Computer Networks 181 (2020) 107447

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Survey paper

Acoustic-based sensing and applications: A survey

Yang Bai [a], Li Lu [b], Jerry Cheng [c], Jian Liu [d], Yingying Chen [a,*], Jiadi Yu [b]

[a] WINLAB, Department of Electrical and Computer Engineering, Rutgers University, New Brunswick, NJ 08901, United States of America
[b] Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, PR China
[c] Department of Computer Science, New York Institute of Technology, New York, NY 10023, United States of America
[d] Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN 37996, United States of America

ARTICLE INFO

Keywords:
Acoustic-based sensing
Recognition and tracking
Localization
User authentication
Short-range communication

ABSTRACT

With advancements of wireless and sensing technologies, recent studies have demonstrated technical feasibility and effectiveness of using acoustic signals for sensing. In the past decades, low-cost audio infrastructures are widely-deployed and integrated into mobile and Internet of Things (IoT) devices to facilitate a broad array of applications including human activity recognition, tracking, localization, and security monitoring. The technology underpinning these applications lies in the analysis of propagation properties of acoustic signals (e.g., reflection, diffraction, and scattering) when they encounter human bodies. As a result, these applications serve as the foundation to support various daily functionalities such as safety protection, smart healthcare, and smart appliance interaction. The already-existing acoustic infrastructure could also complement RF-based localization and other approaches based on short-range communications such as Near-Field Communication (NFC) and Quick Response (QR) code. In this paper, we provide a comprehensive review on acoustic-based sensing in terms of hardware infrastructure, technical approaches, and its broad applications. First we describe different methodologies and techniques of using acoustic signals for sensing including Time-of-Arrival (ToA), Frequency Modulated Continuous Wave (FMCW), Time-Difference-of-Arrival (TDoA), and Channel Impulse Response (CIR). Then we classify various applications and compare different acoustic-based sensing approaches: in recognition and tracking, we review daily activity recognition, human health and behavioral monitoring hand gesture recognition, hand movement tracking, and speech recognition; in localization and navigation, we discuss ranging and direction finding, indoor and outdoor localization, and floor map construction; in security and privacy, we survey user authentication, keystroke snooping attacks, audio adversarial attacks, acoustic vibration attacks, and privacy protection schemes. Lastly we discuss future research directions and limitations of the acoustic-based sensing.

1. Introduction

Many real-world applications in smart home and office environments, such as security surveillance and protection, smart healthcare, and smart appliance interaction, use sensing technologies. For example, sensing technologies based on camera, Radio-Frequency (RF), and acoustic, have been utilized in human activity recognition and tracking, localization and navigation, and security monitoring, etc. Among them, camera-based sensing [1] is the most popular. For non-intrusiveness and convenience, there are a number of studies in RF-based techniques (e.g., RFID [2], WiFi [3]). RFID is widely applied in business functions, such as storage and supply chain management [4], due to its battery-free feature. Moreover, with the wide availability of indoor WiFi infrastructures, WiFi-based sensing techniques have shown great potential in implementing sensing functions in a smart home or office. Besides of camera- and RF-based solutions, acoustic signal provides another dimension for sensing, because microphones and speakers are now widely equipped in mobile and wearable devices, as well as smart appliances. For example, Galaxy Note 3 and Amazon Echo are integrated with multiple advanced microphones for noise elimination and far-field voice pickup. Galaxy S9 and Google Home have multiple speakers for stereo playback. Moreover, many mobile devices have begun to support high recording capability (e.g., 192 kHz) targeted at audiophiles, resulting in a significant enhancement of acoustic-based sensing capabilities.

Surveillance cameras may raise privacy concerns, especially at home and office environments. Moreover, camera-based sensing is highly dependent on the environmental lighting conditions, making it hard to work under conditional of low illumination, smoke, and opaque

* Corresponding author.
E-mail addresses: yang.bai.ece@rutgers.edu (Y. Bai), luli_jtu@sjtu.edu.cn (L. Lu), jcheng18@nyit.edu (J. Cheng), jliu@utk.edu (J. Liu), yingche@scarletmail.rutgers.edu (Y. Chen), jiadiyu@sjtu.edu.cn (J. Yu).

https://doi.org/10.1016/j.comnet.2020.107447
Received 11 April 2020; Received in revised form 21 June 2020; Accepted 25 July 2020
Available online 1 August 2020
1389-1286/Published by Elsevier B.V.

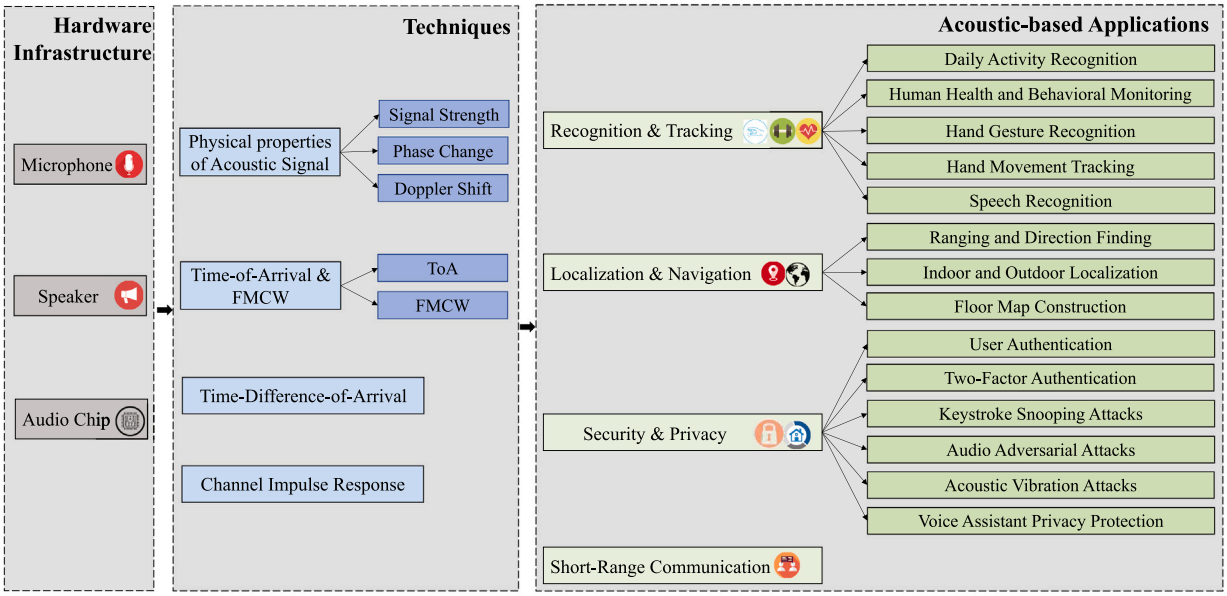

Fig. 1. A general structure for acoustic-based sensing and applications.

obstructions. In contrast, acoustic-based sensing does not have privacy issues with images and is deployable in less favorable conditions.

RFID-based solutions usually require high-cost RFID readers (i.e., >$100) and RFID tags that need to be attached to human bodies, thereby largely limiting their application scenarios. In comparison, acoustic-based sensing is more accessible due to the wide deployment of speakers and microphones on off-the-shelf mobile devices, without additional cost or rapid deployment request. In addition, due to the long-range and good penetration properties of WiFi signals, WiFi-based sensing can capture human activities in a large area and can even detect people moving behind walls [5]. However, due to the long sensing range, WiFi-based sensing can be easily interfered by surrounding environments (e.g., people in motion or change of furniture's positions). Although the coverage area of acoustic-based sensing is limited due to the fast attenuation of acoustic signals, it is more resilient to surrounding movements and the environmental change. This makes acoustic-based sensing a valuable complement to other sensing methodologies in many application scenarios.

Acoustic-based sensing and its applications have a long history. In 1900s, people started using SOund NAvigation Ranging (Sonar) to detect objects under water [6]. Acoustic-based sensing has been applied in the B-ultrasonic examination [7] to create images of internal body structures such as tendons, muscles, and internal organs. Furthermore, acoustic-based sensing solutions can transmit messages [8]. In recent years, with the omnipresence of mobile and wearable devices and the advancement in audio chips, acoustic signals have become easily accessible with high quality and for a wide range of sensing and communication functions. In this paper, we provide a comprehensive survey of acoustic-based sensing technologies and their application domains.

Some existing surveys reviewed different applications of acoustic-based sensing, such as localization [9] and communication [10]. Some other surveys concentrated on analyzing the applications enabled by different kinds of signals (e.g., camera, RF, acoustic). The examples are activity recognition [11,12], localization [13], and user authentication [14]. Different from the existing surveys, our survey aims to provide a more comprehensive understanding of acoustic-based sensing, including hardware infrastructure, main sensing techniques, and its broad applications, as shown in Fig. 1.

1.1. *Hardware infrastructure*

Mobile devices and smart appliances are equipped with audio infrastructures with high definition audio capabilities. The hardware infrastructure, including microphones, speakers, and audio chips, is used to record, generate, and process acoustic signals, respectively. With recent audio-based virtual control software (e.g., Siri, Google Assistant, Alexa Voice Service), there is an increasing demand for noise cancellation and far-field voice pickup. As a result, off-the-shelf smartphones and smart appliances are equipped with several microphones. Generally there are two speakers in a smartphone: an internal one for regular phone calls, and an external one for louder playback. Several smart devices (e.g., Galaxy S9, Google Home) even have stereo external speakers to improve our hearing experience in watching films or listening to music. Moreover, with the development of audio chips, many mobile devices began to support high recording and playback capability (e.g., 192 kHz) targeted at audiophiles. The wide availability of microphones and speakers provide a great chance for acoustic-based sensing, while audio chips with better quality result in a significant enhancement of sensing accuracy.

1.2. *Techniques*

After acoustic signals are captured, the next step is to characterize them with various sensing techniques, such as signal strength variation, phase change, Doppler shift, Time-of-Arrival (ToA)/ Frequency Modulated Continuous Wave (FMCW), Time-Difference-of-Arrival (TDoA), and Channel Impulse Response (CIR). An acoustic signal can be decomposed into a series of sine waves with different amplitudes, phases, and frequencies. By measuring the differences among these basic properties, such as signal strength variation, phase change, and Doppler shift, the movements of objects (e.g., human, phone) can be identified for various sensing applications, such as daily activity recognition [15–20], hand gesture recognition [30–39], hand gesture tracking [42–44,46,47], localization [59,60,70,71], user authentication [97–103], and keystroke snooping attacks [117–119]. In addition, environmental sounds (e.g., ambient noise, acoustic emanations of human activities) can be characterized by these basic properties to enable two-factor authentication [108,109] and keystroke snooping attacks [114–116]. In communication systems [145–162], signal strength, phase, and frequency are also used for data modulation, i.e., amplitude, phase, and frequency shift keying.

Table 1
Acoustic-based applications and their main techniques.

Section #	Applications	Main Techniques				
		Physical properties of Acoustic Signal	ToA & FMCW	TDoA	CIR	Other Techniques
Section III	Daily Activity Recognition	[15–20]	[21]	–	–	[22,23]
	Human Health & Behavioral Monitoring	[24,25]	[26,27]	[28,29]	–	–
	Hand Gesture Recognition	[30–39]	[40]	–	[41]	–
	Hand Movement Tracking	[42–47]	[48–50]	[51]	[52]	–
	Speech Recognition	–	–	–	–	[53–58]
Section IV	Ranging & Direction Finding	[59,60]	[61–67]	[68,69]	–	–
	Indoor & Outdoor Localization	[70,71]	[72–86]	[87–90]	–	–
	Floor Map Construction	–	[91–95]	–	–	–
Section V	User Authentication	[96–103]	–	–	[104]	[105–107]
	Two-Factor Authentication	[108–111]	–	–	–	[112,113]
	Keystroke Snooping Attacks	[114–120]	[121,122]	[123]	–	[124,125]
	Acoustic Adversarial Attacks	–	–	–	[126]	[127–131]
	Audio Vibration Attacks	–	–	–	–	[132–135]
	Voice Assistant Privacy Protection	–	–	–	–	[136–144]
Section VI	Short-Range Communication	[145–162]	–	–	–	–

Among sensing technologies, ToA measures the propagation time of a signal traveling between a transmitter and a receiver. Combined with the propagating velocity of acoustic signals (e.g., 344 m/s for most conditions), the propagation distance can be derived. Therefore, ToA can serve as a key enabler of localization applications [72–86]. To get an accurate measurement, ToA needs precise synchronization between transmitters and receivers. Such a requirement is difficult to achieve, and hence hinders the deployment of ToA on commercial devices. In contrast, FMCW leverages the geometry relationship of chirp signals and derives the frequency difference between transmitted and received signals to quantify the ToA. Without the need for synchronization of commercial devices, FMCW becomes more attractive in practice and has been applied in many applications requiring fine-granularity, such as health monitoring [26,27], human movement tracking [50], and floor-map construction [95].

With the goal to relax the necessity of device synchronization, TDoA is proposed to measure the time difference in arrival times between the signals received by a pair of microphones. Presently smart devices, such as smartphones, smart appliances, and IoT devices, are usually equipped with multiple microphones for dual recording, noise cancellation, etc. This facilitates the usage of TDoA in various domains, such as driving behavior monitoring [28,29], hand movement tracking [51], localization [68,69,87–90], and user authentication [104]. Moreover, instead of directly quantifying received acoustic signals, recent investigations [41,52] analyze the CIR using channel estimation (e.g., the least square channel estimation) to capture subtle variations caused by human body movements. These techniques have been used in various functionalities, such as hand gesture recognition [41] and hand gesture tracking [52]. We will discuss them in detail in Section 2.

1.3. Acoustic-based applications

With these fundamental technologies, a broad range of applications are developed to improve the quality of our daily life and work efficiency, as summarized in Fig. 1. We classify these applications into four categories and discuss them in later sections: *Recognition and Tracking* in Section 3, *Localization and Navigation* in Section 4, *Security and Privacy* in Section 5, and *Short-range Communication* in Section 6.

Recognition and Tracking includes daily activity recognition, human health and behavioral monitoring, hand gesture recognition, hand movement tracking, and speech recognition. By tracking daily activities (e.g., walking, sitting, and cooking) of a person [17,20–23], it is possible to evaluate his or her daily routine and life style, and lead to other related application areas such as elder care and fitness tracking. In comparison, vital signs of human (e.g., breathing, heartbeat) [24,26,27] are finer-grained and can serve as an important indicator of a person's sleep quality, health conditions, and stress level. Besides daily and health-related applications, inattentive driving behavior detection (e.g., phone usage, eating) [15,16,28,29] is important for driving safety. In hand gesture recognition (i.e., hand gesture [31–34], finger gesture [36]), interaction interfaces on smart and mobile devices are enabled. Compared with hand gesture recognition, hand movement tracking is more flexible for human–computer interaction applications, such as remote controllers [50] and virtual keyboards [44,45]. Furthermore, speech recognition have been widely applied in healthcare, marketing, and Internet of Things (IoT), etc. We will review speech recognition process and the commonly used models [53–55].

Localization and Navigation discusses related functions and applications in ranging and direction finding [59–69], indoor and outdoor localization [70–90], and floor map construction [91–95]. Ranging and direction finding serve as the basis of localization and navigation applications, such as face-to-face multi-user gaming and in-car phone use detection. Indoor and outdoor localization are critical for location-based applications such as target localization, store advertisement, inventory management, animal habitat tracking, etc. Moreover, Floor map construction is needed for indoor navigation, VR/AR, construction, facility management, etc.

Security and Privacy investigates both defense and attack mechanisms leveraging acoustic-based sensing techniques in security and privacy applications. Security-related applications include user authentication [96–104,108–113], keystroke snooping attacks [114–125], and audio adversarial attacks [126–131]. Acoustic-based user authentication functions are proposed to safeguard security in mobile devices. In addition to the first round of defense in smart and mobile devices by user authentication, other studies also employ acoustics (e.g., ambient noise, pre-defined acoustic signals) as the second round of the authentication process. On the other hand, acoustic signals could be used to snoop keyboard keystrokes [114–116,120–124] and eavesdrop touch-screen patterns and keystrokes [117–119]. These activities raise security concerns in users. This section also introduces privacy-related applications, such as acoustic vibration attacks [132–135] and voice assistant privacy protection [136–143].

In addition to sensing applications, there are studies in acoustic-based *short-range communication* including audible communication [145–150], near-ultrasonic communication [151–153], ultrasonic communication [154,162], and liquid-based medium acoustic communication [155–158]. Active studies employ audible signals for communication. Acoustic signals are noisy and unpleasant for human. Near-ultrasonic signal (i.e., 18 ∼ 20 kHz) and ultrasonic signal (i.e., > 20 kHz) are outside of the range of most human perceptions. As a result, researchers develop near-ultrasonic and ultrasonic communication for a better user experience. Due to the satisfactory propagation property,

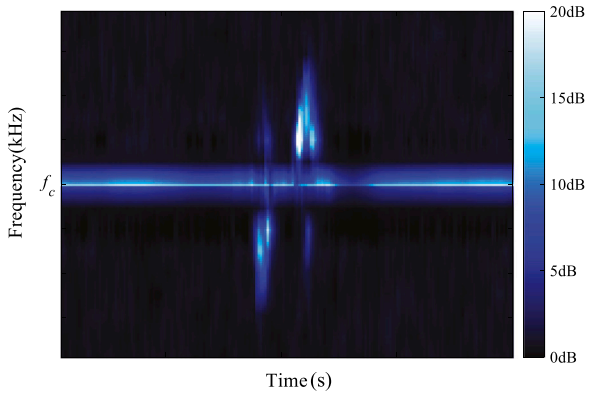

Fig. 2. Illustration of Doppler shift.

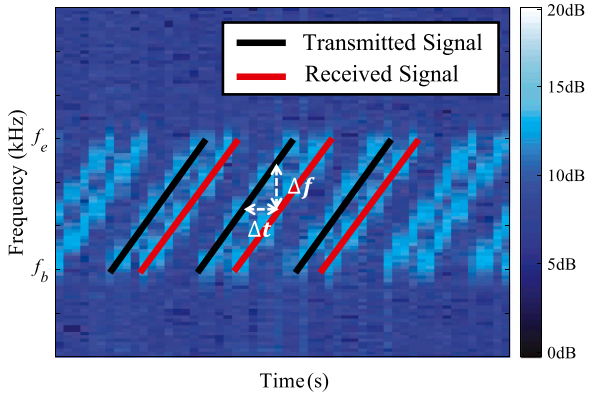

Fig. 3. Illustration of FMCW technique.

acoustic communication is one of the most effective communication techniques in liquid-based environments.

The related work for each application category is summarized in Table 1. The rest of this paper is organized as follows. Section 2 reviews three key techniques in acoustic applications. Section 3 introduces four categories of acoustic-based applications with an emphasis on activity recognition and tracking. Section 4 describes investigations in localization and navigation. Section 5 covers work in security and privacy. Section 6 discusses acoustic-based short-range communication. Section 7 presents the limitations of existing work and future research direction. Section 8 concludes this survey paper.

2. Techniques for acoustic-based sensing and applications

In this section, we introduce key techniques used in acoustic-based sensing and applications, including signal strength variation, phase change, Doppler shift, Time-of-Arrival (ToA) & Frequency Modulated Continuous Wave (FMCW), Time-Difference-of-Arrival (TDoA), and Channel Impulse Response (CIR).

2.1. Physical properties of acoustic signal

According to the principle of Fourier series, any signal can be decomposed into a series of sine waves. Therefore, the physical properties of an acoustic signal can be represented by three parameters in a sine wave, i.e., amplitude, phase, and frequency. Researchers have exploited these properties of acoustic signals for sensing.

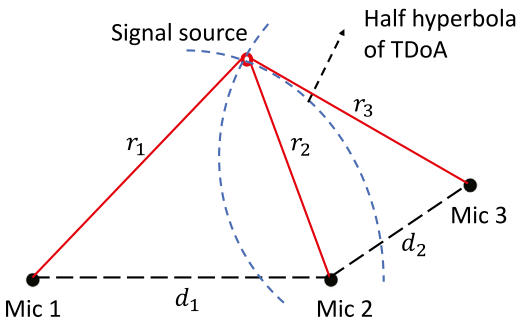

Fig. 4. Illustration of TDoA technique.

2.1.1. Signal strength

The signal strength in an acoustic signal is associated with the power level and easy to measure. Due to propagation dispersions and medium absorptions, the signal strength of acoustic signals would attenuate as the signal propagates through a distance. Assume an acoustic signal is first emitted with a signal strength I_e. After propagating over a distance d, the final signal strength I_r is given by

$$I_r = I_e \frac{k}{d} e^{\alpha d}, \tag{1}$$

where k is a normalization coefficient, α is the attenuation coefficient affected by the frequency of acoustic signals, temperature, relative humidity, atmosphere, etc. Based on Eq. (1), the distance d can be calculated from I_r, I_e, k, and d.

Measured by commercial devices as the amplitude of acoustic signals, signal strengths serve as the basis of many acoustic-based sensing applications, such as daily activity recognition [19], health monitoring [24,26], hand gesture recognition [37], indoor localization [71,83], and keystroke snooping [114–116]. However, the measurements can be greatly impacted by ambient noises. As a result, applications only using signal strength cannot achieve robust and satisfactory performance. To handle this, several studies combine signal strength with ToA [24,26, 83] or phase change [71] for practical applications.

2.1.2. Phase change

In addition to the signal strength, the phase change induced by the propagation of acoustic signals can also support the acoustic-based sensing. For a single-frequency acoustic signal, the phase shift is related to the time difference. Specifically, the phase shift $\Delta\phi$ can be represented as

$$\Delta\phi = 2\pi f \Delta t, \tag{2}$$

where f and Δt are the frequency and time shifts of acoustic signal, respectively. The phase value is periodic from 0 to 2π, so that Δt can only be accurately derived within the cycle time $1/f$. In other words, only the distance within the wavelength of acoustic signal λ (i.e., c/f) can be precisely measured, where c is the propagation speed of an acoustic signal.

Early studies [43,47] show the feasibility of using phase changes modulated by Orthogonal Frequency-Division Multiplexing (OFDM) and Continuous Wave (CW) to carry out fine-grained finger movement tracking. After that, another work [118] employs the phase changes of acoustic signals induced by finger movements to eavesdrop keystrokes on touch screens. Due to the periodicity of a sine wave, the phase change is probable to be the same when the moving distance exceeds a wavelength of the acoustic signal. As a result, without prior knowledge of starting and ending points of a movement, using phase changes alone can hardly achieve precise tracking.

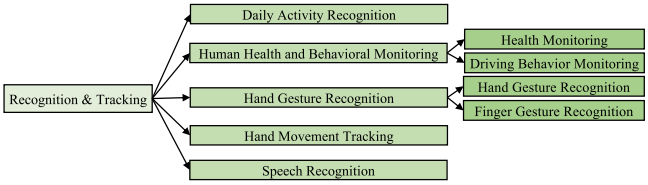

Fig. 5. Structure for acoustic-based recognition and tracking applications.

2.1.3. Doppler shift

The Doppler effect is the frequency shift detected by a moving observer relative to a signal source. Specifically, when an object moves at a velocity v along with a direction θ from the receiver, it will result in a Doppler shift Δf as

$$\Delta f = \frac{2v\cos(\theta)}{v_s} f_c, \tag{3}$$

where v_s is the velocity, and f_c is the pilot frequency of acoustic signals. From Eq. (3), v can be calculated from the measurements of Δf, f_c, θ, and v_s. Fig. 2 shows an example of the Doppler shift associated with a hand moving back and forth in front of a smartphone equipped with both a speaker and a microphone. It can be seen that when the hand moves back to the microphone (i.e., the receiver), the frequency of acoustic signal shifts toward the negative direction, and vice versa.

Also from Eq. (3), with the estimated Δf and the measured v_s, the distance and location of objects can be derived. Researchers utilize Doppler shifts to recognize hand gestures [31–34], monitor driving behavior [15,16], and track hand movements [42,44,50].

2.2. Time-of-Arrival (ToA) and Frequency Modulated Continuous Wave (FMCW)

ToA is the propagating time of a signal traveling between a transmitter and a receiver. Combined with the propagating velocity v of acoustic signals, the propagating distance d can be calculated as, $d = v \times ToA$, where $v = 344$ m/s in most situations. This method needs the transmitter and receiver to be precisely synchronized to avoid measurement error.

ToA has become a key enabler of measuring distances between objects, and is widely implemented to support various applications, such as gesture recognition [40], hand movement tracking [48], localization [72–86], and floor map construction [91–95]. Usually transmitters and receivers are not installed on the same device. To synchronize them, researchers use customized infrastructures (e.g., RF [74], onboard vision sensor [76], LED light [77], and specialized sonar [75]) with increased deployment cost.

FMCW maps time difference to frequency shift in order to measure the ToA without the synchronization requirement. The transmitted and received signals in frequency domain under a FMCW-based application are shown in Fig. 3. The transmitter keeps transmitting a chirp signal sweeping the frequency from f_b to f_e. By comparing the frequencies of the transmitted and the received signals, frequency shift Δf could be measured. The slope of the chirp signal can be determined with the frequency band B (i.e., $f_e - f_b$) and unit duration τ, i.e., $slope = B/\tau$. From the geometry similarity principle of a triangle, ToA can be derived as

$$ToA = \frac{\Delta f \times \tau}{B}. \tag{4}$$

Based on the ToA, the propagating distances of acoustic signals can be measured.

Without the necessity of synchronization, FMCW could achieve satisfactory performance in detecting minute movements. Many investigations employ FMCW in various compelling applications, such as monitoring sleep quality [26] and heartbeats [27], and constructing floor map [95].

2.3. Time-Difference-of-Arrival (TDoA)

Due to measurement errors in commercial devices, it is difficult to realize perfect clock synchronization for accurate measurements of ToA. Other than FMCW, TDoA measures the time difference in arrival times of signals received by a pair of receivers, without the synchronization need. At least two microphones are equipped on off-the-shelf mobile and IoT devices. This can facilitate the implementation of TDoA. TDoA locates a target at intersections of hyperbolas or hyperboloids that are generated with foci at each fixed receiver of a pair. As shown in Fig. 4, Mic 1 and Mic 2 is a pair of foci with distance d_1. By measuring the TDoA between these two microphones, the difference in the distances to the microphones, $r_1 - r_2$, can be obtained. The signal source is located on the half hyperbola defined by these parameters. Similarly, the signal source is on the other half hyperbola defined by the associated pair of microphones (i.e., Mic 2 and Mic 3). Therefore, the signal source is on the intersection of these two half hyperbolas.

Efforts have been put into leveraging TDoA to realize driving behavior monitoring [28,29], hand movement tracking [51], and indoor localization [88,89]. Another work [123] even employs TDoA to accomplish mm-level audio ranging for the side-channel attack on eavesdropping user inputs on keyboards.

2.4. Channel Impulse Response (CIR)

CIR represents an acoustic signal's propagation in response to the combined effect of scattering, fading, and power decay of transmitted signals. When an acoustic signal $S(t)$ is transmitted with a speaker, it propagates through multipath and is received by a microphone as $R(t)$. Let $h(t)$ denote the CIR of the acoustic signal's propagating channel. The relationship between the transmitted and received signals is

$$R(t) = S(t) * h(t), \tag{5}$$

where $*$ is a convolution operator. Due to the discrete representation of received acoustic signals, Eq. (5) is represented as $R[n] = S[n] * h[n]$ in real application scenarios. To resolve $h[n]$, Least Square (LS) channel estimation [163] is commonly used because of the low computational complexity. Specifically, the speaker first transmits a known signal $m = [m_1, m_2, \ldots, m_N]$, and the microphone receives the reflected signals $y = [y_1, y_2, \ldots, y_N]$, where N is the length of transmitted acoustic signal. A circulant training matrix M is generated by the vector m. The dimension of M is $P \times L$, i.e., the vector m circulates P times and constructs the P rows of M. The CIR h is estimated as

$$h = (M^H M)^{-1} M^H y_L, \tag{6}$$

where $y_L = [y_{L+1}, y_{L+2}, \ldots, y_{L+P}]$. In channel estimation, the length of CIR L and the reference P should be determined manually to satisfy the constraint of $N = P + L$. Note that increasing L will estimate more channel states but reduce the reliability of estimation.

Object movements can be captured from the analysis of the changes in CIR. However, similar to directly using received signals, CIR also introduces the surrounding information (e.g., people walking by) of the sensing areas. The fine-grained sensing capability of CIR-based approaches magnifies the interference of device movements [41].

3. Recognition and tracking

Human activity recognition and tracking are the critical functions to support a broad range of applications including security monitoring, vital signs management, human–computer interaction, and elder care. Acoustic-based approaches can be divided into five categories: daily activity recognition, human health and behavioral monitoring, hand gesture recognition, hand movement tracking, and speech recognition. In this section, we discuss related research in these categories, as shown in Fig. 5.

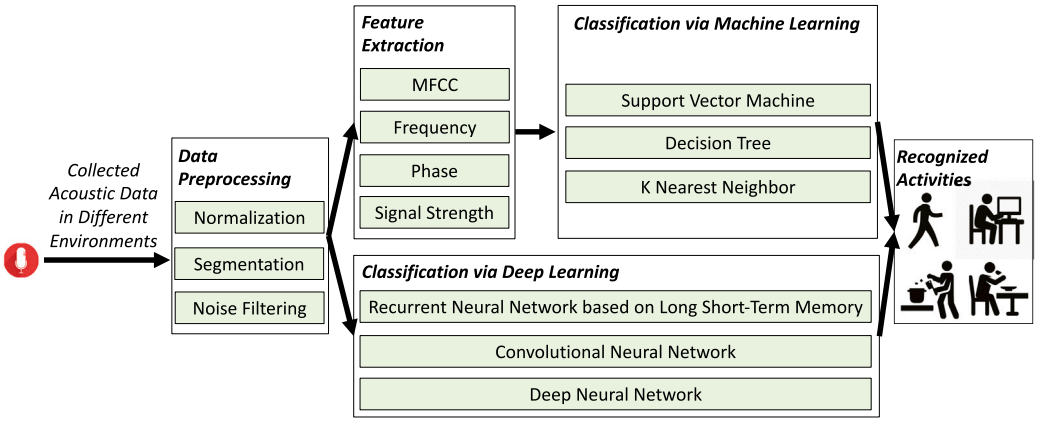

Fig. 6. Workflow of daily activity recognition.

Table 2
Comparison of acoustic-based daily activity recognition work.

Work	Feature	Accuracy	Device	Applied algorithm
Stork et al. [20]	MFCC	94%	Robot	Random Forest
BodyScope [19]	MFCC	71.5%	Modified headset	SVM
SoundSense [17]	Phase, Signal Strength, Frequency, and Bandwidth	around 90%	Commercial devices	Decision Tree & Markov Model
EI [22]	N/A	around 78%	Commercial devices	CNN
Akira et al. [23]	N/A	around 94%	Commercial devices	RNN-LSTM

3.1. Daily activity recognition

Our daily activities include walking, sitting, cooking, eating, etc. Tracking these activities aims to help a person live safer and healthier. As shown in Fig. 6, daily activity recognition systems are built to solve classification problems. A unique source of information for these recognition systems is the sounds produced by daily physical activities. We present two typical flowcharts of daily activity recognition systems, which are machine-learning and deep-learning based respectively. For both approaches, the first step is data preprocessing, which is designed for eliminating the environmental interference and normalizing the data for future use. For traditional machine learning approaches, manually extracted features are required as inputs for training a robust model. BodyScope [19] is developed as a wearable activity recognizer based on commercial headsets to monitor mouth movements (e.g., eating, laughing, and speaking). This work extracts acoustic features (i.e., Mel-Frequency Cepstrum Coefficient (MFCC)) from captured sounds and classifies them into specific categories by feeding the features into Support Vector Machine (SVM). MFCC is a representation of the short-term power spectrum of an acoustic signal. Stork et al. [20] employ MFCC feature and Random Forest algorithm for twenty two categories in daily activities. However, these systems utilize customized infrastructure, i.e., modified headsets [19] and robots [20], thus have difficulties in deployments. SoundSense [17] monitors daily activities (e.g., walking, driving, riding a bus) using off-the-shelf mobile devices. This work extracts acoustic features (i.e., phase, signal strength, frequency, and bandwidth) from captured sounds and classifies them into specific categories using decision trees and the Markov models. However, feature extraction highly relies on human knowledge and experience. Deep learning tends to overcome these limitations, which automatically learn the features through the network for model training. EI [22] leverages acoustic signals reflected by surrounding objects to implement an environment-independent activity recognition method. This work builds a novel adversary network to extract a representation of received signals. It can remove the uniqueness of different environments and individuals to predict activities under unseen environments. Another work [23] uses acoustic and acceleration signals as input, and employs Recurrent Neural Network (RNN) based on Long Short-Term Memory (LSTM) for human activity classification. Acoustic-based daily activity recognition studies are summarized in Table 2.

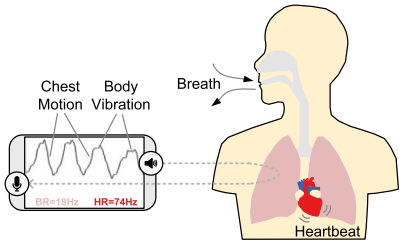

Fig. 7. Illustration of vital sign monitoring by sensing the body movements leveraging acoustic signals [27].

3.2. Human health and behavioral monitoring

Health-related activities and driving behavior monitoring are two vital topics for evaluating a person's health conditions (e.g., sleep quality, stress level) and protecting his/her safety. We introduce acoustic-based studies in these two areas.

Health Monitoring. Health-related activities (e.g., breathing, heartbeats) result in minute body movements. Due to the non-intrusiveness nature, acoustic-based monitoring has attracted much research attention. Specifically, FMCW has been used to extract acoustic reflections from human bodies to capture minute movements. With this technology, a few existing studies (e.g., [26,27]) can capture human's breathing and heartbeats. For instance, Nandakumar et al. [26] can detect sleep apnea events. Qian et al. [27] design a heartbeat monitoring system, Acoustic cardiogram, as shown in Fig. 7, which extracts signal phases of received FMCW signals to capture more fine-grained movements induced by heartbeats. Subsequent studies incorporate machine learning techniques to classify various activities. For example, Ren et al. [24] extract acoustic features (i.e., MFCC) and employ SVM to differentiate sleep events. BreathListener [25] extracts energy spectrum

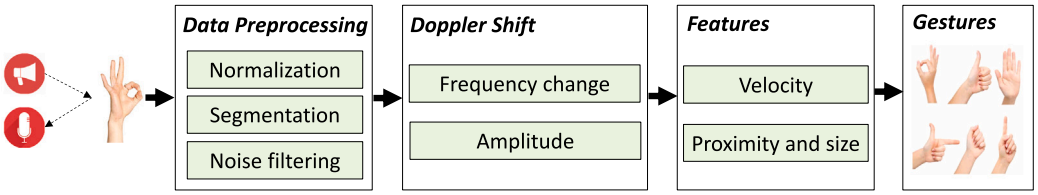

Fig. 8. Illustration of Doppler shift-based hand gesture recognition.

density of acoustic amplitude and uses Convolutional Neural Network (CNN) to recover fine-grained breathing waveforms in driving environments. All of these studies employ just commercial smart devices without the need for any supplemental infrastructures.

Driving Behavior Monitoring. With no wireless signal (e.g., WiFi) coverages inside moving vehicles, it is natural to use acoustic-based technologies to monitor driving behaviors. Yang et al. [28,29] take advantage of speakers in vehicles to detect phone usage of drivers. To do this, the two studies measure the TDoA with assistance of counting the number of acoustic samples between two beeps to detect the location of a phone in a vehicle. Besides making phone calls, inattentive driving behaviors (e.g., leaning forward to pick up from floor, eating, or drinking) are another type of common dangerous driving behaviors. Recently, ER [15,16] reveals that these activities can be captured through Doppler shifts of acoustic signals. ER sets up a gradient model forest, which contains multiple classifiers to recognize different driving behaviors and detect inattentive driving behaviors before they are 50% finished, in order to issue earlier warning alerts.

3.3. Hand gesture recognition

Hand gestures of humans, such as hand movements and finger motions, are crucial interactions between users and smart devices. To improve users' experience, many device-free gesture recognition approaches are proposed. Among them, acoustic-based gesture recognition attracts enormous attentions, because inexpensive audio equipment is widely available.

Hand Gesture Recognition. Human interacting with smart devices is primarily by hands, which are often in constant motion. Doppler shift is one of the most natural and straightforward methods to recognize hand gestures. This type of systems work in four steps: data preprocessing, Doppler extraction, physical features exhibition, and gesture recognition, as shown in Fig. 8. In addition to the frequency shift used to derive the velocity as described in Section 2.1.3, the amplitude of the received signal can be combined to detect the proximity between target object and acoustic source, as well as the size of the object. Early work [30] develops a customized device consisting of three microphones and one speaker to process Doppler shifts of ultrasonic signals. Inspired by this work, many Doppler-based systems [31–34] are proposed. SoundWave [31] first explores audio components in most commercial off-the-shelf devices to recognize in-air hand gestures. Doplink [32] and Airlink [33] extend Doppler shifts of acoustic signals to realize multi-device interactions (e.g., rapid device pairing and file transferring). AudioGest [34] uses only one speaker and one microphone to carry out a multi-level gesture recognition and derives extensive information, such as in-air time, average moving speed and waving range.

In addition, Point & Connect (P&C) [40] implements a device pairing system leveraging chirp signals, in which a user only needs to make a simple hand gesture toward the target for device pairing. P&C measures the distance change between a user and candidate devices by estimating the ToAs of the acoustic signals with synchronized clocks. And the target device is the one with the maximum distance deduction. However, P&C requires an initial wireless channel of communication to enable the connection between the source device and the target device. To overcome this limitation, Spartacus [35] exploits Doppler shift to initiate interactions. When the user makes a gesture pointing directly to the target device, the maximum frequency shift can be observed.

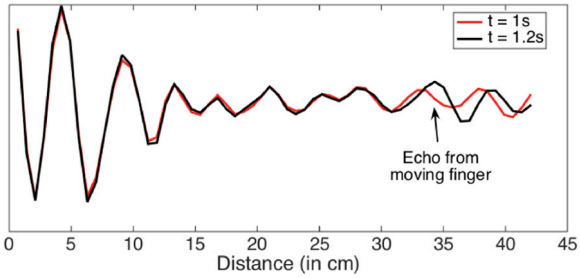

Fig. 9. OFDM echo profiles at two time points [47].

Finger Gesture Recognition. In comparison to relatively distant and coarse-grained hand gestures, closer and finer-grained finger movements are becoming popular in human–computer interactions. Hence, there is much active research in this area. UbiK [36] enables ubiquitous surfaces as keyboards for mobile devices by identifying finger gestures from received acoustic signals. Specifically, UbiK transmits chirp signals and receives signals reflected by keystroke fingers. From the received signals, pattern-related information (i.e., the amplitude spectrum density) is extracted from the frequency domain of acoustic signals, and finger gestures can be recognized by comparing the Euclidean distance of amplitude spectrum densities between extracted features and profiles. Other systems (e.g., [37–39]) use acoustic signals in recognizing handwriting to extend the input interface of human–computer interactions. SoundWrite [37] and SoundWrite II [38] leverage amplitude spectrum density and acoustic features (i.e., MFCC) to characterize handwriting features, and use K-Nearest Neighbors (KNN) to match the captured features with labeled features in the database. Furthermore, WordRecorder [39] extracts the spectrogram of acoustic signals reflected by a writing hand and feeds it to a designed CNN to identify written words. More recently, UltraGesture [41] utilizes CIR together with a deep learning method, i.e., CNN network, to achieve a mm-level gesture recognition, which outperforms Doppler-based and FMCW-based solutions.

3.4. Hand movement tracking

Hand movement tracking provides more flexible capability to support various human–computer interaction applications, such as remote controller, virtual keyboard, and Virtual Reality (VR) gaming. Acoustic-based hand movement tracking systems are summarized in Table 3.

Echoloc [48] produces two-channel chirp signals using a smartphone plugged with a stereo speaker to estimate the ToAs between a hand and two speakers. In this way, the position of the hand can be determined. Another work, EchoTrack [51], estimates a hand trajectory with the assistance of measuring the ToAs of acoustic signals received by two microphones in a smartphone, without any special hardware. This work combines Doppler shifts and ToAs to optimize the tracking accuracy. By calculating the speed of hand movement according to Eq. (3), the hand locations could be accurately estimated.

Table 3
Comparison of acoustic-based hand movement tracking work.

Work	Technique	Resolution	Specific hardware	Device
Echoloc [48]	ToA	5 cm for 2D	Yes	Smartphone, Tablet
EchoTrack [51]	TDoA & Doppler shift	3 cm for 2D	No	Smartphone, Tablet
AAMouse [42]	Doppler Shift	1.4 cm for 2D	Yes	Smartphone, Tablet
CAT [50]	FMCW & Doppler Shift	5–7 mm for 2D, 8–9 mm for 3D	Yes	Smartphone, Tablet
LLAP [43]	Phase Change	3.5 mm for 1D	No	Smartphone, Tablet
FingerIO [47]	OFDM	8 mm for 2D	No	Smartphone, Tablet
Strata [52]	CIR	1.0 cm for 2D	No	Smartphone, Tablet
VPad [44,45]	Amplitude & Doppler Shift	1.55 cm for 2D	No	Smartphone, Tablet, Laptop
Vernier [46]	Phase Change	4 mm for 2D	No	Smartphone

AAMouse [42] applies Doppler shifts of acoustic signals to track hand movements in real time. It sends inaudible signals at different frequencies and uses Doppler shifts to estimate the speeds and distances of hand movement. To improve the robustness, the system combines the estimations from different frequencies to perform outlier removal. CAT [50] further enhances the tracking accuracy by analyzing both FMCW and Doppler shift of acoustic signals. FMCW maps time difference to frequency shift, without the need for precise synchronization. Another two systems, LLAP [43] and FingerIO [47], only use commercial off-the-shelf smartphones to achieve mm-level motion tracking, without additional devices and synchronized clocks. LLAP [43] tracks hand movements via measuring the phase changes in received acoustic signals. Specifically, LLAP first measures the phase change of acoustic signal reflected by static hand as a reference. When the hand moves toward/backwards the sound source, the phase changes proportional to the ToA, according to Eq. (2). With the known ToA and wavelength l of acoustic signal (i.e., $l = \frac{c}{f}$, where c is the speed and f is the frequency of sound), the distance of hand movement could be derived. FingerIO [47] leverages the OFDM technique to realize fine-grained finger tracking. Two OFDM echo profiles at two different time instances are shown in Fig. 9. It can be seen that a shift in the wave peak due to the change in the arrival time of the echo when the finger moves from 34 cm to 35 cm to the microphone. Recently, Strata [52] employs CIR instead of raw received signals to take the multipath propagation into account. It first finds out which channel states give the best performance and then tracks the hand movement with phase changes of these certain channels. All of these studies are designed for mobile devices, such as smartphones and tablets. Due to the different layout of audio devices, most of them cannot be directly deployed on traditional laptops. To handle this issue, VPad [44,45] designs a tracking system for traditional laptops without touchscreens. It utilizes an energy feature and Doppler shift to track horizontal and vertical hand movements, respectively, and can put real-time and highly accurate tracking into effect. A more recent work, Vernier [46], introduces a novel method to calculate the phase change based on very few samples and improves the real-time tracking capability. Specifically, instead of calculating the phase change using FFT, Vernier calculates the phase change with a small window of signal, and the number of local maximum corresponds to the number of cycles of phase change.

3.5. Speech recognition

With speech recognition, machines can understand requests from human voices. Many emerging services enabled by this technique have been deployed to bring benefits to our daily life. For example, in workplaces, people can schedule meetings and print documents on voice requests. Drivers use speech recognition to control radios and make phone calls while keeping their sights on the roads. Furthermore, patients can ask for common symptoms of diseases and quickly find information from medical records. Many research efforts have been made in this area with the wide deployment of speech recognition. Speech recognition mainly requires several steps of processing, including signal processing, feature extraction, and recognition [164,165]. Signal processing mainly includes sampling (e.g., aliasing, filtering) and spectral analysis (e.g., framing, windowing). After that, the commonly used features, such as Linear Predictive Cepstral Coefficients (LPCC), MFCC, and Perceptual Linear Prediction (PLP), are extracted for recognition. The recognition step links the observed features of the speech signals with the expected phonetics of the hypothesis sentence.

Widely used recognition approaches are Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) [53,54] and Deep Neural Network-Hidden Markov Model (DNN-HMM) [55,56]. HMM is an extension of Markov chain that can provide a probability function of the potential results [166]. GMM, represented as a weighted sum of Gaussian probability density functions (PDFs), is used to determine how well each state of the HMM corresponds to the voice input. Over the past few years, advances in both machine learning algorithms and computer hardware have led to more efficient methods for training DNN. Thus, many studies (e.g., [55,56]) have been exploring DNN for speech recognition. Instead of using GMM, these studies use DNN to produce posterior probabilities over HMM states as outputs. Comparing DNN-HMM with GMM-HMM, the outputs are expanded from a small number of phonemes into a large number of them. Several studies [57,58] confirm that DNN-HMM model outperforms GMM-HMM model.

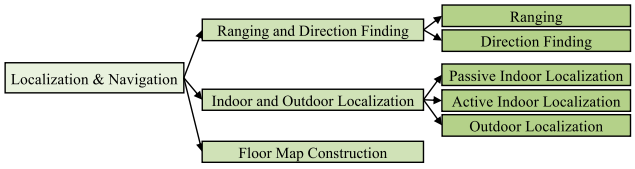

Fig. 10. Structure for acoustic-based localization and navigation applications.

4. Localization and navigation

Localization and navigation support a broad range of pervasive applications (e.g., digital map construction, mobile robot) and attract a lot of research efforts in past decades. With the properties of omni-directional propagation and strong diffraction, acoustic signals are reliable for locating people in low visibility environments (e.g., dark, foggy, or dusty conditions). In this section, we discuss related research in three categories: ranging and direction finding, indoor and outdoor localization, and floor map construction, as shown in Fig. 10.

4.1. Ranging and direction finding

Ranging and direction finding focus on measuring distance and angle between a signal source and a target object. They are the basic techniques of localization and navigation. Many efforts have been put into developing device-to-device ranging and direction finding solutions to support various applications, such as navigation, face-to-face multi-user gaming, and the driver's phone detection.

Ranging. Early studies such as Scott et al. [61] and Lopes et al. [62] carry out the ranging functions with ToA and TDoA techniques. Both research demand accurate synchronized clocks for ToA measurement and can only realize limited ranging resolutions (i.e., 15 cm [61]

Table 4
Comparison of acoustic-based localization work.

Work	Category	Technique	Resolution	Indoor/Outdoor	Infrastructure	Frequency Band
Guoguo [74]	Passive	ToA	0.25 m	Indoor	Infrastructure-based	15–20 kHz
CondioSense [75]	Passive	ToA	N/A	Indoor	Infrastructure-based	12–23 kHz
Haddad et al. [78]	Passive	ToA & TDoA	0.2 m	Indoor	Infrastructure-free	14–20 kHz
Sundar et al. [79]	Passive & Active	TDoA	0.3 m	Indoor	Infrastructure-free	N/A
WalkieLockie [80,81]	Passive	ToA	0.63 m	Indoor	Infrastructure-based	17–24 kHz
Sanchez et al. [88]	Passive	TDoA	0.1 m	Indoor	Infrastructure-based	N/A
Swadloon [60]	Passive	Doppler Shift	0.92 m	Indoor	Infrastructure-based	20 kHz
SITE [70]	Passive	Doppler Shift	0.42 m	Indoor	Infrastructure-based	17–19.5 kHz
Beep [82]	Active	ToA	0.6 m	Indoor	Infrastructure-free	4.01 kHz
Qiu et al. [83]	Active	ToA	0.14 m	Indoor	Infrastructure-free	2–4 kHz
Liu et al. [84]	Active	ToA	1–2 m	Indoor	Infrastructure-free	16–20 kHz
EchoTag [71]	Active	Signal Strength & Phase Change	0.01 m	Indoor	Infrastructure-free	11–22 kHz
AMIL [87]	Active	TDoA	0.5 m	Indoor	Infrastructure-based	18–20 kHz
Murakami et al. [89]	Passive & Active	TDoA	0.13 m	Indoor	Infrastructure-based	14.75–18.25 kHz
Auto++ [90]	Passive	ToA & Doppler Shift	meters	Outdoor	Infrastructure-free	0–4 kHz
Pinna et al. [86]	Passive	ToA	1.5 m	Outdoor	Infrastructure-based	3–7 kHz

and 40 cm [62]). Several investigations [64–67] develop acoustic-based ranging systems and put centimeter-level resolution into effect. However, the need for customized hardware infrastructures restricts the wide deployment of these systems. Without clock synchronization and specialized hardware, BeepBeep [63] designs an algorithm called Elapsed time between the two Time Of Arrivals (ETOA) to precisely measure the distance between devices. Each of the two peer devices transmits an acoustic signal, and starts recording from its microphone simultaneously. ETOAs are calculated by two devices individually, and the distance between devices can be derived by exchanging the time duration information with its peer. This work can accomplish a ranging resolution of 4 cm. We refer the update rate as the number of times in a second that the system updates its distance measurements. The update rate of BeepBeep is arbitrarily low because the two devices are not synchronized so that a time window should be allocated to separate sound signals from different devices. Subsequent explorations [68,69] apply the similar technique. Whistle [68] enables several receivers to measure the ToAs of transmitted signals from a target device so that they can optimize the update rate for real-time ranging. This work achieves 10 $\sim$ 20 cm ranging resolution from measuring the TDoA of different receivers. And it relies upon coordination among the receiving devices to estimate the location of the source device. RF-Beep [69] takes advantage of both RF and acoustic signals to measure the TDoA for ranging. Because the RF signal has different propagation time from the acoustic signal, RF-Beep uses the propagation time difference to measure the TDoAs, and the ranging resolution is 50 cm. Both studies address the timing uncertainty of sending and receiving acoustic signals to speed up the low update rate. These two can only achieve semi-meter ranging resolutions, which is far from satisfaction for indoor localization.

Direction Finding. Direction finding is another important research area for localization and navigation. Early work [59] reveals that Doppler shift of acoustic signals is able to perform direction finding. The mean error of direction finding is 18.0°, which is not satisfactory. Swadloon [60] follows this work and reduces the error to 2.1° by combining velocities captured from the Doppler shift and the inertial sensors (i.e., accelerometer and gyroscope). The experimental setup in a vertical view of this work is shown in Fig. 11. Specifically, an acoustic source keeps playing single-frequency signals to a receiver (i.e., phone), where the orientation angle of the acoustic source is β and the distance between these two is L. The reference objects A, B, and C are used for evaluating the accuracy of direction measurements. When a user shakes the phone, the velocity of movement v is calculated from the readings of inertial sensors, and the Doppler shift Δf is measured by microphone. According to Eq. (3), the direction of phone α_r is calculated with the known frequency of acoustic signal f and the velocity of acoustic signal v_s. Supported by varying reference objects, L and β are changed so that α_r can be measured for each configuration.

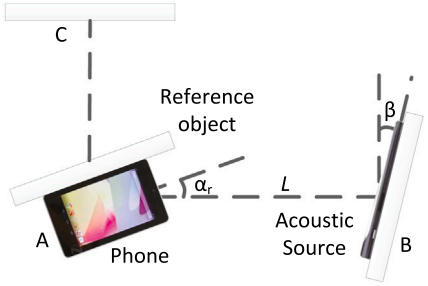

Fig. 11. Illustration of device-to-device direction finding [60].

4.2. Indoor and outdoor localization

Indoor and outdoor localization has been well-studied in the mobile computing community. Locations of users are essential for various applications, such as targeted advertising, tour guides, navigation aids, and social networking. Acoustic-based localization systems are summarized in Table 4.

Passive Indoor Localization. In contrast to outdoor localization, indoor localization can be compromised by complex environments. For example, GPS signals can be blocked or weakened by walls or nearby buildings. This makes well-studied GPS-based localization approaches ineffective. Hence, many research efforts have been made to utilize easily attainable acoustic signals for indoor localization. Many studies deploy additional sensors to transmit acoustic signals received by mobile devices, so as to put acoustic-based localization into effect. This kind of acoustic-based localization is called passive localization [89].

Two early approaches [72,73] capture acoustic fingerprints of background spectrum to determine an indoor location. However, these solutions may be vulnerable to noises and incur high energy costs. To solve this problem, ToA becomes a straightforward option for accurate indoor localization. Guoguo [74] localizes target users via measuring the ToA of acoustic signals. This work can provide sufficient coverage with advanced signal processing techniques, and improve the location update rate by increasing the transmission speed of acoustic signals with a symbol-interleaved signal structure. This study can accomplish an average localization accuracy of 0.25 m in normal environments. More recent investigations [75–77] employ ToA to localize targets in indoor environments. In addition to the ToA, another study SITE [70] combines Doppler shifts of acoustic signals with vision-based techniques for indoor localization, whose median localization error is 0.42 m. SITE deploys a group of speakers transmitting signals with different frequencies. By measuring the Doppler shifts, relative

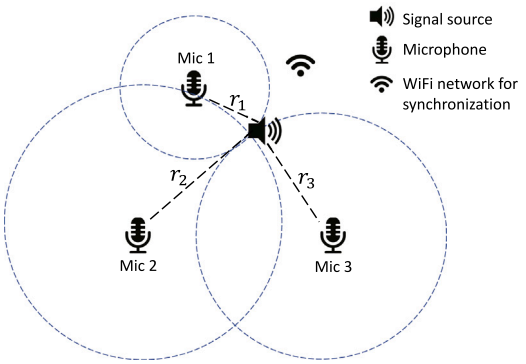

Fig. 12. Illustration of acoustic-based active indoor localization.

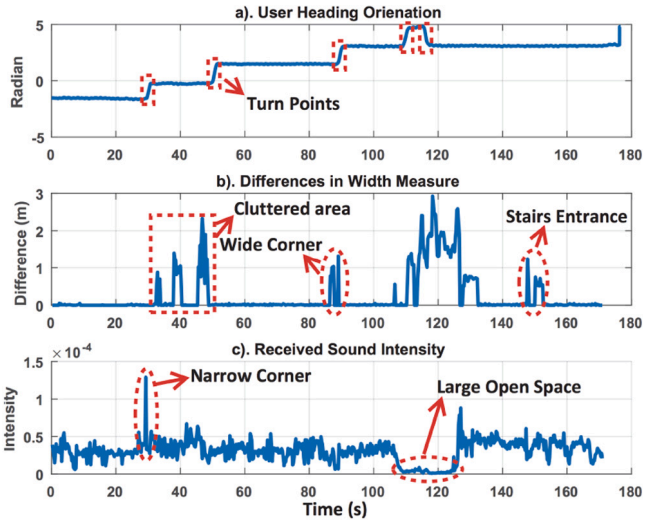

Fig. 13. Different radian, distance, and amplitude patterns of different kinds of indoor areas [94].

directions between the phone and the speakers can be derived. Afterward, SITE uses the angle differences to compute a set of locations, and the vision-based techniques define the final location. Although these studies provide satisfactory localization performance, the need for extra infrastructures (i.e., RF [74], onboard vision sensor [70,76], LED light [77], and specialized sonar [75]) limits the wide adoption of these passive localization methods. More recently, WalkieLokie [80,81] localizes a target by measuring the relative position between a smartphone and an acoustic speaker installed in the target object to estimate ToA with specially designed pulse signals. The WalkieLokie's approach dramatically decreases the deployment cost with only smartphones and low-cost speakers. However, its mean error of ranging is 0.63 m, falling behind the infrastructure-based methods.

Active Indoor Localization. Another body of work localizes users via actively transmitting acoustic signals from speakers in smart devices. This is referred as the active localization [89], an example of which is illustrated in Fig. 12. The smart device hold by a user (i.e., signal source) continuously transmits acoustic signals when the user moves in the indoor environment. Three microphones (i.e., Mic 1, Mic 2, and Mic 3) are placed around the sound source to measure the three ToAs, with which the distances r_1, r_2, r_3 can be calculated by multiplying ToAs by the speed of sound. Then the signal source's location can be determined from r_1, r_2, and r_3 by geometry.

Many studies use this technique for acoustic-based indoor localization. For instance, Beep [82] employs a group of distributed acoustic sensors to localize users equipped with roaming devices synchronized by WiFi. The user transmits acoustic signals using the equipped roaming device, while the acoustic sensors receive the signal and estimate the ToAs with their processing unit. Then the position of roaming device is determined by the distances from those sensors. This work has an accuracy of 0.6 m in more than 97% cases. Qiu et al. [83] use ToA and signal strength measurements to implement high-speed 3-D continuous localization for phone-to-phone scenarios. Since each phone has two microphones, four distances for each pair of mic-speaker combinations are calculated using the ToAs and the angle between two smartphones is derived by the law of cosines. They realize localization resolution within 0.14 m in 90% cases. Liu et al. [84] develop a peer-assist indoor localization system, which combines the acoustic-based localization with WiFi-based approach to enhance the accuracy. Specifically, WiFi-based localization information can be optimized by recruiting other ambient smartphones using distances between smartphones. The localization error can be significantly reduced from 6 $\sim$ 8 m (only WiFi-based) to 1 $\sim$ 2 m. With the cost of continuously transmitting or receiving acoustic signals, these three systems are constrained by the limited power supply of hand-held mobile devices. EchoTag [71] solves the power problem with the help of previous locations of a smartphone. Specifically, this work first determines the coarse-grained location with WiFi SSID and tilt. If this information is matched, the system captures the acoustic signature in the frequency domain to estimate the fine-grained position. All of the four studies [71,82–84] concentrate on self-localization, i.e., localizing the position of a smartphone with its own hardware. In contrast, AMIL [87] allows a smartphone to localize its neighboring smartphones. The smartphone transmits beeps while it is moved in the air, while the listeners measure the TDoAs between beeps. Assuming the listeners are static, their locations can be derived from the TDoAs and the movement trail of smartphone. This approach can estimate the positions of 12 people in a room within an average margin of error of 0.5 m. More recent explorations [88,89] combine both passive and active localizations to improve the localization performance. In these studies, two specialized speakers are pre-deployed, and a smartphone measures TDoA from its two microphones to locate the position of the smartphone, which constructs the passive localization. Then, the smartphone transmits a chirp signal, and measures the ToAs to estimate the distances to side walls, which performs active indoor localization. The two results are combined to improve the localization performance.

Outdoor Localization. There are a few systems extending acoustic sensing to outdoor localization. For instance, ENSBox [85] is a distributed and self-calibrating localization system for outdoor environments, including custom-built hardware with each node consisting of a 4-microphone array. However, it might be impractical to install a large number of outdoor sensors. Other investigations localize emergency sound signals in an outdoor environment. Auto++ [90] localizes surrounding vehicles leveraging TDoA and measures the motion of sound source using Doppler shift. Pinna et al. [86] use acoustic sensor network to localize gun shots. When shooting happens, ToAs between the sound source and acoustic sensors are evaluated to measure the distances between them.

4.3. Floor map construction

Existing studies [91,92] detect the obstacles (e.g., walls and furniture) in the area of interest and construct floor maps through measuring ToA and Angle-of-Arrival (AoA) of acoustic signals, respectively. These approaches however require basic infrastructures (i.e., robots), making them infeasible in some of practical application scenarios. To design a low-cost floor map construction approach, subsequent studies utilize regular smartphones as the signal transmitter and receiver. For example, Kashimoto et al. [93] integrate an ultrasonic gadget into a smartphone for floor map construction. This work first estimates the

Table 5
Comparison of acoustic-based user authentication work.

Work	Category	Technique	Accuracy	Frequency Band	Activity
Zhou et al. [96]	Passive	MFCC	89%	Audible range	Keystroke dynamics
BreathPrint [97]	Passive	GFCC	94%	Audible range	Human breathing gestures
VoiceLive [101]	Active	ToA	99%	Inaudible range	Vocal position inner mouth
VoiceGesture [98]	Active	Doppler Shift	99%	20 kHz	Mouth movement
LipPass [102,103]	Active	Doppler Shift	93.1%	20 kHz	Mouth movement
SilentKey [99]	Active	Doppler Shift	86.7%	Inaudible range	Mouth movement
Wang et al. [100]	Active	Doppler Shift	97%	96 kHz	Gait gesture
EchoPrint [104]	Active	FMCW	93.8%	Inaudible range	Face recognition

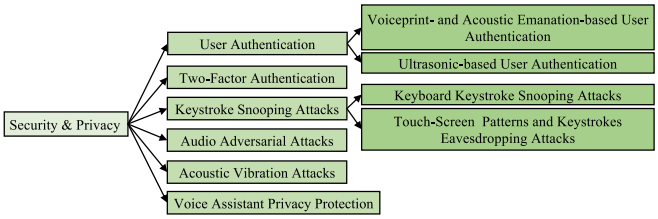

Fig. 14. Structure for acoustic-based security and privacy applications.

room size and shape via measuring ToAs of acoustic signals, then elaborates the details of the floor map with motion sensors in the smartphone. Without any supplemental device for users, BatMapper [94] implements a multi-modal solution, i.e., employing inertial sensors and acoustic signals, to construct floor maps. Specifically, BatMapper adopts inertial sensors to measure the radian of smartphone. After that, the system transmits and receives chirp signals using speakers and microphones of a commercial smartphone. Through analyzing the acoustic signals, BatMapper derives the amplitudes and ToAs to detect different indoor space configurations (e.g., cluttered area, corner, wide-open space). Fig. 13 illustrates the detection using inertial sensors and acoustic signals. There are four turn points in Fig. 13(a) indicating heading orientation changes, which are detected by the inertial sensors. When detecting the orientation changes, the system activates acoustic-based sensing. In Fig. 13(b), different orientations and power changes can exhibit different kinds of areas such as cluttered area, wide corner, and stairs entrance. In Fig. 13(c), a peak in received sound intensity indicates a narrow corner, while a drop in the intensity represents a large open space. However, this may need great training effort in estimating parameters. To do without training, SAMS [95] estimates indoor contours by a person moving around with a smartphone. This work leverages FMCW to measure critical and structural information (e.g., corners and clusters) with less training efforts than BatMapper. Specifically, the FMCW module is customized by exploring its design parameters (i.e., bandwidth, chirp duration), studying its sensitivity under various environments, and extracting FFT features from its profiles to measure the distance and construct the floor map.

5. Security and privacy

As smart and mobile devices are becoming a big part of our daily life, they are often used to store important information, including personal identity and other sensitive data. This trend raises many potential security concerns. In this section, we discuss the issues of security and privacy in mobile device's audio infrastructures (i.e., microphones and speakers), which are among the most integrated components in mobile devices. The processes of acoustic-based security and privacy applications are shown in Fig. 14.

5.1. User authentication

As the first layer of defense in smart and mobile devices, user authentication mechanisms play a crucial role. Among all the user authentication approaches, biometrics-based solutions perform authentication through measuring physiological or behavioral factors, such as fingerprint, face, iris, gait, and voice. All these modalities are made possible by various types of sensors (e.g., camera [167,168], microphone [96,101], and Inertial Measurement Unit (IMU) [169,170]) and RF signal interferences (e.g., WiFi [171,172] or RFID [173,174]). IMU (e.g., accelerometers and gyroscopes)- and camera-based approaches provide precise authentication results. However, they might either inflict intrusiveness or bring privacy concerns. RF-based approaches could provide authentication service in a wider area. However, RFID-based solutions require sensor deployment, and WiFi-based solutions rely on exhaustive signal profiles of surrounding environments, which largely limits their application scenarios. As a complement, acoustic-based authentication has been well explored because of its low-cost properties and the wide deployment of speakers and microphones in mobile devices. We summarize the acoustic-based user authentication systems in Table 5 and discuss them in this section.

Voiceprint- and Acoustic Emanation-based User Authentication. Using voice biometrics for authentication has been well explored in the past decades. The pitch, tone, and volume in a person's voice are associated with his/her physiological characteristics (i.e., size and shape of throat and mouth) and behavioral patterns (i.e., voice pitch, speaking style). Such a relationship is called voiceprint widely employed in the user authentication area to support various applications (e.g., voice assistant) [175,176]. To realize the voiceprint-based user authentication, a number of methods have come up, including Gaussian Mixture Model-Universal Background Model (GMM-UBM) [177], GMM-supervector [178], i-vector model [105] and Deep Neural Network (DNN)-based models [106,107], etc. The basic idea of GMM-UBM is to utilize a combination of Gaussian probability density functions (PDFs) to characterize the voice for modeling the individual uniqueness. Applying GMM to speaker modeling provides the user with a specific PDF, from which a probability score can be obtained. And the UBM, i.e., a pre-trained GMM, is incorporated as the basis of raw GMM to reduce the enormous training data requirements. Though GMM-UBM achieves satisfactory performance in voiceprint authentication, its requirement for aligned voice samples introduces significant computational complexity. Thus, GMM-supervector [178] is developed to obtain a fixed-dimensional vector from a variable-duration utterance. Combining with traditional classification methods (e.g., SVM), GMM-supervector can authenticate individuals in a lightweight manner. Different from GMM-UBM and GMM-supervector models considering speaker factors only, i-vector model [105] involves channel factors (e.g., noises), which facilitates the authentication under an unseen channel distortion. The popularity of DNN is rising due to easily accessible software and affordable hardware solutions. The recent rapid development of DNN also prompts new methods for voiceprint authentication, such as d-vector model [107] and x-vector model [106]. However, due to the open propagation properties of sound, the voices of legitimate users can be easily recorded by attackers, making most of the voiceprint-based authentication models vulnerable to replay attacks [98,102].

Existing studies [96,97] use audible acoustic emanations from specific human behaviors for user authentication. Zhou et al. [96] use the acoustic emanations from keystroke dynamics to authenticate a

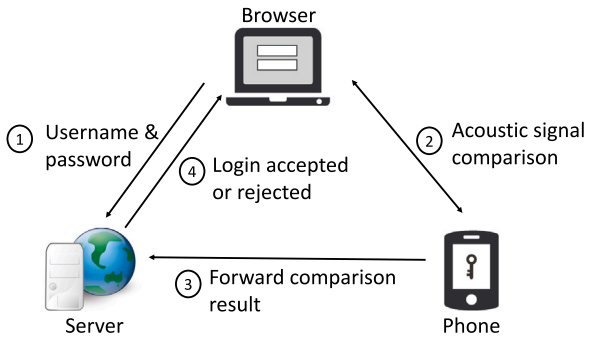

Fig. 15. Illustration of acoustic-based two-factor authentication.

user's identity. This work extracts several acoustic features (i.e., signal strength of acoustic signal, MFCC) and further applies a machine learning approach (i.e., SVM) for authentication. Also, BreathPrint [97] extracts acoustic features (i.e., Gammatone Frequency Cepstral Coefficients (GFCC)) from audible sounds of human breathing in three levels, i.e., sniff, normal, and deep breathing, for user authentication. Similar to voiceprint authentication, these studies are based on the extracted features of acoustic emanations, and they thus may be vulnerable to replay attacks. Moreover, since both keystroke acoustic emanations and breathing sounds are in the audible frequency range, these approaches would be easily interfered by ambient environmental noises.

Ultrasonic-based User Authentication. There have been studies in utilizing inaudible acoustic signals to sense human behavior for user authentication. These studies determine whether a login subject is a human or pre-recorded evidence (e.g., video and sound record) to replay attacks. Zhang et al. [101] measure the TDoA of received acoustic signals to distinguish the sounds of people from those of mechanical speakers. When a person is speaking, the TDoAs of two microphones change in a sequence of phoneme sounds, while TDoA is static under replay attacks. This system requires the user to place audio devices next to a specific position of his/her mouth. To do without this requirement, their following work [98] measures Doppler shifts of acoustic signals to recognize articulatory gestures with a password spoken in a user's habitual ways of speech. The collected password is used for the conventional user authentication, while the Doppler shift of the audio file is extracted for liveness detection. LipPass [102,103] and SilentKey [99] explore lip movements and Wang et al. [100] study gait patterns induced by gait gestures of humans as unique behaviors for user authentication. These systems process the Doppler shifts produced by body motions with the need of specialized device for A/D and D/A conversions. EchoPrint [104] utilizes audio devices in smartphones and frontal camera to extract facial contour for user authentication. In particular, it collects face echos using FMCW, and extracts acoustic features employing CNN to allow for phone-holding pose changes. Meanwhile, facial landmarks are detected with the camera. As joint features, acoustic features and facial landmarks are fed to SVM classifier for training and classification.

5.2. Two-factor authentication

Acoustic-based sensing has also been applied in the Two-Factor Authentication (2FA) field, where a user is granted access only after two kinds of factors are successfully checked by a system. Factors can be knowledge (e.g., a password), possessions (e.g., a smart card), or inherence (e.g., fingerprints). Part of research has focused on novel acoustic-based approaches to detect the proximity between a mobile device and a browser as the second factor, as shown in Fig. 15. As the first of such studies, Sound-Proof [108] uses ambient noises as the second factor in authentication to enhance the security and privacy of smart devices. It measures the cross-correlation between ambient noises received by the user's smartphone and the browser. This system might have security vulnerabilities. Either notification or alarm sounds triggered by APPs on smartphones could be mistreated as ambient noises to defeat the second factor authentication in Sound-Proof [109]. Subsequent studies, Home Alone [110] and Listening Watch [111], are proposed to use randomly-selected acoustic signals for authentication. Home Alone employs active notification sounds generated by user's smartphone to measure the proximity to browser, while Listening Watch uses human speech as the sound factor to detect the proximity between smartwatch and browser. Similar to Sound-Proof, they use cross-correlation to measure the similarity of the sounds recorded by the device on user's body and the one being used to log in. On the basis of previous studies, Proximity-Proof prevents man-in-the-middle and co-located attacks. Proximity-Proof transmits the two-factor authentication response to the browser over OFDM-modulated acoustic signals instead of WiFi or other Internet connections. During the transmission, unique fingerprints of audios are extracted to authenticate the smartphone that transmits the signals. This step serves as the first proximity check that can resist man-in-the-middle attack. Such acoustic fingerprint techniques are also applied in the Internet of Things (IoT) device authentication protocol [113]. Subsequently, a two-way ranging approach [63] measures the distance between two devices to resist the co-located attack, so that the attacker sits beside the user cannot illegally log in.

5.3. Keystroke snooping attacks

In addition to the aforementioned authentication, other investigations reveal potential attacks on mobile devices, laptops, and speech recognition systems using acoustic signals. Keystroke snooping attacks are described and discussed in this section to raise security and privacy awareness of typing using keyboard and performing touch-screen operations.

Keyboard Keystroke Snooping Attacks. Two systems [114,115] reveal that acoustic emanations of keyboard would leak typing information. They extract the signal processing primitives (i.e., FFT, cepstrum) of acoustic emanations and apply machine learning techniques (i.e., neural network, linear classification) for keystroke snooping. Berger et al. [116] combine a dictionary with acoustic emanations to design a more practical keystroke snooping attack without any training. Zhu et al. [121] perform a context-free keystroke snooping attack without any dictionary. This work measures the TDoA of acoustic signals to localize a potential keystroke area with the requirement of three collaborated phones. Liu et al. [123] put only one smartphone close to the keyboard to capture the keystroke sounds. The measurement of TDoAs are grouped and then clustered based on MFCC features. A parallel work [120] proposes to combine acoustic emanations with accelerometer readings to extract typing information. This work explores the FFT power of received acoustic signals, i.e., the amplitude of acoustic signals in frequency-domain, to accurately find the start and end points of a keystroke. Another work [124] focuses on recognizing combined keystrokes instead of a single keystroke since signal fragments are overlapped with each other when two keys are pressed simultaneously. A recent work [122] designs a position-free keystroke snooping attack, in which both TDoA and MFCC are used to determine the relative position between a keyboard and a smartphone, and the potential keystroke area. All of these systems are based on audible keystroke acoustic signals and subject to interference of ambient noises in the surroundings.

Touch-Screen Patterns and Keystroke Eavesdropping Attacks. Touch-screen operations such as pattern drawing and typing rarely produce audible sounds. Patterns and keystrokes eavesdropping attacks are only made possible with the assistance of actively transmitting inaudible acoustic signals. PatternListener [117] leverages phase change of such signals to track finger movements, in order to eavesdrop unlock patterns in Android smartphones. This work however requires victims'

Table 6
Comparison of in-air acoustic-based communication work.

Work	Technique	Operating Range	Frequency Band	Audibility	Throughput
Dhwani [146]	OFDM	10 cm	0–22 kHz	Audible	2.4 kbps
Priwhisper [147]	FSK	30 cm	8–10 kHz	Audible	1 kbps
Matsuoka et al. [149]	OFDM	3 m	5–10 kHz	Audible	Hundreds bps
Dolphin [150]	OFDM	8 m	8–20 kHz	Audible	50 bps
Ka et al. [151]	Chirp Quaternary Orthogonal Keying	2.7 m	18–20 kHz	Near-Ultrasonic	15 bps
Chirp [152]	Chirp Binary Orthogonal Keying	25 m	19.5–22 kHz	Near-Ultrasonic	166 bps
U-Wear [153]	Gaussian Minimum-Shift Keying	20 m	17–19 kHz	Near-Ultrasonic	2.7 kbps
Backdoor [154]	N/A	N/A	40 kHz	Ultrasonic	4 kbps
Dolphin Attack [130]	N/A	15 cm	40 kHz	Ultrasonic	N/A
Bai and Liu et al. [162]	AM & OFDM	10 cm	48 kHz	Ultrasonic	47.49 kbps

smartphones to provide acoustic-based sensing, which might be impractical in the real attack scenario. On the contrary, KeyListener [118, 119] actively transmits inaudible acoustic signals and infers human keystrokes by analyzing the signals reflected back. This work exploits energy attenuation of acoustic signals to determine a coarse-grained keystroke area, and develops a geometric-based approach combining with finger movements between two keystrokes to enhance the recognition performance.

5.4. Audio adversarial attacks

Adversarial attacks were originally studied in image recognition. The adversarial images include small perturbations of less noticeable background pixels in addition to the original images. The subtle modifications are undetectable, but the images can be misclassified by classification models. Similar to the image adversarial attacks, some studies [126–131] found they can also generate adversarial examples against DNN-based speech recognition models, named as audio adversarial attacks. These attacks are imperceptible to humans, but the voice assistant systems can be covertly jammed, mistakenly recognize commands, or secretly controlled. For instance, CommanderSong [127] embeds voice commands into a piece of music to attack voice assistant systems without the awareness of victims. This type of attack can make malicious commands spread on the Internet (e.g., YouTube) and radio, potentially affecting millions of users. Similar to CommanderSong, Adversarial Music [131] jams the voice assistants with inconspicuous adversarial music. This work targets the attack on wake-word detection systems and creates a real-time Denial-of-Service (DoS) attack that can be launched physically over the air. Instead of embedding malicious commands into music, an existing study [128] injects unnoticeable perturbations into voice commands, making the voice assistant recognize the commands as any adversary-desired phrase. However, this study does not consider the impact of over-the-air propagation, such as device distortion, channel effect, and ambient noise. Li et al. [126] first measure the CIR and integrate it into the adversarial example training process toward practical audio examples, which can make the generated adversarial examples remain effective while being played over the air in the physical world. Instead of directly measuring the CIR, Metamorph [129] captures the core distortion's impact from a small set of perturbation measurements and then uses a domain adaptation algorithm to refine the perturbation to improve the attack accuracy and range. Different from previous studies that attack voice assistant systems through software design, Dolphin Attack [130] modulates voice commands from audible frequency bands to ultrasonic frequency bands employing Amplitude Modulation (AM) and utilizes non-linearity of microphones in mobile devices to demodulate received signals and launch attacks.

5.5. Acoustic vibration attacks

Speech privacy is vital in various daily scenarios, such as private meetings, phone conversations, or listen to radio. Traditional methods such as sound-proof walls have been used to prevent eavesdropping.

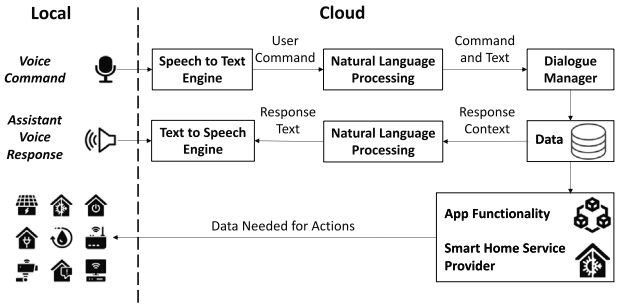

Fig. 16. Illustration of a voice assistant system.

In addition, to prevent potential snooping via built-in microphones, users can simply deny the microphone usage permission. However, studies found that freely accessible motion sensors could leak the private speech information. For example, Gyrophone [132] confirms that the inbuilt gyroscope of smartphones can be used to identify speaker's gender by listening to acoustic signals from an external loudspeaker. Accelword [133] uses the inbuilt accelerometer of mobile devices to extract hotwords (e.g., "Okay Google", "Hi Galaxy") from human voices. Taking advantage of higher sampling rate, Pitchln [134] uses motion sensors in IoT devices to eavesdrop speech. The authors employ a group of sensors in parallel with temporal offset to further improve the sampling rate, named as Time Interleaved Analog-Digital-Conversion (TI-ADC). Different from the aforementioned studies, Spearphone [135] explores the feasibility of revealing the voices played by the smartphone's loudspeakers using the phone's inbuilt motion sensors. Spearphone can perform gender classification and speaker identification, as well as speech recognition and even reconstruction.

5.6. Voice assistant privacy protection

Obtaining informatic messages while protecting user's privacy is an important part of voice assistant systems. As shown in Fig. 16, a voice assistant system involves a local device (e.g., smartphone, smart speaker) and a cloud service provider. When a user sends a voice command to the local device, it then is captured by the microphone and transmitted to the cloud for processing using Natural Language Processing (NLP) to understand the content. After that, the interpreted content is further sent to the data centers to trigger the actions, such as playing music, controlling other smart home devices, and even perform shopping online. However, most users do not use voice assistant systems for shopping due to two privacy concerns: (1) their voice recordings are uploaded to the cloud instead of saved in the local; (2) the system may make authentication mistakes to leak their sensitive information. While voice assistant systems let users review and delete voice recordings, a recent study shows that users are unaware (or do not use) those privacy controls [179].

With the development of edge computing, a related study [136] builds a decentralized voice processing system to reduce the dependency on the cloud, so that it is not necessary to upload sensitive voices

into the cloud. You Talk Too Much [137] and Spreech [138] locally sanitize voice inputs before they are transmitted to the cloud for processing. Other studies [139,140,144] use machine learning techniques to detect malicious commands being spoken to voice assistant systems in order to help users configure the permissions they grant to third parties.

A continuous authentication scheme, VAuth [141] for voice assistant systems, is proposed to defend against authentication mistakes. Specifically, Vauth collects the body-surface vibrations using wearable devices and compares them with the voice commands received by the voice assistant system. Another work [142] combines multi factors, including voice, video, head, and body movements, for secure authentication. Blue et al. [143] propose a two-factor authentication system using mobile devices and IoT devices besides of voice biometrics. They use the microphones of both mobile and IoT devices to measure the Direction-of-Arrival (DoA) in order to localize the sound source of a command. Only the sound sources close to the mobile device can be accepted by the voice assistant system.

6. Short-range communication

Apart from acoustic-based sensing applications, acoustic-based short-range communication also attracts considerable attention, because of the wide availability of audio equipment in mobile devices. The structure of such applications is shown in Fig. 17, and the related explorations are summarized in Table 6.

Audible Communication. There is active research [145–147] using audible acoustic signals for wireless communications. Early work [145] evaluates the impact of standard modulation techniques (i.e., amplitude and frequency shift keying, spread-spectrum modulation) on human perceptions, and builds a communication system with good user experience by selecting the appropriate one. Dhwani [146] implements a near field communication with off-the-shelf phones and achieves a data rate up to 2.4 kbps leveraging the full audible frequency band. This system uses OFDM and the phase shift keying to mitigate the interference of ambient noise and multipath effect. PriWhisper [147] uses the $8 \sim 10$ kHz audible frequency band to enable short-range communication with 1 kbps throughput. PriWhisper adopts frequency shift keying to modulate the signal and estimate the background noise level to assist the transmitter to determine the signal strength. Moreover, Dhwani and PriWhisper both implement friendly jamming techniques to carry out secure communication. Specifically, a receiver transmits a random jamming signal while a sender is transmitting the data signal. Since only the receiver knows the jamming signal, the legitimate receiver could decode the record signals whereas the attackers cannot. However, the audible sounds used in the communication are within the human-perceptible frequency range, resulting in unpleasant user experience.

Communication Through Embedding Message in Common Audio. To improve user experience, another body of work [148–151] leverages the information-concealing technique for audible acoustic communication. Two early systems [148,149] realize the imperceptible acoustic communication with modulating data using OFDM and embedding the signal in regular audio information (e.g., audio sounds from the TV program). To avoid significant quality degrading, the proposed method replaces high-frequency band, i.e., the band out of most human perceptions, of regular audio information with modulated data. These systems transmit short messages from speakers to mobile devices at a medium distance (i.e., around 3 m), and the throughput can be several hundred bps. Other than replacing high frequency band signals, Dolphin [150] takes advantage of masking effects of human auditory system to transmit data-carrying signals and daily sounds simultaneously without human perception. However, all these explorations can only achieve a limited throughput lower than 1 kbps. And their communication process still induces audible sounds.

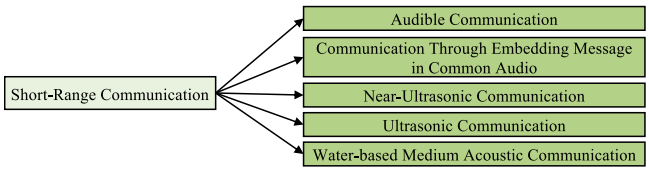

Fig. 17. Structure for acoustic-based short-range communication applications.

Near-Ultrasonic Communication. Near-ultrasonic (i.e., $18 \sim 20$ kHz) acoustic signal is just out of the range of human perceptions, but can still be recorded by microphones of a smartphone and used for wireless communications. Chirp [152] leverages near-ultrasonic band chirp signals to support a longer communication range (i.e., 25 m). This work applies the expected auto-correlation characteristic of chirp signal, i.e., chirp binary orthogonal keying, to eliminate frequency and time selective fading and improve transmission quality of acoustic signals. Based on Chirp, Ka et al. [151] leverage near-ultrasonic chirp signals for wireless communications between TVs and mobile devices, which achieves meter-level range with a very low volume. To be specific, they develop novel synchronization and carrier sensing algorithms to differentiate chirp signal from ambient noise and perform communication. However, these studies can only carry out $15-16$ kbps throughput. U-Wear [153] implements an ultrasonic communication for wearable medical devices, including a physical layer, a data link layer, and a network layer. This work employs Gaussian minimum-shift keying and OFDM for narrowband and wideband signaling schemes respectively, and can achieve 2.7 kbps throughput leveraging a limited 2 kHz bandwidth. Due to the narrow frequency bandwidth, the throughput of this kind of approach is limited.

Ultrasonic Communication. Typically, due to the low-pass filter (whose cut-off frequency is around 24 kHz) of audio infrastructures, commercial mobile devices cannot record the ultrasonic acoustic signals. Two research studies [130,154] discover the non-linearity property of microphones, which contributes to wireless communications on ultrasonic frequency band of acoustic signals. Specifically, Backdoor [154] modulates data on ultrasound signal using FM to achieve inaudibility. Through signal design and non-linearity of microphone, an audible band signal could be reconstructed and demodulated for communication. The throughput they achieved is up to 4 kbps. Dolphin Attack [130] modulates voice commands on an ultrasonic carrier signal through AM to issue an inaudible voice command attack. The voice commands can be recorded by the microphone with its non-linearity. A more recent study [162] innovatively use the OFDM-multiplexing technique together with a non-linearity model and AM to transmit data over multiple narrow-band channels in an ultrasound frequency band in order to achieve high-throughput (i.e., 47.49 kbps) and inaudibility simultaneously. All of the three approaches need extra speakers because regular mobile devices cannot transmit ultrasonic signals.

Water-based Medium Acoustic Communication. Besides aerial wireless communications, acoustic signals can also propagate well in water-based medium. Based on this property, studies [155–158] leverage OFDM to increase the throughput in underwater acoustic communications. A more recent work [159] designs specialized hardware combined with quadrature phase shift keying modulation to further achieve a throughput of 250 kbps for communications in mineral oil. There are also studies [160,161] of transmitting very short ultrasonic pulses in body tissues with an adaptively controllable duty cycle following a pseudo-random adaptive time-hopping pattern to realize communications.

7. Limitations and discussions

While acoustic-based systems have been proven to be powerful and versatile with a broad range of applications, there still exist several limitations and open issues for future research.

Interference of Environmental Noises. Environmental noises can significantly intrude acoustic-based systems for sensing and communications. Many acoustic-based sensing systems need training profiles for activity recognition [180], gesture recognition [37] and indoor localization [71], etc. When such profiles are mixed with noises from conversations or physical activities of people nearby, the performance of those systems will be degraded. A few studies [22,38] incorporate machine learning methods to generate noise-independent profiles to solve this problem. On the other hand, with ubiquitous acoustic noises, acoustic communication systems have to narrow their operation bandwidths to mitigate adverse impacts of the noises. This will reduce communication throughputs. More efforts, such as intelligent profile calibration with multiple acoustic signals and advanced data filtering/machine learning techniques, are needed to improve the noise resistance of acoustic-based applications in future research work.

Impact of Location and Orientation. In addition to environmental changes, a user's location and orientation are critical for the performance of acoustic-based systems. Due to the omni-directional sensing in acoustic signals, the location and orientation are usually recorded by the patterns of acoustic signals, affecting most sensing and recognition applications. For example, an earlier system [15] recognizes driving activities leveraging the Doppler effects of acoustic signals, which requires users to keep the same location and orientation each time to keep the recognition performance. Also, for most human-involved activities, different locations and orientations could induce different variations on the ToA/TDoA patterns sensed by acoustic signals. To deal with this, a recent work [22] utilizes an adversary network to eliminate the location and orientation information from received signals, and achieve a more ubiquitous acoustic-based sensing. However, similar to other machine learning-based approaches, this work requires additional training efforts, which hinders its wide deployment. Eliminating the location and orientation information from acoustic sensed patterns remains an open issue.

Impact of Multi-user in the Sensing Area. In a sensing application, it is highly possible that there are multiple users in a specific area. However, when multiple users are moving around in the sensing area, it is difficult to monitor their movements simultaneously with acoustic signals. Thus most acoustic-based sensing approaches only focus on a single user. To solve this problem, systems [41,102] treat the movements of surrounding people other than the target one as the interference, and apply various methods (e.g., differential CIR, signal gradient) to mitigate them. However, these systems do not intrinsically handle the multi-user sensing problem. Therefore, more research is needed to develop advanced signal separation and differentiation methods. Potential future directions might lie on fusing other sensors (e.g., camera, WiFi), instead of merely using acoustic signals.

Concern of Security and Privacy. The rapid advancement of acoustic-based sensing techniques raises serious security and privacy concerns. For example, human conversations near the sensing devices can be secretly recorded and subsequently leaked to malicious adversaries. Studies [116–118] illustrate various potential side-channel attacks on eavesdropping keystrokes or unlocking patterns with the assistance of acoustic signals. More recent investigations [127,130] demonstrate the possibility of utilizing acoustic signals to launch audio adversarial attacks on popular voice assistant systems (e.g., Apple Siri and Google Now). Some other studies [132,133] show that the readings of smartphones' inertial sensors can be used to reveal speech information (e.g., gender of the speaker, content of the speech). All of these elevate the awareness of potential security and privacy infringements. For countermeasures, some studies [117,130] suggest users enhance and control microphones to defend against keystroke snooping attacks [120,124] and acoustic vibration attacks [132,133]. Another study [144] uses machine learning techniques to defend against malicious hidden voice commands. These can reduce the privacy concern for future acoustic-based application deployment. The defense strategy of the speech-induced vibration attacks [134,135] using motion sensors remains an open question.

8. Conclusion

Leveraging widely-deployed and low-cost audio infrastructures, many research studies have shown the feasibility and effectiveness of using acoustic signals in performing sensing and communication applications. In this paper, we survey up-to-date studies on acoustic-based applications and relative techniques. We first describe key acoustic-based techniques, such as signal strength variation, phase change, Doppler shift, ToA and FMCW, TDoA, and CIR. With these techniques, a broad array of emerging applications are developed using acoustic signals. We review these applications following four categories: recognition and tracking, localization and navigation, security and privacy, short-range communication. These compelling acoustic-based studies have shown promising performance in various application domains. Moreover, we discuss the limitations and challenges of current acoustic-based approaches, and point out a few possible future directions to address these limitations and further extend the acoustic-based applications.

CRediT authorship contribution statement

Yang Bai: Methodology, Investigation, Writing - original draft. **Li Lu:** Methodology, Writing - original draft, Writing - review & editing. **Jerry Cheng:** Methodology, Writing - review & editing, Funding acquisition. **Jian Liu:** Methodology, Writing - review & editing. **Yingying Chen:** Conceptualization, Methodology, Writing - review & editing, Funding acquisition, Project administration, Supervision. **Jiadi Yu:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the anonymous reviewers for the constructive comments to improve the paper. This work was partially supported by the National Science Foundation, United States of America Grants CNS-1814590, CNS-1954959, and CCF-1909963.

References

[1] S. Soro, W. Heinzelman, A survey of visual sensor networks, Adv. Multimedia 2009 (2009).
[2] V. Potdar, A. Sharif, E. Chang, Wireless multimedia sensor networks: A survey, in: Proc. IEEE. AINA, Bradford, UK, 2009.
[3] J. Liu, H. Liu, Y. Chen, Y. Wang, C. Wang, Wireless sensing for human activity: A survey, IEEE Commun. Surv. Tutor. (2019).
[4] R. Angeles, RFID technologies: Supply-chain applications and implementation issues, Inf. Syst. Manag. 22 (1) (2005) 51–65.
[5] K. Chetty, Graeme E. Smith, K. Woodbridge, Through-the-wall sensing of personnel using passive bistatic wifi radar at standoff distances, IEEE Trans. Geosci. Remote Sens. 50 (4) (2011) 1218–1226.
[6] Wikipedia, Sonar, 2019, [Online]. Available: https://en.wikipedia.org/wiki/Sonar. (Accessed 15 June 2020).
[7] Wikipedia, Medical ultrasound, 2019, [Online]. Available: https://en.wikipedia.org/wiki/Medical_ultrasound. (Accessed 15 June 2020).
[8] Wikipedia, Underwater acoustic communication, 2019, [Online]. Available: https://en.wikipedia.org/wiki/Underwater_acoustic_communication. (Accessed 15 June 2020).
[9] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, B. Lee, A survey of sound source localization methods in wireless acoustic sensor networks, Wirel. Commun. Mob. Comput. 2017 (2017).
[10] S. Chen, Z. Qin, G. Xing, K. Ren, Securing acoustics-based short-range communication systems: An overview, J. Commun. Inf. Netw. 1 (4) (2016) 44–51.
[11] J. Morales, D. Akopian, Physical activity recognition by smartphones, a survey, Biocybern. Biomed. Eng. 37 (3) (2017) 388–400.

[12] J. Wang, Y. Chen, S. Hao, X. Peng, L. Hu, Deep learning for sensor-based activity recognition: A survey, Pattern Recognit. Lett. 119 (2019) 3–11.

[13] J. Wang, Ratan K. Ghosh, Sajal K. Das, A survey on sensor localization, J. Control Theory Appl. 8 (1) (2010) 2–11.

[14] C. Wang, Y. Wang, Y. Chen, H. Liu, J. Liu, User authentication on mobile devices: Approaches, threats and trends, Comput. Netw. 170 (2020) 107118.

[15] X. Xu, H. Gao, J. Yu, Y. Chen, Y. Zhu, G. Xue, M. Li, ER: Early recognition of inattentive driving leveraging audio devices on smartphones, in: Proc. IEEE INFOCOM, Atlanta, GA, USA, 2017, pp. 1–9.

[16] X. Xu, J. Yu, Y. Chen, Y. Zhu, S. Qian, M. Li, Leveraging audio signals for early recognition of inattentive driving with smartphones, IEEE Trans. Mob. Comput. 17 (7) (2018) 1553–1567.

[17] H. Lu, W. Pan, Nicholas D. Lane, T. Choudhury, Andrew T. Campbell, SoundSense: Scalable sound sensing for people-centric applications on mobile phones, in: Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services, 2009, pp. 165–178.

[18] J.-S. Hu, C.-C. Cheng, W.-H. Liu, A robust statistical-based speaker's location detection algorithm in a vehicular environment, EURASIP J. Appl. Signal Process. 2007 (1) (2007) 181.

[19] K. Yatani, K.N. Truong, BodyScope: A wearable acoustic sensor for activity recognition, in: Proc. ACM UbiComp, Pittsburgh, PA, USA, 2012, pp. 341–350,

[20] J.A. Stork, L. Spinello, J. Silva, K.O. Arras, Audio-based human activity recognition using non-markovian ensemble voting, in: IEEE RO-MAN, IEEE, Paris, France, 2012, pp. 509–514.

[21] B.A. Swerdlow, T. Machmer, K. Kroschel, Speaker position estimation in vehicles by means of acoustic analysis, in: Proc. DAGA, Stuttgart, Germany, 2007.

[22] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, L. Su, Towards environment independent device free human activity recognition, in: Proc. ACM MobiCom, New Delhi, India, 2018, pp. 289–304.

[23] A. Tamamori, T. Hayashi, T. Toda, K. Takeda, An investigation of recurrent neural network for daily activity recognition using multi-modal signals, in: 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE, 2017, pp. 1334–1340.

[24] Y. Ren, C. Wang, J. Yang, Y. Chen, Fine-grained sleep monitoring: Hearing your breathing with smartphones, in: Proc. IEEE INFOCOM, Hong Kong, China, 2015, pp. 1194–1202.

[25] X. Xu, J. Yu, Y. Chen, Y. Zhu, L. Kong, M. Li, BreathListener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals, in: Proc. ACM Mobisys, Seoul, Republic of Korea, 2019, pp. 54–66.

[26] R. Nandakumar, S. Gollakota, N. Watson, Contactless sleep apnea detection on smartphones, in: Proc. ACM Mobisys, Florence, Italy, 2015, pp. 45–57.

[27] K. Qian, C. Wu, F. Xiao, Y. Zheng, Y. Zhang, Z. Yang, Y. Liu, Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart devices, in: Proc. IEEE INFOCOM, Honolulu, HI, USA, 2018, pp. 1574–1582.

[28] J. Yang, S. Sidhom, G. Chandrasekaran, T. Vu, H. Liu, N. Cecan, Y. Chen, M. Gruteser, R. Martin, Detecting driver phone use leveraging car speakers, in: Proc. ACM MobiCom, Las Vegas, NV, USA, 2011, pp. 97–108.

[29] J. Yang, S. Sidhom, G. Chandrasekaran, T. Vu, H. Liu, N. Cecan, Y. Chen, M. Gruteser, R. Martin, Sensing driver phone use with acoustic ranging through car speakers, IEEE Trans. Mob. Comput. 11 (9) (2012) 1426–1440.

[30] K. Kalgaonkar, B. Raj, One-handed gesture recognition using ultrasonic Doppler sonar, in: Proc. IEEE ICASSP, Taipei, Taiwan, 2009, pp. 1889–1892.

[31] S. Gupta, D. Morris, S. Patel, D. Tan, Soundwave: Using the doppler effect to sense gestures, in: Proc. ACM CHI, Austin, Texas, USA, 2012, pp. 1911–1914.

[32] M. Aumi, S. Gupta, M. Goel, E. Larson, S. Patel, DopLink: Using the doppler effect for multi-device interaction, in: Proc. ACM UbiComp, Zurich, Switzerland, 2013, pp. 583–586,

[33] K.-Y. Chen, D. Ashbrook, M. Goel, S.-H. Lee, S. Patel, AirLink: Sharing files between multiple devices using in-air gestures, in: Proc. ACM UbiComp, Seattle, WA, USA, 2014, pp. 565–569,

[34] W. Ruan, Q. Sheng, L. Yang, T. Gu, P. Xu, L. Shangguan, AudioGest: Enabling fine-grained hand gesture detection by decoding echo signal, in: Proc. ACM UbiComp, Heidelberg, Germany, 2016, pp. 474–485,

[35] Z. Sun, A. Purohit, R. Bose, P. Zhang, Spartacus: Spatially-aware interaction for mobile devices through energy-efficient audio sensing, in: Proc. ACM Mobisys, Taipei, Taiwan, 2013, pp. 263–276.

[36] J. Wang, K. Zhao, X. Zhang, C. Peng, Ubiquitous keyboard for small mobile devices: Harnessing multipath fading for fine-grained keystroke localization, in: Proc. ACM Mobisys, Bretton Woods, NH, USA, 2014, pp. 14–27.

[37] M. Zhang, P. Yang, C. Tian, L. Shi, S. Tang, F. Xiao, SoundWrite: Text input on surfaces through mobile acoustic sensing, in: Proc. ACM SmartObjects, 2015, pp. 13–17.

[38] G. Luo, M. Chen, P. Li, M. Zhang, P. Yang, SoundWrite II: Ambient acoustic sensing for noise tolerant device-free gesture recognition, in: Proc. IEEE ICPADS, Shenzhen, China, 2017, pp. 121–126,

[39] H. Du, P. Li, H. Zhou, W. Gong, G. Luo, P. Yang, WordRecorder: Accurate acoustic-based handwriting recognition using deep learning, in: Proc. IEEE INFOCOM, Honolulu, HI, USA, 2018, pp. 1448–1456.

[40] C. Peng, G. Shen, Y. Zhang, S. Lu, Point&Connect: Intention-based device pairing for mobile phone users, in: Proc. ACM Mobisys, Kraków, Poland, 2009, pp. 137–150.

[41] K. Ling, H. Dai, Y. Liu, A. X. Liu, UltraGesture: Fine-grained gesture sensing and recognition, in: Proc. IEEE SECON, Hong Kong, China, 2018, pp. 1–9.

[42] S. Yun, Y.-C. Chen, L. Qiu, Turning a mobile device into a mouse in the air, in: Proc. ACM Mobisys, Florence, Italy, 2015, pp. 15–29.

[43] W. Wang, A. Liu, K. Sun, Device-free gesture tracking using acoustic signals, in: Proc. ACM MobiCom, New York, NY, USA, 2016, pp. 82–94.

[44] L. Lu, J. Liu, J. Yu, Y. Chen, Y. Zhu, X. Xu, M. Li, VPad: Virtual writing tablet for laptops leveraging acoustic signals, in: Proc. IEEE ICPADS, Singapore, 2018, pp. 244–251.

[45] Li Lu, Jian Liu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, Minglu Li, Enable traditional laptops with virtual writing capability leveraging acoustic signals, Comput. J. 1 (1) (2020) 1–18.

[46] Y. Zhang, J. Wang, W. Wang, Z. Wang, Y. Liu, Vernier: Accurate and fast acoustic motion tracking using mobile devices, in: Proc. IEEE INFOCOM, Honolulu, HI, USA, 2018, pp. 1709–1717.

[47] R. Nandakumar, V. Iyer, D. Tan, S. Gollakota, Fingerio: Using active sonar for fine-grained finger tracking, in: Proc. ACM CHI, San Jose, CA, USA, 2016, pp. 1515–1525.

[48] H. Chen, F. Li, Y. Wang, EchoLoc: Accurate device-free hand localization using COTS devices, in: Proc. IEEE ICPP, Philadelphia, PA, USA, 2016, pp. 334–339.

[49] X. Xu, J. Yu, Y. Chen, Y. Zhu, M. Li, SteerTrack: Acoustic-based device-free steering tracking leveraging smartphones, in: Proc. IEEE SECON, Hong Kong China, pp. 1–9.

[50] W. Mao, J. He, L. Qiu, CAT: high-precision acoustic motion tracking, in: Proc. ACM MobiCom, New York, NY, USA, 2016, pp. 69–81.

[51] H. Chen, F. Li, Wang Y., EchoTrack: Acoustic device-free hand tracking on smart phones, in: Proc. IEEE INFOCOM, Atlanta, GA, USA, 2017, pp. 1–9.

[52] S. Yun, Y.-C. Chen, H. Zheng, L. Qiu, W. Mao, Strata: Fine-grained acoustic-based device-free tracking, in: Proc. ACM Mobisys, Niagara Falls, NY, USA, 2017, pp. 15–28.

[53] E. Rodríguez, B. Ruíz, Á. García-Crespo, F. García, Speech/speaker recognition using a HMM/GMM hybrid model, in: International Conference on Audio-and Video-Based Biometric Person Authentication, Springer, 1997, pp. 227–234.

[54] P. Swietojanski, A. Ghoshal, S. Renals, Revisiting hybrid and GMM-hmm system combination techniques, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2013, pp. 6744–6748.

[55] J. Pan, C. Liu, Z. Wang, Y. Hu, H. Jiang, Investigation of deep neural networks (DNN) for large vocabulary continuous speech recognition: Why DNN surpasses GMMs in acoustic modeling, in: 2012 8th International Symposium on Chinese Spoken Language Processing, IEEE, 2012, pp. 301–305.

[56] Andrew L. Maas, P. Qi, Z. Xie, A.Y. Hannun, Christopher T. Lengerich, D. Jurafsky, Andrew Y. Ng, Building DNN acoustic models for large vocabulary speech recognition, Comput. Speech Lang. 41 (2017) 195–213.

[57] L. Li, Y. Zhao, D. Jiang, Y. Zhang, F. Wang, I. Gonzalez, E. Valentin, H. Sahli, Hybrid deep neural network–hidden Markov model (DNN-HMM) based speech emotion recognition, in: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, IEEE, 2013, pp. 312–317.

[58] M. Shahin, B. Ahmed, J. McKechnie, K. Ballard, R. Gutierrez-Osuna, A comparison of gmm-hmm and dnn-hmm based pronunciation verification techniques for use in the assessment of childhood apraxia of speech, in: Fifteenth Annual Conference of the International Speech Communication Association, 2014.

[59] Y. Nishimura, N. Imai, K. Yoshihara, A proposal on direction estimation between devices using acoustic waves, in: Proc. ACM MobiQuitous, Copenhagen, Denmark, 2011, pp. 25–36.

[60] W. Huang, Y. Xiong, X.-Y. Li, H. Lin, X. Mao, P. Yang, Y. Liu, Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones, in: Proc. IEEE INFOCOM, Toronto, Canada, 2014, pp. 370–378.

[61] J. Scott, B. Dragovic, Audio location: Accurate low-cost location sensing, in: Proc. Springer Pervasive, Munich, Germany, 2005, pp. 1–18.

[62] C. Lopes, A. Haghighat, A. Mandal, T. Givargis, P. Baldi, Localization of off-the-shelf mobile devices using audible sound: Architectures, protocols and performance assessment, ACM SIGMOBILE Mob. Comput. Commun. Rev. 10 (2) (2006) 38–50.

[63] C. Peng, G. Shen, Y. Zhang, Y. Li, K. Tan, Beepbeep: A high accuracy acoustic ranging system using cots mobile devices, in: Proc. ACM Sensys, Sydney, NSW, Australia, 2007, pp. 1–14.

[64] L. Girod, M. Lukac, V. Trifa, D. Estrin, The design and implementation of a self-calibrating distributed acoustic sensing platform, in: Proc. ACM Sensys, Boulder, Colorado, USA, 2006, pp. 71–84.

[65] M. Hazas, C. Kray, H. Gellersen, H. Agbota, G. Kortuem, A. Krohn, A relative positioning system for co-located mobile devices, in: Proc. ACM Mobisys, Seattle, Washington, USA, 2005, pp. 177–190.

[66] G. Kortuem, C. Kray, H. Gellersen, Sensing and visualizing spatial relations of mobile devices, in: Proc. ACM UIST, Seattle, WA, USA, 2005, pp. 93–102.

[67] S. Schörnich, A. Nagy, L. Wiegrebe, Discovering your inner bat: Echo–acoustic target ranging in humans, J. Assoc. Res. Otolaryngol. 13 (5) (2012) 673–682.

[68] B. Xu, R. Yu, G. Sun, Z. Yang, Whistle: Synchronization-free TDOA for localization, in: Proc. IEEE ICDCS, Minneapolis, Minnesota, USA, 2011, pp. 760–769.

[69] M. Uddin, T. Nadeem, Rf-beep: A light ranging scheme for smart devices, in: Proc. IEEE PerCom, San Diego, CA, USA, 2013, pp. 114–122.

[70] R. Xi, D. Liu, M. Hou, Y. Li, J. Li, Using acoustic signal and image to achieve accurate indoor localization, Sensors 18 (8) (2018) 2566.

[71] Y.-C. Tung, K.G. Shin, EchoTag: Accurate infrastructure-free indoor location tagging with smartphones, in: Proc. ACM MobiCom, Paris, France, 2015, pp. 525–536,

[72] P.S. Tarzia, A.P. Dinda, P.R. Dick, G. Memik, Indoor localization without infrastructure using the acoustic background spectrum, in: Proc. ACM Mobisys, Bethesda, MD, USA, 2011, pp. 155–168.

[73] H. Satoh, M. Suzuki, Y. Tahiro, H. Morikawa, Ambient sound-based proximity detection with smartphones, in: Proc. ACM Sensys, Roma, Italy, 2013, pp. 58:1–58:2.

[74] K. Liu, X. Liu, X. Li, Guoguo: Enabling fine-grained indoor localization via smartphone, in: Proc. ACM Mobisys, Taipei, Taiwan, 2013, pp. 235–248.

[75] F. Li, H. Chen, X. Song, Q. Zhang, Y. Li, Y. Wang, Condiosense: High-quality context-aware service for audio sensing system via active sonar, Pers. Ubiquitous Comput. 21 (1) (2017) 17–29.

[76] C. Jiang, M. Fahad, Y. Guo, J. Yang, Y. Chen, Robot-assisted human indoor localization using the kinect sensor and smartphones, in: Proc. IEEE IROS, Chicago, IL, USA, 2014, pp. 4083–4089.

[77] T. Akiyama, M. Sugimoto, H. Hashizume, Time-of-arrival-based indoor smartphone localization using light-synchronized acoustic waves, IEICE Trans. Fundam. Electron. Commun. Comput. Sci. 100 (9) (2017) 2001–2012.

[78] B.D. Haddad, VS.M. Lima, A.W. Martins, WP.L. Biscainho, O.L. Nunes, B. Lee, Acoustic sensor self-localization: Models and recent results, Wirel. Commun. Mob. Comput. 2017 (2017) 1–13.

[79] H. Sundar, V.T. Sreenivas, C.S. Seelamantula, TDOA-based multiple acoustic source localization without association ambiguity, IEEE/ACM Trans. Audio Speech Lang. Process. 26 (11) (2018) 1976–1990.

[80] W. Huang, X.-Y. Li, Y. Xiong, P. Yang, Y. Hu, X. Mao, F. Miao, B. Zhao, J. Zhao, WalkieLokie: Sensing relative positions of surrounding presenters by acoustic signals, in: Proc. ACM UbiComp, Heidelberg, Germany, 2016, pp. 439–450.

[81] W. Huang, X.-Y. Li, Y. Xiong, P. Yang, Y. Hu, X. Mao, F. Miao, B. Zhao, et al., Stride-in-the-loop relative positioning between users and dummy acoustic speakers, IEEE J. Sel. Areas Commun. 35 (5) (2017) 1104–1117.

[82] A. Mandal, C. Lopes, T. Givargis, A. Haghighat, R. Jurdak, P. Baldi, Beep: 3D indoor positioning using audible sound, in: Proc. IEEE CCNC, Las Vegas, NV, USA, 2005, pp. 348–353.

[83] J. Qiu, D. Chu, X. Meng, T. Moscibroda, On the feasibility of real-time phone-to-phone 3d localization Proc. ACM Sensys, Seattle, WA, USA, 2011, pp. 190–203.

[84] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, F. Ye, Push the limit of WiFi based localization for smartphones, in: Proc. ACM MobiCom, Istanbul, Turkey, 2012, pp. 305–316.

[85] L. Girod, M. Lukac, V. Trifa, D. Estrin, A self-calibrating distributed acoustic sensing platform, in: Proc. ACM Sensys, Boulder, Colorado, USA, 2006, pp. 335–336.

[86] S.A. Pinna, G. Portaluri, S. Giordano, Shooter localization in wireless acoustic sensor networks, in: Proc. IEEE ISCC, IEE, Heraklion, Greece, 2017, pp. 473–476.

[87] H. Han, S. Yi, Q. Li, G. Shen, Y. Liu, E. Novak, Amil: Localizing neighboring mobile devices through a simple gesture, in: Proc. IEEE INFOCOM WKSHPS, San Francisco, CA, USA, 2016, pp. 1027–1028.

[88] H. Sánchez-Hevia, D. Ayllón, R. Gil-Pita, M. Rosa-Zurera, Indoor self-localization and orientation estimation of smartphones using acoustic signals, Wirel. Commun. Mob. Comput. 2017 (2017) 1–11.

[89] H. Murakami, M. Nakamura, S. Yamasaki, H. Hashizume, M. Sugimoto, Smartphone localization using active-passive acoustic sensing, in: Proc. IEEE IPIN, Nantes, France, 2018, pp. 206–212.

[90] S. Li, X. Fan, Y. Zhang, T. Wade, J. Lindqvist, R. E Howard, Auto++ detecting cars using embedded microphones in real-time, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1 (3) (2017) 1–20.

[91] J.D. Tardós, J. Neira, P.M. Newman, J.J. Leonard, Robust mapping and localization in indoor environments using sonar data, Int. J. Robot. Res. 21 (4) (2002) 311–330.

[92] C. Schymura, D. Kolossa, Potential-field-based active exploration for acoustic simultaneous localization and mapping, in: Proc. IEEE ICASSP, Calgary, AB, Canada, 2018, pp. 76–80,

[93] Y. Kashimoto, Y. Arakawa, K. Yasumoto, A floor plan creation tool utilizing a smartphone with an ultrasonic sensor gadget, in: Proc. IEEE CCNC, Las Vegas, NV, USA, 2016, pp. 131–136.

[94] B. Zhou, M. Elbadry, R. Gao, F. Ye, BatMapper: Acoustic sensing based indoor floor plan construction using smartphones, in: Proc. ACM Mobisys, Niagara Falls, NY, USA, 2017, pp. 42–55.

[95] S. Pradhan, G. Baig, W. Mao, L. Qiu, G. Chen, B. Yang, Smartphone-based acoustic indoor space mapping, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2 (2) (2018) 75:1–75:26.

[96] Q. Zhou, Y. Yang, F. Hong, Y. Feng, Z. Guo, User identification and authentication using keystroke dynamics with acoustic signal, in: Proc. Springer MSN, Hefei, China, 2016, pp. 445–449.

[97] J. Chauhan, Y. Hu, S. Seneviratne, A. Misra, A. Seneviratne, Y. Lee, BreathPrint: Breathing acoustics-based user authentication, in: Proc. ACM Mobisys, Niagara Falls, NY, USA, 2017, 278–291.

[98] L. Zhang, S. Tan, J. Yang, Hearing your voice is not enough: An articulatory gesture based liveness detection for voice authentication, in: Proc. ACM CCS, Dallas, TX, USA, pp. 57–71.

[99] J. Tan, X. Wang, C.-T. Nguyen, Y. Shi, Silentkey: A new authentication framework through ultrasonic-based lip reading, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2 (1) (2018) 36:1–36:18.

[100] Y. Wang, Y. Chen, M. Bhuiyan, Y. Han, S. Zhao, J. Li, Gait-based human identification using acoustic sensor and deep neural network, Future Gener. Comput. Syst. 86 (2018) 1228–1237.

[101] L. Zhang, S. Tan, J. Yang, Y. Chen, VoiceLive: A phoneme localization based liveness detection for voice authentication on smartphones, in: Proc. ACM CCS, Vienna, Austria, 2016, pp. 1080–1091.

[102] L. Lu, J. Yu, Y. Chen, H. Liu, Y. Zhu, Y. Liu, M. Li, LipPass: Lip reading-based user authentication on smartphones leveraging acoustic signals, in: Proc. IEEE INFOCOM, Honolulu, HI, USA, 2018, pp. 1466–1474.

[103] L. Lu, J. Yu, Y. Chen, H. Liu, L. Kong, M. Li, Lip reading-based user authentication through acoustic sensing on smartphones, IEEE/ACM Trans. Netw. 27 (1) (2019) 1–14.

[104] B. Zhou, J. Lohokare, R. Gao, F. Ye, EchoPrint: Two-factor authentication using acoustics and vision on smartphones, in: Proc. ACM MobiCom, New Delhi, India, 2018, pp. 321–336.

[105] D. Garcia-Romero, Carol Y. Espy-Wilson, Analysis of i-vector length normalization in speaker recognition systems, in: Twelfth Annual Conference of the International Speech Communication Association, 2011.

[106] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, S. Khudanpur, X-vectors: Robust dnn embeddings for speaker recognition, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 5329–5333.

[107] L. Wan, Q. Wang, A. Papir, Ignacio L. Moreno, Generalized end-to-end loss for speaker verification, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 4879–4883.

[108] N. Karapanos, C. Marforio, C. Soriente, Sound-proof: Usable two-factor authentication based on ambient sound, in: Proc. USENIX Security, Washington, D.C., USA, 2015, pp. 483–498.

[109] B. Shrestha, M. Shirvanian, P. Shrestha, N. Saxena, The sounds of the phones: Dangers of zero-effort second factor login based on ambient audio, in: Proc. ACM CCS, Vienna, Austria, 2016, pp. 908–919.

[110] P. Shrestha, B. Shrestha, N. Saxena, Home alone: The insider threat of unattended wearables and a defense using audio proximity, in: Proc. IEEE CNS, Beijing, China, 2018, pp. 1–9.

[111] P. Shrestha, N. Saxena, Listening watch: Wearable two-factor authentication using speech signals resilient to near-far attacks, in: Proc. ACM WiSec, Stockholm, Sweden, 2018, pp. 99–110.

[112] D. Han, Y. Chen, T. Li, R. Zhang, Y. Zhang, T. Hedgpeth, Proximity-proof: Secure and usable mobile two-factor authentication, in: Proc. ACM MobiCom, New Delhi, India, 2018, pp. 401–415.

[113] D. Chen, N. Zhang, Z. Qin, X. Mao, Z. Qin, X. Shen, X. Li, S2M: A lightweight acoustic fingerprints-based wireless device authentication protocol, IEEE Internet Things J. 4 (1) (2017) 88–100.

[114] D. Asonov, R. Agrawal, Keyboard acoustic emanations, in: Proc. IEEE S&P, Berkeley, California, USA, 2004, pp. 3–11.

[115] L. Zhuang, F. Zhou, J.D. Tygar, Keyboard acoustic emanations revisited, in: Proc. ACM CCS, Alexandria, VA, USA, 2005, pp. 373–382.

[116] Y. Berger, A. Wool, A. Yeredor, Dictionary attacks using keyboard acoustic emanations, in: Proc. ACM CCS, Alexandria, VA, USA, 2006, pp. 245–254.

[117] M. Zhou, Q. Wang, J. Yang, Q. Li, F. Xiao, Z. Wang, X. Chen, PatternListener: Cracking android pattern lock using acoustic signals, in: Proc. ACM CCS, Toronto, ON, Canada, 2018, pp. 1775–1787.

[118] L. Lu, J. Yu, Y. Chen, Y. Zhu, X. Xu, G. Xue, M. Li, KeyListener: Inferring keystrokes on QWERTY keyboards of touch screen through acoustic signals, in: Proc. IEEE INFOCOM, Paris, France, 2019, pp. 1–9.

[119] J. Yu, L. Lu, Y. Chen, Y. Zhu, Kong L., An indirect eavesdropping attack of keystrokes on touch screen through acoustic sensing, IEEE Trans. Mob. Comput. (2019).

[120] X. Liu, Z. Zhou, W. Diao, Z. Li, K. Zhang, When good becomes evil: Keystroke inference with smartwatch, in: Proc. ACM CCS, Denver, CO, USA, 2015, pp. 1273–1285.

[121] T. Zhu, Q. Ma, S. Zhang, Y. Liu, Context-free attacks using keyboard acoustic emanations, in: Proc. ACM CCS, Scottsdale, AZ, USA, 2014, pp. 453–464.

[122] Y. Fang, Z. Zhao, Z. Wang, G. Min, Y. Cao, H. Huang, H. Yin, Eavesdrop with PoKeMon: Position free keystroke monitoring using acoustic data, Future Gener. Comput. Syst. 87 (2018) 704–711.

[123] J. Liu, Y. Wang, G. Kar, Y. Chen, J. Yang, M. Gruteser, Snooping keystrokes with mm-level audio ranging on a single phone, in: Proc. ACM MobiCom, Paris, France, 2015, pp. 142–154.

[124] J. Wang, R. Ruby, L. Wang, K. Wu, Accurate combined keystrokes detection using acoustic signals, in: Proc. Springer MSN, Hefei, China, 2016, pp. 9–14.
[125] T. Halevi, N. Saxena, Acoustic eavesdropping attacks on constrained wireless device pairing, IEEE Trans. Inf. Forensics Secur. 8 (3) (2013) 563–577.
[126] Z. Li, C. Shi, Y. Xie, J. Liu, B. Yuan, Y. Chen, Practical adversarial attacks against speaker recognition systems, in: Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications, 2020, pp. 9–14.
[127] X. Yuan, Y. Chen, Y. Zhao, Y. Long, X. Liu, K. Chen, Shengzhi Zhang, Heqing Huang, Xiaofeng Wang, Carl A. Gunter, CommanderSong: A systematic approach for practical adversarial voice recognition, in: Proc. USENIX Security, Baltimore, MD, USA, 2018, pp. 49–64.
[128] N. Carlini, D. Wagner, Audio adversarial examples: Targeted attacks on speech-to-text, in: 2018 IEEE Security and Privacy Workshops (SPW), IEEE, 2018, pp. 1–7.
[129] T. Chen, L. Shangguan, Z. Li, K. Jamieson, Metamorph: Injecting inaudible commands into over-the-air voice controlled systems, in: Proceedings of the Network and Distributed Systems Security (NDSS) Symposium 2020, 2020.
[130] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, W. Xu, DolphinAttack: Inaudible voice commands, in: Proc. ACM CCS, Dallas, TX, USA, 2017, pp. 103–117.
[131] J. Li, S. Qu, X. Li, J. Szurley, J Zico Kolter, F. Metze, Adversarial music: Real world audio adversary against wake-word detection system, in: Advances in Neural Information Processing Systems, 2019, pp. 11908–11918.
[132] Y. Michalevsky, D. Boneh, G. Nakibly, Gyrophone: Recognizing speech from gyroscope signals, in: 23rd USENIX Security Symposium (USENIX Security 14), San Diego, CA, USA, 2014, pp. 1053–1067.
[133] L. Zhang, Parth H Pathak, M. Wu, Y. Zhao, P. Mohapatra, Accelword: Energy efficient hotword detection through accelerometer, in: Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, 2015, pp. 301–315.
[134] J. Han, A.J. Chung, P. Tague, Pitchln: Eavesdropping via intelligible speech reconstruction using non-acoustic sensor fusion, in: Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks, 2017, pp. 181–192.
[135] S. Abhishek Anand, C. Wang, J. Liu, N. Saxena, Y. Chen, Spearphone: A speech privacy exploit via accelerometer-sensed reverberations from smartphone loudspeakers, 2019, arXiv preprint arXiv:1907.05972.
[136] A. Coucke, A. Saade, A. Ball, T. Bluche, A. Caulier, D. Leroy, C. Doumouro, T. Gisselbrecht, F. Caltagirone, T. Lavril, et al., Snips voice platform: An embedded spoken language understanding system for private-by-design voice interfaces, 2018, arXiv preprint arXiv:1805.10190.
[137] T. Vaidya, M. Sherr, You talk too much: Limiting privacy exposure via voice input, in: 2019 IEEE Security and Privacy Workshops (SPW), IEEE, 2019, pp. 84–91.
[138] S. Ahmed, Amrita R. Chowdhury, K. Fawaz, P. Ramanathan, Spreech: A system for privacy-preserving speech transcription, 2019, arXiv preprint arXiv:1909.04198.
[139] K. Olejnik, I. Dacosta, Joana S. Machado, K. Huguenin, Mohammad E. Khan, J.-P. Hubaux, Smarper: Context-aware and automatic runtime-permissions for mobile devices, in: 2017 IEEE Symposium on Security and Privacy (SP), IEEE, 2017, pp. 1058–1076.
[140] G. Misra, Jose M. Such, Pacman: Personal agent for access control in social media, IEEE Internet Comput. 21 (6) (2017) 18–26.
[141] H. Feng, K. Fawaz, Kang G. Shin, Continuous authentication for voice assistants, in: Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking, 2017, pp. 343–355.
[142] V. Kepuska, G. Bohouta, Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home), in: 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, 2018, pp. 99–103.
[143] L. Blue, H. Abdullah, L. Vargas, P. Traynor, 2ma: Verifying voice commands via two microphone authentication, in: Proceedings of the 2018 on Asia Conference on Computer and Communications Security, 2018, pp. 89–100.
[144] C. Wang, S. Abhishek Anand, J. Liu, P. Walker, Y. Chen, N. Saxena, Defeating hidden audio channel attacks on voice assistants via audio-induced surface vibrations, in: Proceedings of the 35th Annual Computer Security Applications Conference, 2019, pp. 42–56.
[145] V.C. Lopes, MQ.P. Aguiar, Aerial acoustic communications, in: Proc. IEEE WASPAA, New Platz, NY, USA, 2001, pp. 219–222.
[146] R. Nandakumar, K. Chintalapudi, V. Padmanabhan, R. Venkatesan, Dhwani: Secure peer-to-peer acoustic NFC, ACM SIGCOMM Comput. Commun. Rev. 43 (4) (2013) 63–74.
[147] B. Zhang, Q. Zhan, S. Chen, M. Li, K. Ren, C. Wang, D. Ma, Priwhisper: Enabling keyless secure acoustic communication for smartphones, IEEE Int. Things J. 1 (1) (2014) 33–45.
[148] H. Matsuoka, Y. Nakashima, T. Yoshimura, T. Kawahara, Acoustic OFDM: Embedding high bit-rate data in audio, in: Proc. Springer MMM, Kyoto, Japan, 2008, pp. 498–507.
[149] H. Matsuoka, Y. Nakashima, T. Yoshimura, Acoustic communication system using mobile terminal microphones, NTT DoCoMo Tech. J. 8 (2) (2006) 2–12.
[150] Q. Wang, K. Ren, M. Zhou, T. Lei, D. Koutsonikolas, L. Su, Messages behind the sound: Real-time hidden acoustic signal capture with smartphones, in: Proc. ACM MobiCom, New York, NY, USA, 2016, pp. 29–41.
[151] S. Ka, T. H. Kim, J. Y. Ha, S. H. Lim, S. C. Shin, J. W. Choi, C. Kwak, S. Choi, Near-ultrasound communication for TV's 2nd screen services, in: Proc. ACM MobiCom, New York, NY, USA, 2016, pp. 42–54.
[152] H. Lee, T. H. Kim, J. W. Choi, S. Choi, Chirp signal-based aerial acoustic communication for smart devices, in: Proc. IEEE INFOCOM, Hong Kong, China, 2015, pp. 2407–2415.
[153] G.E. Santagati, T. Melodia, A software-defined ultrasonic networking framework for wearable devices, IEEE/ACM Trans. Netw. 25 (2) (2017) 960–973.
[154] N. Roy, H. Hassanieh, R. Roy Choudhury, Backdoor: Making microphones hear inaudible sounds, in: Proc. ACM Mobisys, Niagara Falls, NY, USA, 2017, pp. 2–14.
[155] B. Li, J. Huang, S. Zhou, K. Ball, M. Stojanovic, L. Freitag, P. Willett, MIMO-OFDM for high-rate underwater acoustic communications, IEEE J. Ocean. Eng. 34 (4) (2009) 634–644.
[156] M. Stojanovic, OFDM for underwater acoustic communications: Adaptive synchronization and sparse channel estimation, in: Proc. IEEE ICASSP, Las Vegas, Nevada, USA, 2008, 5288–5291.
[157] A. Radosevic, R. Ahmed, M.T. Duman, G.J. Proakis, M. Stojanovic, Adaptive OFDM modulation for underwater acoustic communications: Design considerations and experimental results, IEEE J. Ocean. Eng. 39 (2) (2014) 357–370.
[158] S. Zhou, Z. Wang, OFDM for Underwater Acoustic Communications, John Wiley & Sons, West Sussex, United Kingdom, 2014.
[159] L.M. Wang, A. Arbabian, Exploiting spatial degrees of freedom for high data rate ultrasound communication with implantable devices, Appl. Phys. Lett. 111 (13) (2017) 1–4.
[160] G.E. Santagati, T. Melodia, L. Galluccio, S. Palazzo, Ultrasonic networking for e-health applications, IEEE Wirel. Commun. 20 (4) (2013) 74–81.
[161] L. Galluccio, S. Milardo, E. Sciacca, A feasibility analysis on the use of ultrasonic multihop communications for E-health applications, in: Proc. IEEE ICC, Paris, France, 2017, pp. 1–6.
[162] B. Yang, J. Liu, Y. Chen, L. Lu, J. Yu, Inaudible high-throughput communication through acoustic signals, in: Proc. ACM MobiCom, Mexico, 2019.
[163] M. Pukkila, Channel estimation modeling, Nokia Res. Center (2000).
[164] H. Jiang, Confidence measures for speech recognition: A survey, Speech Commun. 45 (4) (2005) 455–470.
[165] H. Jiang, Discriminative training of HMMs for automatic speech recognition: A survey, Comput. Speech Lang. 24 (4) (2010) 589–608.
[166] Lawrence R Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (2) (1989) 257–286.
[167] A. Sankaran, A. Malhotra, A. Mittal, M. Vatsa, R. Singh, On smartphone camera based fingerphoto authentication, in: 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), IEEE, 2015, pp. 1–7.
[168] M. Azimpourkivi, U. Topkara, B. Carbunar, Camera based two factor authentication through mobile and wearable devices, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1 (3) (2017) 1–37.
[169] T. Feng, Z. Liu, K.-A. Kwon, W. Shi, B. Carbunar, Y. Jiang, N. Nguyen, Continuous mobile authentication using touchscreen gestures, in: 2012 IEEE Conference on Technologies for Homeland Security (HST), IEEE, 2012, pp. 451–456.
[170] N.T. Trung, Y. Makihara, H. Nagahara, Y. Mukaigawa, Y. Yagi, Performance evaluation of gait recognition using the largest inertial sensor-based gait database, in: 2012 5th IAPR International Conference on Biometrics (ICB), IEEE, 2012, pp. 360–366.
[171] Z. Zhang, Richard W Pazzi, A. Boukerche, Design and evaluation of a fast authentication scheme for wifi-based wireless networks, in: 2010 IEEE International Symposium on" a World of Wireless, Mobile and Multimedia Networks"(WoWMoM), IEEE, 2010, pp. 1–6.
[172] C. Shi, J. Liu, H. Liu, Y. Chen, Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT, in: Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 1–10.
[173] Y.-P. Liao, C.-M. Hsiao, A secure ECC-based RFID authentication scheme integrated with ID-verifier transfer protocol, Ad hoc Netw. 18 (2014) 133–146.
[174] Y. Chen, J.-S. Chou, H.-M. Sun, A novel mutual authentication scheme based on quadratic residues for RFID systems, Comput. Netw. 52 (12) (2008) 2373–2380.
[175] Z. Saquib, N. Salam, R. Nair, N. Pandey, Voiceprint recognition systems for remote authentication-a survey, Int. J. Hybrid Inf. Technol. 4 (2) (2011) 79–97.
[176] Z. Saquib, N. Salam, R.P. Nair, N. Pandey, A. Joshi, A survey on automatic speaker recognition systems, in: Signal Processing and Multimedia, Springer, 2010, pp. 134–145.
[177] R. Zheng, S.u Zhang, B. Xu, Text-independent speaker identification using GMM-UBM and frame level likelihood normalization, in: 2004 International Symposium on Chinese Spoken Language Processing, IEEE, 2004, pp. 289–292.

[178] William M. Campbell, Douglas E. Sturim, Douglas A. Reynolds, Support vector machines using GMM supervectors for speaker verification, IEEE Signal Process. Lett. 13 (5) (2006) 308–311.
[179] J. Lau, B. Zimmerman, F. Schaub, Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers, Proc. ACM Human-Comput. Interact. 2 (CSCW) (2018) 1–31.
[180] G. Laput, K. Ahuja, M. Goel, C. Harrison, Ubicoustics: Plug-and-play acoustic activity recognition, in: Proc. ACM UIST, Berlin, Germany, 2018, pp. 213–224.

Yang Bai is currently pursuing the Ph.D. degree at Wireless Information Network Laboratory (WINLAB), Department of Electrical and Computer Engineering, Rutgers University. She got her M.S. degree at Stevens Institute of Technology in 2018. Her research interests include mobile and ubiquitous computing, and mobile communication and sensing. She is currently working in the Data Analysis and Information SecuritY (DAISY) Lab under the supervision of Prof. Yingying Chen.

Li Lu received the B.E. degree in Computer Science and Technology from Xi'an Jiaotong University, Xi'an, China, in 2015. He is currently a Ph.D. candidate in Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. He was also a visiting research student in Wireless Information Network Laboratory (WINLAB) and Department of Electrical and Computer Engineering at Rutgers University during 2018–2019. His research interests include mobile and ubiquitous computing, cyber security and privacy, human–computer interaction.

Jerry Cheng is an Assistant Professor of Computer Science at New York Institute of Technology. His research interests include big data analytics, statistical learning, Bayesian statistics, and their applications in computer systems and smart healthcare. He was an Assistant Professor in Robert Wood Johnson Medical School at Rutgers University. He was formerly a Postdoc Researcher in the Department of Statistics at Columbia University and had extensive industrial experiences as a Member of Technical Staff at AT&T Labs. His background is a combination of Computer Science, Statistics and Physics. His work has been reported by many new media including MIT Technology Review, Yahoo News, Digital World, FierceHealthcare, and WTOP Radio.

Jian Liu is an Assistant Professor in the Department of Electrical Engineering and Computer Science at the University of Tennessee, Knoxville. He received his Ph.D. degree in Wireless Information Network Laboratory (WINLAB) at Rutgers University. His current research interests include mobile sensing and computing, cybersecurity and privacy, intelligent systems and machine learning. He is the recipient of the Best Paper Awards from IEEE SECON 2017 and IEEE CNS 2018. He also received Best-in-session Presentation Award from IEEE INFOCOM 2017, and two Best Poster Award Runner-up from ACM MobiCom 2016 and 2018.

Yingying (Jennifer) Chen is a Professor of Electrical and Computer Engineering at Rutgers University and the Associate Director of Wireless Information Network Laboratory (WINLAB). She also leads the Data Analysis and Information Security (DAISY) Lab. She is an IEEE Fellow. Her research interests include mobile sensing and computing, cyber security and privacy, Internet of Things, and smart healthcare. She has co-authored three books, published over 150 journals and referred conference papers and obtained 8 patents. Her background is a combination of Computer Science, Computer Engineering and Physics. Prior to joining Rutgers, she was a tenured professor at Stevens Institute of Technology and had extensive industry experiences at Nokia (previously Alcatel-Lucent). She is the recipient of the NSF CAREER Award and Google Faculty Research Award. She also received NJ Inventors Hall of Fame Innovator Award. She is the recipient of multiple Best Paper Awards from EAI HealthyIoT 2019, IEEE CNS 2018, IEEE SECON 2017, ACM AsiaCCS 2016, IEEE CNS 2014 and ACM MobiCom 2011. She is the recipient of IEEE Region 1 Technological Innovation in Academic Award 2017; she also received the IEEE Outstanding Contribution Award from IEEE New Jersey Coast Section each year 2005–2009. Her research has been reported in numerous media outlets including MIT Technology Review, CNN, Fox News Channel, Wall Street Journal, National Public Radio and IEEE Spectrum. She has been serving/served on the editorial boards of IEEE Transactions on Mobile Computing (IEEE TMC), IEEE Transactions on Wireless Communications (IEEE TWireless), IEEE/ACM Transactions on Networking (IEEE/ACM ToN) and ACM Transactions on Privacy and Security.

Jiadi Yu received the Ph.D. degree in Computer Science from Shanghai Jiao Tong University, Shanghai, China, in 2007. He is currently an Associate Professor in Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. Prior to joining Shanghai Jiao Tong University, he was a postdoctoral fellow in the Data Analysis and Information Security (DAISY) Laboratory at Stevens Institute of Technology from 2009 to 2011. His research interests include cyber security and privacy, mobile and pervasive computing, cloud computing and wireless sensor networks. He is a member of the IEEE and the IEEE Communication Society.