclidean distance as a metric. For DT we used Gini's diversity index as slip criterion and the maximum number of splits was set to 100 which corresponds to the Fine Tree classifier type. SVM was used with a linear kernel function and the default heuristic procedure to select the kernel scale. The Ensemble classifier had a subspace dimension of 2048 (equal with the number of samples of the power spectrum), 30 learning cycles and used the discriminant learner.

A. PROCEDURE OVERVIEW

For each signal we extract the power spectrum which is used as input for the classifiers. The power spectrum is an array with 4096 elements, so for each audio signal the input for the classifiers will consist in the 4096 features.

Further, for each dataset we analyze the impact of distinct types of ambient noise in our fingerprinting procedure. That is, to the original signal we add noise at distinct SNR levels, i.e.,

$$\mathsf{SigWithNoise} = \mathsf{Sig} + \mathsf{NoiseAmp}.$$

Here, SigWithNoise is the signal with noise, Sig is the original signal (recorded with the tested microphones in the time domain) and NoiseAmp consists in traffic or market noise as retrieved from SoundArchive. The NoiseAmp is actually the noise amplified by a specific SNR factor computed by us as

$$\mathsf{NoiseAmp} = \mathsf{Noise} \times \mathsf{Fac} \times \frac{\mathsf{MaxNoise}}{\mathsf{MaxSig}}.$$

Here, Noise is the noise signal from SoundArchive in the time domain, MaxNoise is the maximum absolute value of the noise, MaxSig is the maximum absolute value of the recorded signal and Fac is the scalar amplification factor. The SNR is computed as:

$$\mathsf{SNR} = 10 \times \log_{10} \frac{\mathsf{OrigBandPower}}{\mathsf{NoiseBandPower}}[\mathsf{dB}].$$

Where OrigBandPower is the average power of the original signal (the signal recorded by the microphones) and NoiseBandPower is the average power of the noise (market or traffic noise from the MP3 file on SoundArchive).

B. FINGERPRINTING MICROPHONES BASED ON HUMAN SPEECH

For fingerprinting smartphones based on human speech we used the MOBIPHONE dataset [6] which contains 21 smartphones from distinct brands and models. For each smartphone there are 24 audio files from 24 speakers, 12 males and 12 females. For each audio file we compute the power spectrum, i.e., the frequency response, which is used as input for the machine learning classifiers.

Since the primary application scenario that we target is device identification, we evaluate the classifier's performance

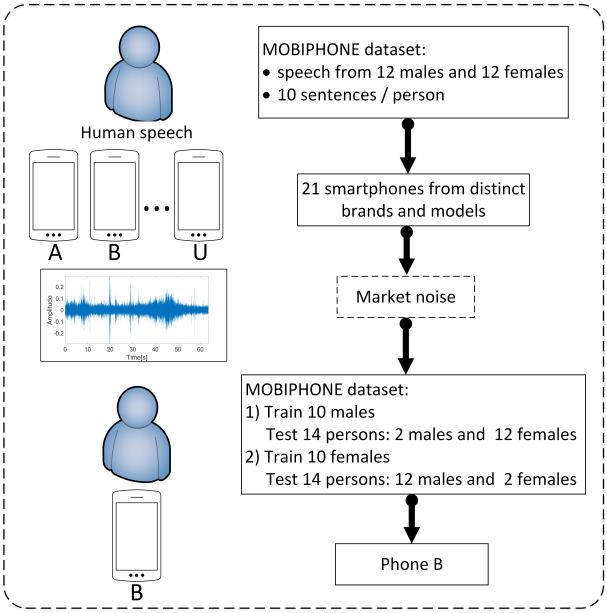


FIGURE 9. Overview of the method for smartphone recognition based on human speech (Mobiphone Dataset)

TABLE 3. Precision, recall and accuracy for five classifiers (MOBIPHONE dataset)

training	metrics	LD	ENS	DT	KNN	SVM
mobiphone (10 males)	precision	0.98	0.96	0.69	0.83	0.81
	recall	0.98	0.96	0.76	0.85	0.84
	accuracy	0.98	0.97	0.74	0.76	0.79
mobiphone (10 females)	precision	0.96	0.95	0.72	0.91	0.87
	recall	0.97	0.96	0.74	0.92	0.90
	accuracy	0.99	0.99	0.64	0.86	0.79

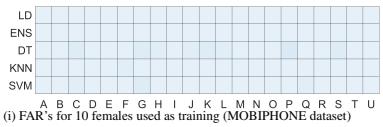
in what follows in terms of False Acceptance Rate (FAR) and False Rejection Rate (FRR). FAR is the probability of an unauthorized microphone to be accepted as legitimate and FRR is the probability of an authorized microphone to be rejected. The FAR and FRR coefficients are computed as follows:

$$FAR = \frac{FP}{TN + FP}, \quad FRR = \frac{FN}{TP + FN}.$$

Here TP are the true positives, TN the true negatives, FP the false negatives and FP the false positives.

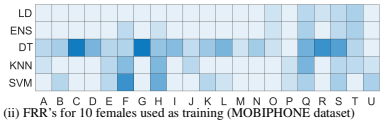
1) Fingerprinting microphones based on human speech (clean recordings)

In order to fingerprint distinct smartphones based on human speech from MOBIPHONE dataset, we use the five classifiers previously mentioned. To make the identification process more challenging we consider two cases. First we use as training the power spectrum from the speech of 10 male speakers and as test the speech of 12 females and the rest of 2 male speakers. Secondly, we use as training the power spectrum from the speech of 10 female speakers and as test



(i) FAR's for 10 females used as training (MOBIPHONE dataset)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
LD	0.	0.	0.007	0.	0.	0.	0.	0.007	0.	0.007	0.	0.	0.	0.003	0.	0.003	0.003	0.	0.	0.	0.
ENS	0.	0.003	0.010	0.	0.	0.	0.003	0.007	0.	0.	0.	0.	0.	0.007	0.	0.010	0.	0.	0.	0.	0.
DT	0.003	0.003	0.032	0.007	0.007	0.010	0.028	0.017	0.007	0.021	0.024	0.021	0.	0.007	0.007	0.037	0.010	0.	0.031	0.007	0.003
KNN	0.	0.007	0.010	0.	0.003	0.014	0.010	0.010	0.003	0.003	0.	0.	0.003	0.003	0.	0.010	0.007	0.	0.003	0.	0.
SVM	0.	0.007	0.021	0.	0.003	0.007	0.027	0.003	0.007	0.003	0.007	0.	0.003	0.003	0.	0.010	0.017	0.003	0.003	0.	0



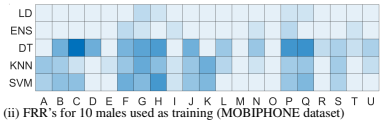
(ii) FRR's for 10 females used as training (MOBIPHONE dataset)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
LD	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.066	0.	0.133	0.066	0.125	0.176	0.
ENS	0.	0.	0.	0.	0.	0.125	0.	0.076	0.	0.	0.	0.	0.	0.	0.066	0.	0.176	0.	0.125	0.176	0.
DT	0.133	0.187	0.642	0.333	0.142	0.153	0.647	0.250	0.294	0.111	0.222	0.333	0.066	0.200	0.076	0.	0.352	0.517	0.444	0.142	0.
KNN	0.066	0.	0.	0.	0.133	0.285	0.	0.312	0.	0.133	0.	0.	0.	0.	0.	0.083	0.333	0.	0.133	0.125	0.066
SVM	0.	0.142	0.	0.	0.133	0.520	0.	0.381	0.	0.	0.142	0.066	0.	0.	0.	0.083	0.181	0.	0.187	0.	0.176

FIGURE 10. FARs (up) and FRRs (down) as heatmap (left) and numerical values (right) for the LD, ENS, DT, KNN and SVM classifiers for 10 females used as training (MOBIPHONE dataset)

(i) FAR's for 10 males used as training (MOBIPHONE dataset)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
LD	0.	0.	0.	0.	0.	0.010	0.	0.	0.	0.003	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.003
ENS	0.	0.003	0.	0.	0.007	0.	0.	0.	0.	0.	0.	0.003	0.	0.	0.	0.003	0.003	0.	0.	0.	0.007
DT	0.003	0.003	0.019	0.007	0.017	0.034	0.024	0.024	0.014	0.017	0.021	0.024	0.010	0.014	0.003	0.007	0.014	0.021	0.017	0.014	0.
KNN	0.003	0.021	0.010	0.007	0.	0.014	0.010	0.003	0.003	0.003	0.007	0.007	0.	0.003	0.010	0.007	0.014	0.	0.007	0.017	0.017
SVM	0.	0.021	0.021	0.010	0.017	0.007	0.	0.	0.	0.007	0.010	0.007	0.	0.	0.010	0.014	0.017	0.	0.014	0.024	0.007



(ii) FRR's for 10 males used as training (MOBIPHONE dataset)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
LD	0.	0.	0.	0.	0.	0.	0.125	0.066	0.	0.	0.	0.	0.	0.	0.	0.066	0.066	0.	0.	0.	0.
ENS	0.	0.	0.	0.	0.	0.125	0.066	0.	0.066	0.	0.	0.	0.	0.066	0.	0.133	0.	0.	0.	0.066	0.
DT	0.	0.350	0.709	0.368	0.	0.333	0.416	0.461	0.	0.357	0.	0.222	0.	0.230	0.	0.500	0.523	0.111	0.181	0.090	0.176
KNN	0.235	0.200	0.083	0.	0.066	0.230	0.388	0.277	0.071	0.235	0.368	0.142	0.	0.071	0.	0.200	0.230	0.	0.200	0.	0.100
SVM	0.176	0.272	0.272	0.	0.	0.368	0.300	0.481	0.066	0.250	0.312	0.	0.	0.	0.	0.230	0.307	0.	0.166	0.	0.

FIGURE 11. FARs (up) and FRRs (down) as heatmap (left) and numerical values (right) for the LD, ENS, DT, KNN and SVM classifiers for 10 males used as training (MOBIPHONE dataset)

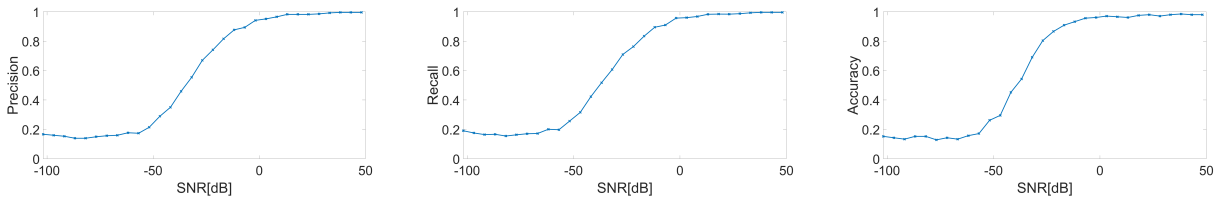


FIGURE 12. Precision (left), recall (middle) and accuracy (right) obtained with linear discriminant classifier with noise between -80dB to 20dB with 5dB increment

the speech of 12 male and the rest of 2 female speakers. In Figure 9 we depict the flowchart of these scenarios. Table 3 shows the mean precision, mean recall and accuracy for each classifier for the both cases from this scenario. It is obvious that the LD classifier has the best results, followed closely by the ENS, while the KNN and SVM have much poor results and the worst results are obtained with DT classifier likely due to its tendency for over-fitting. As an additional metric for performance, more specifically focused on the authentication/impersonation success rate, in Figure 10 we depict the FAR (False Acceptance Rate) and FRR (False Rejection Rate) as heatmaps (left) and as numerical values (right) for each classifier and microphone for the 10 females used as training samples. It is again obvious that the DT classifier has the worst results followed by KNN, while the best results are obtained with LD and ENS classifiers. However, the FAR is very low for all classifiers, the maximum value is 3.7% for the DT classifier on microphone P. The FRR however reaches 64% for the DT classifier on microphones C and G which is too high. For the LD classifier, the maximum value for FAR is only 0.7% on microphones C, H and J and the maximum value for FRR is only 17% on microphone T. In Figure 11 we depict the FAR (up) and FRR (down) as heatmap (left) and numerical values (right) for each classifier and microphone for the 10 males used as training. Again, it is obvious that the DT classifier gives the worst results followed by KNN, while the best results are obtained with the LD and ENS classifiers. However, overall the FAR is very low for all classifiers, the maximum values is 3.4% for the DT classifier on microphone F. The FRR reaches 70% for the DT classifier on microphone C. For the LD classifier, the maximum value for FAR is 1% on microphone F and the maximum value for FRR is 12.5% on microphone G.

2) Fingerprinting microphones based on human speech with market noise

To make the fingerprinting process more challenging and comparable to a real-life scenario in which ambient noise is present, we add market noise to the signals at different SNRs, i.e., from -80dB to 20dB with an increment step of 5dB. In order to obtain different SNRs, we simply amplified the amplitude of the noise in the prerecorded signal. Since the noise comes from external recordings, e.g., indoors and outdoors noise, we needed a gradual analysis of the noise impact, for which amplifying the amplitude was the only option. For this scenario we use only the LD classifier because for this classifier we obtained the best results and nonetheless because it is faster than the others, i.e., $\sim$ 87 samples/second

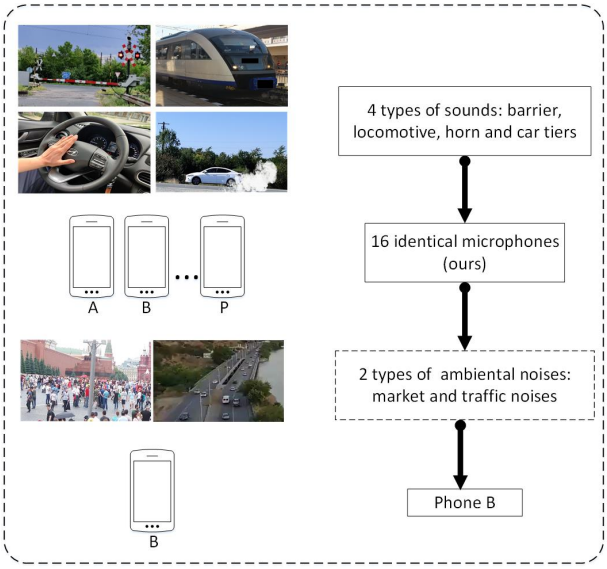


FIGURE 13. Overview of the method for smartphone recognition based on environmental noise

prediction speed with 18.364 seconds training time. The training and the test dataset have the same SNR level. In Figure 12 we plot the mean precision (left), mean recall (middle) and accuracy (right) for different levels of noise. At a SNR lower than -50dB the identification no longer works while at a SNR between -50 and 0dB the precision, recall and accuracy are increasing from -0.2 to 0.9 and finally at a SNR greater than 0dB the precision, recall and accuracy are close to 1.

C. FINGERPRINTING IDENTICAL MICROPHONES BASED ON ENVIRONMENTAL NOISE (INDOOR EXPERIMENTS)

The dataset which we build contains 16 microphones from the same smartphone model. We use four prerecorded environmental noises from the SoundArchive: (i) a locomotive signaling departure, (ii) barriers closing with two bells jingling, (iii) car arriving with screeching tiers and the (iv) two tone horn sound of a Mercedes-Benz. With each sound played in the background we did 50 measurements on each microphone, i.e., resulting in 800 measurements for each sound and a total of 3200 measurements. Again, to get closer to a real-life scenario, we add the two previous types of noise, i.e., market and traffic, at distinct levels. In Figure 13 we depict the flowchart of this test scenario.

1) Fingerprinting microphones based on environment sounds (clean recordings)

We use 20 measurements for training and 30 measurements for testing. In Table 4 we depict the mean precision, mean recall and accuracy for each classifier on each type of sound. It is obvious that the LD and ENS classifiers have the best results, followed closely by the SVM. The KNN has poor results and as expected the worst results are again obtained with the DT classifier.

TABLE 4. Precision, recall and accuracy for five classifiers (this paper dataset)

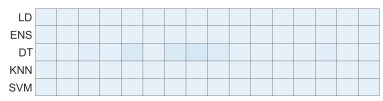
sound type	metrics	LD	ENS	DT	KNN	SVM
locomotive	precision	1.00	1.00	0.79	0.99	0.99
	recall	1.00	1.00	0.79	0.99	0.99
	accuracy	1.00	1.00	0.91	0.98	0.99
barrier	precision	1.00	1.00	0.89	0.89	0.98
	recall	1.00	1.00	0.90	0.90	0.98
	accuracy	1.00	1.00	0.91	0.92	0.98
car	precision	1.00	1.00	0.79	0.88	0.99
	recall	1.00	1.00	0.82	0.88	0.99
	accuracy	1.00	0.99	0.89	0.83	0.98
horn	precision	1.00	1.00	0.82	0.98	0.99
	recall	1.00	1.00	0.83	0.98	0.99
	accuracy	1.00	1.00	0.90	0.96	1.00

In Figures 14, 15, 16 and 17 we depict the FAR and FRR for each classifier and microphone on the four testing sounds as heatmaps and numerical values. In Figure 14, we depict the FAR (up) and FRR (down) for the locomotive sound and the overall FAR is very low for all classifiers, while the maximum values is 4.5% for the DT classifier for microphone H. The FRR reaches 62% for the DT classifier on microphone E. For the LD and ENS classifiers the FAR and FRR are zero for all microphones. In Figure 15 we depict the FAR (up) and FRR (down) for the barrier sound. Again, the overall FAR is very low for all classifiers, the maximum values is 2.6% for the DT classifier on microphone F. The FRR reaches 36% for the DT classifier on microphone O. Again, the LD and ENS classifiers have a FAR and FRR equal to zero for all microphones. In Figure 16 we depict the FAR (up) and FRR (down) for the car tiers sound. The FAR is very low for all classifiers with a maximum value of 4.2% for the DT classifier on microphone P. The FRR reaches 48% for the KNN classifier on microphone H. For the LD and ENS classifiers the FAR and FRR are again zero for all microphones. In Figure 17 we depict the FAR (up) and FRR (down) for the car horn sound. The maximum value for the FAR is 4% for the DT classifier on microphone C, the FRR reaches 45% for the DT classifier on microphone F. For the LD and ENS classifiers the FAR and FRR are again zero for all microphones.

Overall, as can be seen from these results, the LD and ENS classifiers have the FAR and FRR equal to zero on all microphones. For the other classifiers, the barrier and horn sounds resulted in the highest values for the FAR and FRR, i.e., worst identification rates, than the locomotive and car tiers sounds.

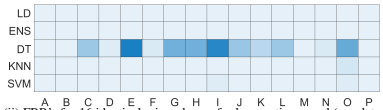
2) Fingerprinting microphones based on environment sounds with ambient noise

Since in real-life scenarios ambient noise is present, we again add two types of noise (traffic and market noise) at distinct SNR, i.e., from -80dB to 20dB with a increment step of 5dB, to the clean signals. For this scenario we use



(i) FAR's for 16 identical microphones for locomotive sound (our dataset)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
LD	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
ENS	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
DT	0.002	0.	0.019	0.	0.042	0.002	0.041	0.045	0.031	0.	0.008	0.006	0.	0.011	0.002	0.004
KNN	0.	0.	0.	0.	0.004	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
SVM	0.	0.	0.	0.	0.	0.002	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.002



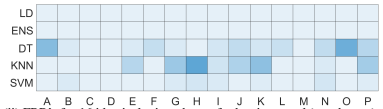
(ii) FRR's for 16 identical microphones for locomotive sound (our dataset)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
LD	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
ENS	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
DT	0.	0.	0.222	0.032	0.620	0.033	0.352	0.357	0.578	0.230	0.133	0.228	0.	0.038	0.383	0.
KNN	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.062	0.
SVM	0.	0.	0.	0.	0.	0.	0.	0.	0.032	0.	0.	0.	0.	0.	0.032	0.

FIGURE 14. FAR (up) and FRR (down) as heatmap (left) and numerical values (right) for the LD, ENS, DT, KNN and SVM classifiers for 16 identical microphones for locomotive sound (our dataset)

(i) FAR's for 16 identical microphones for barrier sound (our dataset)

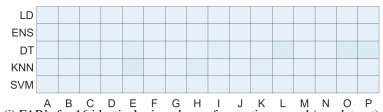
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
LD	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
ENS	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
DT	0.002	0.011	0.002	0.008	0.002	0.026	0.013	0.006	0.	0.	0.	0.021	0.002	0.015	0.	0.002
KNN	0.	0.006	0.002	0.002	0.004	0.002	0.013	0.024	0.004	0.	0.006	0.015	0.002	0.008	0.002	0.015
SVM	0.	0.002	0.002	0.004	0.	0.002	0.002	0.	0.002	0.	0.	0.	0.002	0.002	0.	0.



(ii) FRR's for 16 identical microphones for barrier sound (our dataset)

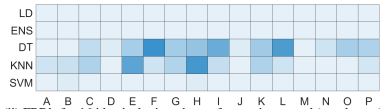
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
LD	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
ENS	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
DT	0.292	0.038	0.	0.	0.033	0.100	0.040	0.	0.062	0.117	0.062	0.090	0.	0.115	0.361	0.147
KNN	0.	0.	0.	0.	0.151	0.	0.225	0.406	0.066	0.166	0.270	0.	0.	0.037	0.	0.206
SVM	0.	0.033	0.	0.	0.	0.	0.033	0.062	0.033	0.	0.032	0.	0.	0.064	0.	0.032

FIGURE 15. FAR (up) and FRR (down) as heatmap (left) and numerical values (right) for the LD, ENS, DT, KNN and SVM classifiers for 16 identical microphones for barrier sound (our dataset)



(i) FAR's for 16 identical microphones for car tiers sound (our dataset)

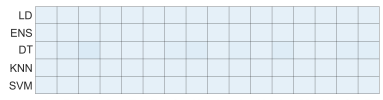
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
LD	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
ENS	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
DT	0.	0.019	0.017	0.	0.028	0.004	0.004	0.004	0.011	0.	0.	0.033	0.002	0.008	0.036	0.042
KNN	0.004	0.004	0.013	0.	0.035	0.004	0.006	0.024	0.008	0.	0.011	0.004	0.	0.006	0.	0.002
SVM	0.002	0.	0.	0.	0.002	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.



(ii) FRR's for 16 identical microphones for car tiers sound (our dataset)

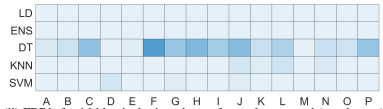
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
LD	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
ENS	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
DT	0.	0.	0.120	0.032	0.190	0.517	0.200	0.243	0.375	0.032	0.210	0.464	0.	0.071	0.187	0.166
KNN	0.066	0.096	0.172	0.	0.416	0.066	0.156	0.486	0.103	0.	0.166	0.066	0.	0.	0.	0.064
SVM	0.	0.032	0.	0.	0.	0.	0.	0.032	0.	0.	0.	0.	0.	0.	0.	0.

FIGURE 16. FAR (up) and FRR (down) as heatmap (left) and numerical values (right) for the LD, ENS, DT, KNN and SVM classifiers for 16 identical microphones for car tiers sound (our dataset)



(i) FAR's for 16 identical microphones for car horn sound (our dataset)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
LD	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
ENS	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
DT	0.002	0.013	0.040	0.002	0.	0.009	0.015	0.022	0.006	0.013	0.004	0.021	0.002	0.004	0.011	0.013
KNN	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.004	0.	0.004	0.002	0.002
SVM	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.004	0.002	0.002	0.



(ii) FRR's for 16 identical microphones for car horn sound (our dataset)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
LD	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
ENS	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
DT	0.033	0.076	0.266	0.	0.032	0.458	0.233	0.310	0.205	0.294	0.096	0.166	0.	0.096	0.074	0.250
KNN	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.062	0.032	0.066	0.	0.034	0.	0.
SVM	0.	0.	0.	0.062	0.	0.	0.032	0.	0.	0.032	0.	0.	0.	0.	0.	0.

FIGURE 17. FAR (up) and FRR (down) as heatmap (left) and numerical values (right) for the LD, ENS, DT, KNN and SVM classifiers for 16 identical microphones for car horn sound (our dataset)

only the LD classifier because it gave the best results and had a fast prediction speed. In Figure 18 we depict the mean precision (left), mean recall (middle) and the accuracy (right) for distinct noise levels. The noise level influences the classification distinctly on the four sounds type. For the barrier sound, at a SNR lower than -70dB the identification is not working, at a SNR between -70 and -63dB the precision, recall and accuracy are increasing from -0.2 to 0.9, while at a SNR greater than -63dB the precision, recall and accuracy are close to 1. For the horn sound, at a SNR lower than -60dB the identification is not working, at a SNR between -60 and -40dB the precision, recall and accuracy are increasing from -0.2 to 0.9, while at a SNR greater than -40dB the precision, recall and accuracy are close to 1. For the locomotive sound the influence of the noise is similar as in case of horn sound. For the car tiers sound, at a SNR lower than -35dB the identification is not working, at a SNR between -35 and -8dB the precision, recall and accuracy are increasing from -0.2 to 0.9, while at a SNR greater than -8dB the precision, recall and accuracy are close to 1.

As a partial conclusion, the least influence of the ambient noise is on the screeching tiers, while fingerprinting based on

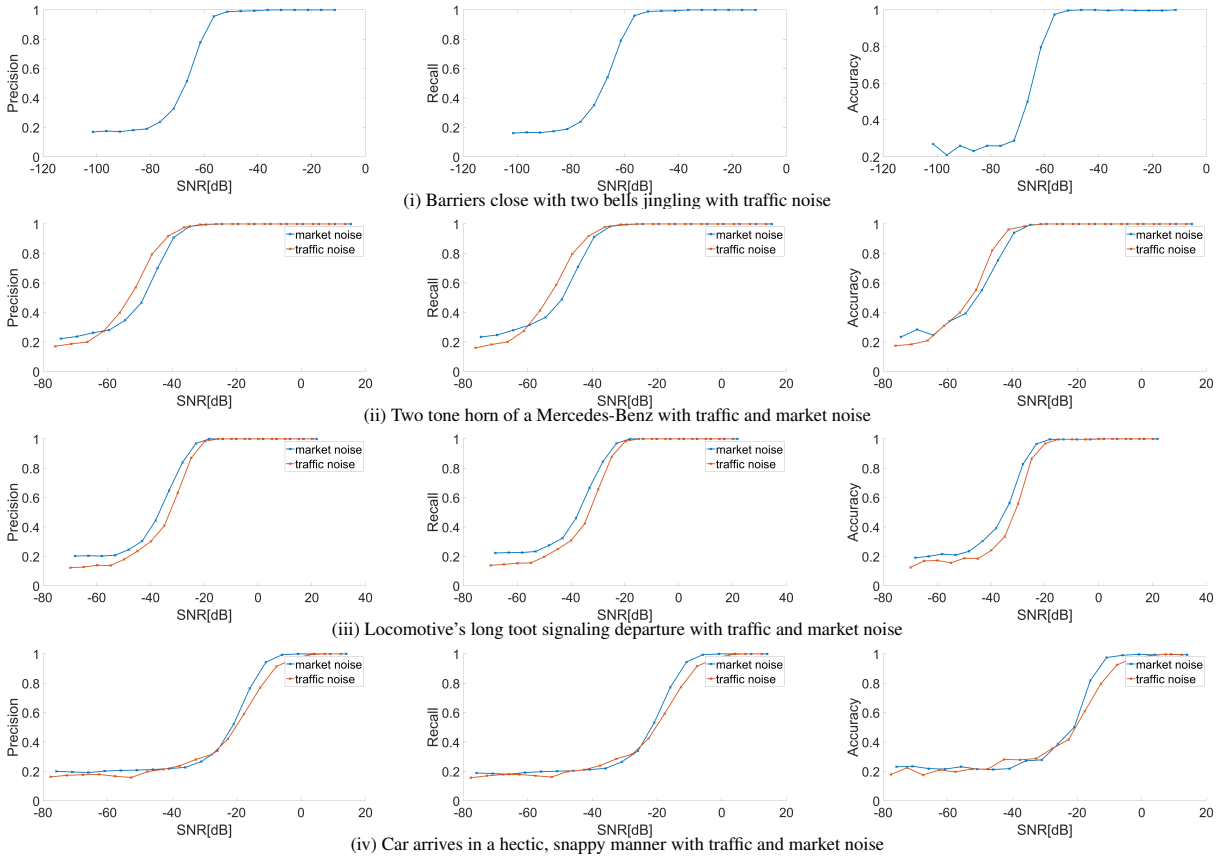


FIGURE 18. Precision (left), recall (middle) and accuracy (right) obtained with linear discriminant classifier with noise between -80dB to 20dB with 5dB increment

the barriers, horn and locomotive sounds are influenced to a higher degree.

V. FINGERPRINTING MICROPHONES BASED ON LIVE RECORDINGS

In this section we evaluate microphone fingerprinting in the more challenging scenario with live recordings. As an additional comparison, we also add a deep-learning CNN architecture and compare it with the traditional machine learning algorithm LD (we keep only the LD algorithm since it gave the best results in all the previous tests).

A. DETAILS ON THE TEST SCENARIO

In this section we analyze scenario C for which we build our outdoor recordings with 16 smartphones that record: i) a car honk for which we did 400 measurements for each smartphone, totaling 6400 measurements, ii) in-vehicle hazard lights blinking for which we did 300 measurements for each smartphone, totaling 4800 measurements and iii) vehicle wipers running at low speed for which we did 300 measurements for each smartphone, totaling 4800 measurements. For each signal we extract the power spectrum which is used as input for the classifiers. The power spectrum is an array with 4096 elements, so for each audio signal the input for the classifiers will be the 4096 features. For both, LD and CNN we chose random 55% of measurements for training

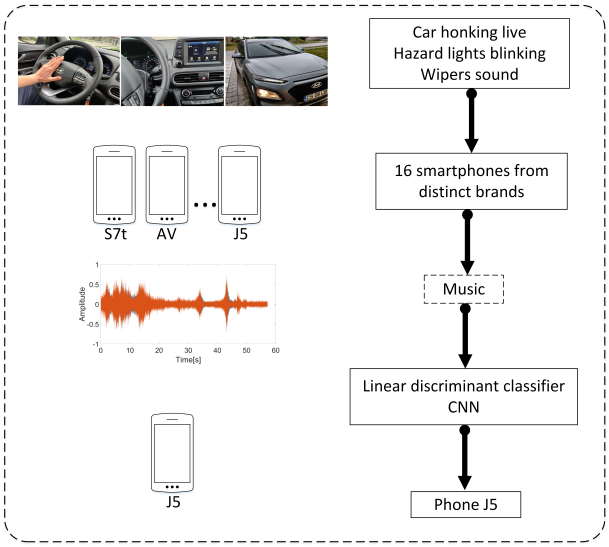
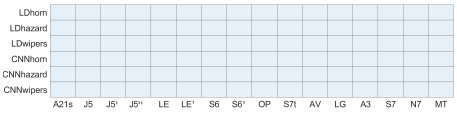


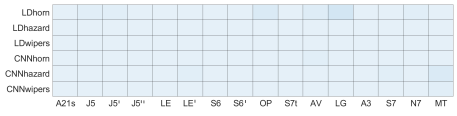
FIGURE 19. Smartphone recognition based on car honk

and the remaining 45% of measurements are used for testing. To make the identification even more challenging, we also add background music noise to the recordings. For this we use the first 3 songs from Spotify top 10 list available in 2021. In Figure 19 we depict this scenario.



(i) FAR's for 16 microphones for horn, hazard lights and wipers sounds

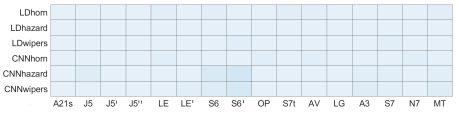
	A21s	J5	J5'	J5''	LE	LE'	S6	S6'	OP	S7t	AV	LG	A3	S7	N7	MT
LDhorn	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.	0.001	0.001	0.001	0.001	0.001	0.	0.001	0.001
LDhazard	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
LDwipers	0.	0.	0.	0.001	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
CNNhorn	0.	0.001	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
CNNhazard	0.	0.	0.002	0.	0.001	0.	0.004	0.001	0.	0.001	0.	0.	0.	0.	0.	0.
CNNwipers	0.	0.	0.	0.	0.	0.001	0.	0.001	0.	0.	0.	0.	0.	0.	0.	0.



(ii) FRR's for 16 microphones for horn, hazard lights and wipers sounds

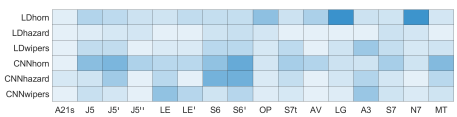
	A21s	J5	J5'	J5''	LE	LE'	S6	S6'	OP	S7t	AV	LG	A3	S7	N7	MT
LDhorn	0.	0.016	0.016	0.	0.	0.	0.	0.	0.033	0.	0.027	0.055	0.005	0.	0.016	0.
LDhazard	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
LDwipers	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.006	0.	0.	0.	0.	0.
CNNhorn	0.	0.	0.005	0.	0.	0.	0.	0.	0.	0.	0.016	0.	0.005	0.	0.	0.
CNNhazard	0.	0.007	0.	0.	0.	0.014	0.	0.007	0.007	0.	0.	0.	0.007	0.014	0.	0.037
CNNwipers	0.	0.	0.	0.007	0.	0.	0.	0.	0.007	0.	0.007	0.	0.007	0.	0.	0.014

FIGURE 20. FAR (up) and FRR (down) as heatmap (left) and numerical values (right) for the linear discriminant classifier and CNN for 16 microphones for horn, hazard lights and wipers sounds



(i) FAR's for 16 microphones for horn, hazard lights and wipers sounds affected by music

	A21s	J5	J5'	J5''	LE	LE'	S6	S6'	OP	S7t	AV	LG	A3	S7	N7	MT
LDhorn	0.002	0.010	0.008	0.004	0.017	0.008	0.013	0.006	0.006	0.011	0.020	0.010	0.008	0.009	0.008	0.020
LDhazard	0.001	0.001	0.006	0.001	0.001	0.001	0.003	0.007	0.002	0.003	0.001	0.001	0.	0.006	0.001	0.001
LDwipers	0.001	0.009	0.007	0.	0.003	0.003	0.006	0.014	0.001	0.007	0.	0.004	0.006	0.020	0.001	0.002
CNNhorn	0.001	0.007	0.001	0.004	0.012	0.008	0.001	0.001	0.017	0.001	0.017	0.003	0.010	0.001	0.015	0.001
CNNhazard	0.004	0.026	0.008	0.	0.008	0.001	0.042	0.061	0.004	0.010	0.001	0.008	0.015	0.014	0.001	0.021
CNNwipers	0.001	0.006	0.009	0.001	0.008	0.015	0.023	0.045	0.	0.002	0.	0.001	0.025	0.014	0.001	0.029



(ii) FRR's for 16 microphones for horn, hazard lights and wipers sounds affected by music

	A21s	J5	J5'	J5''	LE	LE'	S6	S6'	OP	S7t	AV	LG	A3	S7	N7	MT
LDhorn	0.016	0.155	0.127	0.077	0.066	0.116	0.100	0.100	0.266	0.083	0.177	0.516	0.083	0.072	0.511	0.061
LDhazard	0.010	0.025	0.050	0.055	0.010	0.005	0.070	0.095	0.055	0.009	0.	0.005	0.005	0.080	0.065	0.
LDwipers	0.	0.113	0.113	0.	0.040	0.040	0.133	0.140	0.053	0.122	0.006	0.066	0.226	0.106	0.053	0.073
CNNhorn	0.	0.283	0.322	0.172	0.138	0.138	0.255	0.388	0.055	0.188	0.161	0.077	0.072	0.188	0.066	0.305
CNNhazard	0.044	0.074	0.214	0.	0.140	0.	0.348	0.355	0.037	0.103	0.	0.059	0.163	0.059	0.	0.140
CNNwipers	0.007	0.096	0.059	0.	0.259	0.148	0.066	0.148	0.022	0.044	0.	0.022	0.237	0.074	0.037	0.074

FIGURE 21. FARs (up) and FRRs (down) as heatmap (left) and numerical values (right) for the linear discriminant classifier and CNN for 16 microphones for horn, hazard lights and wipers sounds affected by music

B. DEEP-LEARNING APPROACH WITH CNN

For each smartphone i, we induce a CNN-based binary classifier that is responsible for authenticating it (i.e., return '1' if a given input sample is associated with smartphone i, and '0' otherwise). The dataset for training binary classifier i consists of positive and negative examples. The positive examples are associated with smartphone i, and the negative examples are associated with other smartphones.

To induce the best binary classifier for authenticating smartphone i, we first execute a hyperparameter tuning procedure which is based on a random search over 50 trials. On each trial, the best hyperparameters are chosen using a stratified 3-fold cross-validation procedure [45]. On each fold iteration, to address the data unbalance, we use the cost-sensitive learning method described in [46]. We choose the known cross-entropy score [47] to measure the best set of hyperparameters.

Finally, given the best set of hyperparameters for binary classifier i, we generate the model. We first divide the training dataset, which represents 55% of the entire dataset as the remaining 45% was used for testing, into 70% for training and 30% for validation. Then we train the model until the loss function reaches its minimum on the validation set. Also, at this step, we use the same cost-sensitive learning method as we used during the hyperparameter tuning [46]. The loss function that we pick to minimize is the *binary cross-entropy*, and the optimizer that we use for this mission is the RMSProp. The learning rate is optimized over the set of [0.001, 0.0001]

Regarding the CNN architecture, it consists of filter layers followed by fully connected layers. We vary the number of filter layers in the set of [1, 2, 3]. All the filter layers are applied with a kernel size in the set of [3, 4, 5] and a filters count in the set of [16, 32, 64]. To avoid overfitting, the filter layers are followed by a dropout in the set of [0, 0.1, 0.2, 0.3, 0.4]. Regarding the fully connected layers, we vary the number of them in the set of [2, 3, 4, 5], and the number of neurons for each fully connected layer in the set of [16, 32, 64, 128]. Finally, another dropout is attached, varied in the set of [0, 0.1, 0.2, 0.3, 0.4].

C. RESULTS ON CLEAN RECORDINGS

In Figure 20 we depict the FAR (up) and FRR (down) as heatmap (left) and numerical values (right) for the linear discriminant classifier and CNN for 16 microphones for horn, hazard lights and wipers sounds. For the horn sound, recorded outdoors, it is obvious that for both algorithms the FAR and FRR are very low. But even so, the best results are obtained with CNN, i.e., only for the J5 the FAR is 0.1% and for the rest of the smartphones the FAR is zero. In case of the LD classifier, the FAR is between 0% and 0.1%. Also, the FRR are more accurate for the CNN, i.e., only for three smartphones the FRR are not zero, but it is kept very low between 0.5% and 1.6%. For the linear discriminant classifier, seven smartphones have non-zero FRR that are between 0.5% and 5.5%. For hazard lights, the blinking sound was recorded inside the vehicle. Again, for both algorithms the FAR and FRR are very low, but in this scenario the best results are obtained with the LD classifier. Both the FAR and FRR are zero in case of the LD classifier while in case of the CNN the FAR is between 0% and 0.4% and the FRR are between 0% and 3.7%. For wipers, the sound was recorded

inside the vehicle. Again, for both algorithms the FAR and FRR are very low, but the best results are obtained with the LD classifier. The FAR and FRR are zero in case of the LD classifier, except for a smartphone where the FAR is 0.1% and the FRR is 0.6%. In case of CNN the FAR is zero, except for two smartphones where the FAR is 0.1%, while the FRR is not zero for four smartphones with values between 0.7% and 1.4%.

D. RESULTS ON RECORDINGS INFLUENCED BY NOISE

In Figure 21 we depict the FAR (up) and FRR (down) as heatmap (left) and numerical values (right) for the linear discriminant classifier and CNN for 16 microphones for horn, hazard lights and wipers sounds when affected by background music. In case of the horn sound, the results are close between the CNN and the LD classifier. For the CNN, the FAR is between 0.1% and 1.7% and the FRR between 0% and 38.8% while for LD the FAR is between 0.2% and 2% with the FRR are between 1.6% and 51.6%. In case of the hazard lights affected by noise, the best results are obtained with the LD classifier. The FAR for the LD classifier is between 0% and 0.6% and the FRR is between 0% and 9.5%. For the CNN, the FAR is between 0% and 6.1% and the FRR is between 0 and 35%. In case of the wipers sound affected by music, the FAR is little bit lower for the LD classifier, while the FRR is close between the CNN and LD. The FAR for the LD classifier is between 0% and 2% and the FRR is between 0% and 22%. For the CNN, the FAR is between 0% and 4.5% and the FRR between 0% and 25.9%.

VI. DISCUSSION AND CONCLUSION

In this work, we explored smartphone microphone fingerprinting based on microphone data by using the power spectrum of the recorded signal with distinct supervised machine learning algorithms, i.e., Linear Discriminant, Ensemble-Subspace Discriminant, Decision Tree, Fine KNN and Linear SVM. We tested three major use cases of fingerprinting based on human speech, synthetically reproduced environmental sound and finally live recordings. In all the scenarios, noise was added to make identification more challenging. For the first two scenarios the LD classifier behaved almost perfect. The last scenario was more demanding and we added a CNN deep-learning architecture to serve as a comparison. There was no clear cut between the accuracy of the LD and CNN, on the recordings unaffected by noise they performed similar. When noise was added, the LD gave poor identification results for 2 phones (the LG and Nexus 7), while the CNN had no particular problem with these 2 phones but the identification was slightly poorer for the rest of the phones. Since the LD classifier has a fast prediction speed and uses small amounts of memory, it may be still preferable to the CNN architecture. The rest of the traditional machine learning classifiers gave poorer results.

As expected, separating between identical smartphones, i.e., same model and manufacturer, is more challenging than separating smartphones of different types or brands. This is visible as for clean recordings, in case of different smartphones, in Figure 8, the maximum FAR is 0.004 for Samsung S6 with CNN for the sound produced by the hazard lights, while the maximum FRR is 0.055 for LG with the LD classifier in case of the car honk. While in case of identical smartphones, the FAR reaches 0.042 for smartphone E in the case of DT on the locomotive sound, and the FRR reaches 0.62 for several smartphones and classifiers – this means a one order of magnitude higher FARs and FRRs when identical phones are used.

It may also be expected that some smartphones may be easier to identify than others due to specifics related to the manufacturing process or quality control, etc. Our results do not necessarily indicate that this is so. By inspecting Figures 20 and 21 which show the heatmaps and numerical results for the batch of 16 distinct phones, the FAR and FRR are small and comparable between different smartphones. It seems that the results are much more influenced by the type of sounds that are used as some sounds contain frequencies that are reproduced differently by the smartphones, making identification easier. Concretely, in the case of the live recordings with 16 identical smartphones, the best results were obtained for the locomotive and car honk sounds which gave better results than the barrier and hazard lights sounds.

Possible applications of such fingerprints are manifold: security minded use cases could include attestation of possession of a particular phone to act as second, unclonable factor token; however, such fingerprinting could also be abused by apps to fingerprint devices without otherwise having access to device-unique identifiers. While this could indeed be a powerful fingerprint, we argue that malicious apps (or libraries embedded within) with high-fidelity access to microphone sampling already has more serious security and privacy impact [48] without the added device fingerprint. Nonetheless, on-device countermeasures to this particular method — such as adding noise or lowering sampling fidelity — are still subject to future work.

ACKNOWLEDGMENT

This work was supported by a grant of the Romanian Ministry of Research, Innovation and Digitalization, project number PFE 26/30.12.2021, PERFORM-CDI@UPT100- The increasing of the performance of the Polytechnic University of Timișoara by strengthening the research, development and technological transfer capacity in the field of "Energy, Environment and Climate Change" at the beginning of the second century of its existence, within Program 1 - Development of the national system of Research and Development, Subprogram 1.2 - Institutional Performance - Institutional Development Projects - Excellence Funding Projects in RDI, PNCDI III".

REFERENCES

[1] K. Lofstrom, W. R. Daasch, and D. Taylor, "Ic identification circuit using device mismatch," in *2000 IEEE International Solid-State Circuits Conference. Digest of Technical Papers (Cat. No. 00CH37056)*. IEEE, 2000, pp. 372–373.

[2] B. Gassend, D. Clarke, M. Van Dijk, and S. Devadas, "Silicon physical random functions," in *Proceedings of the 9th ACM conference on Computer and communications security*, 2002, pp. 148–160.
[3] B. Groza, A. Berdich, C. Jichici, and R. Mayrhofer, "Secure accelerometer-based pairing of mobile devices in multi-modal transport," *IEEE Access*, vol. 8, pp. 9246–9259, 2020.
[4] H. Malik, "Acoustic environment identification and its applications to audio forensics," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1827–1837, 2013.
[5] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.
[6] C. Kotropoulos and S. Samaras, "Mobile phone identification using recorded speech signals," in *2014 19th International Conference on Digital Signal Processing*. IEEE, 2014, pp. 586–591.
[7] J. Garofolo, "Getting started with the darpa timit cd-rom: An acoustic phonetic continuous speech database, national institute of standards and technology (nist), gaithersburg," *Gaithersburgh, MD, USA*, 1988.
[8] G. Baldini, I. Amerini, and C. Gentile, "Microphone identification using convolutional neural networks," *IEEE Sensors Letters*, vol. 3, no. 7, pp. 1–4, 2019.
[9] G. Baldini and I. Amerini, "Smartphones identification through the built-in microphones with convolutional neural network," *IEEE Access*, vol. 7, pp. 158 685–158 696, 2019.
[10] ——, "An evaluation of entropy measures for microphone identification," *Entropy*, vol. 22, no. 11, p. 1235, 2020.
[11] H. Bojinov, Y. Michalevsky, G. Nakibly, and D. Boneh, "Mobile device identification via sensor fingerprinting," *arXiv preprint arXiv:1408.1416*, 2014.
[12] A. Hafeez, K. M. Malik, and H. Malik, "Exploiting frequency response for the identification of microphone using artificial neural networks," in *Audio Engineering Society Conference: 2019 AES International Conference on Audio Forensics*. Audio Engineering Society, 2019.
[13] S. Ikram and H. Malik, "Microphone identification using higher-order statistics," in *Audio engineering society conference: 46th international conference: audio forensics*. Audio Engineering Society, 2012.
[14] H. Q. Vu, S. Liu, X. Yang, Z. Li, and Y. Ren, "Identifying microphone from noisy recordings by using representative instance one class-classification approach," *Journal of networks*, 2012.
[15] R. Buchholz, C. Kraetzer, and J. Dittmann, "Microphone classification using fourier coefficients," in *International Workshop on Information Hiding*. Springer, 2009, pp. 235–246.
[16] Y. Jiang and F. H. Leung, "Source microphone recognition aided by a kernel-based projection method," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 11, pp. 2875–2886, 2019.
[17] Ö. Eskidere and A. Karatutlu, "Source microphone identification using multitaper mfcc features," in *9th International Conference on Electrical and Electronics Engineering (ELECO)*. IEEE, 2015, pp. 227–231.
[18] D. Luo, P. Korus, and J. Huang, "Band energy difference for source attribution in audio forensics," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2179–2189, 2018.
[19] X. Lin, J. Zhu, and D. Chen, "Subband aware cnn for cell-phone recognition," *IEEE Signal Proc. Letters*, vol. 27, pp. 605–609, 2020.
[20] C. Jin, R. Wang, and D. Yan, "Source smartphone identification by exploiting encoding characteristics of recorded speech," *Digital Investigation*, vol. 29, pp. 129–146, 2019.
[21] X. Li, D. Yan, L. Dong, and R. Wang, "Anti-forensics of audio source identification using generative adversarial network," *IEEE Access*, vol. 7, pp. 184 332–184 339, 2019.
[22] V. Verma and N. Khanna, "Cnn-based system for speaker independent cell-phone identification from recorded audio." in *CVPR Workshops*, 2019, pp. 53–61.
[23] V. A. Hadoltikar, V. R. Ratnaparkhe, and R. Kumar, "Optimization of mfcc parameters for mobile phone recognition from audio recordings," in *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*. IEEE, 2019, pp. 777–780.
[24] F. Kurniawan, M. S. M. Rahim, M. S. Khalil, and M. K. Khan, "Statistical-based audio forensic on identical microphones," *Intl. Journal of Electrical and Computer Eng.*, vol. 6, no. 5, p. 2211, 2016.
[25] L. Cuccovillo and P. Aichroth, "Open-set microphone classification via blind channel analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2016, pp. 2074–2078.
[26] C. Kraetzer, A. Oermann, J. Dittmann, and A. Lang, "Digital audio forensics: a first practical evaluation on microphone and environment classification," in *Proceedings of the 9th workshop on Multimedia & security*, 2007, pp. 63–74.
[27] D. Bykhovsky, "Recording device identification by enf harmonics power analysis," *Forensic Science International*, vol. 307, p. 110100, 2020.
[28] S. Qi, Z. Huang, Y. Li, and S. Shi, "Audio recording device identification based on deep learning," in *2016 IEEE International Conference on Signal and Image Processing (ICSIP)*. IEEE, 2016, pp. 426–431.
[29] A. Das, N. Borisov, and M. Caesar, "Do you hear what i hear?: Fingerprinting smart devices through embedded acoustic components," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2014, pp. 441–452.
[30] ——, "Fingerprinting smart devices through embedded acoustic components," *arXiv preprint arXiv:1403.3366*, 2014.
[31] Z. Zhou, W. Diao, X. Liu, and K. Zhang, "Acoustic fingerprinting revisited: Generate stable device id stealthily with inaudible sound," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2014, pp. 429–440.
[32] T. Qin, R. Wang, D. Yan, and L. Lin, "Source cell-phone identification in the presence of additive noise from cqt domain," *Information*, vol. 9, no. 8, p. 205, 2018.
[33] A. Berdich, B. Groza, R. Mayrhofer, E. Levy, A. Shabtai, and Y. Elovici, "Sweep-to-unlock: Fingerprinting smartphones based on loudspeaker roll-off characteristics," *IEEE Transactions on Mobile Computing, Early Access*, 2021.
[34] Z. Ding and M. Ming, "Accelerometer-based mobile device identification system for the realistic environment," *IEEE Access*, vol. 7, pp. 131 435–131 447, 2019.
[35] G. Baldini, G. Steri, I. Amerini, and R. Caldelli, "The identification of mobile phones through the fingerprints of their built-in magnetometer: An analysis of the portability of the fingerprints," in *2017 International Carnahan Conference on Security Technology*. IEEE, 2017, pp. 1–6.
[36] I. Amerini, R. Becarelli, R. Caldelli, A. Melani, and M. Niccolai, "Smartphone fingerprinting combining features of on-board sensors," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 10, pp. 2457–2466, 2017.
[37] W.-C. Yang, J. Jiang, and C.-H. Chen, "A fast source camera identification and verification method based on prnu analysis for use in video forensic investigations," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 6617–6638, 2021.
[38] R. Deka, C. Galdi, and J.-L. Dugelay, "Hybrid g-prnu: Optimal parameter selection for scale-invariant asymmetric source smartphone identification," *Electronic Imaging*, vol. 2019, no. 5, pp. 546–1, 2019.
[39] D. Chen, N. Zhang, Z. Qin, X. Mao, Z. Qin, X. Shen, and X.-Y. Li, "S2m: A lightweight acoustic fingerprints-based wireless device authentication protocol," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 88–100, 2016.
[40] D. Schürmann and S. Sigg, "Secure communication based on ambient audio," *IEEE Transactions on mobile computing*, vol. 12, no. 2, pp. 358–370, 2011.
[41] N. Nguyen, S. Sigg, A. Huynh, and Y. Ji, "Using ambient audio in secure mobile phone communication," in *IEEE Intl. Conf. on Pervasive Computing and Communications Workshops*. IEEE, 2012, pp. 431–434.
[42] S. Sigg, Y. Ji, N. Nguyen, and A. Huynh, "Adhocpairing: Spontaneous audio based secure device pairing for android mobile devices," in *Proceedings of the 4th Intl. Workshop on Sec. and Priv. in Spontaneous Interaction and Mobile Phone Use, IWSSI/SPMU*, vol. 12, 2012.
[43] B. Zhang, Q. Zhan, S. Chen, M. Li, K. Ren, C. Wang, and D. Ma, "Priwhisper: Enabling keyless secure acoustic communication for smartphones," *IEEE internet of things journal*, vol. 1, no. 1, pp. 33–45, 2014.
[44] M. Wang, W.-T. Zhu, S. Yan, and Q. Wang, "Soundauth: Secure zero-effort two-factor authentication based on audio signals," in *IEEE Conf. on Communications and Network Security (CNS)*. IEEE, 2018, pp. 1–9.
[45] S. Purushotham and B. Tripathy, "Evaluation of classifier models using stratified tenfold cross validation techniques," in *Intl. Conf. on Computing and Communication Systems*. Springer, 2011, pp. 680–690.
[46] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy, "Training deep neural networks on imbalanced data sets," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 4368–4374.
[47] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Annals of operations research*, vol. 134, no. 1, pp. 19–67, 2005.
[48] I. Shumailov, L. Simon, J. Yan, and R. Anderson, "Hearing your touch: A new acoustic side channel on smartphones," 2019.

Adriana Berdich is a PhD student at Politehnica University of Timisoara (UPT). She received the Engineer title in 2017 and MsC. degree in 2019, both from UPT. She has a 7-year background in the automotive industry as a Function Software Developer with main focus on torque structure and vehicle motion functions, through all phases of the V-model. Between 2015-2018 she has been working as a software developer in the automotive industry for Continental Corporation in Timisoara. Currently, she continues as a function developer in the automotive industry for Vitesco Technologies focusing on vehicle power-train applications. She was also a research student in the PRESENCE project (2019-2020) focusing on environment-based device association inside cars.

Bogdan Groza is Professor at Politehnica University of Timisoara (UPT). He received his Dipl.Ing. and Ph.D. degree from UPT in 2004 and 2008 respectively. In 2016 he successfully defended his habilitation thesis having as core subject the design of cryptographic security for automotive embedded devices and networks. He has been actively involved inside UPT with the development of laboratories by Continental Automotive and Vector Informatik. Besides regular participation in national and international research projects in information security, he lead the CSEAMAN (2015-2017) and PRESENCE (2018-2020) projects, two research programs dedicated to the security of vehicular ecosystems funded by the Romanian National Authority for Scientific Research and Innovation.

Efrat Levy is a senior security researcher at Intel Corporation and a Ph.D. student at Ben-Gurion University of the Negev (BGU). She has been actively leading significant security innovative projects in the industry and academia for more than a decade. She holds an M.Sc. degree in the field of Quantum Computing from the Hebrew University of Jerusalem, Israel. Her primary research interests are computer and network security, machine learning, cryptography and side-channel attacks.

Asaf Shabtai is a Professor in the Department of Software and Information Systems Engineering at Ben-Gurion University of the Negev. His main areas of interest are computer and network security, machine learning, security of the IoT and smart mobile devices, and security of avionic and operational technology systems.

Yuval Elovici is the director of the Telekom Innovation Laboratories at Ben-Gurion University of the Negev (BGU), head of BGU Cyber Security Research Center, Professor in the Department of Software and Information Systems Engineering at BGU. He holds B.Sc. and M.Sc. degrees in Computer and Electrical Engineering from BGU and a Ph.D. in Information Systems from Tel-Aviv University. His primary research interests are computer and network security, cyber security, web intelligence, information warfare, social network analysis, and machine learning. Prof. Elovici also consults professionally in the area of cyber security and is the co-founder of Morphisec, startup company that develop innovative cyber-security mechanisms that relate to moving target defense.

René Mayrhofer is head of the Institute of Networks and Security at Johannes Kepler University Linz (JKU), Austria and continues to be involved with Android platform security as a domain expert. His research interests include computer security, mobile devices, network communication, and machine learning, which he currently brings together in his research on securing mobile devices and digital identity. René has contributed to over 100 peer-reviewed publications and is a reviewer for numerous journals and conferences.