Detection and Suppression of Intentional EMI Attacks to Smart Speakers

Detection and Suppression of Intentional EMI Attacks to Smart Speakers

Ahmed Abdullah, *Member, IEEE,* Francesco Musolino, *Member, IEEE* and Paolo Stefano Crovetti, *Senior Member, IEEE*

***Abstract*—A new digital technique to detect and suppress Intentional Electromagnetic Interference (IEMI) attacks on Smart Speakers (SSs) without disrupting their nominal operation during the attack is proposed in this paper. The effectiveness of the proposed approach is verified by experiments on a proof-of-concept SS prototype equipped with an Artificial Intelligence speech recognition algorithm. Thanks to the proposed suppression technique, the power of malicious signal in the audio band is suppressed by $19\,\mathrm{dB}$, thus effectively preventing the execution of malicious commands in more than $99.6\,\%$ of the cases under IEMI attacks of up to $5\,\mathrm{dBm}$ injected RF power in $10\,\mathrm{MHz}$-$100\,\mathrm{MHz}$ bandwidth. Under the same test conditions, the proposed detection technique, operating independently from the suppression method, makes it possible to successfully identify IEMI attacks in more than $99.8\,\%$ of the cases. Experimental results also show the proper recognition of nominal commands while IEMI attacks are performed.**

***Index Terms*—Smart speaker, intentional electromagnetic interference (IEMI), inaudible attack on smart speaker, artificial intelligence (AI) speech recognition algorithm.**

I. Introduction

Smart speakers (SSs) are electronic systems designed to respond to voice commands which are recognized by an Artificial Intelligence (AI)-based natural language processing algorithm. They are used as virtual assistants to control smart home devices, search the web, set reminders, and perform other tasks based on input voice commands [1].

In spite of their high functionality and convenience [2], SSs raise some security-related concerns, since their operation can be impaired by several types of interference attacks, like dolphin or ultrasound attacks [3], laser pointer attacks [4], and intentional electromagnetic interference (IEMI) attacks [5]. In all these cases, an attack signal consisting of a carrier modulated according to a malicious modulating signal in the audio band is demodulated within the SS due to the nonlinearity of the SS components. This leads to the malicious audio content being added to the nominal signal, making it difficult to distinguish between the two as their spectral characteristics are similar [6]. Since the attack signal can be designed to be audible to a SS but not to human ears, attackers carry out their attack undetected [7].

Since this approach can be adopted to execute malicious commands and gain the control of a SS-based system, several techniques intended to strengthen SSs against ultra-sound and laser pointer attacks have been proposed in recent literature.

Manuscript created August, 2023; (corresponding author: Ahmed Abdullah).

A. Abdullah, F. Musolino and P. Crovetti are with the Department of Electronics and Telecommunication (DET), Politecnico di Torino, 10129 Turin, Italy (emails: ahmed.abdullah@polito.it; paolo.crovetti@polito.it; francesco.musolino@polito.it)

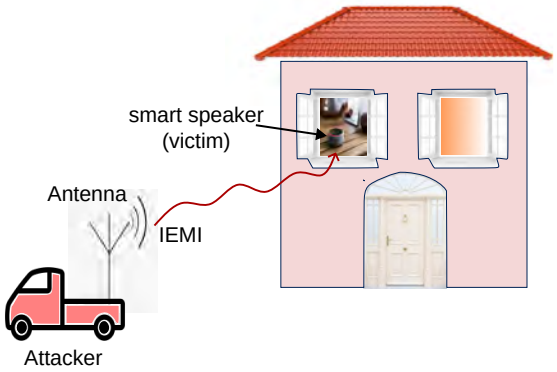

Fig. 1. Example of a Smart Speaker remotely controlled by an intentional EM interference generated by an RF attacker signal radiated from an antenna placed far away.

In particular, aiming to detect ultrasound attacks, references [3], [8], [9] propose to use ultrasound receivers to search for malicious content in the ultrasound band. In details, Mao et al. use voice activity detection (VAD) to identify malicious voice signals [8], Zhang et al. suggest to adopt low-pass filter (LPF) to eliminate the attack signal band [3], and Iijima et al. employ the presentation attack detection (PAD) method to identify audio hotspots [9]. Another method described in [10] introduces an active inaudible-voice-command cancellation (AIC) technique. It employs a custom ”guard” signal transmitter to create a suspicious spectrum within the microphone’s passband by interfering with the attack signal, which is then neutralized using software means.

In the case of laser attacks, Sugawara et al. introduce sensor fusion and sensor-based intrusion detection techniques, and also propose using barriers or diffracting films to reduce the intensity of light beam reaching the microphone [11].

On the other hand, IEMI attacks exploit the demodulation of RF signals in baseband analog integrated circuits to inject malicious content into the audio acquisition chain [12]. While ultrasound attacks and laser pointer attacks have inherent limitations, since the first require a significant power to penetrate obstacles [13] and the latter need the target to be in line-of-sight [11], IEMI attacks represent a more serious security threat, since EM waves can penetrate windows and walls with a minimum loss and do not require the direct visibility of the target [5] as illustrated in Fig. 1.

Various countermeasures have been proposed against IEMI attacks. Standard hardware techniques, such as shielding,

differential signaling, and filters, can be employed to attenuate incoming attack signals [14]. The IEMI coupling to SS can be reduced by strategically placing floating metal structures farther away from the microphone and also by improving the high-frequency connection of the microphone's can structure to the PCB return plane [15]. However, they do not entirely eliminate interference, nor can they provide the system with the capability to detect incoming attack signals. Additionally, in [14], a digital adaptive noise-cancelling technique is proposed. This technique requires a specialized hardware component in the victim device to capture incoming attack signals and subsequently suppress them, thus resulting in increased cost and complexity. The detection method described in [16] for sensor systems relies on the concept that when the sensor is deactivated, the signal received by the microcontroller connected to the sensor output should be zero. To make the sensor output unpredictable to adversaries, sensor is modulated secretly. Consequently, any deviations in the sensor output detected by the microcontroller can indicate the presence of an attacking signal. Although effective, it requires complex design, suspends operation during an attack, and lacks suppression capability. Another technique described in [17] utilizes primary and reference nominal signals processed through a differential comparator, which triggers an output when an input exceeds a certain threshold. Subsequently, a microcontroller analyzes this output for detecting attacks. While this method is effective in detection, it lacks suppression capabilities.

Furthermore, several application-level countermeasures to SS attacks, which are potentially effective regardless of the physical attack mechanism, involve the authentication of legitimate voices only. Nonetheless, artificial voice synthesis poses a challenge to these countermeasures [18]. One can replicate legitimate voices using text-to-speech-based brute force, concatenative synthesis, voice morphing [19], voice masquerading [20], or voice squatting [21] tools. To protect SS to these attacks, several improved countermeasures are proposed. VAuth technique [22] matches the body-surface vibrations of the user with the speech signal received by the microphone. VoiceLive utilizes the individual characteristics of the user's vocal system along with the stereo recording capabilities of smartphones [23]. VoiceGesture identifies users by utilizing distinct articulatory gestures made while speaking a passphrase [24]. Additionally, machine learning models, such as context-sensitive detector [20] and Resemblyzer could aid in differentiating fake and natural speech. However, all these methods are designed either for wearable devices [23] or can be bypassed by training a generative adversarial network (GAN) using samples of the target user's voice [18]. Further challenges of application-level countermeasures include the potential misinterpretation of real speech as fake due to a limited number of voice samples [18], the requirement for extensive datasets, and the limitation of addressing all real-world attack cases [20], ultimately leading to reduced overall recognition rates.

In this paper, a simple and cost-effective IEMI attack suppression and detection digital technique to be adopted in next-generation smart microphone integrated circuits (ICs) for SSs is proposed, which takes the advantage of the method presented in [25] within the context of EMI susceptibility of analog amplifiers. The new approach effectively enhances the robustness of SSs againts IEMI attacks without affecting the SS nominal operation, ensuring no disruption in the SS operation even during attacks. It overcome challenges posed by existing software countermeasures. The Tab. I summarizes the comparison of proposed method with existing methods. The effectiveness of the proposed method is demonstrated through experiments on a proof-of-concept prototype, specifically, a microphone acquisition front end equipped with an AI speech recognition algorithm.

This paper is organized as follows. In Sect. II, the block diagram of a SS highlighting the nominal signal path and IEMI attack coupling path is introduced. The spectral characteristics of the nominal signal and of the IEMI attack signal are analyzed, and the amplifier susceptibility to EMI is discussed at the integrated circuit level. The proposed digital methods to detect and suppress the IEMI attacks on SS are then explained in Sect. III and Sect. IV presents a proof-of-concept SS prototype implementing the proposed methods, and the experimental test setup. The experimental results are presented and discussed in Sect.V, where the effectiveness of the proposed approach in accurately detecting and suppressing the IEMI attacks over a wide range of EMI amplitudes and carrier frequencies is verified. Finally, some concluding remarks are drawn in Sect. VI.

II. Vulnerability of smart speakers to IEMI

In this section, the general architecture of a SS and the mechanisms which make it vulnerable to IEMI attacks are described and traced back to the susceptibility to EMI of the conditioning amplifier.

A. Smart Speaker Architecture

A smart microphone IC is composed of five essential blocks, as illustrated in Fig. 2(a): a microelectromechanical system (MEMS) microphone, an anti-aliasing low-pass filter (LPF), a conditioning amplifier, which is implemented by operational amplifier (opamp)-based circuits, an analog-to-digital converter (ADC) and a microcontroller. The microphone generates electrical signals proportional to the mechanical displacement induced by acoustic waves in a miniaturized piezoelectric element. Since microphones are designed to detect speech commands in the audible frequency range, which is concentrated below $4\,\mathrm{kHz}$ [26], the MEMS captured nominal signal, $v_{\mathrm{mic}}(t)$, is filtered by the anti-aliasing filter, whose cut-off frequency is $f_{\mathrm{max}} = 4\,\mathrm{kHz}$. The amplitude of the filtered signal, $v_{\mathrm{in}}(t)$ having frequency f_{in}, is too weak to be directly processed by the ADC, so a conditioning amplifier, with a gain A_{v_1} and a bandwidth $f_{\mathrm{B,amp}}$ in the audio range, is used to amplify it, that is $v_{\mathrm{out}}(t) = A_{v_1} \cdot v_{\mathrm{in}}(t)$. The N-bit ADC, whose reference voltage is V_{REF}, samples and quantizes the amplified signal, yielding the sequence of samples $y_{\mathrm{out}}[k] = v_{\mathrm{out}}(kT_s) \cdot 2^N / V_{\mathrm{REF}}$ where $T_s = 1/f_s$, is the sampling period and k is an integer. The signal at each digital output line of the ADC is represented as $v_{\mathrm{ADC,n}}$, where $n = 0, \cdots, N-1$. The sequence $y_{\mathrm{out}}[k]$

TABLE I
COMPARISON OF THE PROPOSED COUNTERMEASURE WITH EXISTING COUNTERMEASURES AGAINST ATTACKS ON SMART SPEAKERS.

Method	Architecture type	Attack type	Detection	Suppression	Operation continuity during attack	Complexity	Cost	Processing speed
Proposed	Hybrid	IEMI	✓✓	✓✓	✓✓	✓	✓	✓✓
Voice activity detection [8]	Hybrid	Ultrasound	✓	××	××	✓	✓×	✓✓
Low-pass filter [3]	Hybrid	Ultrasound	××	✓	✓✓	××	××	✓
Presentation attack detection (PAD) [9]	Hybrid	Ultrasound	✓	××	××	✓	✓×	✓✓
Active inaudible-voice command cancellation [10]	Hybrid	Ultrasound	✓✓	✓✓	✓✓	×	××	×
Laser detectors, & diffractors [11]	Hardware	Laser	✓	✓	××	××	××	×
EMI coupling reduction [15]	Hardware	IEMI	××	✓	✓✓	×	✓×	✓✓
EMI shielding [14]	Hardware	IEMI	××	✓	✓✓	✓✓	×	✓✓
EMI filter [14]	Hardware	IEMI	××	✓	✓✓	××	×	×
EMI hardening [14]	Hardware	IEMI	××	✓	✓✓	✓✓	×	✓✓
Adaptive noise cancellation [14]	Hybrid	IEMI	✓✓	✓✓	✓✓	××	××	××
ADC sampling-based detection [16]	Hybrid	IEMI	✓✓	××	××	××	✓✓	×
Differential comparator-based detection [17]	Hybrid	IEMI	✓✓	××	××	✓	✓	✓✓
VAuth [22]	Software	Any	✓	××	✓✓	✓×	××	××
VoiceLive [23]	Software	Any	✓	××	✓✓	✓	✓✓	××
VoiceGesture [24]	Software	Any	✓	××	✓✓	✓	✓✓	××
Machine Learning Resemblyzer [18]	Software	Any	✓	××	✓✓	✓	✓×	××

Note: (✓✓: Best), (✓: Better), (×: Worse), (××: Worst).

is first recognized by AI-based speech recognition algorithms and then processed in the microcontroller to perform various operations based on the user's speech input.

B. IEMI Coupling paths in smart speakers

The operations of the SS could be compromised by the occurrence of IEMI attacks. These attacks involve an RF carrier signal, $v_c(t)$, that is modulated by a malicious audio-band signal, $v_a(t)$. This modulated signal is then amplified using an RF power amplifier and transmitted with power P_{EMI} through an antenna towards the SS, as illustrated in Fig. 2(a). The IEMI field is coupled to the smart speaker at various points through multiple mechanisms, e.g., via PCB signal and power supply tracks, via silicon substrate, via package frame, and via non-ideal grounding [15]. All IEMI coupling mechanisms can potentially result in IEMI-related voltages being superimposed onto the nominal audio signals at virtually any point in the audio acquisition chain, where they can be demodulated by the nonlinearities of active devices. This may lead to malicious audio-band content being added to the legitimate signal.

In practice, the effects of IEMI at different points of the acquisition chain can be very different, in consideration of the different amplitude of the nominal signal on which it is superimposed, and of the different degree of nonlinearity of the different blocks: In the MEMS transducer (node 1 in Fig. 2(a)), the amplitude of the nominal signal is small (millivolts to tens of millivolts) and easily comparable with the attacker signal, but there is no effective demodulation mechanism due to the substantial linearity of the transducer itself. In contrast, IEMI that reaches the input of the opamp-based conditioning amplifier (node 2 in Fig. 2(a)), can be effectively demodulated due to the non-linearity of active devices such as bipolar or MOS transistors [12], so that relevant audio-band demodulated components are added at this point to the nominal audio signal, whose amplitude is still small.

IEMI that reaches the input of the ADC (node 3 in Fig. 2(a)) can be also demodulated due to the ADC nonlinearities and the aliasing effect, however, the demodulated components originated at this point are easily negligible compared to the amplitude of the nominal signal, which is quite high (in the Volt range) after the conditioning amplifier. Lastly, IEMI may be injected directly into the digital output lines of the ADC (node 4 in Fig. 2(a)). However, RF voltages at these nodes are not critical due to the inherent robustness to noise of digital circuits [27].

In view of these considerations, the effects of IEMI in SS will be discussed in the rest of the paper focusing on IEMI that reaches the inputs of the conditioning amplifier, since only at this point it can be effectively demodulated and can corrupt the nominal audio signal in a relevant way.

C. IEMI on the conditioning amplifier input

The amplitude of IEMI on the conditioning amplifier input can be expressed in terms of the forward power of the attacker RF amplifier as $v_{\text{EMI}} \cdot H(s)$, as illustrated in the equivalent circuit diagram in Fig. 2(b), where $v_{\text{EMI}} = \sqrt{P_{\text{EMI}} \cdot R_o}$, $R_o = 50\Omega$ and $H(s)$ is the transfer function that captures the overall coupling paths from the attacker to the input of the conditioning amplifier.

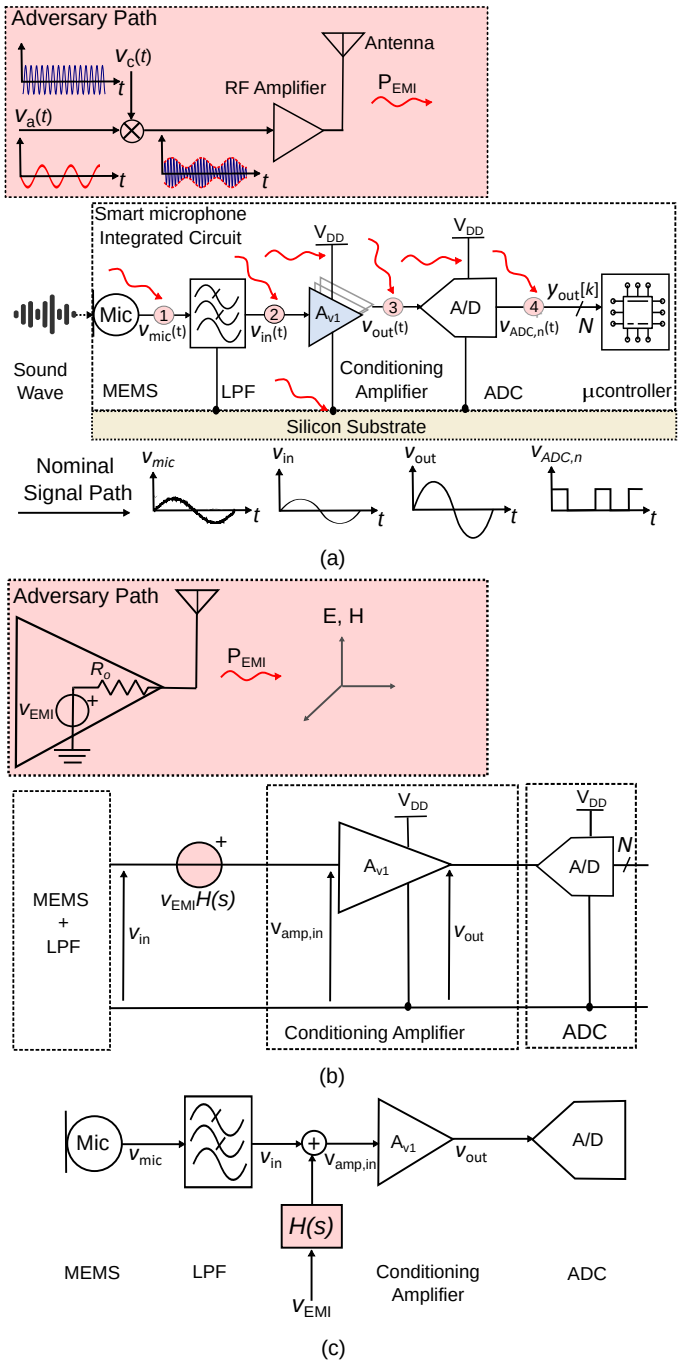

Fig. 2. Smart Speakers architecture (a) nominal signal path and IEMI coupling points (b) equivalent circuit diagram and (c) simplified block diagram.

Applying Kirchhoff's Voltage Law (KVL) in Fig. 2(b) gives the input of the conditioning amplifier as

$$V_{\text{amp,in}} = V_{\text{in}} + V_{\text{EMI}} \cdot H(s). \tag{1}$$

Based on (1), the simplified block diagram of the SS under attack is shown in Fig. 2(c).

The coupling function could be coursely estimated considering the situation depicted in Fig. 3, where a transmitting antenna radiating electromagnetic field is located at a distance d from a conductor. The power density, S, of a transmitter

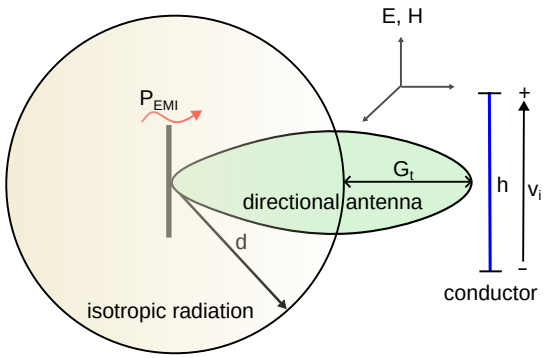

Fig. 3. Voltage induced in a circuit by electromagnetic field emitted through antenna.

antenna with an antenna gain G_{t} and power P_{EMI} at a distance d is given by

$$S = \frac{P_{\text{EMI}} G_{\text{t}}}{4\pi d^2}. \tag{2}$$

Furthermore, it is known that the power density of a plane wave is defined as the product of the E-field and H-field. In a free-space it will be

$$S = |E \times H| = \frac{|E|^2}{Z_o} \tag{3}$$

where $Z_o = 376.7\,\Omega$ is a free-space impedance. By equating (2) and (3), we get

$$|E| = \sqrt{\frac{Z_o P_{\text{EMI}} G_{\text{t}}}{4\pi d^2}}. \tag{4}$$

The EMI voltage v_{i} induced in a conductor of length h is given by

$$v_{\text{i}} = \sqrt{\frac{Z_o P_{\text{EMI}} G_{\text{t}}}{4\pi d^2}} \cdot h. \tag{5}$$

It can be inferred, using equation (5), that an attacker has the capability to induce peak RF voltage in the order of several hundred millivolts into a victim circuit with a size typically around $10^{-2}\,\text{m}$. This can be achieved by employing an antenna positioned at a distance of $5\,\text{m}$ - $6\,\text{m}$, while transmitting power levels around $50\,\text{dBm}$.

Various research investigations have demonstrated the feasibility of IEMI attacks. According to [28], a highly directive Vivaldi antenna transmitted an RF signal in the MHz range at a power of $32.6\,\text{dBm}$ $(1.82\,\text{W})$ towards an analog integrated circuit containing a sensor and an ADC. The experimental results showed that at distances of $0.1\,\text{m}$ and $1\,\text{m}$, the induced peak RF voltage within the circuit exceeded $1\,\text{V}$ and $100\,\text{mV}$, respectively. Another test in [12], involved a microphone circuit exposed to modulated EMI from a biconical antenna with a $100\,\text{MHz}$ RF carrier frequency. The induced RF voltages were measured to be in the range of several hundred millivolts peak-to-peak, a significant level according to the European standard EN55020 [29].

D. IEMI Demodulation

To gain a better understanding of baseband AM demodulation of IEMI-based attacks, a sine wave RF carrier signal

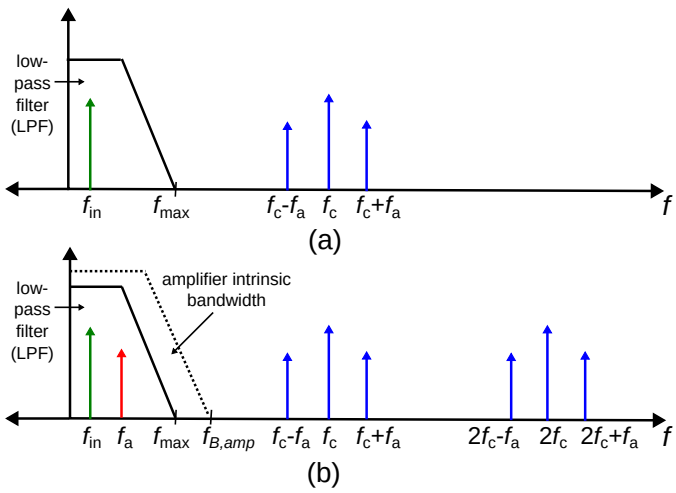

Fig. 4. Frequency Spectrum of (a) nominal signal and RF-modulated signal, (b) nominal signal, attack signal and RF-modulated signal filtered.

$[v_c(t) = A_c \sin(2\pi f_c t)]$ and a sinusoidal baseband modulating attack signal $[v_a(t) = M\cos(2\pi f_a t)]$ having frequencies f_c and f_a, respectively, with relation: $f_a < f_{\max} \ll f_c$, where A_c is the amplitude of carrier signal and M is the amplitude of modulating attack signal are considered in what follows. In practice, $v_a(t)$ is a vocal signal and needs not to be a sinewave, but here we are just considering a sinewave for the sake of illustration. The $v_{\text{EMI}}(t)$ is given by [30]

$$\begin{aligned} v_{\text{EMI}}(t) &= A_c \sin(2\pi f_c t)[1 + m \cdot \cos(2\pi f_a t)] \\ v_{\text{EMI}}(t) &= \frac{A_c m}{2}[\sin(2\pi(f_c + f_a)t) + \sin(2\pi(f_c - f_a)t)] + \\ &+ \; A_c \sin(2\pi f_c t) \end{aligned} \tag{6}$$

where $m = M/A_c$ is the modulation index.

The nominal signal, $v_{\text{in}}(t)$, is taking the nominal path, while $v_{\text{i}} = |H(f_c)| \cdot v_{\text{EMI}}(t)$ is injected through the adversary path illustrated in Fig. 2(c). The spectra of both legitimate and EMI modulated attack signals are shown in Fig. 4(a), which reveals that the nominal spectral content, f_{in}, lies within the LPF bandwidth, while the EMI spectral components, i.e., $f_c \pm f_a$ and f_c exist at significantly higher frequencies than $f_{\max}$. Due to the nonlinear effects of EMI in an integrated opamps [31], it not only generates the expected linear component but also introduces a non-linear EMI-induced distortion component expressed in (7),

$$v_{\text{out}}(t) = A_{v_1} v_{\text{amp,in}}(t) + A_{v_2} v^2_{\text{amp,in}}(t) + A_{v_3} v^3_{\text{amp,in}}(t) + \ldots. \tag{7}$$

By inserting (1) in (7), we get

$$\begin{aligned} v_{\text{out}}(t) = A_{v_1}[v_{\text{in}}(t) + |H(f_c)| \cdot v_{\text{EMI}}(t)] + A_{v_2}[v_{\text{in}}(t) + |H(f_c)| \cdot \\ \cdot v_{\text{EMI}}(t)]^2 + A_{v_3}[v_{\text{in}}(t) + |H(f_c)| \cdot v_{\text{EMI}}(t)]^3 + \ldots. \end{aligned} \tag{8}$$

In the above expression, the most critical term is $v^2_{\text{EMI}}(t)$, which can be expanded as follows:

$$\begin{aligned} v^2_{\text{EMI}}(t) = \frac{A_c^2 m^2}{8}\{2 + 2\cos(2\pi \cdot 2f_a t) - 2\cos(2\pi \cdot 2f_c t) - \\ - \cos[2\pi \cdot 2(f_c + f_a)t] - \cos[2\pi \cdot 2(f_c - f_a)t]\} - \\ - \frac{A_c^2 m}{2}\{\cos[2\pi(2f_c + f_a)t] - 2\cos(2\pi f_a t) + \\ + \cos[2\pi(2f_c - f_a)t]\} - \frac{A_c^2}{2}[1 - \cos(2\pi \cdot 2f_c t)]. \end{aligned} \tag{9}$$

Eqn. (9) reveals the presence of baseband spectral components at frequency f_a in addition to f_{in}, and additional high-frequency components, i.e., $2f_c \pm f_a$, $2f_c$, $2f_a$and $2f_c \pm 2f_a$ in the output. The intrinsic filtering effect of the conditioning amplifier due to its bandwidth limitation removes the high frequency components leaving the baseband ($B.B.$) components as depicted in Fig. 4(b), i.e.,

$$v^2_{\text{EMI}}(t)|_{B.B.} = A_c^2 m \cos(2\pi f_a t) = A_c v_a(t). \tag{10}$$

Based on (8), the amplifier output voltage can be therefore expressed as:

$$v_{\text{out}}(t) = A_{v_1} v_{\text{in}}(t) + A_{v_2} A_c |H(f_c)|^2 v_a(t). \tag{11}$$

which reveals that the attack signal is demodulated within the nominal spectral bandwidth and can be executed.

E. Amplifier Susceptibility to EMI

The susceptibility to EMI of audio amplifiers based on opamp integrated circuits has been extensively studied in the literature [32] and will be briefly discussed in what follows, considering the circuit model of the operational amplifier input stage depicted in Fig. 5(a).

In the presence of EMI, both differential and common-mode RF voltage components are present. Due to the parallel admittance of the tail current source (M3 in Fig. 5(a)), which is high at RF due to the presence of parasitic capacitance C_T , common-mode disturbances are translated into RF fluctuations i_s of the differential pair bias current I_o. In these conditions, the differential pair functions as an RF mixer, resulting in the multiplication of the EMI components of v_D and i_s, and the output differential current can be expressed as

$$i_D = g_{\text{m}} v_d + g_{\text{p}} v_d i_s \tag{12}$$

where g_{m} is the linear transconductance of the differential pair and $g_{\text{p}} = g_{\text{m}}/2I_o$.

This mechanism results in even-order non-linearity, that demodulates high-frequency EMI giving rise to an equivalent EMI-induced baseband error, Δv_{EMI}, that is superimposed onto the nominal opamp input [33] as shown in Fig. 5(b).

Based on (11), the EMI-induced output voltage of the conditioning amplifier in closed-loop, as depicted in Fig. 5(b), can be expressed in terms of Δv_{EMI} as:

$$v_{\text{out}}(t) = A_{v_1}(v_{\text{in}}(t) + \Delta v_{\text{EMI}}(t)) \tag{13}$$

where

$$\Delta v_{\text{EMI}}(t) = \frac{A_{v_2} A_c}{A_{v_1}} \cdot |H(f_c)|^2 \cdot v_{\text{a}}(t) \tag{14}$$

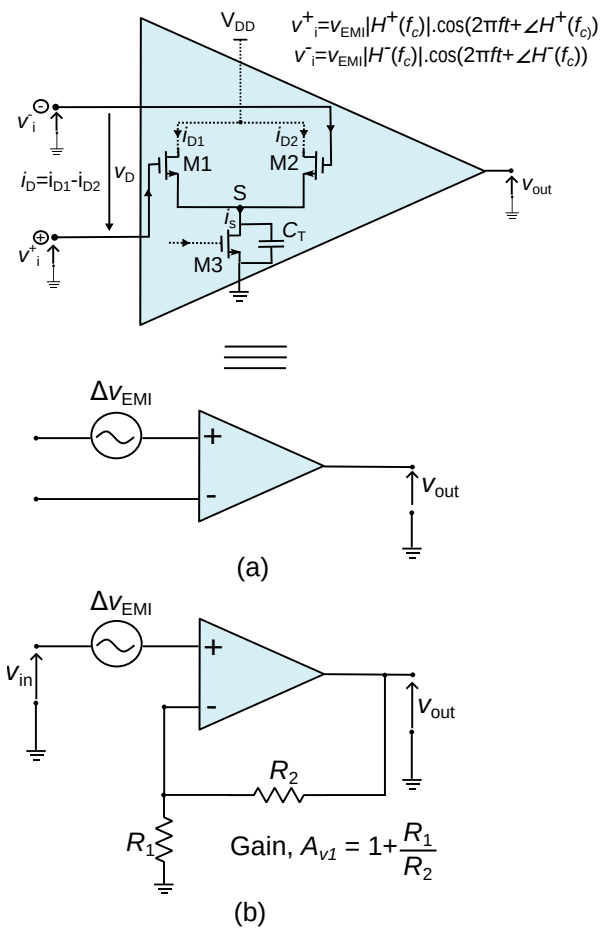

Fig. 5. CMOS Operational amplifier (a) simplified schematic view, (b) closed-loop circuit.

and $\frac{A_{v_2}A_c}{A_{v_1}}$ is a constant that depends on opamp circuit parameters and parasitics as explained in [33]. Most importantly, it has been shown in [34] that A_{v_2} exhibits a positive (negative) value in an nMOS (pMOS) input differential pair opamps.

III. IEMI Attack Suppression and Detection

Taking advantage of the insight in the nonlinear mechanisms which give rise to EMI demodulation in opamp-based circuits, which have been revised in Sect. II, a novel smart microphone IC architecture and a digital processing technique offering inherent IEMI attack detection and suppression features is presented in what follows.

A. Digital suppression method to mitigate IEMI attacks

According to the proposed method, the architecture of the SS, as depicted in Fig. 2, is modified by replacing a single conditioning amplifier with two conditioning amplifiers having the same gain, bandwidth, and feedback network. However, these amplifiers are based on opamps with different input stages: one with an nMOS input stage and the other with a pMOS input stage. Thus, they exhibit similar gain characteristics but opposite susceptibilities to EMI. Based on the analysis presented in Sect. II, the opamp output voltages in the proposed structure (see Fig. 6) can be expressed as:

$$v_{\text{out,i}}(t) = A_{v_1}\left(v_{\text{in}}(t) + \Delta v_{\text{EMI,i}}(t)\right) \tag{15}$$

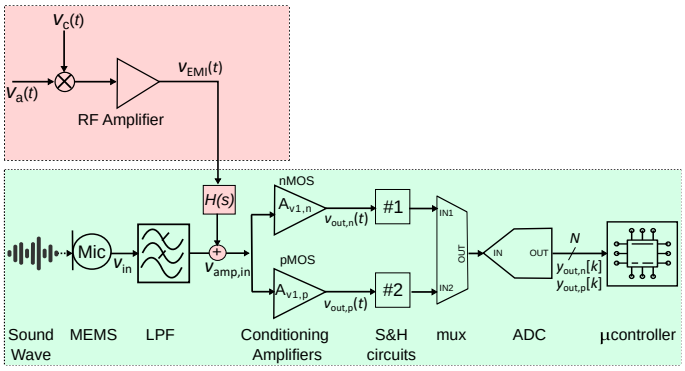

Fig. 6. Proposed architecture of smart speaker.

where

$$\Delta v_{\text{EMI,i}}(t) = \frac{A_{v_{2,i}}A_c}{A_{v_1}} \cdot |H(f_c)|^2 \cdot v_{\text{a}}(t) \tag{16}$$

in which, $A_{v_{2,i}}(i = \text{n(p)})$ is positive (negative) for an nMOS (pMOS) input differential pair [34] as discussed above.

In the proposed solution, the output voltages $v_{\text{out,n}}(t)$ and $v_{\text{out,p}}(t)$ of these two amplifiers are simultaneously sampled at instant kT_s using a sample-and-hold (S&H) circuit and quantized by the same ADC via a multiplexer (mux), such that $y_{\text{out,n}}[k] = v_{\text{out,n}}(kT_s) \cdot \frac{2^N}{V_{\text{REF}}}$ and $y_{\text{out,p}}[k] = v_{\text{out,p}}(kT_s) \cdot \frac{2^N}{V_{\text{REF}}}$. Based on (15), they can be represented in the digital domain as:

$$y_{\text{out},i}[k] = x_{\text{in}}[k] + n_{\text{EMI},i}[k] \tag{17}$$

where

$$x_{\text{in}}[k] = \frac{A_{v_1}v_{\text{in}}(kT_s)}{V_{\text{REF}}}2^N \tag{18}$$

is the input digital stream in an EMI-free environment, and

$$n_{\text{EMI},i}[k] = \frac{A_{v_1}\Delta v_{\text{EMI},i}(kT_s)}{V_{\text{REF}}}2^N. \tag{19}$$

Considering the weighted-sum of $y_{\text{out,n}}[k]$ and $y_{\text{out,p}}[k]$ with weights α and $(1-\alpha)$ summing to one, that is

$$y_{\text{out}}[k] = \alpha \cdot y_{\text{out,n}}[k] + (1-\alpha) \cdot y_{\text{out,p}}[k] \tag{20}$$

and considering (17), the equation (20) can be expanded as

$$y_{\text{out}}[k] = x_{\text{in}}[k] + [\alpha \cdot n_{\text{EMI,n}}[k] + (1-\alpha) \cdot n_{\text{EMI,p}}[k]] \tag{21}$$

which highlights the legitimate signal $x_{\text{in}}[k]$ and the EMI-induced error term:

$$\varepsilon_{\text{EMI}}[k] = \alpha \cdot n_{\text{EMI,n}}[k] + (1-\alpha) \cdot n_{\text{EMI,p}}[k] \tag{22}$$

which can be cancelled-out for a specific value of $\alpha = \alpha^*$, i.e.,

$$\varepsilon_{\text{EMI}}[k] = \alpha^* \cdot n_{\text{EMI,n}}[k] + (1-\alpha^*) \cdot n_{\text{EMI,p}}[k] = 0. \tag{23}$$

For $n_{\text{EMI,n}}[k] \neq n_{\text{EMI,p}}[k]$, in particular:

$$\alpha^*[k] = \frac{n_{\text{EMI,p}}[k]}{n_{\text{EMI,n}}[k] - n_{\text{EMI,p}}[k]}. \tag{24}$$

By replacing (16) in (19) and then in (24), we get

$$\alpha^* = \frac{\frac{A_{v_1}2^N}{V_{\text{REF}}}\frac{A_{v_{2,\text{p}}}A_c}{A_{v_1}}|H(f_c)|^2 v_{\text{a}}[k]}{\frac{A_{v_1}2^N}{V_{\text{REF}}}\frac{A_{v_{2,\text{n}}}A_c}{A_{v_1}}|H(f_c)|^2 v_{\text{a}}[k] - \frac{A_{v_1}2^N}{V_{\text{REF}}}\frac{A_{v_{2,\text{p}}}A_c}{A_{v_1}}|H(f_c)|^2 v_{\text{a}}[k]} = \frac{A_{v_{2,\text{p}}}}{A_{v_{2,\text{n}}} - A_{v_{2,\text{p}}}} \tag{25}$$

which reveals that the value of α^*, resulting in a full cancellation of the EMI-induced error, solely depends on the constants $A_{v_2,p}$ and $A_{v_2,n}$ and is independent of the EMI incident power amplitudes and frequencies, and of the discrete time instant k.

It follows that α^* can be obtained by performing one-point calibration in the presence of continuous wave EMI. The corresponding $n_{\text{EMI},i}[k]$, which are demodulated as a single DC offset voltages, are computed based on the acquired $y_{\text{out},i}$ and for the known x_{in} as

$$n_{\text{EMI},i}[k] = y_{\text{out},i}[k] - x_{\text{in}}[k] \tag{26}$$

are directly applied in (24) to get α^*. For a more effective IEMI cancellation, α^* can be evaluated by minimizing the sum of square errors computed by (22) over the calibration set encompassing various EMI incident power amplitudes and frequencies, and nominal signal values within the input range of the conditioning amplifiers [25], so that to have an optimal EMI cancellation over the all practical IEMI amplitude and frequency ranges even in the presence of higher-order effects not captured by (22).

Inadequate calibration leads to insufficient suppression of EMI-induced attacks. Although calibration remains valid indefinitely, it is advisable to perform regular background recalibration to track the effects of temperature variations and aging. This can be achieved by minimizing total noise power, assuming that the attack signal and nominal signal are not correlated [35]. Furthermore, since the signal path from the microphone is split into two paths towards the input of the amplifiers, it is advisable to keep the lengths of these paths very short, exactly the same, and close to each other to minimize the probability of unbalanced IEMI.

By utilizing α^* in equation (20), we can suppress the EMI-induced attack signal by digitally processing the corrupted signals without impacting the SS nominal operation, thus ensuring no disruption during attacks, even when both attack and nominal signals are present simultaneously.

B. Attack detection

The same hardware architecture in Fig. 6, which has been introduced in Sect. III-A for IEMI suppression, can be conveniently exploited to detect IEMI attacks in SSs.

For this purpose, it can be observed that the quantity

$$z[k] = y_{\text{out},n}[k] - y_{\text{out},p}[k] \tag{27}$$

which can be easily calculated in the digital domain from the ADC output stream, can be expressed from (17) as:

$$z[k] = n_{\text{EMI,n}}[k] - n_{\text{EMI,p}}[k] = \frac{A_c|H(f_c)|^2 v_{\text{a}}[k]\cdot(A_{v_{2,\text{n}}} - A_{v_{2,\text{p}}})}{V_{\text{REF}}}2^N \tag{28}$$

which is independent of the nominal signal $x_{\text{in}}[k]$ and in which the demodulated IEMI contributions $n_{\text{EMI,n}}[k]$ and $n_{\text{EMI,p}}[k]$ are effectively summed due to the opposite sign of $A_{v_{2,\text{n}}}$ and $A_{v_{2,\text{p}}}$.

Based on that, a value of $|z[k]|$ exceeding the noise floor is a clear indicator of an ongoing IEMI attack, and the Boolean flag

$$F[k] = |z[k]| > \epsilon \tag{29}$$

where the threshold ϵ is chosen so that to be slightly larger than the noise floor of the system, can be introduced as an IEMI warning.

If no attack is being performed, the attack signals are zero, resulting in $F[k] = 0$. It should be noted that some levels of noise may be present during practical use, but the output amplitude is usually negligible and considered as zero. However, if an attack occurs, the amplifiers will pick up the attack signal, resulting in a non-zero signal greater than ϵ, which allows the microcontroller to detect the attack ($F[k] = 1$).

It can be observed that detection method operates independently of the suppression method and do not necessitate calibration. This ensures that an attack will be detected even if it cannot be completely suppressed by the suppression method described in Sect.III-A for any reason.

IV. Experimental Validation

The effectiveness of the proposed technique, that is intended to be embedded in next-generation smart microphones ICs, has been experimentally validated by a PCB proof-of-concept prototype. The prototype includes the main building blocks of an audio acquisition front-end, similar to the one adopted in [12] for EMI susceptibility analysis, and has been tested under IEMI injection with reference to amplitude and frequency ranges similar to those expected in a real SS microphone undergoing IEMI attacks based both on the IEMI coupling analysis outlined in Sect. II-B and on experimental results on EMI coupling presented in [12] and also in [5], [14] and [15].

Our prototype includes an EMI-robust PCB test circuit and a peripheral board. The PCB test circuit is designed as in Fig. 6, and its detailed schematic can be found in [25]. The circuit is specifically designed by using two different precision opamp circuits available in market, i.e., the OPA2277 opamp [36] and the TLC272 [37]. These feature an nMOS and a pMOS input differential pair, respectively, and show opposite EMI induced offset voltage under continuous wave EMI excitation. The PCB test circuit contains an LC network so that to superimpose RF-modulated attack signal onto the legitimate signal that needs to be processed. The simplified schematic diagram of the test PCB considered in this work is shown in Fig. 7(a), while Fig. 7(b) presents a photograph of the PCB.

A Texas instruments c2000 microcontroller peripheral board containing an onboard 16-channel 12-bits ADC module with a 3 V input range, equipped with two S&H units and a 150 MHz TMS320F2833X processor [38], is chosen for digitization and to implement the digital EMI-suppression and detection method presented in sect. III.

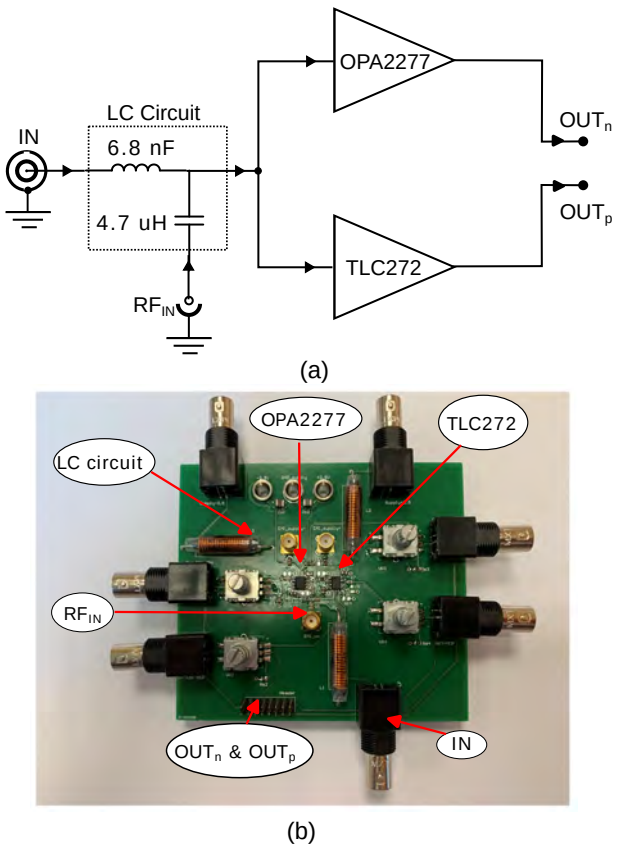

Fig. 7. PCB Test Circuit (a) part of schematic diagram (b) photograph of the PCB.

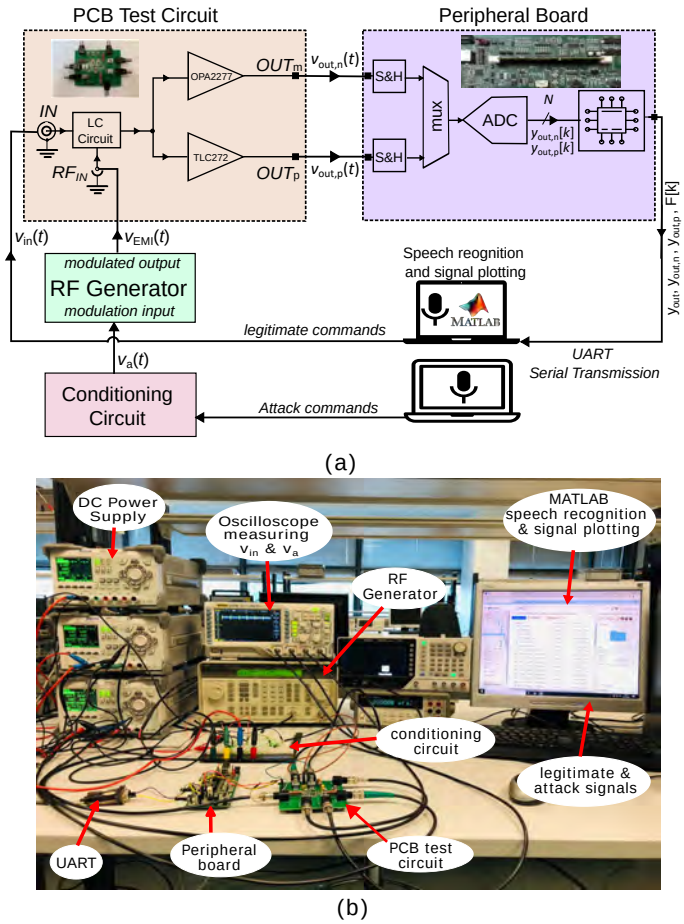

Fig. 8. Experimental setup (a) block diagram and (b) picture in the lab.

A. Experimental test setup

The block diagram of the whole test setup is depicted in Fig. 8(a) while its photograph is presented in Fig. 8(b). The audio signals, both legitimate ($v_{\text{in}}(t)$) and attack ($v_{\text{a}}(t)$), are generated by a PC sound card and retrieved from the PC audio output (line-out) by a jack connector. The nominal input $v_{\text{in}}(t)$ is fed to the PCB test circuit at IN, by a stereo-to-BNC cable. Similarly, $v_{\text{a}}(t)$ is generated by a different PC sound card and is fed to the modulation channel input connector of the RF signal source HP-8648C [39], through a conditioning circuit to ensure the compatibility with the input swing of the RF modulator. The RF signal source is capable of producing a modulated RF signal with a carrier in $1\,\text{MHz}$ -$3.2\,\text{GHz}$ bandwidth and generating an incident power of up to $20\,\text{dBm}$ for EMI injection. The RF signal, $v_{\text{EMI}}(t)$, which is modulated by $v_{\text{a}}(t)$, is then fed to the PCB test circuit at RF_{IN}, by an SMA cable. The DC power supply is used to provide a supply voltage of +$2.5\,\text{V}$ and -$2.5\,\text{V}$, relative to the reference voltage of the input signal ($0\,\text{V}$) to PCB test circuit.

The outputs of the conditioning amplifiers $v_{\text{out,n}}(t)$ and $v_{\text{out,p}}(t)$ taken at OUT_{n} and OUT_{p}, respectively, of the PCB test circuit are connected to two independent input channels of ADC module, which are configured to simultaneously sample both signals at $f_s = 8\,\text{kHz}$, $N =$ 12-bits and $V_{\text{REF}} = 3\,\text{V}$. Consequently, the digitized signals $y_{\text{out,n}}[k]$ and $y_{\text{out,p}}[k]$ are made available to a processor where the attack detection ($F[k]$) and suppression ($y_{\text{out}}[k]$) strategy is implemented. The calibration process in sect. III-A yielded an optimal coefficient of $\alpha^* = 0.37$, and attack detection flag threshold value described in sect. III-B specifies the $\epsilon = 6$ LSB for the conditioning amplifiers considered in this test.

B. Speech Recognition Algorithm

To validate the proposed IEMI detection and suppression strategy, the digital output streams $y_{\text{out,n}}[k]$ and $y_{\text{out,p}}[k]$ obtained converting the OPA2277 and the TLC272 output voltages, respectively, and the stream $y_{\text{out}}[k]$ obtained applying the proposed IEMI suppression as in (20) are transmitted to a PC through UART and are processed by a speech recognition algorithm, which enables a SS to accurately identify different words. A comparison has been made between proposed EMI-suppressed output by setting $\alpha = \alpha^*$ and the output obtained in a traditional acquisition front-end featuring a conditioning amplifier based on the OPA2277 opamp only [without suppression ($\alpha = 1$), OPA2277] or based on the TLC272 opamp only [without suppression ($\alpha = 0$), TLC272], with reference to the nominal signal and attack signal.

The speech recognition has been performed using an open-source algorithm based on a Convolutional Neural Network (CNN) implemented within MATLAB environment [40]. The CNN was pre-trained on a publicly available large dataset containing 2000 training files and 300 validation files for each word, such as *down*, *go*, *up*, *no*, *yes*, *left*, *right*,

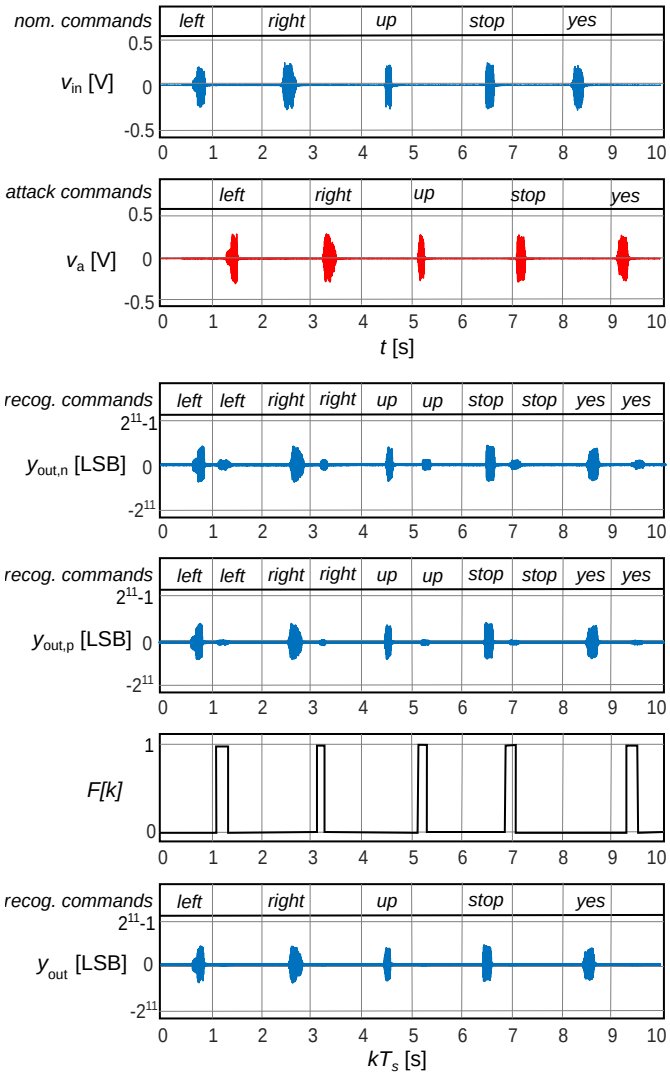

Fig. 9. Results obtained under attack command modulated with EMI excitation of 50 MHz and incident power of 0 dBm.

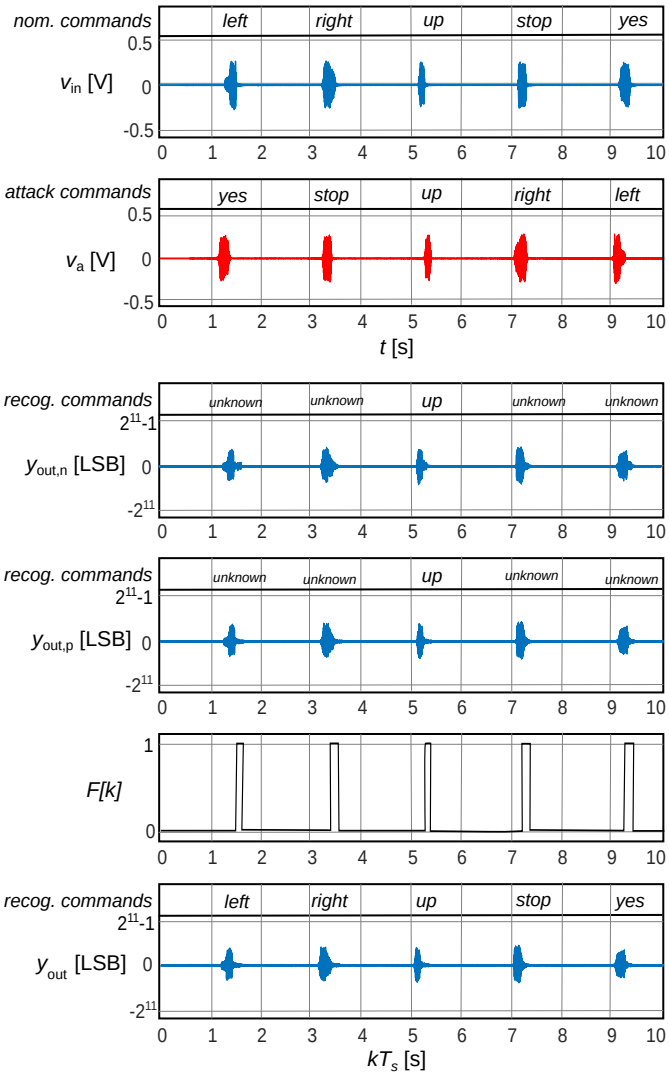

Fig. 10. Results obtained when both legitimate and attack commands (modulated with EMI excitation of 50 MHz and incident power of 0 dBm) occur simultaneously.

stop, and *on*, [41] with no customization made by us. As reported by [40], the training enables the algorithm to process audio signals, extract relevant features like the centroid and kurtosis of the Bark spectrum, as well as the pitch of the audio signal. These features are then matched with accurately identified speech commands, resulting in accuracy rates of 97.28 % on the training data and 93.61 % on the validation data. This ensures that the SS can effectively recognize and respond to speech commands with a high level of accuracy. The misclassified speech is categorized as $unknown$.

V. Results and Analysis

The results obtained to validate the proposed method for detecting and suppressing IEMI attacks in SSs are presented and analyzed in what follows.

A. IEMI Attack Detection and Suppression

The architecture has been tested by applying 10 s-long audio sequences as a nominal signal v_{in}, and the same, just shifted by 1 s stream, is applied as an attack signal v_{a}. The audio streams consist of five commands: $left$, $right$, up, $stop$ and yes in a sequence of one command after each two seconds as shown in Fig. 9. The attack signal has been modulated with an EMI excitation frequency of 50 MHz and an incident power of 0 dBm (equivalent to a peak-to-peak voltage of 630 mV).

The results in Fig. 9 show that the signals obtained by directly digitizing the output voltages of OPA2277 ($y_{\text{out,n}}[k]$) and TLC272 ($y_{\text{out,p}}[k]$) contain both legitimate and attack commands, since their audio contents are interpreted using a recognition algorithm. The amplitude of the attack commands differs, having a root mean square (rms) amplitudes of 14.33 LSB and 7.16 LSB for OPA2277 and TLC272 opamps, respectively, due to their different susceptibility to EMI.

These attack commands are accurately detected by the proposed detection method, denoted as $F[k]$ in Fig. 9.

When the proposed digital suppression method is applied, the attack commands are effectively suppressed, represented as y_{out} in Fig. 9, resulting in a rms value of 4.09 LSB, equivalent to the noise floor level. This translates to a power suppression of more than 19 dB compared to OPA2272. Thanks to the

TABLE II
PROPOSED METHOD ACCURACY RATE COMPUTED FOR ATTACK SIGNAL MODULATED WITH EMI 50 MHz AND 0 dBm.

Architecture	rms [LSB]	Std. dev. [LSB]	Attack success rate	Detection rate (proposed)
with digital supp. (proposed)	4.09	0.16	0.4 %	>99.8 %
without digital-supp. OPA2277	14.06	1.81	88 %	>99.8 %
without digital-supp. TLC272	6.96	0.61	61.3 %	>99.8 %

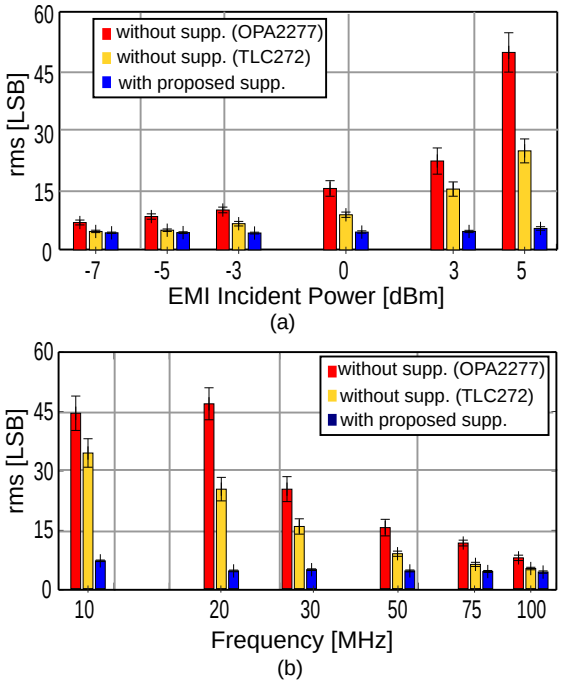

Fig. 11. Computed rms values of attack commands obtained by system with and without digital suppression applied for (a) attack commands modulated with EMI excitation frequency of 50 MHz and incident power ranging from −7 dBm to 5 dBm and (b) with incident power of 0 dBm and frequency ranging from 10 MHz to 100 MHz.

proposed suppression algorithm, the amplitude of the attackers is reduced below the noise floor and the malicious vocal commands are no longer recognized by the SS. This confirms the effectiveness of the proposed suppression method.

Furthermore, when both attack and legitimate signals are applied simultaneously, the commands in $y_{\text{out,n}}$ and $y_{\text{out,n}}$ cannot be recognized, thus resulting in their categorization as $unknown$, as depicted in Fig. 10. In such case, our approach allows for the suppression of only the attack signal, as shown in y_{out}, ensuring that the legitimate signal remains unaffected and thereby preserving the nominal functioning of the system.

B. Accuracy Rate of Attack Detection and Suppression

To conduct a statistical comparison, the test presented in Fig. 9 is replicated 100 times. The architecture is characterized with attack suppression and detection, considering parameters such as the mean rms value, standard deviation, attack success rate, and detection rate. The attack success rate represents the rate of attacks successfully recognized by the recognition algorithm, while the detection rate indicates the rate at which attacks are detected by the detection method.

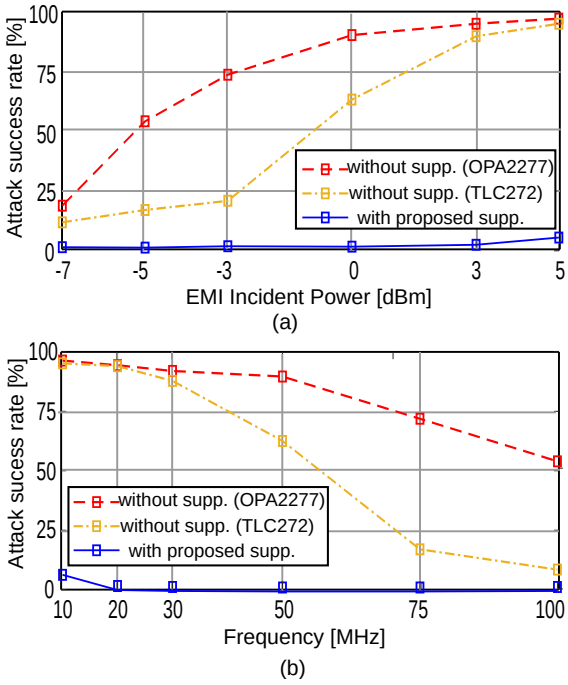

Fig. 12. Success rate of attack commands recognized by system with and without digital suppression applied for (a) attack commands modulated with EMI excitation frequency of 50 MHz and incident power ranging from −7 dBm to 5 dBm and (b) with incident power of 0 dBm and frequency ranging from 10 MHz to 100 MHz.

The computed results are presented in Tab. II. Without digital suppression, the rms amplitude for attack commands in the outputs of both opamps corresponds to 14.06 LSB(6.96 LSB) with a standard deviation of 1.81 LSB (0.61 LSB) for OPA2277 (TLC272) opamp, respectively. Correspondingly, the attack success rate is found to be 88 % (61.3 %).

After applying the digital-suppression method, attack commands are successfully suppressed, resulting in rms amplitude of 4.09 LSB with a standard deviation of 0.16 LSB. Consequently, the attack success rate is significantly reduced to 0.4 %.

The proposed detection method, operating independently, detects all the attacks in our 100 test cases, i.e., it is effective in more than 99.8 % of the cases, irrespective of whether they are recognized by the recognition system or not. This capability allows identification of malicious commands in the system, even in the rare cases (0.4 % in our experiment) when attacks are not completely suppressed and have succeeded.

The architecture underwent extensive testing using the same signals under various conditions. The attack signal is modulated with EMI excitation at a frequency of 50 MHz, with incident power ranging from −7 dBm to 5 dBm (corresponding to a peak-to-peak voltage of 0.283 V to 1.1247 V). Additionally, it is modulated with an incident power of 0 dBm and a frequency ranging from 10 MHz to 100 MHz.

The corresponding rms values with standard deviations are reported in the barplot shown in Fig. 11. The attack success rates are plotted in Fig. 12. These results indicate that the proposed digital suppression method effectively reduces the components of the malicious signal to a maximum rms of 4.91 LSB and a minimum of 3.82 LSB, compared to a maximum

of 45.61 LSB (32.91 LSB) and a minimum of 7.64 LSB (4.08 LSB) for the OPA2277 (TLC272) opamp. This reduction leads to a maximum attack success rate of 6.2 % and a minimum of 0.2 %, compared to a maximum of 96.2 % (93.4 %) and a minimum of 20.2 % (12.4 %) for the OPA2277 (TLC272) opamp, across the entire range of EMI incident powers and frequencies.

Based on the experimental results, the proposed IEMI suppression method provides satisfactory results over a wide range of EMI incident power amplitudes up to 5 dBm. At higher IEMI amplitudes, the suppression method may not completely cancel the malicious content, but, at least, such high-power attacks are very effectively detected by the proposed detection method.

VI. Conclusion

In this paper, a technique has been proposed to detect and suppress the IEMI-modulated attacks appearing at a particularly critical node within the audio acquistion front-end of a smart microphone integrated circuit. The approach has been experimentally validated using a proof-of-concept smart speaker prototype equipped with an AI speech recognition algorithm to interpret input speech commands. The effectiveness of the approach is assessed by testing multiple attack commands modulated with EMI incident powers amplitudes ranging from -7 dBm to 5 dBm and frequencies ranging from 10 MHz to 100 MHz. The proposed technique is able to suppress the power of the malicious signal in the audio band down to 19 dB and detect malicious commands in more than 99.8 % of cases. The results indicate significant improvements, confirming the efficacy of our approach. The preliminary results obtained from our proof-of-concept prototype can be conveniently exploited to develop new smart microphone ICs. An IC implementation, provided that the effects of parasitics and process tollerances is properly taken into account, is expected to further reduce IEMI coupling and hence to enhance the robustness of the proposed approach. Moreover, the technique can be used synergistically with other methods, such as EMI shielding, ADC sampling-based detection, and various software techniques, to further enhance its effectiveness.

References

[1] R. Aldhafiri, G. Powell, E. Smith, and C. Perera, "Voice-enabled privacy assistant towards facilitating successful ageing in smart homes," in *2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, 2023, pp. 343–345.

[2] K. Baimirov, E. Mergengali, and B. Baimirov, "Overview of the latest research related to smart speakers," in *2022 IEEE 7th International Energy Conference (ENERGYCON)*, 2022, pp. 1–5.

[3] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinatack: Inaudible voice commands," *CoRR*, vol. abs/1708.09537, 2017.

[4] A. Seyitmammet, M. A. Al-Absi, A. A. Al-Absi, and H. J. Lee, "Attack on AI smart speakers with a laser beam," in *Proceedings of 2nd International Conference on Smart Computing and Cyber Security*. Singapore: Springer Nature Singapore, 2022, pp. 32–45.

[5] Z. Xu, R. Hua, J. Juang, S. Xia, J. Fan, and C. Hwang, "Inaudible attack on smart speakers with intentional electromagnetic interference," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 5, pp. 2642–2650, 2021.

[6] C. Yan, X. Ji, K. Wang, Q. Jiang, Z. Jin, and W. Xu, "A survey on voice assistant security: Attacks and countermeasures," *ACM Comput. Surv.*, vol. 55, no. 4, nov 2022. [Online]. Available: https://doi.org/10.1145/3527153

[7] S. S. Alchekov, M. A. Al-Absi, A. A. Al-Absi, and H. J. Lee, "Inaudible attack on AI speakers," *Electronics*, vol. 12, no. 8, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/8/1928

[8] J. Mao, S. Zhu, X. Dai, Q. Lin, and J. Liu, "Watchdog: Detecting ultrasonic-based inaudible voice attacks to smart home systems," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8025–8035, 2020.

[9] R. Iijima, S. Minami, Y. Zhou, T. Takehisa, T. Takahashi, Y. Oikawa, and T. Mori, "Audio hotspot attack: An attack on voice assistance systems using directional sound beams and its feasibility," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 4, pp. 2004–2018, 2021.

[10] Y. He, J. Bian, X. Tong, Z. Qian, W. Zhu, X. Tian, and X. Wang, "Canceling inaudible voice commands against voice control systems," ser. MobiCom '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3300061.3345429

[11] T. Sugawara, B. Cyr, S. Rampazzi, D. Genkin, and K. Fu, "Light commands: Laser-Based audio injection attacks on Voice-Controllable systems," in *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, Aug. 2020, pp. 2631–2648. [Online]. Available: https://www.usenix.org/conference/usenixsecurity20/presentation/sugawara

[12] J. Gago, J. Balcells, D. Gonzalez, M. Lamich, J. Mon, and A. Santolaria, "EMI susceptibility model of signal conditioning circuits based on operational amplifiers," *IEEE Transactions on Electromagnetic Compatibility*, vol. 49, no. 4, pp. 849–859, 2007.

[13] T. D. Mast, L. Hinkelman, M. Orr, V. Sparrow, and R. Waag, "Simulation of ultrasonic pulse propagation through the abdominal wall," *The Journal of the Acoustical Society of America*, vol. 102, pp. 1177–90, 09 1997.

[14] D. F. Kune, J. Backes, S. S. Clark, D. Kramer, M. Reynolds, K. Fu, Y. Kim, and W. Xu, "Ghost talk: Mitigating EMI signal injection attacks against analog sensors," in *2013 IEEE Symposium on Security and Privacy*, 2013, pp. 145–159.

[15] T. Fokkens, S. Xia, A. Harmon, and C. Hwang, "Coupling path analysis for smart speaker intentional electromagnetic interference attacks," in *2023 IEEE Symposium on Electromagnetic Compatibility & Signal/Power Integrity (EMC+SIPI)*, 2023, pp. 488–494.

[16] Zhang, Youqian and Rasmussen, Kasper, "Detection of electromagnetic interference attacks on sensor systems," in *2020 IEEE Symposium on Security and Privacy (SP)*, 2020, pp. 203–216.

[17] Y. Zhang and K. Rasmussen, "Detection of electromagnetic signal injection attacks on actuator systems," in *Proceedings of the 25th International Symposium on Research in Attacks, Intrusions and Defenses*, ser. RAID '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 171–184. [Online]. Available: https://doi.org/10.1145/3545948.3545949

[18] T. Fokkens, Z. Xu, O. Hoseini Izadi, and C. Hwang, "Machine learning voice synthesis for intention electromagnetic interference injection in smart speaker devices," in *2021 IEEE International Joint EMC/SI/PI and EMC Europe Symposium*, 2021, pp. 673–677.

[19] D. Mukhopadhyay, M. Shirvanian, and N. Saxena, "All your voices are belong to us: Stealing voices to fool humans and machines," in *Computer Security – ESORICS 2015*, G. Pernul, P. Y A Ryan, and E. Weippl, Eds. Cham: Springer International Publishing, 2015, pp. 599–621.

[20] N. Zhang, X. Mi, X. Feng, X. Wang, Y. Tian, and F. Qian, "Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems," in *2019 IEEE Symposium on Security and Privacy (SP)*, 2019, pp. 1381–1396.

[21] D. Kumar, R. Paccagnella, P. Murley, E. Hennenfent, J. Mason, A. Bates, and M. Bailey, "Skill squatting attacks on amazon alexa," ser. SEC'18. USA: USENIX Association, 2018, p. 33–47.

[22] H. Feng, K. Fawaz, and K. G. Shin, "Continuous authentication for voice assistants," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 343–355. [Online]. Available: https://doi.org/10.1145/3117811.3117823

[23] L. Zhang, S. Tan, J. Yang, and Y. Chen, "Voicelive: A phoneme localization based liveness detection for voice authentication on smartphones," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1080–1091. [Online]. Available: https://doi.org/10.1145/2976749.2978296

[24] L. Zhang, S. Tan, and J. Yang, "Hearing your voice is not enough: An articulatory gesture based liveness detection for voice authentication,"

in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 57–71. [Online]. Available: https://doi.org/10.1145/3133956.3133962

[25] P. Crovetti and F. Musolino, "Digital suppression of EMI-induced errors in a baseband acquisition front-end including off-the-shelf, EMI -sensitive operational amplifiers," *Electronics*, vol. 10, no. 17, 2021.

[26] D. R. Calvert, "Clinical measurement of speech and voice," *The Laryngoscope*, vol. 98, no. 8, pp. 905–906, 1988.

[27] X. Chen and N. A. Touba, *Fundamentals of CMOS design*, L.-T. Wang, Y.-W. Chang, and K.-T. T. Cheng, Eds. Boston: Morgan Kaufmann, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123743640500096

[28] J. Selvaraj, G. Y. Dayanıklı, N. P. Gaunkar, D. Ware, R. M. Gerdes, and M. Mina, "Electromagnetic induction attacks against embedded systems," in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, ser. ASIACCS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 499–510. [Online]. Available: https://doi.org/10.1145/3196494.3196556

[29] C. Wei, "Electromagnetic compatibility of broadcast receivers," in *1997 Proceedings of International Symposium on Electromagnetic Compatibility*, 1997, pp. 57–59.

[30] A. R. Relay and A. R. R. League, *The ARRL Handbook for Radio Communications: The Comprehensive RF Engineering Reference*, ser. ARRL Handbook for Radio Communications Series. American Radio Relay League, 2010.

[31] F. Fiori and P. Crovetti, "Nonlinear effects of radio-frequency interference in operational amplifiers," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 3, pp. 367–372, 2002.

[32] ——, "Prediction of EMI effects in operational amplifiers by a two-input volterra series model," *IEE Proceedings - Circuits, Devices and Systems*, vol. 150, pp. 185–193(8), June 2003.

[33] Fiori, Franco and Crovetti, Paolo, "Demodulation of radio frequency interference in CMOS operational amplifiers," *IEICE Trans. Electron., C*, vol. 86, no. 11, pp. 2309–2319, 11 2003.

[34] F. Fiori and P. Crovetti, "Complementary differential pair with high immunity to RFI," *Electronics Letters*, vol. 38, pp. 1663–1664(1), December 2002.

[35] Widrow, B. and Glover, J.R. and McCool, J.M. and Kaunitz, J. and Williams, C.S. and Hearn, R.H. and Zeidler, J.R. and Eugene Dong, Jr. and Goodlin, R.C., "Adaptive noise cancelling: Principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692–1716, 1975.

[36] "OPAx277 high-precision operational amplifiers," Accessed: (February 19, 2024). [Online]. Available: https://www.ti.com/lit/ds/symlink/opa2277.pdf?ts=1707750128299&ref_url=https%253A%252F%252Fwww.google.com%252F

[37] "TLC272, TLC272A, TLC272B, TLC272Y, TLC277 LinCMOS PRECISION DUAL OPERATIONAL AMPLIFIERS," 1987, Accessed: (February 19, 2024). [Online]. Available: https://www.ti.com/lit/ds/symlink/tlc272.pdf

[38] "TMDSPREX28335 evaluation board," www.ti.com, Accessed: (February 19, 2024). [Online]. Available: https://www.ti.com/tool/TMDSPREX28335

[39] "Signal sources," Accessed: (February 19, 2024). [Online]. Available: https://www.sglabs.it/public/HP_8647A-8648series_Datasheet.pdf

[40] "Speech command recognition using deep learning - MATLAB & simulink," www.mathworks.com, Accessed: (February 19, 2024). [Online]. Available: https://www.mathworks.com/help/audio/ug/speech-command-recognition-using-deep-learning.html

[41] Pete Warden, "Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition," *Computing Research Repository (CoRR)*, vol. abs/1804.03209, 2018. [Online]. Available: {http://arxiv.org/abs/1804.03209}

Ahmed Abdullah was born in Abbottabad, Pakistan in 1993. He received his Bachelor's degree in electrical engineering from National university of Computer and Emerging Sciences, Pakistan in 2015, Master's and Ph.D. degree in electronics engineering from the Politecnico di Torino, Torino, Italy, in 2020 and 2024, respectively. He is currently a researcher at the Department of Electronics and Telecommunications (DET), Politecnico di Torino, Torino. His research interests include digital-controlled power conversion, and analog and mixed-signal integrated circuits.

Francesco Musolino (S'01-M'03) was born in Torino, Italy in 1972. He received the Laurea and Ph.D. degrees in electronic engineering from the Politecnico di Torino, Torino, Italy, in 1999 and 2003, respectively. He is currently a Researcher with the Department of Electronics and Telecommunications (DET), Politecnico di Torino, Torino where he teaches courses on power electronics and electronics for electric drives. His research interests include electronics for power conversion and motor drive applications, mixed-signal circuits, electromagnetic compatibility at the system levels and the analysis, modeling,and experimental characterization of electromagnetic compatibility problems at the printed circuit board and package level.

Paolo S. Crovetti (S'00-M'04-SM'20) was born in Turin, Italy, in 1976. He received the Laurea (summa cum laude) and Ph.D. degrees in electronic engineering from the Politecnico di Turin, Turin, Italy, in 2000 and 2003, respectively. He is currently an Associate Professor with the Department of Electronics and Telecommunications (DET), Politecnico di Torino, Turin where he leads a research group in Analog, Mixed-Signal and Power microelectronics with several international collaborations and teaches courses on basic and automotive electronics. He has co-authored more than 100 papers appearing in journals and international conference proceedings. In 2009 and in 2019 he was co-recipient of the excellent paper award of the EMC'09 Kyoto Symposium and of the Best Student Paper Award of the International Symposium of Circuits and Systems ICECS 2019. His main research interests are in the fields integrated circuit design and electromagnetic compatibility. His recent research activities are focused on non-conventional digital-based information processing techniques and ultra-low-voltage, ultra-low-power IC design for the Internet of Things. Prof. Crovetti is the co-Chair of the EMC Society Italy Chapter and serves as the Editor-in-Chief of IET Electronics Letters and as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS - Part I Regular Papers and of the IEEE TRANSACTIONS ON VERY LARGE SCALE OF INTEGRATION .