Mobile Communication Among COTS IoT Devices via a Resonant Gyroscope with Ultrasound

Feng Lin, *Senior Member, IEEE,* Ming Gao, *Student Member, IEEE,* Lingfeng Zhang, Yimin Li, Weiye Xu, Jinsong Han, *Senior Member, IEEE,* Xian Xu, Wenyao Xu, *Senior Member, IEEE,* and Kui Ren, *Fellow, IEEE*

***Abstract*—Incompatible protocols and electromagnetic interference obstruct the realization of an everything-connected Internet of Things (IoT) communication network. Our system, *Deaf-Aid*, utilizes a stealthy speaker-to-gyroscope channel to build robust communication. Compared with existing solutions adopting physical covert channels, *Deaf-Aid* is free from the limitations of manual receiver distinction, additional hardware, conditional placement, or physical contact. It exploits ultrasounds to force gyroscopes embedded in receivers to resonate, so as to convey information. We investigate the relationship among axes in a gyroscope to deal with frequency offset and support multi-channel communication. Meanwhile, receivers are identified automatically via device fingerprints consisting of diversity of gyroscopes' resonant frequency ranges. Furthermore, we enable *Deaf-Aid* the capability of mobile communication, which is an essential demand for IoT devices. We address the challenge of recovering accurate signals from motion interference. Extensive evaluations, including that on the commercial off-the-shelf devices, demonstrate that *Deaf-Aid* yields 47 bps with BER below 1%. To our best knowledge, *Deaf-Aid* is the first work to enable stealthy mobile IoT communication based on inertial sensors.**

***Index Terms*—MEMS gyroscopes, covert channel, mobile IoT communication.**

I. Introduction

INTERNET of Things (IoT) has attracted increasing attention in recent years [1]. It connects various electronic appliances for people's convenience. It is predicted that the expenditure on the deployment of IoT will continue to grow and increase, rising to $726.5 billion worldwide annually [2]. Artificial intelligence and 5G communication technology also help to combine various devices, aiming at building a comprehensive IoT network.

This paper is partially supported by the National Key RD Program of China (2021QY0703), National Natural Science Foundation of China under grant U21A20462, 61872285, 62032021, 61772236, 62172359, and 61972348, Research Institute of Cyberspace Governance in Zhejiang University, and Zhejiang Key R&D Plan (Grant No. 2019C03133).

Feng Lin, Ming Gao, Lingfeng Zhang, Yimin Li, Weiye Xu, and Jinsong Han are with Zhejiang University, China, and with ZJU-Hangzhou Global Scientific and Technological Innovation Center, China. Email: flin@zju.edu.cn, gaomingppm@zju.edu.cn, lingfengzhang@zju.edu.cn, ninalym13@gmail.com, xuweiye@zju.edu.cn, hanjinsong@zju.edu.cn.

Wenyao Xu is with University at Buffalo, the State University of New York, US. Email: wenyaoxu@buffalo.edu.

Xian Xu is with College of Civil Engineering and Architecture, Zhejiang University, China. Email: xianxu@zju.edu.cn.

Kui Ren is with Zhejiang University, China, and with ZJU-Hangzhou Global Scientific and Technological Innovation Center, China, and with Key Laboratory of Blockchain and Cyberspace Governance of Zhejiang Province, China, and with Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies, China. Email: kuiren@zju.edu.cn.

Feng Lin and Ming Gao contribute equally to the article.

Jinsong Han is the corresponding author.

A shorter conference version of this work were presented at the ACM MobiCom in 2020, "Deaf-Aid: mobile IoT communication exploiting stealthy speaker-to-gyroscope channel" [1].

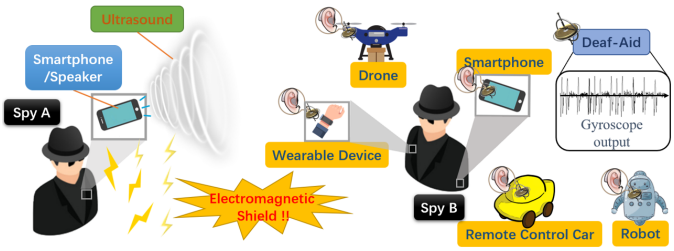


Fig. 1. An application scenario for *Deaf-Aid* in the case of espionage. Spy A sends private information to Spy B, who can receive it stealthily via various IoT devices even under the electromagnetic shield.

However, creating such an everything-connected IoT network involves abundant obstacles. Incompatible communication standards have aggravated the problem of information exchange via IoT devices. Requirements in different scenarios promote various protocols, while devices usually support one or a few of them, making the cross-protocol IoT communication hard. Wi-Fi and Bluetooth are widely used in mobile communication. ZigBee and MQTT are suitable for small-streamed data transmission, especially with resource constraints. Furthermore, there are EnOcean, 6LowPan in the field of smart home and AMQP, COAP in industrial IoT. To make matters worse, manufacturers develop their own protocols and build distinctive systems. These methods rely only on the electromagnetic wave and would fail upon the electromagnetic interference and shielding, as shown in Fig. 1.

To address aforementioned problems, researchers take advantage of physical characteristics to build a covert channel between nodes that are physically and logically separated [3], [4], [5], so that devices can communicate in spite of the protocols. Nevertheless, these systems are confronted with several hindrances, such as the need of additional hardware, confined placement, or physical contact. For instance, *Ripple* [6], [7] demands specialized vibration motors and physical contact; *BitWhisper* [8] can only be applied between two desktop PCs in a fixed position. Moreover, they are dependent on manual receiver identification, which is impractical in a comprehensive and mobile IoT network. More feasible and robust communication among IoT devices is needed urgently [9].

We turn attention to the channel of speaker-to-gyroscope. It has been widely reported that micro-electro-mechanical systems (MEMS) inertial sensors are vulnerable to the ultrasonic injection [9], [10], [11], [12]. Ultrasound with the right frequencies can couple to MEMS gyroscopes and make them produce low-frequency angular rate readings [13], [14]. However, little attention was drawn to the potential benefits of its sensibility. Inspired by this, we explore the gyroscope resonance from a communicative perspective. Despite the lim-

itation of protocols, a robust system is proposed for bridging a stable transmission in an IoT network, transmitting via speakers, and decoding them through gyroscopes. The channel frequency is selected according to the receivers as each gyroscope has its own unique resonant frequency. Such a non-contact speaker-to-gyroscope channel in IoT communication is feasible. Ultrasonic signals can be easily obtained through commodity high-sampling speakers and most modern phones without any peripherals. Moreover, gyroscopes have become an indispensable part of intelligent devices, e.g., smartphones, VR sets, vehicles, wearable devices, and drones.

Communication among mobile IoT devices should be robust against motion interference. Movement introduces noise, masking characteristic signals, especially on inertial-based systems [15]. Moreover, unpredictable frequency offset confuses the frequency features, preventing signal recovery using spectrum analysis. It is certainly a key issue for gyroscope-based communication to work stably in a dynamic environment.

For robust gyroscope-based communication, we need to specifically address several practical challenges: (1) **Capability**: How to leverage gyroscopes to build a channel of high quality with precise receiver identification. (2) **Mobility**: How to accurately recover signals in a mobile communication scenario. (3) **Drift**: How to deal with the frequency offset caused by drift to ensure communication stability.

We present a convenient and robust system that exchanges data over the air, namely *Deaf-Aid*. It is free from restrictions including peripherals, position requirements, and artificial receiver identification. It provides an alternative and complementary communication channel for current IoT devices. We model the instantaneous and steady-state responses of a resonant gyroscope, analyze the frequency offset caused by sampling rate drift, and exploit the inter-axial relation for noise cancellation. The compositions of *Deaf-Aid* include receiver identification, encoding, channel selection, and threshold. It supports simultaneous communication on double channels, even with two different transmitters. Movement influence is also taken into account. Multiple technologies are employed to adjust our system to a mobile IoT network. We build prototypes and perform a comprehensive evaluation on several kinds of commercial off-the-shelf (COTS) IoT devices, including smartphones, smartwatches and drones. We exert various movement on them involving 22 participants to validate the effectiveness under real-world scenarios. We release our source code [16] to facilitate the gyroscope-based research.

The contribution of *Deaf-Aid* can be summarized as follows:

- We investigate the possibility of communicating through a gyroscope and realize a stealthy channel without the restriction of peripherals, physical contact, fixed placement, and especially the manual receiver identification.
- We comprehensively analyze the relationship among axes in a resonant gyroscope. Accordingly, noise is eliminated, which is introduced by intrinsic errors, unstable instantaneous response, frequency offset, and motion.
- We develop an accurate communication system for a mobile IoT network. In particular, we take the initiative in excavating the potential of inertial sensors applied for the robust communication against motion interference.

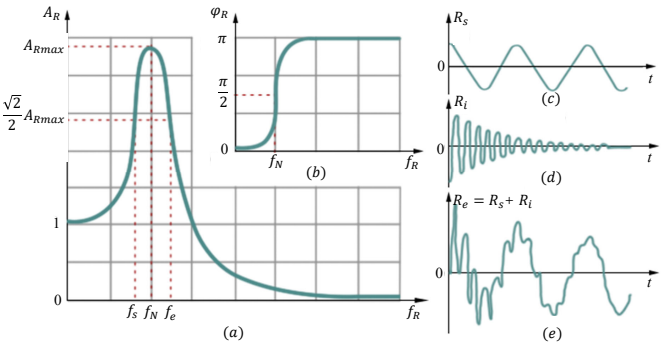


Fig. 2. Illustrations of the (a) amplitude-frequency and (b) phase-frequency characteristic of the steady-state response and the (c) steady-state, (d) instantaneous, and (e) total response under the forced vibration.

II. Preliminaries

A. A Gyroscope and Its Resonance Principle

A MEMS gyroscope is implemented with Coriolis force [1] where its Coriolis acceleration a_x is proportionate to the angular rate ω, according to $a_x = -2\omega\dot{y}$, where $\dot{y}$ is a linear velocity predetermined during manufacture. Coriolis acceleration leads to capacitance change, and then the gyroscope reading is ultimately obtained after processing – amplifying, filtering, and converting analog signals to digital ones.

The structure of MEMS gyroscopes can be described as a single-degree-of-freedom system. Damping is ignored at a low frequency, and gyroscopes retain linear outputs. As frequency increases, damping gradually dominates, and oscillation occurs. The steady-state characteristics of the forced vibration [17] are shown in Fig. 2 (a) & (b), indicating that resonance will introduce an extra phase ϕ_R that changes dramatically from 0 to π and equals $\pi/2$ at the natural frequency f_N, where the gain coefficient A_R reaches the peak. As a result, gyroscopes respond to acoustic injection whose frequency is close to or coincident with such a frequency.

The damping architecture is typically designed to share the same natural frequency with resonating mass. However, inevitable errors may bring about natural frequency alternation. This implies the diversity in f_N among gyroscopes.

B. Feasibility Investigation

We demonstrate the feasibility and stealth of the speaker-to-gyroscope channel in respect of noiselessness, availability, and inaudibility. Eight mainstream gyroscope models are selected as representatives and the speaker-to-gyroscope channel is tested on four chips of each model. The resonant frequency bands of 32 tested gyroscope chips are shown in Fig. 3.

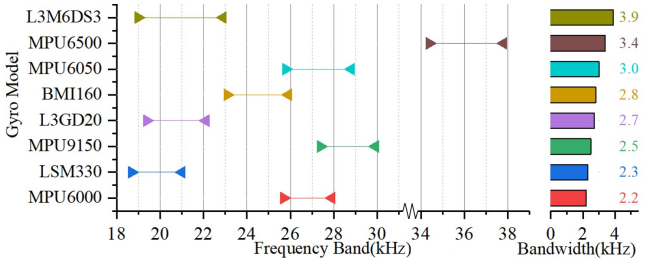


Fig. 3. Gyroscopes have resonant frequency ranging beyond human audibility.

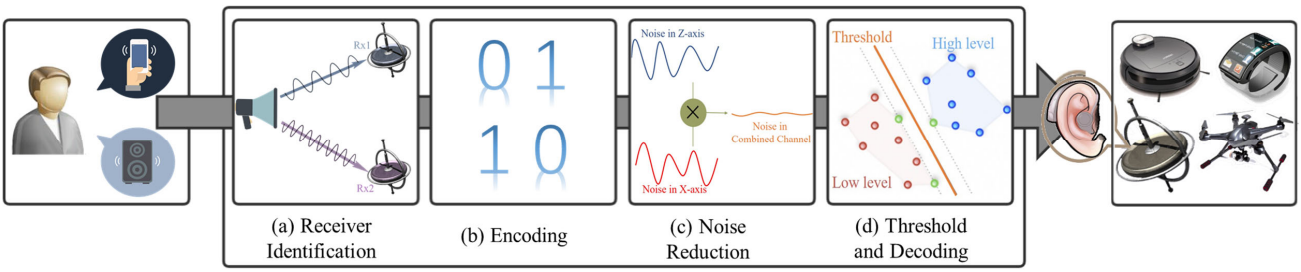


Fig. 4. *Deaf-Aid*, a speaker-to-gyroscope channel for mobile IoT communication, where the transmitter is realized by a smartphone or a commercial speaker and the receiver can be any IoT device equipped with a gyroscope.

Inaudibility. Resonant frequencies of gyroscopes typically exceed 18 kHz. They are inaudible to humans [10], [11], [13], [14] and are always ignored by speech recognition systems, whose sampling rates are always below 16 kHz (As these systems are generally focused on humans' fundamental vocal band of below 8 kHz and this measure is also beneficial to reducing the requirement of storage and computing overhead) [18], [19], [20]. Although accelerometers also resonate with acoustic injection [21], we do not adopt them as the receivers in this study due to their audible resonant frequency (usually below 10 kHz) [12]. Otherwise, accelerometer-based communication with sound will disturb people nearby.

No peripheral. According to the sampling theorem, speakers induce sound within 24 kHz with a 48 kHz sampling rate. Therefore, current speakers can support communication with gyroscopes like L3M6DS3, L3GD20, LSM330, and BMI160. As for the ultrasound in the higher band, Hi-Fi speakers and advanced mobiles perform better. For example, Samsung Galaxy S8 is manufactured with a sound card up to 32 bit/384 kHz. Smartphones, e.g., Vivo X9s, MI 10, Honor 9, OnePlus 5, and their subsequent models have the sampling rates of at least 96 kHz. All Reno smartphones support Hi-Res Audio [22], which means that they can play sound within 40 kHz without distortion. More modern smartphones (especially those new or high-end models) are able to support a high sampling rate of over 48 kHz [23]. Those devices can emit the required ultrasonic signals for transmission, which is inaudible to humans. Our pilot experiment on 15 smartphones (as listed in Tab. II) shows that modern smartphones can serve as the transmitters of *Deaf-Aid*, indicating that most commercial high-sampling speakers and new smartphones can cover the resonant frequency band of main off-the-shelf gyroscopes and adopt *Deaf-Aid* without peripherals.

Little environment interference. Common application scenarios of ultrasound, e.g. medical examination, prefer frequency bands of above 40 kHz. Most gyroscopes resonate in the frequency band between 18 kHz and 40 kHz, where few devices work. Therefore, *Deaf-Aid* is shielded from environmental noise. Meanwhile, the transmission will not affect the normal operation of surrounding devices.

C. Our Vision

We propose a novel communication system that utilizes the sensibility of a gyroscope to ultrasound. It will involve combined efforts from four modules, as shown in Fig. 4.

In a communication channel, it is fundamental to remove noise for error-free transmission. With the mathematical model of the ultrasonic resonance mechanism of gyroscope in Sec. III, we analyze the source of noise and the cause of frequency offset, and accordingly propose a novel offset-independent noise cancellation method in Sec. IV. Afterward, we design *Deaf-Aid* with four modules including receiver identification, encoding, noise reduction, and decoding, as detailed in Sec. V. Furthermore, we suppress the influence of motion in Sec. VI.

III. Resonance Model

We develop a physics-based model to quantitatively analyze the resonant outputs of a gyroscope.

Ultrasound waves impose force of the same frequency on one axis in a gyroscope. The force can be described as follows,

$$F(t) = F \cdot sin(2\pi f_R t + \phi_0), \tag{1}$$

where F is the magnitude decided by intensity and position of the sound source, f_R is the frequency of the sound source, and ϕ_0 indicates the initial phase. The force will produce the resulting steady-state oscillation R_S [17] as follows,

$$R_S(t) = A_R F \cdot sin(2\pi f_R t + \phi_0 + \phi_R), \tag{2}$$

where the gain coefficient A_R and phase ϕ_R are determined by the acoustic frequency f_R and the natural frequency f_N.

Typical MEMS architecture in a gyroscope comprises three parts: amplifier, filter, and analog-to-digital conversion (ADC).

Amplifier and Filter: The rotation will be converted into analog signals in a gyroscope. They are processed by an amplifier and a low-pass filter (LPF) for removing noise. An ideal LPF can completely remove the high-frequency noise beyond the cut-off frequency. Nevertheless, filters are less effective in handling noise whose frequency is much higher than the cut-off frequency in gyroscopes. Instead, filters may introduce amplitude alteration besides slight frequency changes and phase shifts. In general, the analog signal in gyroscopes follows this formula,

$$R(t) = A \cdot sin(2\pi f_R t + \Phi), \tag{3}$$

where A is the amplitude of the resonant data after processed by the amplifier and filter, and $\Phi = \phi_0 + \phi_R + \phi'$ is the total phase shift while ϕ' is introduced during processing. Though the filter might introduce a slight frequency alteration, it can be regarded as the constant in a given gyroscope.

ADC: The frequency of ultrasonic input is over 18 kHz, much higher than gyroscopes' sampling rate (within 1 kHz). This leads to aliasing, where the high-frequency signal fails to maintain the original spectrum characteristics [9] and would

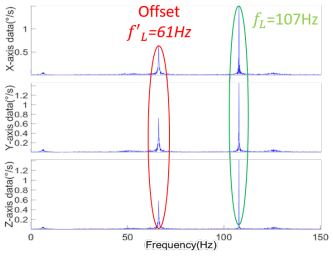


Fig. 5. A sample of the frequency offset caused by the sampling rate drift.

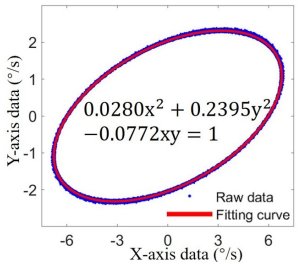


Fig. 6. A scatter plot between X and Y axis and its fitting curves.

fall into the low-frequency band. Given the sampling rate F_s, the sampled signal $R[k]$ can be expressed as follows,

$$f_R = n_f \times F_s + f_L, \quad (|f_L| < F_s/2,\ n_f \in \mathbb{N}_+)$$
$$R[k] = A \cdot sin(2\pi f_L k/F_s + \Phi). \tag{4}$$

where f_L is the frequency of digitized gyroscope outputs, and n_f is a constant. For example, an ultrasonic signal of 23820 Hz will make a BMI160 gyroscope chip with F_s =200 Hz to output the readings of 20 Hz (=23820-200×119) experimentally. The final readings are dependent on the input frequency and sampling rate, where aliasing induces low-frequency readings.

IV. Offset-independent Noise Cancellation

We analyze the influence of noise and the frequency offset. Here, we exploit the relationship among axes in a gyroscope and thus propose the offset-independent noise cancellation.

A. Instantaneous Response

The actual oscillation consists of steady-state and instantaneous responses, as illustrated in Fig. 2(e). We adopt the steady-state responses as the signals for communication, while the instantaneous response R_I [17] would act as the noise,

$$R_I(t) = e^{-2\pi f_N \xi t} A \cdot [c \cdot cos(2\pi f_N t) + d \cdot sin(2\pi f_d t)], \tag{5}$$

where f_d is the frequency of the instantaneous response, c and d are the gain coefficients, and $e^{-2\pi f_N \xi t}$ is the attenuation factor which approaches zero over a short period of time because of the high damping ratio ξ. f_d, c, and d are determined compositely by the input frequency f_R, the natural frequency f_N, and the damping ratio ξ. Since the latter two items are definite in a given gyroscope, these parameters depend merely on f_R. The instantaneous response $R_I(t)$ cannot be ignored especially in the initial period of the resonance.

The inherent noise of a gyroscope also introduces additional channel noise. This inherent noise is regarded as independent Gaussian white noise [24]. Therefore, the gyroscope's total response $R_T(t)$ under acoustic injection comprises the targeted steady-state response R_S, the undesirable instantaneous response R_I, and the inherent noise $n(t)$ as follows,

$$R_T(t) = R_S(t) + R_I(t) + n(t). \tag{6}$$

Traditional denoising methods use various filters based on the spectrum analysis. However, the steady-state response's frequency is not always stable due to the sampling rate drift, detailed in Sec. IV-B. Therefore, these frequency-based methods might not work. As a countermeasure, we analyze the law of the frequency offset caused by the sampling rate drift and explore offset-independent characteristics to remove instantaneous responses and inherent noise (see Sec. IV-D).

B. Sampling Rate Drift

A severe weakness of sampling rate drift is that it leads to obvious but unpredictable deviations of output frequency [9], making the outputs unstable. This is an issue that remains to be resolved, especially in the mobile communication system. We assume ΔF_s to be the sampling rate drift and substitute it into Eq. 4. The output frequency alters as

$$f_R = n_f \times (F_s + \Delta F_s) + f'_L, \quad (|f'_L| < F_s/2)$$
$$R[k] = A \cdot sin(2\pi f'_L \frac{k}{F_s + \Delta F_s} + \Phi). \tag{7}$$

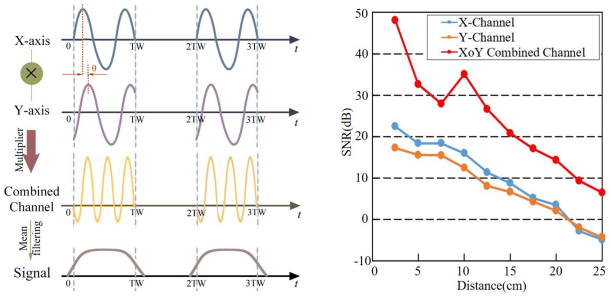
where f'_L is the frequency of gyroscope readings affected by the drift. Since f_R is usually hundreds of times more than F_s, slight fluctuations in sampling rates may initiate a significant frequency offset. We inject ultrasonic signals of 21107 Hz into an L3M6DS3 gyroscope with a sampling rate of 300 Hz for 1 hour, with the spectrum shown in Fig. 5. Following Eq. 4, this gyroscope produces the signals of 107 Hz. However, due to the sampling rate drift (merely about 0.6 Hz occurring at the 46th minute in this measurement), the output frequency is only 61 Hz. Such a frequency offset makes it difficult to distinguish signals from noise via spectrum analysis.

We have conducted an experiment to corroborate the randomness and universality of sampling rate drifts. Eight models of gyroscopes as listed in Fig. 3 run continuously for 24 hours with an initialized sampling rates of 200 Hz. Their sampling rates drift from 199.6 Hz to 200.9 Hz. The drift varies completely randomly over time and the value of the drift is also random. The inherent hardware defect causes the drift and hence there is no law to predict it. The receiver cannot learn and correct its own sampling rate drift. In addition, the drift has little influence on the measurement of inertia (though it badly affects the communication via ultrasound), and thus the manufacturers pay no attention nor take no action on the drift. Fortunately, in multi-axis inertial sensors, the sampling operation of each axis occurs synchronously as each axis in a sensor shares the identical internal clock. Therefore, the drift of sampling rates (i.e. the drift of the internal clock) occurs on all axes simultaneously and thus the variation values of the sampling intervals ΔF_ss on all axes are equal.

C. Inter-axial Characteristics

Previous studies focused on the resonance of only one axis and neglected relations among axes. We thoroughly investigate these inherent inter-axial characteristics and exploit the offset-independent characteristics to correct the frequency offset.

Frequency synchronization. The steady-state responses occur simultaneously on all axes in a gyroscope. They originate from the same ultrasonic input, undergo the identical process of digitization, and thus share the same response frequency. The sampling rates shift simultaneously if any, and accordingly, the frequency offsets are equal. On the contrary,



(a) (b)

Fig. 7. An illustration of (a) the coherence-based offset-independent channel and (b) its performance on noise reduction.

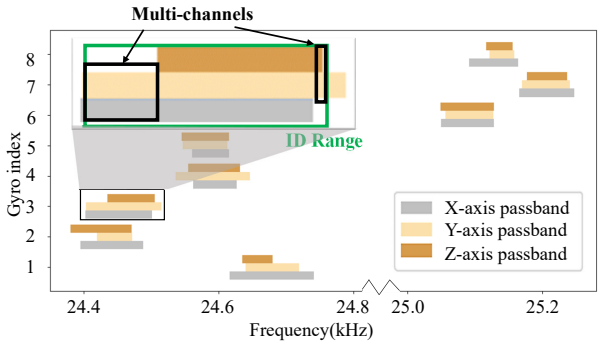


Fig. 8. The bandwidth for 8 identified BMI160 chips.

the instantaneous responses (according to Eq. 5), as well as the inherent noise [17], among different axes are independent and identically distributed due to the production diversity.

Fixed phase difference. Because of the synchronous resonance and the identical digitization process, the phase difference is only introduced at sensing and resonance stages. We experimentally discover that each axis oscillates at the same frequency with a fixed phase difference. Axes differ in the natural frequency f_N owing to production. When subjected to the same frequency of vibration, the ratios f_R/f_N in multiple axes are unequal, bringing about a disparity in amplitude coefficient A_R and phase ϕ_R. The scatter plot in Fig. 6 exemplifies the relationship between each variable pair from a resonant BMI160 gyroscope. The curve fits an ellipse, which reflects these variables are coherent, sharing the same frequency and following a fixed phase difference.

Particularly, the amplitude coefficient and phase difference are decided only by f_R and will not be affected by motion.

D. Coherence-based Correction

Considering the existence of synchronous frequency and fixed phase difference, we employ the coherent demodulation. It leverages a multiplier to convert two coherent signals into constants and filter out the high-frequency non-coherent components. Noting that instantaneous responses and the inherent noise among axes are independent and identically distributed (i.e. non-coherent), they can be filtered by coherent demodulation. For example, according to Eq. 6, the combined channel of the X- and Y-axes is represented as follows,

$$\begin{aligned} U_{\times}^{noise}[k] &= R_{Tx}[k] \times R_{Ty}[k] \\ &= (R_{Sx}[k] + R_{Ix}[k] + n_x[k]) \times (R_{Sy}[k] + R_{Iy}[k] + n_y[k]), \end{aligned} \tag{8}$$

where $R_{Tj}[k]$, $R_{Sj}[k]$, and $R_{Ij}[k]$ are the total, steady-state, and instantaneous responses in turn, and $n_j[k]$ is inherent noise on the j-axis ($j = x, y$). We have

$$\begin{aligned} U_{\times}[k] &= R_x[k] \times R_y[k] \\ &= \frac{A_x A_y}{2}[cos(\Phi_x - \Phi_y) - cos(4\pi f_R \frac{k}{F_s} + \Phi_x + \Phi_y)], \end{aligned} \tag{9}$$

where $R_x[k]$ and $R_y[k]$ are the readings on the two axes and A_x, A_y, Φ_x and Φ_y are their amplitudes and phases respectively. Through filtering, the high-frequency non-coherent noise (including the instantaneous response and inherent noise) and the harmonic component $cos(4\pi f_R \frac{k}{F_s} + \Phi_x + \Phi_y)$ are cancelled. We have a combined channel as follows,

$$U_{bias}[k] = \frac{1}{2} A_x A_y cos(\Phi_x - \Phi_y). \tag{10}$$

Such a constant result is independent of the signal frequency, and thus is free from offset caused by the sampling rate drift.

Briefly speaking, we remove noise based on the coherent inter-axial characteristics for signal extraction. Two axes are combined as a channel by a multiplier, with an average filter for high-frequency components and noise removal, as elaborated in Fig. 7. The combined channel has a higher signal-to-noise ratio and extends communication distance excellently.

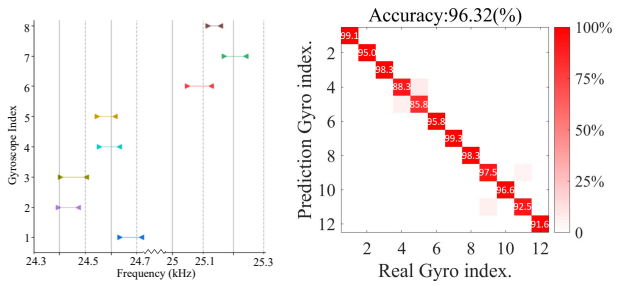
V. System Design

We design *Deaf-Aid* that utilizes the resonate characteristics of a gyroscope for communication among IoT devices. It is composed of four modules, as illustrated in Fig. 4.

A. Receiver Identification

Receiver identification is fundamental for mobile communication, while traditional methods are not suitable in a mobile IoT network. We propose a novel device fingerprint using the resonant frequency diversity for receiver identification.

Motivation. The motivation lies in the drawbacks associated with the use of traditional methods in a mobile IoT network. Recognizing devices manually is widely used in existing covert channels. Traditionally, existing covert-channel communication approaches [6], [7], [8], [25] require users to manually place the Rx-Tx devices in a specific layout such that no other devices are in the coverage of communication, because they cannot distinguish different receivers. Clearly, such treatment is impractical, especially for mobile IoT networks, where the networks usually contain extensive IoT devices. In the mobile scenario, it is possible that an IoT device accidentally enters the communication range of a pair of communicating Rx-Tx devices. In this case, the message is likely to be sent to the wrong receiver. Meanwhile, the routing protocol and address resolving demand an excessive configuration, especially in mobile scenarios. However, IoT devices usually have weak computation ability [26] and cannot support the complex address resolving in most cases. Hence, it is challenging to balance between the overheads and automation in designing the identification mechanism for Deaf-Aid. To solve this problem, we design a novel identifier based on device fingerprint.



(a) (b)

Fig. 9. (a) The ID ranges for 8 identical BMI160 chips and (b) the confusion matrix of identification among 12 gyroscopes.

Solution. Figure 3 reveals that different kinds of gyroscopes have various resonant frequency ranges. However, these ranges may coincide and it is difficult to further distinguish different gyroscopes of the same model. We perceive the diversity of gyroscopes of the same model in the resonant passband. *Deaf-Aid* leverages this diversity as a device fingerprint to identify receivers in a dynamic position. We measure the accurate resonant passband ranges where a speaker sends chirps from 18 kHz to 38 kHz at a step of 100 Hz to obtain the rough resonate frequencies of target gyroscopes in front of the speaker 5 cm away, and then the speaker sends chirps composed of the rough resonate frequencies at a step of 1 Hz. Conventionally, the frequencies corresponding to $\sqrt{2}/2$ of the peak of gain coefficients are deemed as starting and ending points (f_s and f_e in Fig. 2(a)), as plotted in Fig. 8. Experimental results indicate that each gyroscope of the same model varies in the passband. Additionally, we observe that each axis in one gyroscope may differ slightly. We make a comparison of each axis among gyroscopes, as shown in Fig. 9. The difference of the natural frequency f_N, spawned by the production diversity, accounts for it. We select the frequency range where at least two axes have a response as the ID range (the green frame in Fig. 8). It supports faster authentication than the traversal comparisons in each axis and avoids the confusion where some gyroscopes share similar ranges on one axis. Fig. 9(a) confirms the validity of this fingerprint. In practice, the resonant frequency of a gyroscope is measured in advance and all information is known by users. The whole measurement process is very fast (within several minutes) and multiple devices can be measured simultaneously.

Use case. Before the transmission of messages, the transmitter sends an identifier to recognize the receiver. The identifier is composed of ultrasonic chirps modulated by the target gyroscope's ID range. The time duration of transmitting the identifier is about several seconds (1.5 seconds in our defaulted setting). It pushes the gyroscope in the receiver to oscillate with the homologous chirps, even if there is an offset or movement disturbance. The device that receives a full identifier will be regarded as the communication target and it will receive the following messages. For example, there are eight IoT devices (i.e., the eight gyroscopes in Fig. 9(a)) in the coverage of a speaker (as the transmitter). The user wants to send messages to the 6th gyroscope, whose ID range is (25050,25128). The transmitter sends an identifier, i.e., the ultrasound sweeping from 25050 Hz to 25128 Hz every 0.5 seconds for three times. Only the 6th gyroscope keeps resonating for 1.5 seconds, and then it will decode the following information. Even though there might be other gyroscopes whose resonant bands are overlapped with those of the 6th gyroscope, these gyroscopes will not keep resonating for 1.5 seconds. Thus, the ID range identifies the target receiver, i.e., the 6th gyroscope.

Effectiveness. To verify the stability of the ID range, we prepare an experiment a month after the ranges were first measured. We test 6 speakers and 12 chips of two models, including eight BMI160 chips (1-8th) and four L3GD20 chips (9-12th) referring to the confusion matrix shown in Fig. 9(b). It achieves an accuracy of 96.32%. There is a slight drop in the accuracy of the 4th and 5th chips. We note that these errors are concentrated during the test via a JBL GO2 speaker. Its poor performance (i.e., the bad frequency resolution around 24.6 kHz) is to blame for the mistakes in identification. Even in the worst circumstances with speakers of poor frequency resolution, it still has the capability to distinct multiple devices.

Advantage. In particular, the advantage of the ultrasonic ID range is two-fold. It realizes an automatic recognition among various IoT devices, without the requirement of the manual assistant. On the other hand, the ID range is numeric-address-free and suitable for IoT devices that have limited computation ability. In general, using numeric addresses in the transmission preamble may require the receivers to remember their own addresses and support the address resolution. If the computation ability is extremely limited, the IoT devices can hardly resolve numeric addresses in a timely way or even have no space to store the address. In comparison, the ultrasonic ID range is like a hardware fingerprint and does not need numeric addressing. It only requires that the gyroscopes can detect the resonance. Thus, the ID range is effective to identify receivers in a mobile IoT network in most scenarios.

Discussion on limitations/alternatives. The main limitation of ID ranges is that it requires little computation ability for receiver identification with a penalty of relatively large latency, i.e., 1.5 seconds for transmitting and receiving an identifier. A possible alternative is to combine the ID range with a numeric address in the transmission preamble. Such a solution may reach a trade-off between latency and computing complexity.

B. Encoding

We modulate the ultrasound to transmit messages. We define the high level when the ultrasound is on and lasts for a pulse width (PW), and the low level when the ultrasound is off. With the modulation scheme, we adjust pulse interval encoding (PIE) [27] to serve as the channel coding, as illustrated in Fig. 10. Specifically, we encode the data by defining different time gap widths between the rising edges of the pulses where a short interval (i.e., PW) indicates '1' and a long one (of $3\times PW$) implies '0'. The start of frame (SOF) is defined as two successive intervals of $5\times PW$ and PW respectively, and the end of frame (EOF) is defined as a longer interval of $7\times PW$. In particular, we define that only one rising edge can be detected in a bit. This scheme can use fewer samples to represent one bit while the traditional method adopts the amplitude envelope which needs massive samples per bit at

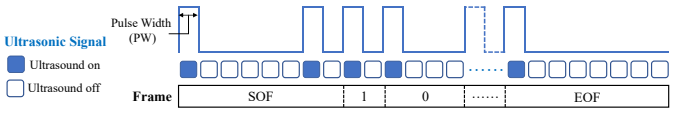


Fig. 10. Encoding scheme of our adjusted PIE.

the sacrifice of speed. This scheme also addresses the issue of the signal energy fluctuation which is not predetermined and varies along with the location and energy of sound sources.

C. Noise Reduction and Channel Selection

Benefiting from the inter-axial coherent characteristics, we remove noise in an offset-independent manner and obtain an error-free channel in Sec. IV. The higher resonance intensities on the selected axes suggest the higher signal intensity in the combined channel according to Eq. 10, resulting in a higher SNR. Rotational energy is a direct metric but is susceptible to motion interference. We observe that motion contributes much less than the resonance to the variance of gyroscope readings. Therefore, we propose the variance-based selection. Two axes with higher variances are chosen to obtain the combined channel. Meanwhile, if the gyroscope is moving in a plane, where at least one axis must have a zero mean, we give preference to these axes. Such a selection also benefit the suppression of the motion influence (see Sec. VI).

The appropriate modulation enables the multi-channel communication. Although resonance on each axis is coherent, it varies in the resonant frequency range. At some frequencies, only some of the axes oscillate. We choose the band where only two axes resonate as multi-channels (the black frames in Fig. 8). The mutual interference on the common axis can be reduced in the same way in Sec. IV-D. Thus, these channels can deliver different messages over different frequency ranges. Taking the 3th gyroscope chip in Fig. 8 as an example, the ultrasound of 24.41 kHz can resonate with the chip's X- and Y-axes, while the ultrasound of 25.5 kHz resonates with the Y- and Z-axes. Therefore, we can modulate different messages over the two ultrasound bands (the two ultrasonic signals can come from one speaker or two different speakers that play the signals of different frequencies respectively). Though the ultrasound of 25.5 kHz interferes in the Y-axis when we decode the messages on the XoY combined channel, the Coherence-based correction method in Sec. IV-D can focus on the coherent signals between the X- and Y-axis (i.e., the components of 24.41 kHz), and similarly, the YoZ combined channel is free from the disturbance from the ultrasound of 24.41 kHz. Therefore, the messages on the two channels can be transmitted and received without mutual interference. In practice, a user can send messages on the two channels to improve the channel capacity. Or two users can send messages on respective channels using different frequencies. **Such multi-channel communication provides double capacity or allows a receiver to listen to two users simultaneously.**

D. Threshold and Decoding

It is insufficient to rely solely on empirical thresholds. The signal amplitude relies on several aspects, including distance, sound source, and resonance intensity. Inspired by the image

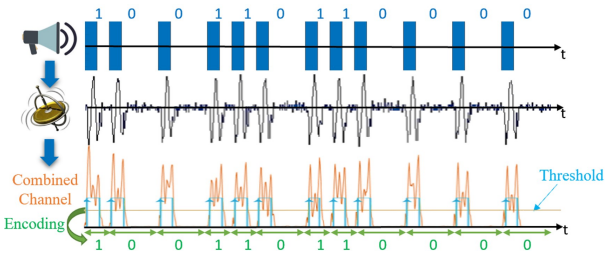


Fig. 11. An example of signal transmission.

threshold, we adopt the maximum entropy threshold method [28]. The basic idea is to find the maximum entropy and take the corresponding threshold as the final one. Concretely, for a channel with resolution r and maximum value $K \cdot r$, we decide threshold $q = k \times r, (k = 1, 2, \ldots, K)$ when the entropy reaches a maximum as follows,

$$H(q) = -\sum_{i=1}^{k} \frac{p(i \times r)}{\sum_{j=1}^{k} p(j \times r)} log \frac{p(i \times r)}{\sum_{j=1}^{k} p(j \times r)} - \sum_{i=k+1}^{K-1} \frac{p(i \times r)}{\sum_{j=k+1}^{K-1} p(j \times r)} log \frac{p(i \times r)}{\sum_{j=k+1}^{K-1} p(j \times r)},$$

where $p(\cdot)$ is the probability density. Then, we decode the signals where the points with a value larger than this threshold are regarded as the high level, or as the low level otherwise.

We establish a gyroscope-based communication channel, with an illustration of transmission in Fig. 11. This channel meets the demand for faster transmission speed with less noise, whereas it still suffers from the motion interference.

VI. Motion Influence Suppression

Motion exerts a huge impact, especially on the gyroscope-based system. It is possible to be either affected by the mixture of motion and resonance during communication or disturbed by obstacles. Fig. 12 illustrates that motion influences can be resolved into three simple forms: transmitter (Tx) motion, the line-of-sight (LOS) blocking, and receiver (Rx) motion, which can combine to form complex motion in practice. We analyze the motion effect and propose solutions accordingly for robust communication in a mobile IoT network.

A. Transmitter Motion

Transmitter motion contributes to an unpredictable variation in transmission distance. It changes the force on gyroscopes and results in signal fluctuation, including amplitude shifting and phase offset. The gyroscope output is rewritten as follows,

$$R_{TM}[k] = A[k] \cdot sin(2\pi f_L k / F_s + \Phi[k]), \qquad (11)$$

where $\Phi[k] = \phi_0[k] + \phi_R + \phi'$. Because of the signal jitters, a fixed threshold promotes the probability of error, which results in the instability of communication.

We assume the effect of distance change maintains stable in a pulse which is roughly measured in millisecond. The communication distance will not sharply change in such a short time, and thus the intensity and phase changes are negligible within a pulse. Inspired by the threshold window in image recognition [29], we calculate threshold in a short time (such as several bits) to achieve adaptive threshold segmentation,

(a) Tx moving (b) LOS blocking (c) Rx moving.

Fig. 12. Three basic kinds of motion interference.

handling the fluctuation of the pulse. Afterwards, we can normalize amplitude on the basis of these thresholds.

B. Line-of-Sight Blocking

Sound transmission is affected by the medium especially on LOS. Though users can avoid obstacles that always block LOS, moving obstacles are likely to appear disorderly in real scenarios. This results in a sudden error denoted as $SE[k]$, and the gyroscope output is rewritten as follows,

$$R_{LOS}[k] = A \cdot sin(2\pi f_L k/F_s + \Phi) + SE[k]. \tag{12}$$

We utilize interleaving technology to reduce these errors. Interleaving allocates the transmission bits in the domain of time, frequency, or both. It changes the information structure to the greatest extent without content alternation. Thus, the decoder can treat these errors as random ones, which indicates that it maximizes the dispersion of concentrated errors during channel transmission. One of the most common ways is block interleaver [30]. It writes the input sequence into an $m \times n$ matrix in the order of rows and then reads by columns. The reading and writing objects are swapped during reordering. The mapping function is expressed as follows,

$$I(i) = [(i-1) \bmod n] + \lfloor (i-1)/n \rfloor + 1, \tag{13}$$

where $I(i)$ is the location of the ith $(i = 1, 2 \ldots, N)$ data in the original line, $\lfloor \cdot \rfloor$ is the floor function, m and n refer to the number of rows and columns, and $N = m \times n$ represents the interleaving length. This function maximizes the dispersion of the burst errors in the process of channel transmission and effectively cuts down the errors aroused by the sudden block.

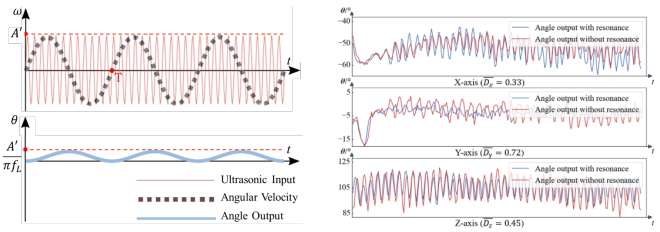
C. Receiver Motion

Receiver motion triggers distance variation and forces gyroscopes to produce additional readings. According to the solved distance variation and normalization based on the adaptive threshold in Sec. VI-A, the gyroscope output is rewritten as

$$R_{RM_j}[k] = A' \cdot sin(2\pi f_L k/F_s + \Phi) + M_j[k], \tag{14}$$

where $M_j (j = x, y, z)$ represents the additional readings of the gyroscope introduced by movement on the corresponding axis, and A' is the normalized amplitude. Particularly, A_R and ϕ_R are immune to the motion interference.

Ruled in Sec. V-B, the length of intervals of rising edges between bits must fall into a definite range. Such a rule benefits the detection of motion interference. Once the interval length exceeds this range, it is judged to be disturbed by the receiver motion during data transmission. Since the initial objective of gyroscopes is to measure movements, we first recover the motion to provide timely movement information to the device's control center. For more accurate communication against motion, we propose signal separation methods in two situations where the receiver is moving in a plane or the space.



(a) (b)

Fig. 13. (a) The aliasing of the angular velocity and accumulative errors on the angle measurement in a resonant gyroscope, and (b) the angle outputs with and without resonance when a BMI160 gyroscope follows the same trajectory.

1) Motion Recovery: Since the resonant data is sinusoidal with a peak A' and the frequency f_L, the accumulative error on the angle measurement is tiny, with the maximum error $\frac{A'}{f_L \pi}$ shown in Fig. 13. The average Euclidean distance between the angle outputs with and without resonance when a BMI160 gyroscope follows the same trajectory is 0.50°. It is just a little higher than 0.46°, the average Euclidean distance when the BMI160 gyroscope follows the same trajectory for ten times without resonance. Besides, the frequency of the sinusoidal oscillations often exceeds that of motion. It could be removed easily by wavelet transform [31] at the cost of an acceptable loss of accuracy in angular rate measure. This method enables *Deaf-Aid* to transfer data to drone-like devices that rely on gyroscope heavily without interference in the normal use of gyroscopes. However, this approach results in an accuracy loss in signal transmission due to the random frequency offset generated by the sampling rate drift. It is essential to separate resonant oscillation from the motion in a more accurate manner for robust communication.

2) Signal Separation from Plane Motion: Plane motion is ubiquitous among various devices, like cars, smart assistants, or cleaning robots. It affects some of axes in a gyroscope, which means that $M_j[k]$s in Eq. 14 do not necessarily exist synchronously. For instance, a motion is concentrated on the XoY plane, which indicates that the $M_z[k]$ is zero.

We employ coherent inter-axial characteristics for signal recovery. Such characteristics are robust to motion. We select Z-axis to multiply another axis as a combined channel according to Sec, V-C. Taking X-axis as an example, we have

$$\begin{aligned} U_{\times}^{motion}[k] &= R_{RMx}[k] \times R_z[k] \\ &= R_x[k] \times R_z[k] + M_x[k] \times R_z[k]. \end{aligned} \tag{15}$$

It introduces an item $M_x[k] \times R_z[k]$, where the energy of the low-frequency components, if any, is low, and the high-frequency components are removed by the mean filter, since $M_x[k]$ is often low-frequency. Therefore, the plane receiver movement induces no alteration in signal transmission.

3) Signal Separation from Spatial Motion: Spatial motion is more widespread and complicated. Its complexity invalidates the Coherence-based signal extraction. We leverage the single-channel blind source separation (BSS) method with the ensemble empirical mode decomposition (EEMD) for error-free channels under spatial motion interference.

A BSS model can be represented by

$$X = A_{N \times M} S, \ (N, M \in \mathbb{N}_+) \tag{16}$$

Algorithm 1: Components Reorganization Based on the Inter-axial Characteristics

Input: The n-dimension matrices $\hat{S}_1$ and $\hat{S}_2$;
Output: The resonant data $D_1[k]$ and $D_2[k]$

Initialize $A_{max} = 0$;
$B_i = [B_{i1}, B_{i2}, \ldots, B_{i(2^n-1)}](i = 1, 2)$ are Power Set of $\hat{S}_i$, where $B_{ij}(j = 1, 2, \ldots, 2^n - 1) \subset \hat{S}_i$;
$T_{ij}[k] = \sum B_{ij}[k]$;
$L \leftarrow$ the length of $T_{ij}[k]$;
for $i \in [1, 2^n - 1]$ **do**
  for $j \in [1, 2^n - 1]$ **do**
    $Q = \emptyset$;
    for $m = 1 : L$ **do**
      $Q = Q \cup (T_{1i}[m], T_{2j}[m])$;
    Q fits an ellipse E ;
    $\xi \leftarrow$ the mean square error of fitting;
    if $\xi < 0.01$ **then**
      $A \leftarrow$ Area of ellipse E;
      if $A > A_{max}$ **then**
        $A_{max} = A$; $i_{max} = i$; $j_{max} = j$;
return $D_1[k] = P_{1i_{max}}[k]$; $D_2[k] = P_{2j_{max}}[k]$;

where $X = [x_1[k], x_2[k], \ldots, x_N[k]]^T$ is the N-dimension observation matrix, $S = [s_1[k], s_2[k], \ldots, s_M[k]]^T$ is the M-dimension source matrix, and $A_{N\times M}$ is an $N \times M$ matrix, where N and M are constants. Typically, it requires that the number of independent observers is not less than the number of sources, that is $N \geq M$. The goal is to search for inverse matrix to estimate $W = A^{-1}$ and obtain the source S. Thanks to the encoding rules in Sec. V-B, resonance must occur in the odd number pulse width where the variance is greater than the mean. We take the subsequent high-frequency parts after wavelet transform and energy normalization, a mix of the resonant data and remnant of motion, as the observation X, with $N = 1$ here. However, because of frequency offset and motion, there are several independent source vectors, that is $M > 2 > N$, where the dimension requirement is unsatisfied.

EEMD [32], a noise-assisted improved empirical mode decomposition (EMD) algorithm, is employed to decompose the single-channel mixed data to fulfill the requirement on the dimension of the observation matrix. Different from FFT, EEMD manages non-stationary signal analysis. It is based on the data itself and does not require any basic function, making it more suitable for arbitrary data. EEMD, improved from EMD, decomposes single-channel data into several intrinsic mode functions (IMFs). Fundamentally, EMD obtains a series of IMFs from a unidimensional input $x[k]$ as follows:

$$x[k] = \sum_{i=1}^{n} imf_i[k] + r_n[k], \tag{17}$$

where $imf_i[k]$ is the ith IMF component, and $r_n[k]$ is the residue component [33]. But it trips up on performance for its sensitivity to noise [34]. As an improvement for mode mixing [33], EEMD adds a series of Gaussian white noise $noise_{num}[k]$ $(num = 1, 2, \ldots, K)$ with the same standard deviation to the original data, and $x[k]$ in Eq. 17 is replaced by $x_{num}[k] = x[k] + noise_{num}[k]$ cyclically. We have

$$x_{num}[k] = \sum_{i=1}^{n} imf_{numi}[k] + r_{num}[k]. \tag{18}$$

The noise will offset each other after multiple average calculations to recover all the decomposition results. After looping K times, the final IMFs $imf_i[k]$s become

$$imf_i[k] = \frac{1}{K} \sum_{num=1}^{K} imf_{numi}[k]. \tag{19}$$

They constitute an n-dimensional matrix as the observation X instead and the dimension requirement is granted.

We utilize Fast ICA [35], a widely used solution for BSS, with an n-dimension matrix $\hat{S} = [\hat{s}_1[k], \hat{s}_2[k], \ldots, \hat{s}_n[k]]$ as a result. Nevertheless, the number of sources is unclear due to the complexity of motion components, and there is no principle on how to combine those vectors into resonant data.

We reorganize components based on the inter-axial characteristics. Due to the fixed phase difference, resonant data on any two axes can fit a circle or an ellipse. We decompose mixed data from all axes and list all possible combinations for fitting. The one with the largest area and the accredited mean square error is considered to contain only resonant data, with detailed flow clarified in Algorithm 1. We reduce the error bit rates (BERs) to below 0.7% experimentally, which is better than only using wavelet transform, as shown in Fig. 14.

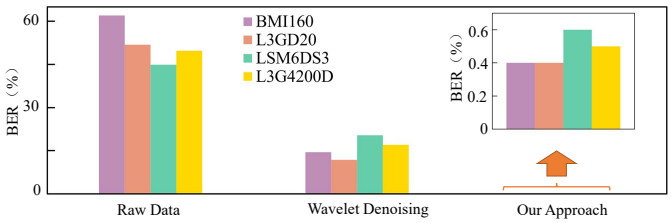


Fig. 14. Performance comparison on separation methods from spatial motion.

To summarize, we analyze the influence of motion interference on gyroscope-based applications and prepare *Deaf-Aid* for the robustness against movement.

VII. Evaluation

We build the prototype of *Deaf-Aid* using COTS devices. We conduct a comprehensive study to evaluate the accuracy and robustness of our system.

A. Experimental Setup and Metrics

We prototype *Deaf-Aid* using COTS speakers as transmitters and gyroscopes as receivers. These devices are fixed into brackets with adjustable distances, as shown in Fig. 15.

Transmitter. We use a JBL GTO 750T speaker [36] as the transmitter. It is supplied by a power amplifier TI LM386. The power supply is set as 5 W or 30 W, both of which are common values for COTS speakers. The speaker is connected to a computer that modulates the ultrasonic signals. COTS laptops (e.g., Dell xps15 9570 in Fig. 15(b)) and smartphones (e.g., Google Pixel 4 in Fig. 15(c)) can also act as the transmitters.

Encoding. We use the adjusted PIE in Sec. V-B as the channel coding. We modulate the ultrasonic signals via denoting

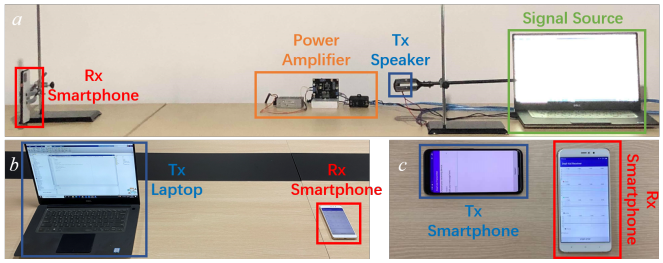
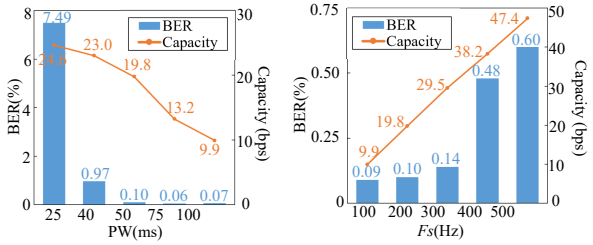


Fig. 15. Experimental setup.



(a) Pulse width (b) Sampling rate

Fig. 16. Performance in different conditions.

the high level as the ultrasound is on and the low level as the ultrasound is off. Before a frame, we send the identifier, consisting of the ultrasound sweeping within the target's ID range every 0.5 seconds for three times. In each frame, there are 256 bits following the SOF. In particular, these bits are interleaved as 8×32, and will be reordered in the receiver.

Receiver. We first test *Deaf-Aid* on gyroscope chips and then apply it to IoT devices represented by smartphones. Gyroscope chips' readings are collected by an Arduino (UNO R3) and an Android APP is developed to record gyroscope readings inside phones (running on Android 11). Our source codes about the detailed Android implementation and Arduino implementation are released in [16]. The threshold values are adaptive as they are determined by the samples in every second using the maximum entropy threshold method in Sec. V-D for decoding. The sampling rate is set as 200 Hz and the pulse width is 50 ms unless otherwise stated.

Shannon channel capacity [37], a theoretically achievable upper bound, is widely used to measure the effectiveness. It is based upon the realized bit error rate, and in a binary symmetric channel, we have the channel capacity as follows,

$$C = \frac{1}{PW}[1 + BER log_2 BER + (1 - BER) log_2 (1 - BER)], \tag{20}$$

where BER is the realized bit error rate and PW is the pulse width. In each experiment, we send 2200 random bits composed of '0' and '1' encoded as Sec. V-B, and thus measure BER. Here, we send messages via the adjusted PIE, which is a typical unequal length coding. Therefore, its actual data rate varies depending on the message to be delivered. In the following evaluation, the measured data rates range from 72% to 81% of the theoretical Shannon channel capacity that could be approached practically with advanced encoding methods like turbo-codes [38].

B. Transmission Capacity

Transmission speed is conditional on sampling rate F_s and pulse width PW. Here, a pair of a speaker (JBL GTO 750T, 5 W) and a gyroscope (BMI160) is placed 15 cm away to analyze the transmission capability with different parameters.

Pulse width. We adjust PW from 25 ms to 100 ms with the F_s of 200 Hz, a widely used default value in mobile devices. The product $PW \times F_s$ decides the amount of samples per bit. A shorter PW means fewer samples are used to form a bit, which possibly causes more errors. In Fig. 16(a), the results demonstrate that BER maintains below 1% when PW is over 40 ms and 0.1% when PW is longer than 50 ms.

TABLE I
VALIDITY OF SIMULTANEOUS MULTI-CHANNEL

Mode	Channels	BER(%)	Capacity(bps)
Case1	XoY	0.23	38.79
	YoZ	0.39	
Case2	XoY	0.27	19.46
	YoZ	0.17	19.64

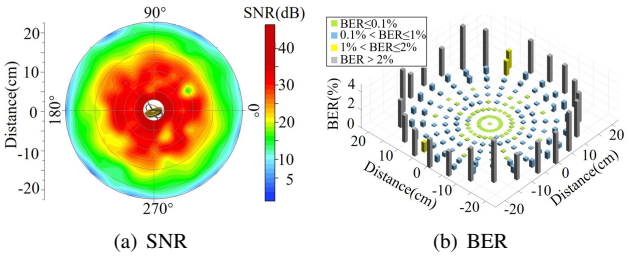
Sampling rate. Similarly, we repeat the experiment where F_s varies evenly between 100 Hz and 500 Hz at 100 Hz intervals. If we keep a constant PW (for example, 100 ms), the capacity would keep around 10 bps whatever the sampling rate F_s is. Therefore, to evaluate the influence of Fs, we keep the samples per bit (i.e., $PW \times F_s$) to be a constant. Here, we set the samples per bit to be 10 and adjust PW according to the value of F_s. As plotted in Fig. 16(b), the channel capacity ascends with the incline of F_s, up to 47.4 bps. Although there is a slight increase in BER, it remains a low level within 0.6%.

Multi-channel. Communication capacity can be doubled, or one receiver can get information simultaneously from two transmitters in different scenarios. The third chip in Fig. 8 is exemplified to bear out the feasibility of multi-channel, where XoY channel works at 24.41 kHz and YoZ channel works at 25.5 kHz. We test *Deaf-Aid* in two cases. In the case 1, one speaker delivers different messages on these two channels simultaneously. In the case 2, two speakers deliver on two different channels, with the performance attached to Tab. I. In both cases, we succeed at the expense of a slight accuracy loss. ***Deaf-Aid* supports simultaneous communication on multiple channels, even from two transmitters.**

Deaf-Aid can flexibly meet the different requirements of transmission speed and error tolerance in various applications.

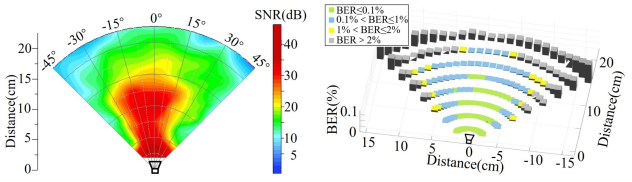
C. Orientation and Distance

We examine the resilience under multiple layouts to further demonstrate the less restriction on the layout of devices. We rotate a speaker (5 W JBL750T) around a fixed gyroscope (BMI160). The performance in the XoY plane is shown in Fig. 17. It reflects that the placement of gyroscopes makes no difference when we keep the speaker face toward them (it is reasonable to ask users to face toward the receivers for communication in practice). Furthermore, we rotate the gyroscope around the fixed speaker. Fig. 18 illustrates the effectiveness in the range of a 22.5° opening angle of speakers, with a BER of 0.1% at 15 cm and 1% at 20 cm. It is practical for users to turn toward the objective, and a slight direction deviation is tolerable. Moreover, we perform arbitrary layouts



(a) SNR (b) BER

Fig. 17. Performance centered on the gyroscope.



(a) SNR (b) BER

Fig. 18. Performance centered on the speaker.

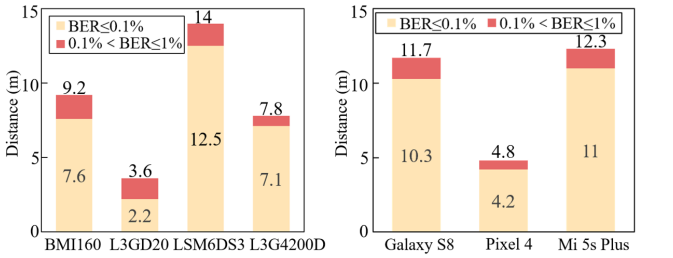
and observe similar performance with gyroscopes in front of speaker (within a 22.5° opening angle).

The above results on the communication range are obtained from a power-limited speaker, whose power is only 5 W. We adopt such a setting with the consideration that some IoT devices are equipped with a similar power-limited speaker for the purpose of reducing power consumption.

For those devices without strict requirements on power consumption, the communication distance can dramatically increase. We raise the power of the speaker to 30 W, which is also very common in existing commodity speakers. **The communication distance can extend up to 14 m**, as shown in Fig. 19(a). In general, different gyroscopes have different resonance peaks A_{Rmax}, and different communication distances correspondingly. Nevertheless, even the L3GD20 chip, which performs the worst among our gyroscopes, supports a communication distance of 3.6 m. Such a communication range is sufficient to cover most of the application scenarios. Furthermore, taking three Android phones (Samsung Galaxy S8, Google Pixel 4, and Mi 5s Plus) as examples, those devices are with encapsulated gyroscopes, corresponding to LSM6DSL, BMI160, and ICG-20660/L, respectively. We measure the transmission distance using a 30 W speaker. Fig. 19(b) reflects that our system is able to communicate with those phones up to 12.3 m away, as their screens are set vertical to the ground. The BMI160 encapsulated in Pixel 4 has the shortest communication distance of 4.4 m. It is still satisfactory in many scenarios, e.g., the indoor environment.

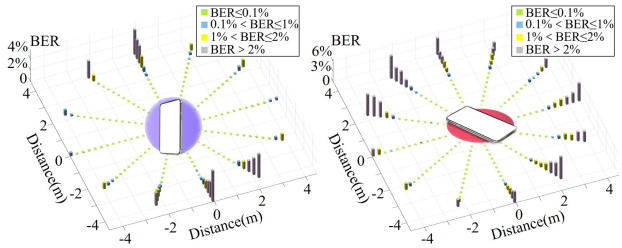
We use a Pixel 4 to validate the flexibility of *Deaf-Aid* in terms of layout. We rotate the speaker around the fixed phone, vertically and parallelly respectively, with the results in Fig. 20. The communication distance fluctuates between 3.1 m and 4.8 m with a BER less than 1%. This enables a smartphone to retrieve messages sent within three meters accurately via *Deaf-Aid*, no matter which orientation it is in. This demonstrates that *Deaf-Aid* is capable of establishing communication among realistic devices in the wild.

The speaker we use is placed 5 cm from an NI USB-4431 sound measuring instrument and a GRAS 46AM free-field microphone. The unweighted sound pressure levels (SPLs) are measured with 68.5 dB/76 dB with the power setting



(a) Gyroscope chips (b) Phones with encapsulated gyroscope

Fig. 19. Communication distance.



(a) Vertical (b) Parallel

Fig. 20. Performance centered on a Pixel 4.

of 5 W/30 W (consistent with other experiments) when the speaker is operated near its maximum amplitude. The experiment is performed in a quiet laboratory with 49.5 dB environment noise. As SPL increases from 68.5 dB to 76 dB, the communication distance is enlarged from 20 cm to at most 14 m, without affecting the channel capacity. We further evaluate the inaudibility of data transmission via *Deaf-Aid*. We apply the speaker-to-gyroscope channel on eight gyroscopes listed in Fig. 3 using a 30 W JBL750T (76 dBSPL). Note that we adopt ultrasound of over 18 kHz that is barely audible to human. We recruit 22 volunteers aged from 18 to 45. Only one volunteer (aged 20) reports being able to vaguely distinguish the existence of modulated ultrasound during the communication with the LSM330 gyroscope, which obtains the lowest resonant frequency of 18.4 kHz.

D. Motion Interference

Motion interference is involved for a better understanding of the robustness of *Deaf-Aid* against movement. We bind a 5 W JBL speaker, an obstacle, and gyroscopes to a manipulator respectively. They move under the control of the program. The experimental distance is set within 15 cm and the gyroscopes keep in front of the speaker with the opening angle below 20 ° by default. Moreover, we recruit 22 participants, and they are asked to hold 30 W speakers and three kinds of phones with encapsulated gyroscopes for further confirmation.

Tx motion and LOS blocking. We move a speaker in three simple directions (transverse, longitude, and verticality), while a BMI160 chip is fixed. As shown in Fig. 21(a), the BER is around 0.1% when the speaker moves and is always below 0.2% even under obstacle disturbance. Then we manipulate 5×5×5 cm^3 wooden and cloth obstacles into moving randomly in LOS between the fixed speaker and gyroscope. We consider the influence of different materials of mobile obstacles on the LOS. We repeat experiments using metal (aluminium), cloth, glass, and silica gel balls (diameter of 5

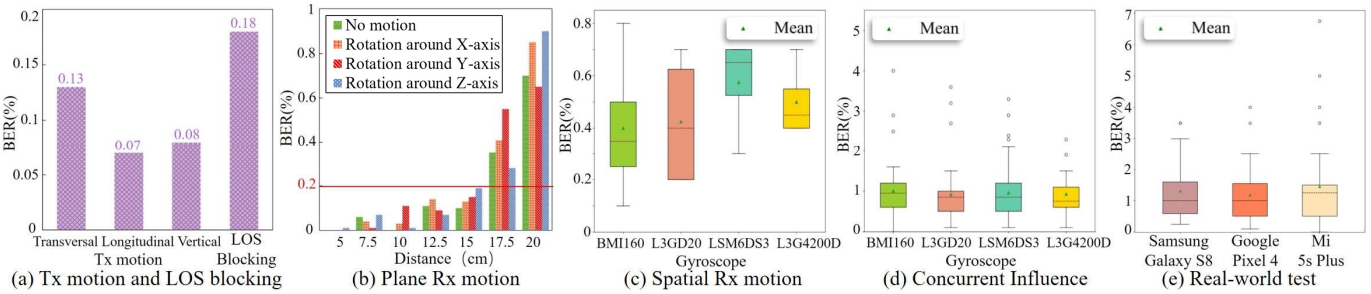


Fig. 21. Performance of *Deaf-Aid* against motion interference.

cm). *Deaf-Aid* obtains robust performance with BERs lower than 0.25%. We also recruit volunteers acting as obstacles. They cross the LOS between a BMI160 gyroscope and a 30 W speaker. At this time, BERs still maintain below 1%. We also test the acoustic damping material, a special material made of fiber materials, which can reduce the intensity of acoustics travelling through it. Though the channel's SNR decreases sharply from over 20 dB to about 4.8 dB and the BER increases from below 0.1% to about 1.1%, mobile obstacles with such damping materials are not commonly used in daily life or industrial scenarios. In practice, users can intentionally avoid obstacles with acoustic damping materials.

Plane Tx motion. Then we fix the speaker and rotate a BMI160 chip around its axes with the comparison at different distances shown in Fig. 21(b). This system maintains a low BER. It is less than 0.2% at a distance of 15 cm and rises to 1% as the distance increases to 20 cm. Thereby, the plane motion has little impact on the stability of our system.

Spatial Tx motion. Here, gyroscopes move in space irregularly within 15 cm from a fixed speaker. This evaluation involves four types of gyroscopes and each type contains 8 chips. Repetitive experiments are conducted on these chips where the manipulator repeats the same trajectory. It has a maximum error of 0.8% with all averages lower than 0.7% in Fig. 21(c). We have prepared *Deaf-Aid* for the robustness against the fundamental movement.

Concurrent influence. We ask 22 volunteers to send information with a speaker in hand where 32 gyroscope chips move in space irregularly under the same conditions as above. As plotted in Fig. 21(d), our system performs well under the multiple concurrent motion interference. The mean of BERs maintains below 1%. Albeit there exists off-group data, the peak is lower than 6%. One possible explanation for those outliers is that volunteers accidentally deflect the orientation of the speaker away from receivers.

Real-world motion. In order to evaluate the robustness of the prototypes, we group volunteers into pairs. In each pair, one volunteer holds a 30 W speaker and the other carries a smartphone. Both of them move freely within a range of 2 m and fiddle with the devices. We evaluate on three kinds of smartphones and find that the mean of BERs is below 1% and the peak is lower than 7% during the entire experiment, as the result shown in Fig. 21(e). This indicates that *Deaf-Aid* facilitates a robust channel among mobile IoT devices.

To sum up, *Deaf-Aid* shows immense potential as a communication bridge in the real-world implementation even under various motion disturbances in a complicated IoT network.

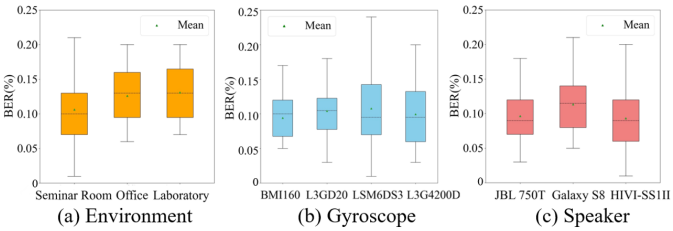


Fig. 22. The impact of environment and devices.

TABLE II
THE RESONANT INFORMATION ABOUT 21 COTS DEVICES

Device	Gyroscope model†	f_N (kHz)	Distance‡ (m)
Samsung Galaxy S8	STM LSM6DSL	19.47	11.7
Google Pixel 4	BS BMI160	24.59	4.8
Mi 5s Plus	IS ICG-20660/L	20.1	12.3
Mi 10	STM LSM6DSO	19.03	4
HUAWEI P20	IS ICM-20690	19.9	14
HONOR 20	BS BMI160	19.45	15
OPPO R15	BS BMI160	19.65	5
VIVO NEX35	STM LSM6DSM	19.65	5
VIVO S6	Unknown	19.98	8
iPhone 6	Unknown	26.9	0.7
iPhone 6s plus	IS MP67B	27.2	1
iPhone XS	BS BMI282	26	1.5
iPhone 11 Pro Max	BS BMI282	24.15	2
iPad Air 3	Unknown	25.8	0.5
iPad Pro 2020	Unknown	26.4	0.6
Apple Watch Series 6	BS BMI282	25.9	2
AMAZFIT Mi	BS BMI160	19.8	3.4
Baidu Apollo D-KIT	IS MPU6050	27.5	3.1
EAIBOT N1 UGV	M R6093U	27.2	3.5
QQL RC UAV	IS IMU3000	27.1	10
DJI Spark UAV	UnKonwn	23.8	5.5

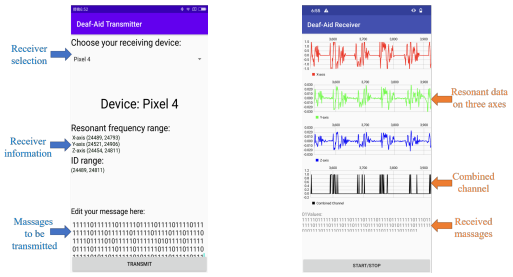
† STM: STMicroelectronics, BS: Bosch, IS: InvenSense, M: Microinfinity.
‡ The maximum distance for communication with BERs below 1%.

E. Universality and COTS Devices Implementation

We take the diversity of devices and environments into account to further verify the universality of *Deaf-Aid*. Here we place the speaker and gyroscope at a distance of 15 cm in three different locations including a large seminar room, a small office, and a crowded laboratory. We test on six speakers of three kinds (including JBL 750T, Samsung Galaxy S8, and HIVI-SS1II), whose supply power is limited within 5 W, and 32 gyroscopes of four models (including BMI160, L3GD20, LSM6DS3, and L3G4200D). The distribution is presented in Fig. 22. In these situations, our system performs satisfactorily, with BERs lower than 0.25% comprehensively. It guarantees the stable communication quality among numerous devices with little deformation due to ambient disturbances.

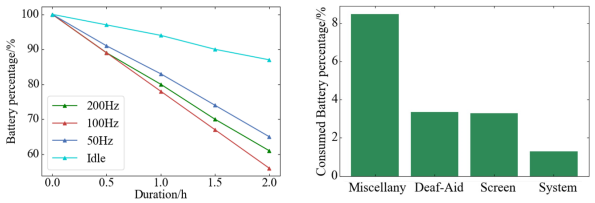
TABLE III
COMPARISON WITH PREVIOUS WORK

System	Basic	Speed	Accuracy	Receiver Identification	Placement	Distance	Motion Robustness
Ripple [6]	Vibra-motor to accelerator	200 bps	BER<1.7%	Manually	Fixed on a plane	0.15 cm	Not allow movement
Ripple II [7]	Vibra-motor to microphone	30 kbps	SNR>15dB	Manually	Physical contact	Touch-based	Just to tiny vibration
BitWhisper [8]	Heat emission to thermal sensor	1-8 bits per hour	Not evaluate	Manually	Fixed position	40 cm	Not allow movement
Dhwani [25]	Speaker to microphone	2.4 kbps	BER<5%	Manually	Remote	10 cm	Yes
BackDoor [39]	Speaker to microphone	4 kbps	PER≈2%	Manually	Remote	1m	Yes
Dolphin [40]	Speaker to microphone	500 bps	BER <5%	Manually	Remote	8m	Yes
Deaf-Aid	Speaker to gyroscope	47 bps	BER<0.6%	Automatic	Remote	14m	Yes



(a) Transmitting terminal (b) Receiving terminal

Fig. 23. User interfaces of our APP.



(a) Speed with different F_s (b) Usage ranking in an hour

Fig. 24. Power consumption of *Deaf-Aid*.

To verify the feasibility of *Deaf-Aid* on devices with encapsulated gyroscopes, we apply it to COTS devices, including 13 phones, 2 iPads, 2 smart watches, and 4 automatic vehicles, as listed in Tab. II. We develop an APP on Android smartphones for practical deployments with the user interfaces displayed in Fig. 23. *Deaf-Aid* is capable of being extended to more COTS devices and achieves half-duplex data transmission in real-world scenarios. The speakers installed on portable devices tend to be more available than the commercial speakers for users. Particularly, to investigate the qualification of using COTS devices as transmitters, a laptop or a smartphone, taking the place of a commercial speaker in the experiments above, sends messages to a smartphone via *Deaf-Aid*, as arranged in Fig. 15(b) and (c). They are qualified for the transmitters, where the communication distance maintains 15 cm or 5 m respectively with a BER below 1%.

F. Power Consumption

Power consumption is a critical issue that should not be overlooked, especially for COTS devices. *Deaf-Aid* demands a speaker to transmit ultrasonic modulated signals and continuous gyroscope readings to acquire the resonant data, and hence consumes extra power. We discuss its power consumption from the perspectives of transmitting and receiving terminals.

1) Transmitting terminal: In the audio components of IoT devices, the power is almost consumed by the speaker. In general, the consumption of a commodity speaker is designed and constricted within an acceptable range for an IoT device. In our experiment, the power supply of the speakers (i.e., the transmitters) is 5 W or 30 W. We have demonstrated the effectiveness of *Deaf-Aid* using a speaker with the restricted power supply of 5 W, with the consideration of the low power consumption of IoT devices in some cases (see Sec. VII-C). In addition, if the transmitter is a mobile phone, the power consumption would not be a big issue with the aid of high-capacity battery and power bank. *Deaf-Aid* prepares itself for occasions with both finite and sufficient power supply.

2) Receiving terminal: To evaluate the amount of extra power being consumed by *Deaf-Aid* on the receiving terminal, we run the APP to obtain the gyroscope data at different sampling rates, and compare the power consumption with that of the smartphone's idle status, whose screen is always on with identical brightness. As plotted in Fig. 24, continuous communication via *Deaf-Aid* for two hours consumes less than 40% of the smartphone's battery power, where *Deaf-Aid* attributes merely about 20% of whole power consumption.

G. Comparison with Previous Systems

Covert channels take advantage of physical phenomenon to transfer data among adjacent devices. We select some typical cases for comparison, listed in Tab. III. Communication through vibration, for example *Ripple* [7] and *Ripple II* [6], is good in speed but weak in fixed position and poor motion robustness. *BitWhisper* [8] delivers messages further but slowly on a covert channel using thermal manipulations. The speaker-to-microphone channel is exploited by *Dhwani* [25] and *Dolphin* [40]. However, with the purpose of recording human voices, the microphones on IoT devices are more likely to filter out 8 kHz [41]. In this case, approaches like *Dhwani* and *Dolphin* require peripherals to utilize ultrasound for stealthy communication, such as high-quality microphones and sound cards with high sampling rates. Otherwise, people nearby will be disturbed. *BackDoor* [39] benefits from the non-linearity of the microphone, but interferes with the normal operation of the microphone for the speech collection and recognition. To leverage the non-linearity, *BackDoor* [39] uti-

lizes the ultrasound of over 40 kHz, while the ultrasounds used in *Deaf-Aid* range from 20 kHz to 38 kHz. The experimental results also demonstrate that our adoption of ultrasonic signals does not induce the non-linearity of microphones in the test devices, nor does it produce additional low-frequency noise. *Deaf-Aid* has no such issues instead, not to mention that it also has other advantages, such as multi-channel communication and automatic receiver identification.

In summary, *Deaf-Aid* enables IoT devices to identify and chat with their neighbors. It provides an alternative and complementary communication channel to existing IoT devices.

VIII. Discussion

A. Implementation Consideration

Communication distance and capacity will soar along with technology. A better speaker, with a wider spectrum of responses or more power, extends the communication range. It is reported that ultrasound is capable to affect gyroscopes 37 m away [11]. This indicates the great potential of *Deaf-Aid* in more scenarios. On the other hand, increasing the sampling rate would result in a higher transmission rate. *Deaf-Aid* will contentiously improve in the transmission rate upon the emergence of new hardware. For example, the gyroscopes MPU6050 and BMI160 [42] support over 1 kHz sampling rates, respectively. If we adopt a gyroscope with a sampling rate over 10 kHz, *Deaf-Aid* can raise the transmission rate to thousands of bps by the conservative estimation. We will obtain a more efficient system as new hardware emerges.

Adaptable power supply can adjust energy consumption according to different scenes. Intuitively, a higher power supply supports a wider communication range. In some cases, the distance between the transmitter and receiver is definite, where the coverage over this distance makes no sense. There is much room for future work to exploit the quantitative relationship between power consumption and coverage, and then find the optimal device setting for energy-saving.

Signal clipping means that the analog voltage exceeds the sensor's input range, and thus distorts. For instance, it occurs in communication via an L3GD20 chip within 5 cm experimentally. Even so, rising edges are still recognizable. Clipping introduces fewer additional errors statistically.

More IoT devices and platforms will be supported in the foreseeable future. In view of proven resonance phenomenon in the 3D mouse, screwdriver, VR device [9], drone [11], and remote control model car [12], our system can be applied in a broader range of devices including those above. This could be an essential step to expand application fields, thus leading to a more comprehensive IoT network based on *Deaf-Aid*.

B. Security

The current resonant frequency band is relatively narrow, determined by the inherent structure of gyroscopes. In this case, some sophisticated communication techniques, such as FDM and OFDM, are not applicable. However, we exploit the potential of the narrow bands from a security perspective.

Jamming: It is difficult for malicious jammers to find out the appropriate band of a gyroscope. To our knowledge, though an attacker can determine the type of gyroscope in a receiver via the product datasheet, it typically provides a rough range about the resonant bands. For example, the datasheet of MPU6050 gyroscopes claims the resonant frequencies in 27 ± 3 kHz. We observe that the range of a given chip's resonant band is less than 50 Hz. In our experiment, an MPU6050 chip's resonant band measures about 24910 to 24960 Hz (only a tiny segment within the resonant frequencies claimed by its datasheet). The attacker cannot pick up this appropriate band used for communication exactly from the 6 kHz band provided by the datasheet. The out-of-band jamming signals thereby are difficult to block communication. Without sufficient knowledge, the attacker has to jam in a broadband spectrum. However, this method demands professional and high-energy-consuming acoustic loudspeakers. We design acoustic jamming noise (24$\sim$30 kHz, 70$\sim$85 dB) to evaluate *Deaf-Aid*'s resilience against interference using the aforementioned MPU6050 gyroscope. It maintains BERs within 0.6% unless the jamming noise overlaps the 50 Hz appropriate band. Meanwhile, such interference can be detected easily. A practicable means of avoiding jamming is to detect it in a timely manner and avoid communicating when it occurs. It is easy for *Deaf-Aid*, as it only requires a microphone.

Eavesdropping: *Deaf-Aid* can prevent replay attacks even if the private key is leaked. Benefiting from gyroscopes' narrow band-pass width, intended non-informative ultrasound signal could broadcast at the nearby frequency to confuse attackers but receivers are impervious to the noise with the help of Coherence-based signal extraction. Furthermore, users can utilize multi-channel with signals on one channel and deceptive data on the other. Prior information is an absolute necessity for eavesdroppers, such as the communication frequency band, which is difficult to pick out the right one from camouflage.

IX. Related Work

Privacy is recorded by inertial sensors. A malicious attacker can easily obtain inertial sensors data inside mobile platforms without access permission, for keystroke inference [15], [43], [44], [45], [46], [47], device identification [48], [26] and speech recognition [49], [50], [51]. Approaches like *Gyrophone* [49] leverage a gyroscope as an eavesdropper to recognize speeches, often lower than 1 kHz. Their intention is to eavesdrop on the context of human conversation via vibration. Different from *Gyrophone*, *Deaf-Aid* benefits from the resonance of gyroscopes and is aiming at transferring modulated information from a speaker to a gyroscope.

Gyroscope is vulnerable to acoustic injection attacks. It has been demonstrated that resonance of gyroscopes could be triggered by acoustic signals [12], [13], [14]. An adversary can impose on outputs of gyroscopes, bringing about control system error. A DoS attack was conducted to incapacitate drones [11]. Tu *et al.* [9] realized a black box switching attack to push victim gyroscope to produce expected outputs.

Covert channels have attracted great interest. They leverage physical phenomena, such as heat [3], [8], [52], light [53], [54], electromagnetic leakage [55], ultrasound [56], and inertia [6], [7], [21]. Nevertheless, these methods demand

physical contact, specialized equipment or artificial assistance, none of which is needed in *Deaf-Aid*.

X. Conclusion

We leverage the speaker-to-gyroscope channel for mobile IoT communication. We probe the inter-axial characteristics in resonant gyroscopes. Such characteristics support error-free and multi-channel communication against frequency offset. As an innovation, the diversity of resonant frequency ranges among gyroscopes is employed as fingerprint for automatic receiver identification. The motion influence suppression and mobile communication have been delicately designed. Our system, *Deaf-Aid*, reaches up to 47 bps with a low BER even under motion interference. It could act as a stepping-stone for an everything-related IoT network.

References

[1] M. Gao, F. Lin, W. Xu, M. Nuermaimaiti, J. Han, W. Xu, and K. Ren, "Deaf-aid: Mobile iot communication exploiting stealthy speaker-to-gyroscope channel," in *Proc. MobiCom*, 2020, pp. 705–717.

[2] IDC, "Worldwide internet of things forecast, 2019–2023," https://www.idc.com/getdoc.jsp?containerId=US45373120, 2019.

[3] Z. Wu, Z. Xu, and H. Wang, "Whispers in the hyper-space: High-speed covert channel attacks in the cloud," in *Proc. Usenix Secur. Symp.*, 2012, pp. 159–173.

[4] K. Lee, H. Wang, and H. Weatherspoon, "PHY covert channels: Can you see the idles?" in *Proc. USENIX NSDI*, 2014, pp. 173–185.

[5] A. Al-Haiqi, M. Ismail, and R. Nordin, "A new sensors-based covert channel on android," *Sci. World J.*, vol. 2014, 2014.

[6] N. Roy, M. Gowda, and R. R. Choudhury, "Ripple: Communicating through physical vibration," in *Proc. USENIX NSDI*, 2015, pp. 265–278.

[7] N. Roy and R. R. Choudhury, "Ripple II: faster communication through physical vibration," in *Proc. USENIX NSDI*, 2016, pp. 671–684.

[8] M. Guri, M. Monitz, Y. Mirski, and Y. Elovici, "Bitwhisper: Covert signaling channel between air-gapped computers using thermal manipulations," in *Proc. IEEE CSF*, 2015, pp. 276–289.

[9] Y. Tu, Z. Lin, I. Lee, and X. Hei, "Injected and delivered: Fabricating implicit control over actuation systems by spoofing inertial sensors," in *Proc. Usenix Secur. Symp.*, 2018, pp. 1545–1562.

[10] Z. Wang, W. Zhu, J. Miao, H. Zhu, C. Chao, and O. K. Tan, "Micromachined thick film piezoelectric ultrasonic transducer array," *Sens. Actuator A-Phys.*, vol. 130-131, pp. 485–490, 2005.

[11] Y. Son, H. Shin, D. Kim, Y. Park, J. Noh, K. Choi, J. Choi, and Y. Kim, "Rocking drones with intentional sound noise on gyroscopic sensors," in *Proc. Usenix Secur. Symp.*, 2015, pp. 881–896.

[12] T. Trippel, O. Weisse, W. Xu, P. Honeyman, and K. Fu, "WALNUT: waging doubt on the integrity of MEMS accelerometers with acoustic injection attacks," in *Proc. IEEE EuroSP*, 2017, pp. 3–18.

[13] R. N. Dean, S. T. Castro, G. T. Flowers, G. Roth, A. Ahmed, A. S. Hodel, B. E. Grantham, D. A. Bittle, and J. P. Brunsch, "A characterization of the performance of a MEMS gyroscope in acoustically harsh environments," *IEEE Trans. Ind. Electron.*, vol. 58, no. 7, pp. 2591–2596, 2011.

[14] R. N. Dean, G. T. Flowers, A. S. Hodel, G. Roth, S. Castro, R. Zhou, A. Moreira, A. Ahmed, R. Rifki, B. E. Grantham, D. Bittle, and J. Brunsch, "On the degradation of mems gyroscope performance in the presence of high power acoustic noise," in *Proc. IEEE ISIE*, 2007, pp. 1435–1440.

[15] L. Cai and H. Chen, "On the practicality of motion based keystroke inference attack," in *Proc. TRUST*, 2012, p. 273–290.

[16] Deaf-aid, "Deaf-aid codes," https://github.com/Deaf-aid/Deaf-aid.git, 2022.

[17] W. T. Thomson, *Theory of vibration with applications*. Prentice Hall, 1981.

[18] Iflytek CO.,LTD, "iflytek open platform- an artificial intelligence platform focusing on intelligent speech interaction wihch provides solutions for global developers," https://global.xfyun.cn/, 2021.

[19] Google Cloud, "Speech-to-text: Automatic speech recognition," https://cloud.google.com/speech-to-text, 2021.

[20] D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishankar, and A. I. Rudnicky, "Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices," in *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, 2006, pp. 185–188.

[21] K. Block, S. Narain, and G. Noubir, "An autonomic and permissionless android covert channel," in *Proc. ACM CCS*, 2017, pp. 184–194.

[22] S. Inc., "Hi-res audio," https://www.highresaudio.com/en, 2011.

[23] Q. Inc., "Qualcomm snapdragon sound: how sound should sound," https://www.qualcomm.com/products/features/snapdragon-sound, 2021.

[24] Y. Stebler, S. Guerrier, J. Skaloud, and M. Victoria-Feser, "A framework for inertial sensor calibration using complex stochastic error models," in *Proc. IEEE/ION PLANS*, 2012, pp. 849–861.

[25] R. Nandakumar, K. K. Chintalapudi, V. Padmanabhan, and R. Venkatesan, "Dhwani: secure peer-to-peer acoustic nfc," *ACM SIGCOMM Comp. Commun. Rev.*, vol. 43, no. 4, pp. 63–74, 2013.

[26] J. Zhang, A. R. Beresford, and I. Sheret, "Sensorid: Sensor calibration fingerprinting for smartphones," in *Proc. SP*, 2019, pp. 638–655.

[27] P. R. Prucnal and P. A. Perrier, "Optical self-routing in a self-clocked photonic switch using pulse-interval encoding," in *Proc. ECOC*, 1988.

[28] J. Kapur, P. Sahoo, and A. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Compute Vision, Graphics, and Image Processing*, vol. 29, pp. 273–285, 1980.

[29] D. Zhou and W. Cheng, "Image denoising with an optimal threshold and neighbouring window," *Pattern Recognit. Lett.*, vol. 29, no. 11, pp. 1694–1697, 2008.

[30] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, 1996.

[31] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[32] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Adv. Comput. Math*, vol. 1, no. 1, pp. 1–41, 2009.

[33] B. Mijovic, M. De Vos, I. Gligorijevic, J. Taelman, and S. Van Huffel, "Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2188–2196, 2010.

[34] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. London*, vol. 454, no. 1971, pp. 903–995, 1998.

[35] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, 1999.

[36] Harman, "Jbl stadium gto750t," https://www.onlinecarstereo.com/CarAudio/p_51143_JBL_STADIUMGTO750T.aspx, 2019.

[37] T. M. Cover, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2017.

[38] P. Sweeney, *Error Control Coding: From Theory to Practice*. John Wiley & Sons, Inc., 2002.

[39] N. Roy, H. Hassanieh, and R. R. Choudhury, "Backdoor: Making microphones hear inaudible sounds," in *Proc. MobiSys*, 2017, pp. 2–14.

[40] Q. Wang, K. Ren, M. Zhou, T. Lei, D. Koutsonikolas, and L. Su, "Messages behind the sound: real-time hidden acoustic signal capture with smartphones," in *Proc. MobiCom*, 2016, pp. 29–41.

[41] H. Abdullah, W. Garcia, C. Peeters, P. Traynor, K. R. B. Butler, and J. Wilson, "Practical hidden voice attacks against speech and speaker recognition systems," in *Proc. NDSS*, 2019.

[42] Bosch, "Bmi160 datasheet," https://www.bosch-sensortec.com/products/motion-sensors/imus/bmi160.html, 2018.

[43] E. Miluzzo, A. Varshavsky, S. Balakrishnan, and R. R. Choudhury, "Tapprints: your finger taps have fingerprints," in *Proc. MobiSys*, 2012, pp. 323–336.

[44] Z. Xu, K. Bai, and S. Zhu, "Taplogger: inferring user inputs on smartphone touchscreens using on-board motion sensors," in *Proc. ACM CCS*, 2012, pp. 113–124.

[45] X. Liu, Z. Zhou, W. Diao, Z. Li, and K. Zhang, "When good becomes evil: Keystroke inference with smartwatch," in *Proc. ACM CCS*, 2015, pp. 1273–1285.

[46] C. Wang, X. Guo, Y. Wang, Y. Chen, and B. Liu, "Friend or foe?: Your wearable devices reveal your personal PIN," in *Proc. ACM CCS*, 2016, pp. 189–200.

[47] E. Owusu, J. Han, S. Das, A. Perrig, and J. Zhang, "Accessory: password inference using accelerometers on smartphones," in *Proc. HotMobile*, 2012, pp. 1–6.
[48] S. Dey, N. Roy, W. Xu, R. R. Choudhury, and S. Nelakuditi, "Accelprint: Imperfections of accelerometers make smartphones trackable," in *Proc. NDSS*, 2014.
[49] Y. Michalevsky, D. Boneh, and G. Nakibly, "Gyrophone: Recognizing speech from gyroscope signals," in *Proc. Usenix Secur. Symp.*, 2014, pp. 1053–1067.
[50] J. Han, A. J. Chung, and P. Tague, "Pitchin: Eavesdropping via intelligible speech reconstruction using non-acoustic sensor fusion," in *Proc. ACM CCS*, 2017, pp. 181–192.
[51] Z. Ba, T. Zheng, X. Zhang, Z. Qin, B. Li, X. Liu, and K. Ren, "Learning-based practical smartphone eavesdropping with built-in accelerometer," in *Proc. NDSS*, 2020.
[52] D. B. Bartolini, P. Miedl, and L. Thiele, "On the capacity of thermal covert channels in multicores," in *Proc. ECCS*, 2016, pp. 24:1–24:16.
[53] A. Maiti and M. Jadliwala, "Light ears: Information leakage via smart lights," *Proc. ACM IMWUT*, vol. 3, no. 3, pp. 98:1–98:27, 2019.
[54] T. Sugawara, B. Cyr, S. Rampazzi, D. Genkin, and K. Fu, "Light commands: Laser-based audio injection attacks on voice-controllable systems," in *USENIX Security Symposium*, 2020, pp. 2631–2648.
[55] M. Guri, G. Kedma, A. Kachlon, and Y. Elovici, "Airhopper: Bridging the air-gap between isolated networks and mobile phones using radio frequencies," in *Proc. MALWARE*, 2014, pp. 58–67.
[56] A. Madhavapeddy, R. Sharp, D. J. Scott, and A. Tse, "Audio networking: the forgotten wireless technology," *IEEE Pervasive Comput.*, vol. 4, no. 3, pp. 55–60, 2005.

Feng Lin received the Ph.D. degree from the Department of Electrical and Computer Engineering, Tennessee Technological University, USA, in 2015. He is currently a Professor with the School of Cyber Science and Technology, College of Computer Science and Technology, Zhejiang University, China. He was an Assistant Professor with the University of Colorado Denver, USA, a Research Scientist with the State University of New York (SUNY) at Buffalo, USA, and an Engineer with Alcatel-Lucent (currently, Nokia). His current research interests include mobile sensing, wireless sensing, Internet of Things security, biometrics, and AI security. Dr. Lin was a recipient of the ACM SIGSAC China Rising Star Award, the Best Paper Awards from ACM MobiSys'20, IEEE Globecom'19, IEEE BHI'17, the Best Demo Award from ACM HotMobile'18, and the Best Paper Award Nomination from SenSys'21 and INFOCOM'21.

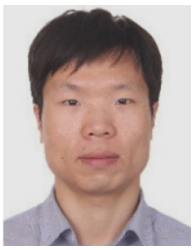
Ming Gao is a Ph.D. candidate at the school of cyber science and technology, Zhejiang University. He received the Master and Bachelor degree from Xi'an Jiaotong University. His research interests include cyber-physical security, mobile computing, and privacy protection. He is a recipient of the Best Paper Award Nomination from SenSys'21.

Lingfeng Zhang is pursuing M.Sc. at the School of Cyber Science and Technology, Zhejiang University, under the supervision of Prof. Jinsong Han. His research interests include cyber-physical security and smart sensing.

Yimin Li received her M.Sc. from University College London in 2021. She was a Visiting Student with Zhejiang University from 2019 to 2020. Her research interests include wireless sensor network and privacy protection.

Weiye Xu received her B.S. degree in computer science and technology from Wuhan University, China in 2019. She is currently a Phd student in Department of Computer Science and Technology at Zhejiang University, under the supervision of Prof. Jinsong Han. Her research interests include IoT security, wireless sensing and mobile computing.

Jinsong Han is now a professor at the School of Cyber Science and Technology, Zhejiang University. He is a senior member of the ACM and IEEE. His research interests focus on IoT security, smart sensing, wireless and mobile computing.



Xian Xu is a professor at the College of Civil Engineering and Architecture, Zhejiang University. His research interests include smart structural health monitoring.

Wenyao Xu received the Ph.D. degree from the University of California at Los Angeles, Los Angeles, USA, and both the Master and Bachelor degree from Zhejiang University, China. Wenyao Xu is an Associate Professor with tenure of the Computer Science and Engineering Department, University at Buffalo (SUNY). His research has focused on exploring novel sensing and computing technologies to build up innovative Internet-of-Things (IoT) systems for high-impact human-technology applications in the fields of Smart Health and Cyber-Security. Results have been published in peer reviewed top research venues across multiple disciplines, including Computer Science conferences (e.g., ACM MobiCom, SenSys, MobiSys, UbiComp, ASPLOS, ISCA, HPCA, Oakland, NDSS and CCS), Biomedical Engineering journals (e.g., IEEE TBME, TBioCAS, and JBHI), and Medicine journals (e.g., LANCET). To date, his group has published over peer-reviewed 180 papers, won nine best paper awards, two best paper nominations and three international best design awards. Currently, Wenyao Xu serves as an Associate Editor of IEEE Transactions on Biomedical Circuits and Systems (TBCAS), the technical program committee of numerous conferences in the field of Smart Health and Internet of Things, and has been a TPC co-chair of IEEE Body Sensor Networks in 2018.

Kui Ren (Fellow, IEEE and ACM) received the Ph.D. degree in electrical and computer engineering from the Worcester Polytechnic Institute. He is currently a Professor and the Associate Dean of the College of Computer Science and Technology, Zhejiang University, where he also directs the Institute of Cyber Science and Technology. Before that, he was the SUNY Empire Innovation Professor of The State University of New York at Buffalo. His H-index is 74 and his total publication citation exceeds 32 000 according to Google Scholar. His current research interests include data security, the IoT security, AI security, and privacy. He has published extensively in peer-reviewed journals and conferences and received the Test-of-Time Paper Award from IEEE INFOCOM and many Best Paper Awards from IEEE and ACM, including MobiSys 2020, Globecom 2019, ASIACCS 2018, and ICDCS 2017. He received the NSF CAREER Award in 2011, the Sigma Xi Research Excellence Award in 2012, the IEEE CISTC Technical Recognition Award in 2017, the SUNY Chancellor's Research Excellence Award in 2017, and the Guohua Distinguished Scholar Award from ZJU in 2020. He is a Clarivate Highly-Cited Researcher. He is a frequent reviewer for funding agencies internationally and serves on the editorial boards of many IEEE and ACM journals. He also serves as the Chair for SIGSAC of ACM China.