Huang JL, Wang YS, Zou YP *et al.* Ubiquitous WiFi and acoustic sensing: Principles, technologies, and applications. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 38(1): 25-63 Jan. 2023. DOI: 10.1007/s11390-023-3073-5

Ubiquitous WiFi and Acoustic Sensing: Principles, Technologies, and Applications

Jia-Ling Huang[1,†] (黄佳玲), Yun-Shu Wang[1,†] (王云舒), Yong-Pan Zou[1,*] (邹永攀), *Member, CCF, ACM, IEEE*
Kai-Shun Wu[1] (伍楷舜), *Fellow, IEEE*, and Lionel Ming-shuan Ni[2,3] (倪明选), *Life Fellow, IEEE*

[1] *The IoT Research Center, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060 China*

[2] *The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511455, China*

[3] *The Hong Kong University of Science and Technology, Hong Kong, China*

E-mail: huangjialing2021@email.szu.edu.cn; wangyunshu2022@email.szu.edu.cn; yongpan@szu.edu.cn
wu@szu.edu.cn; ni@ust.hk

Received January 4, 2023; accepted January 23, 2023.

Abstract With the increasing pervasiveness of mobile devices such as smartphones, smart TVs, and wearables, smart sensing, transforming the physical world into digital information based on various sensing medias, has drawn researchers' great attention. Among different sensing medias, WiFi and acoustic signals stand out due to their ubiquity and zero hardware cost. Based on different basic principles, researchers have proposed different technologies for sensing applications with WiFi and acoustic signals covering human activity recognition, motion tracking, indoor localization, health monitoring, and the like. To enable readers to get a comprehensive understanding of ubiquitous wireless sensing, we conduct a survey of existing work to introduce their underlying principles, proposed technologies, and practical applications. Besides we also discuss some open issues of this research area. Our survey reals that as a promising research direction, WiFi and acoustic sensing technologies can bring about fancy applications, but still have limitations in hardware restriction, robustness, and applicability.

Keywords WiFi sensing, acoustic sensing, human-computer interaction, human activity recognition

1 Introduction

In recent years, a new round of scientific and technological revolution has been booming around the world. Human beings have stepped into the era of Internet of Things (IoT). With increasing pervasiveness of wireless and acoustic hardware, researchers have begun to pay more attention to developing novel sensing technologies based on WiFi and acoustic signals. Numerous studies have demonstrated the technical feasibility and effectiveness of sensing applications using these two types of signals, such as in the human activity recognition, health caring, positioning and navigation, and many other aspects of human life.

Among different types of sensing media, WiFi signals have the advantages of prominent pervasiveness, nearly zero hardware cost, and robustness to environmental conditions, such as light, temperature, and humidity. Bahl and Padmanabhan[1] first proposed a WiFi-based sensing application—indoor localization based on the received signal strength indication (RSSI). RSSI is usually computed by using the energy of signals as a reference to 1 mW, which is expressed in a logarithmic form.

Survey

Special Issue in Honor of Professor Kai Hwang's 80th Birthday

This work is supported by the National Natural Science Foundation of China under Grant Nos. 62172286 and U2001207, the Natural Science Foundation of Guangdong Province of China under Grant Nos. 2022A1515011509 and 2017A030312008, and the Guangdong "Pearl River Talent Recruitment Program" under Grant No. 2019ZT08X603.

†Co-First Authors (Jia-Ling Huang and Yun-Shu Wang are mainly responsible for the survey of WiFi and acoustic sensing, respectively.)

*Corresponding Author

The CSI tool was developed in [2] for extracting the channel state information (CSI) from commercial network cards, which greatly facilitated the acquisition of the CSI from commercial WiFi devices and made the use of more fine-grained CSI for sensing a new trend. The sensing of the CSI has become a new trend. Subsequently, the human behavior sensing technology based on WiFi signal was developed rapidly. The emergence of the CSI read interface makes CSI widely used in the WiFi sensing and sleep monitoring[3–8], fall detection[9–17], gesture detection[18–32], lip language recognition[33, 34], crowd detection[35, 36], daily behavior detection[9, 37–48], respiration and heartbeat detection[7, 32, 49–51], gait recognition[52–55], indoor localization[56–69] and a series of other applications[70–72].

The WiFi signal has long-range and good penetration characteristics, and WiFi-based sensing can capture a large range of human activities and even detect people moving behind obstacles. However, a long sensing range also makes it vulnerable to the surrounding environment. In contrast, although the acoustic signal has limited coverage for surrounding sensing due to its fast decay, it is more sensitive and resilient to changes in the environment. This makes acoustic-based sensing and WiFi-based sensing a valuable complement to each other. Based on the sound wave sensing, the hardware base mainly includes two components: microphone and speaker. With the popularity of mobile devices and wearable devices, widespread deployment of microphones and speakers, and the continuous progress of audio chips and technologies, the acoustic signals have become very easy to acquire and handle with high-quality and extensive sensing and communication capabilities. After obtaining the WiFi or acoustic signal, the next step is to characterize it using various sensing techniques. Similar to WiFi sensing, typical applications based on acoustic sensing include daily actions monitoring[73–78], gesture and hand movements recognition[79–87], health caring[88–92], localization and navigation[93–98] and privacy and security[99–111]. In the following, we will also introduce the basic content of signals and other characterization methods.

Through these basic contents and technologies, a wide range of applications can improve the quality of our daily life and work efficiency, and bring great influence and changes to the human life. These applications include: daily behavior recognition, gesture and hand motion recognition, and tracking in behavior recognition; health-related applications, such as breathing monitoring, heartbeat monitoring, lung monitoring, sleep quality detection, fall detection, and abnormal sleep detection in abnormal events; various positioning and navigation applications based on WiFi and sound waves; and the user in privacy and security authentication, keystroke snooping, voice assistant attacks, and voice assistant protection touch in every aspect of human life. In Section 4, we will introduce applications based on WiFi sensing and acoustic sensing from four aspects: behavior recognition and tracking, health caring, positioning and navigation, and privacy and security. Fig.1 shows the overall structure frame of this paper.

The rest of the paper is organized as follows. Section 2 introduces the background knowledge of WiFi and acoustic signals. Section 3 demonstrates some important key technologies that enable WiFi and acoustic sensing. Section 4 presents WiFi and acoustic sensing applications from four aspects including behavior recognition and tracking, health caring, localization, and privacy and security. Section 5 discusses the limitations of existing work and highlights future research directions. Section 6 summarizes this survey paper.

2 Background

Both WiFi and acoustic signals are wireless signals. They share a common characteristic, namely the multipath effect. It describes the phenomenon that the signal reaches the receiver through different propagation paths, which can affect the performance of many sensing systems.

The wave in the process of transmission inevitably encounters many obstacles due to different material obstructions. The wave incidence angles cause wave refraction and reflection, and make the same waves through different paths to receive node multipath signals in time and to overlap, which causes direct signal distortion and affects the receiving end of signal recognition. The phenomenon of phase inconsistency, caused by the multipath effect and resulting in the fading state of the received signal, is called the multipath fading, which has a great impact on communication, detection, etc.

In the HAR (human action recognition) scenario of WiFi, the WiFi channel includes signals reflected by static objects in some environments, such as furniture or others. CSI represents the signal change from the transmitter to the receiver. These additional reflections caused by different activities can be observed in Fig.2.

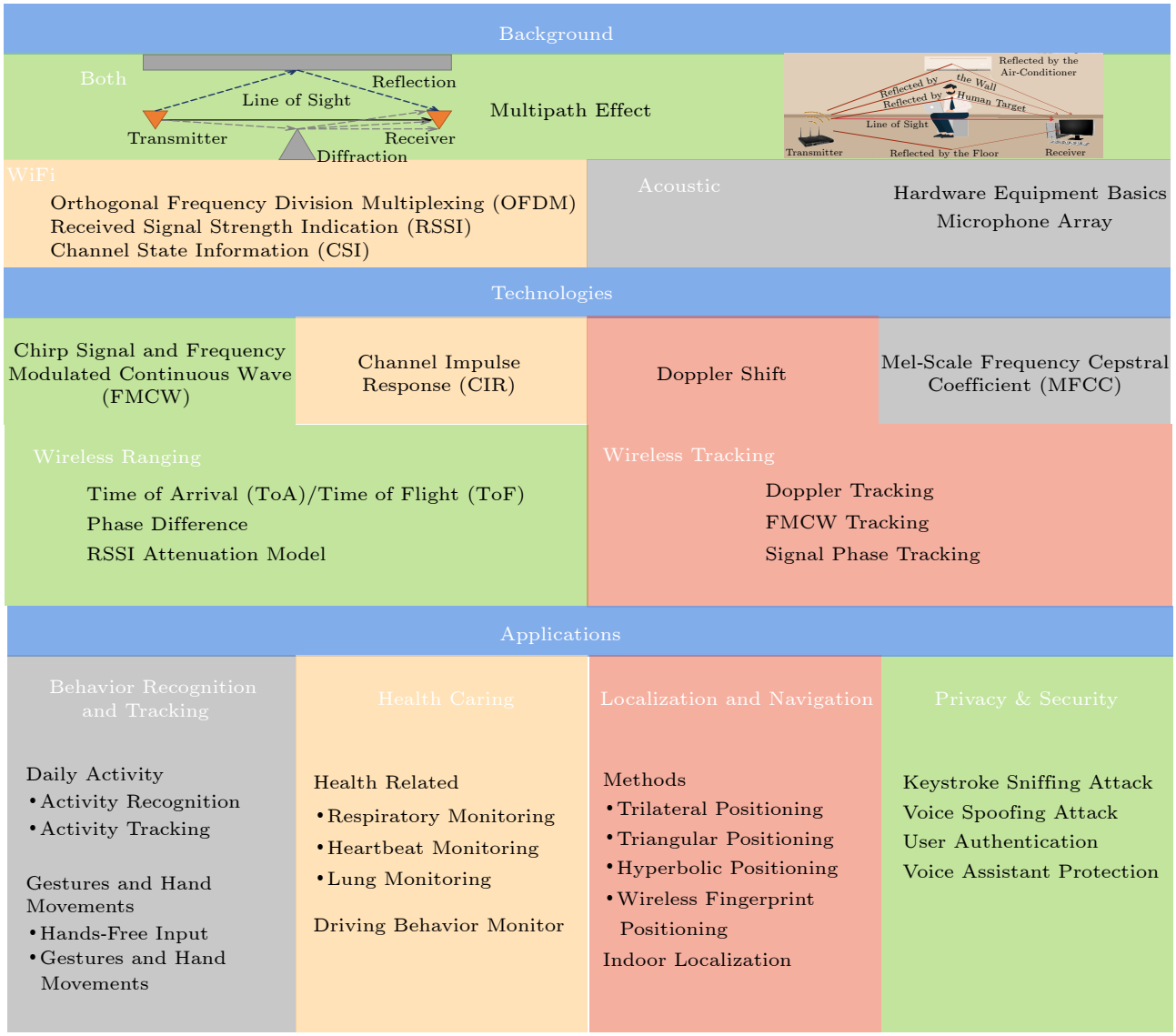


Fig.1. Overall frame diagram of the paper.

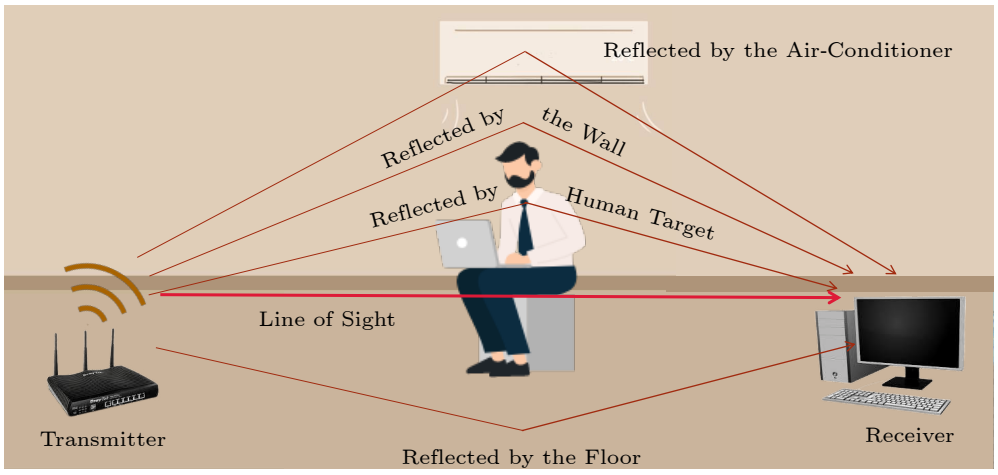


Fig.2. Propagation of Wi-Fi radio waves across a room.

For sound waves, the air is a multipath and time-varying attenuation channel. The propagation of sound waves in air can be regarded as the superposition of multiple signals with different delays and phases like Fig.3.

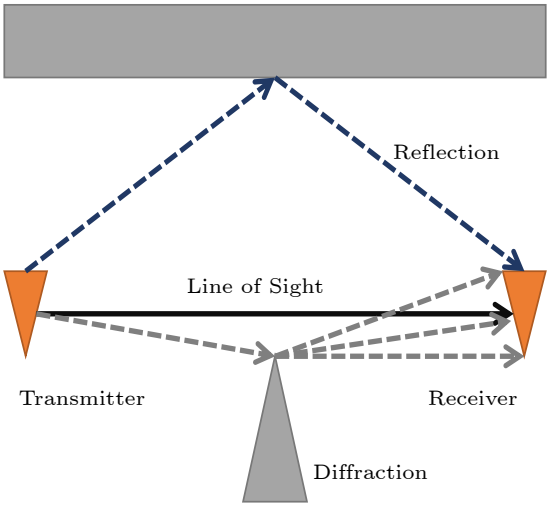


Fig.3. Schematic of multipath effect.

Fig.4 is a roadmap of perception by application, using WiFi and acoustic signals that have evolved over time. In addition to their common features, they also have some unique features, which we will be introduced in two parts.

2.1 WiFi

In this subsection, we mainly introduce Orthogonal Frequency Division Multiplexing (OFDM), as well as channel state information (CSI) and received signal strength indication (RSSI) in the WiFi sensing field. This information will help us understand the WiFi sensing behind it. The physical quantities used by researchers for WiFi sensing are described below.

2.1.1 *Orthogonal Frequency Division Multiplexing (OFDM)*

There are some applications of acoustic sensing which use this technology, except in FingerIO, where the OFDM technology is used to recognize gestures. However, OFDM is mainly manifested in the CSI signal as introduced in Subsection 2.1.3. Compared with continuous wave (CW) signals, OFDM signals have stronger ability to resist environmental interference. Its mathematical formula is:

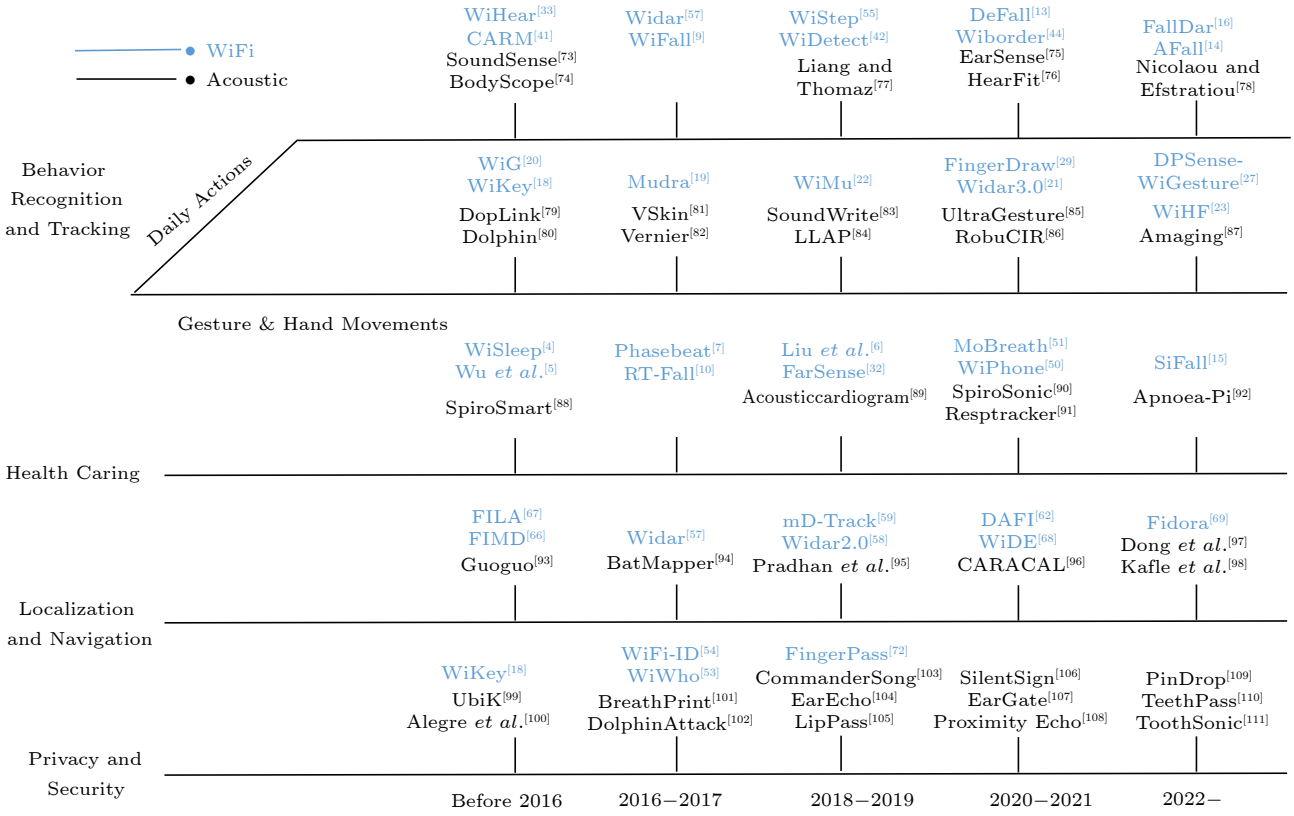


Fig.4. Roadmap of WiFi and acoustic sensing in terms of applications.

$$x_k = \sum_{n=0}^{N-1} X_n \mathrm{e}^{\frac{2\pi k n_i}{N}}, \quad k = 0, ..., N-1,$$

where N represents the number of parts into which the bandwidth is cut. The data bits X_n are transmitted to the n-th channel.

The main idea of OFDM is to divide the channel into several orthogonal sub-channels. A high-speed data signal is converted into a parallel low-speed sub-data stream, and then it is modulated to transmit to each sub-channel. The signal bandwidth on each sub-channel is less than the correlation bandwidth of the channel.

Therefore, it can be regarded as flat fading on each sub-channel, and the inter-symbol interference can be eliminated. Moreover, since the bandwidth of each sub-channel is only a small part of the bandwidth of the original channel, the channel equalization becomes relatively easy.

2.1.2 Received Signal Strength Indication (RSSI)

WiFi signals are widely deployed. Since no sensor needs to be carried, human senseless sensing as well as NLOS (none line of sight) sensing, are not affected by external conditions, such as light, humidity, and temperature. WiFi signals were used for sensing for the first time in 2000 when Bahl and Padmanabhan[1] proposed RADAR as a system for indoor positioning based on WiFi signal strength information (RSS). RSS is an important indicator for the wireless transmission layer to determine link quality. The transmission layer uses RSS to determine whether it is necessary to increase the sending intensity of the sender. Normally, RSS is represented by power in watts (W). However, the power of the wireless signal is weak, usually at the milliwatt (mW) level. Therefore, the signal energy is expressed in logarithmic form on the basis of 1 mw, i.e., RSSI. This is a common practice.

RSS information can be read directly through the program interface on universal devices, such as mobile phones and computers, without requiring any equipment or program modification. It is quick and convenient to obtain. It is also compatible with the advantages of universal devices. But RSSI value sensing accuracy is low and fine-grained sensing cannot be achieved as the RSS information obtained from universal devices is not a real signal strength. Meanwhile, the RSS values are updated slowly, but they cannot be updated in real time. In addition, RSS is susceptible to the environmental interference.

2.1.3 Channel State Information (CSI)

In 2011, Halperin *et al.*[2] released the CSI tool to extract CSI from commercial network cards, which greatly facilitates the acquisition of CSI on commercial WiFi devices, making it a new trend to use finer-grained CSI for sensing. Subsequently, the human behavior sensing technology based on WiFi signals was developed rapidly.

CSI provides information to each transmitter and receiver antenna pair at each carrier frequency based on multiple input multiple output (MIMO) and OFDM. Mathematically a CSI can be expressed as[32]:

$$s(t) = A\cos(2\pi(\frac{f_{\min}t + kt^2}{2}) + \varphi),$$

where L is the number of paths, A_i is the complex attenuation and $d_i(t)$ is the propagation length of the i-th path.

CSI estimates each subcarrier for each transmission link. Therefore, compared with RSS, it has finer granularity and sensitivity, and can sense more subtle changes in the channel. As shown in Fig.5, the time series of the CSI matrix characterizes the MIMO channel variation in different domains[112] (time, frequency, and space). The OFDM technology divides the WiFi channel with MIMO into multiple subcarriers.

The 3D CSI matrix is similar to a digital image with a spatial resolution of $N \times M$ and K color channels. Therefore, this also enables CSI-based WiFi sensing combined with the field of computer vision.

CSI can be obtained in three ways, namely beacon frame, injection frame, and data frame. The beacon frame is transmitted periodically and it has the effect of announcing the existence of WLAN. The injection frames are created in the monitor mode to detect network failures. The data frames appear when data is communicated. Since the injection frame monopolizes a sensing channel, the CSI measurement is more controllable, while the data frame can coexist with data communication and CSI acquisition. Therefore, it becomes the two most commonly-used frames for CSI acquisition in most sensing applications. However, most sensing applications do not use the beacon frame as the sampling rate of CSI measurements in the beacon frame is too low for most sensing applications.

In addition, after the release of the CSI tool in 2011[2], Xie *et al.*[113] released another CSI acquisition tool in 2015, namely the Atheros-CSI-Tool, based on

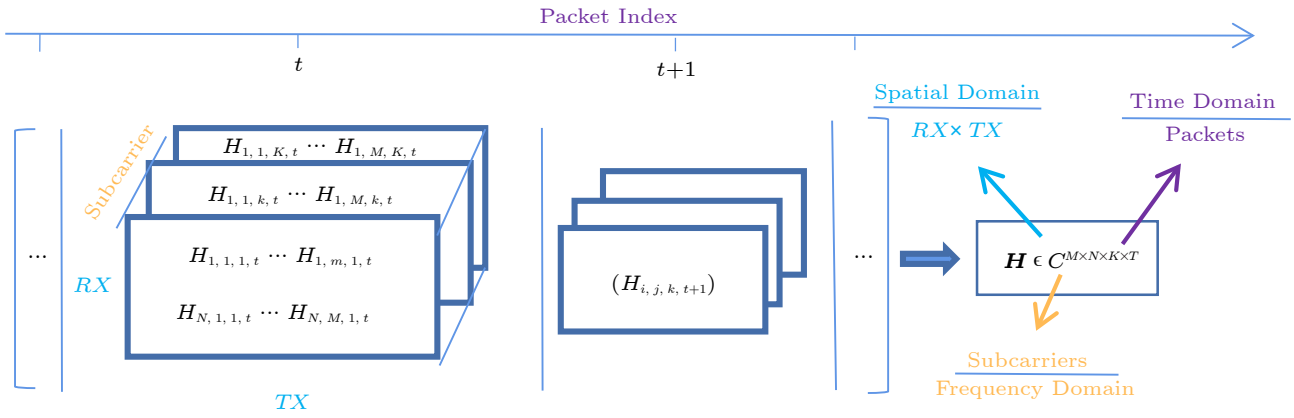


Fig.5. The 4D CSI tensor is the time series of the CSI matrix of the MIMO-OFDM channel. It captures multipath channel changes, including amplitude decay and phase offset, in the spatial, frequency, and time domains. *RX*: the number of receiver antennas. *TX*: the number of transmitter antennas.

Ath9k, a Linux open source network card driver. In 2020, Hernandez and Bulut[31] developed the ESP32 CSI toolkit, which allows researchers to access CSI directly from the ESP32 microcontroller. ESP32 with this toolkit can provide online CSI processing from any computer, smartphone, or even stand-alone device. In 2021, Jiang *et al.*[114] developed the PicoScene platform, which is a versatile and powerful middleware for CSI-based WiFi sensor research. It helps researchers to overcome two barriers in the WiFi sensor research: inadequate hardware functionality and inadequate measurement software functionality. These new developments have made it possible to obtain CSI on more devices and have greatly expanded the range of WiFi-aware applications.

2.2 Acoustic

2.2.1 *Hardware Basics*

The hardware base of the acoustic sensing technology mainly includes microphone and loudspeaker. The microphone is a kind of transducers that can convert the physical sound into analog electrical signals. Most microphones are capacitive in nature, which mainly include two types: electret (ECM) microphones and micro electromechanical (MEMS) microphones. The capacitive microphones are air-gap capacitors with removable membranes and fixed electrodes.

Air pressure due to sound waves can cause the diaphragm to get bent with changes in air pressure. Since the other electrode remains stationary, the movement of the membrane can cause a change in the capacitance value between the membrane and the fixed electrode. Due to their miniature size, low power consumption and excellent temperature characteristics, microphones of micro electromechanical systems have been widely used in mobile devices, including smart phones and wearable devices. Fig.6 shows the schematic diagram of sound signal transmission pathway.

The speaker is a transducer device that converts electrical signals into acoustic signals. It is a sound transmission device that can transmit sound to the mobile phone system. It has the characteristic of causing vibration when it receives current data. At present, many mobile phones are accommodating dual speakers. When a mobile phone is playing sound,

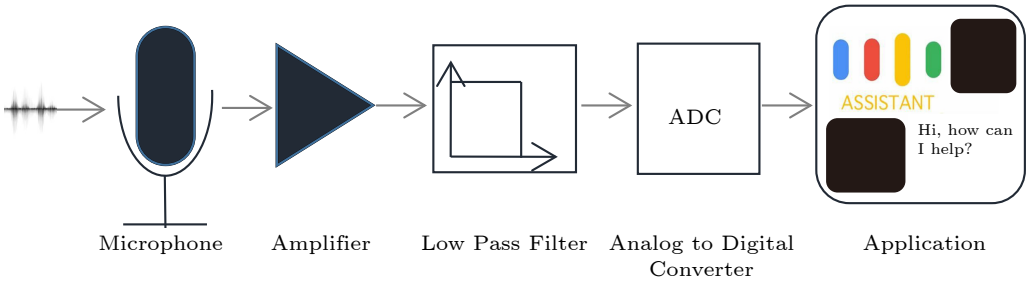


Fig.6. Schematic diagram of sound signal transmission pathway.

the top of the body will also emit sound, and there will be a feeling of surrounding sound. Currently, there are some models whose mainstream configuration is for dual speakers, such as iPhone, Samsung, and so on. In addition, the purpose of the human body with two ears is to hear the sound in three dimensions and identify the position of the sound source. A single speaker cannot transmit three-dimensional sound effects, but dual speakers can do it. The widespread use of microphones and speakers offers great opportunities for acoustic-based sensing, and better-quality hardware can significantly improve the sensing accuracy.

2.2.2 *Microphone Array*

Microphone array is an array formed with a group of omnidirectional microphones, located at different positions in space, according to certain shape rules. It is a device for spatial sampling of spatially propagated sound signals. According to the distance between the sound source and the microphone array, the array can be divided into the near-field model and the far-field model. According to the topology of the microphone array, it can be divided into linear array, cross array, plane array, spiral array, and so on. After the microphones are arranged according to the specified requirements, the corresponding algorithm (arrangement and algorithm) can be used to solve many acoustic problems, such as sound source localization, abnormal sound detection, sound recognition, speech enhancement, whistle capture, and so on.

According to the distance between the sound source and the microphone array, the sound field model can be divided into two groups: the near field model and the far field model. The near-field model regards the sound wave as a spherical wave, which considers the amplitude difference between the received signals of the microphone elements. The far-field model regards the sound wave as a plane wave, which ignores the amplitude difference between the received signals of any array element, and approximately considers the difference between the received signals. Obviously, the far-field model is a simplification of the actual model, which greatly simplifies the processing difficulty. The general speech enhancement method is based on the far-field model. There is no absolute standard for classifying the near-field model and the far-field model. It is generally considered to be a far-field model when the distance between the sound source and the reference point of the center of the microphone array is much greater than the signal wavelength; otherwise, it is a near-field model. Let the distance between the adjacent array elements of the uniform linear array (also known as the array aperture) be d, and the wavelength of the highest frequency speech of the sound source (that is, the minimum wavelength of the sound source) is $\lambda_{\min}$. If the distance from the sound source to the center of the array is greater than $\frac{2d^2}{\lambda_{\min}}$, then it is a far-field model; otherwise it is a near-field model, as shown in Fig.7.

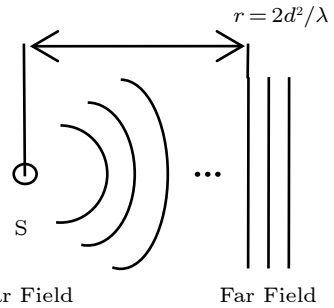


Fig.7. Near-field model/far-field model. S: signal source.

3 Technologies

3.1 Chirp Signal and FMCW

3.1.1 *Chirp Signal*

A chirp signal is a signal whose frequency increases (up-chirp) or decreases (down-chirp) as the signal changes. A linear chirp signal is expressed as

$$s(t) = A\cos(2\pi(\frac{f_{\min}t + kt^2}{2}) + \varphi),$$

where $f_{\min}$ is the initial frequency, A is the maximum amplitude, φ is the initial phase, and k is the modulation coefficient or chirp tweet rate. During the sensing, the chirp signal is transmitted repeatedly, for which it is also called as the frequency modulated continuous wave (FMCW). The time and frequency domains of a chirped signal are shown in Fig.8. An auto-correlated chirped signal produces sharp and narrow peaks whose temporal bandwidth is inversely proportional to the signal bandwidth, a property also known as the pulse compression. Since the energy of the signal does not change during the pulse compression, the concentration of the signal power in a narrower time interval results in a peak signal-to-noise gain proportional to the product of the signal band-

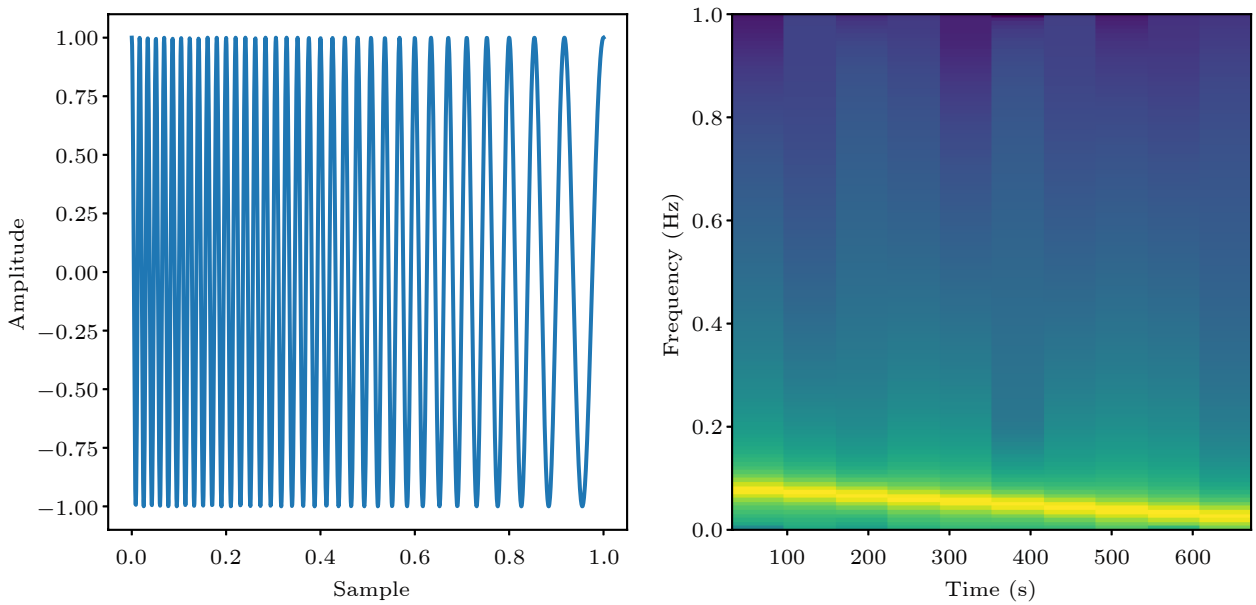


Fig.8. Chirp signal representation in time and frequency domain. (a) Linear chirp in time domain. (b) Spectrum of the chirp signal.

width and duration. Therefore, acoustic sensing systems using chirped signals are robust to dynamic channel conditions, such as Doppler effects, insensitive to Doppler effects, resistant to strong background noise or interference, and resilient to multipath fading.

3.1.2 FMCW

FMCW is a frequency-modulated continuous wave. It transmits a chirp signal. The frequency of the signal increases linearly within a predetermined period. The signal touches the reflector of the environment and returns to the receiving end after a period of delay, so that the delay can be determined by comparing the frequencies of the received and transmitted signals. As shown in Fig.9, if f is the frequency difference, the time delay can be obtained as $t = f/k$, where k represents the slope of the line that can be obtained as $k = (F_2 - F_1)/T$. F_2 and F_1 represent the upper and lower limits of the wave frequency, respectively, and T is the time of one cycle of the wave signal. Accordingly, the displacement of the target unit can be obtained from $s = v_t/2$ and v represents the propagation speed of the wave in air. In the time domain, the FMCW can be formulated as follows:

$$b = A\cos 2\pi\left(\frac{F_1 + F_2}{2}t + \frac{(F_2 - F_1)(t - N \times T)^2}{2T}\right),$$

where A represents the amplitude of the wave and N means the number of cycles.

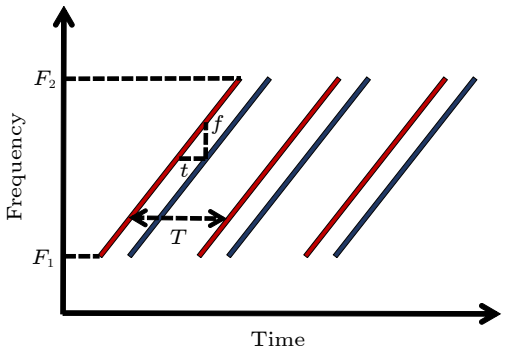


Fig.9. Schematic diagram of FMCW-modulated acoustic wave.

3.2 CIR

In WiFi, CSI is CFR (Channel Frequency Response) in the frequency domain. CIR (Channel Impulse Response) is obtained by IFFT (inverse fast Fourier transform).

In acoustics, CIR represents the propagation of acoustic signals under the combined effects of scattering, fading, and power attenuation of the transmitted signals. When the sound signal $S(t)$ is transmitted through the loudspeaker, it propagates into the microphone through multipath and is accepted as $R(t)$. If $h(t)$ is the CIR of the sound signal propagation channel, then $R(t) = S(t) \times h(t)$, where $\times$ is the convolution operator. Since the received acoustic signal is represented discretely, in the actual application sce-

nario, the formula is $R[n] = S[n] \times h[n]$. In order to solve $h[n]$, the least squares channel estimation method is usually used. It is assumed that the speaker transmits the known signal $\boldsymbol{m} = (m_1, m_2, ..., m_N)$, and the microphone receives the signal $\boldsymbol{y} = (y_1, y_2, ..., y_N)$. For the length of the transmitted signal, a cyclic training matrix $\boldsymbol{M}$ is generated from the vector $\boldsymbol{m}$, where the dimension of $\boldsymbol{M}$ is $P \times L$, i.e., the vector $\boldsymbol{m}$ is repeated P times to form the P rows of $\boldsymbol{M}$, and the CIR is estimated as $h = (\boldsymbol{M}^H \boldsymbol{M})^{-1} \boldsymbol{M}^H \boldsymbol{y_L}$ (H means conjugate transpose), where $\boldsymbol{y}_L = (y_{L+1}, y_{L+2}, ..., y_{L+P}), N = L + P$. In order to satisfy the constraints, the lengths of L and P of the CIR need to be determined manually. Increasing L can estimate more channel states, but it may reduce the reliability of the estimation. For some practical applications of CIR, UltraGesture[85] provides a resolution of 7 mm, which is sufficient to identify slight finger movements. It encapsulates the CIR measurements into images with better accuracy than Doppler-based schemes, and it can run on commercial speakers and microphones that have already existed in most mobile devices without requiring any hardware modification. RobuCIR[86] uses a frequency hopping mechanism to mitigate frequency of the selective fading to avoid any signal interference. This high-precision CIR can recognize 15 gestures.

3.3 Doppler Shift

When there is a relative motion between the wave source and the observer, the frequency of the wave received by the observer is not the same as the frequency emitted by the wave source. The wave here can be a mechanical wave or an electromagnetic wave. As the wavelength is compressed, the frequency increases. Otherwise, the frequency decreases. Suppose that the observed frequency of the observer is f, the frequency of the wave source is f_s, the propagation velocity of the wave in the medium is c, the relative velocity of the wave source is v_r, and the velocity of the wave source is v_s. The formula for computing frequency f of the wave source is (in the following formula, "/" stands for "or"):

$$f = \left(\frac{c +/- v_r}{c -/+ v_s}\right).$$

In specific applications, taking the gesture recognition based on the Doppler effect as an example, both microphone and the speaker are integrated in a smart device, where the speaker acts as a wave source to emit ultrasonic waves. Assume that the velocity of the gesture movement relative to the wave source is v_h, the propagation rate of the sound wave in the air medium is v, and the frequency of the sound wave emitted by the speaker wave source is f_s. When the hand is the receiver of the sound wave, the received sound wave frequency f_h can be expressed as follows (in the following formula, "/" stands for "or"):

$$f_h = \left(\frac{v +/- v_h}{v}\right) \times f_s.$$

The sound waves reflected by the movement of the hand and received by the microphone can be regarded as emitted by the hand. The hand is used here as a new wave source, and the microphone is used as the receiver. At this time, the receiving frequency difference Δf generated by the hand movement can be expressed as (in the following formula, "/" stands for "or"):

$$\Delta f = |f - f_s| = \frac{2v_h}{v -/+ v_h} \times f_s.$$

3.4 MFCC

The Mel-scale Frequency Cepstral Coefficient (MFCC) is one of the most commonly-used audio feature parameters[115]. For audio recognition, the most basic operation is to extract feature parameters from speech information. In other words, it is to extract the identifiable components from the audio information that can characterize the entire audio signal, and then to discard other information, such as background noise, which is the most basic audio feature extraction. The Mel scale is a nonlinear feature that can be used to characterize the human ear's sensory judgment to sound changes at equal distances. Since the human ear's sensing of audio signals has nonlinear correlation characteristics, the relationship between the auditory sensing ability and its frequency satisfies the logarithmic expression when the speech signal exceeds 1 kHz. The corresponding relationship can be expressed as:

$$m = 2\,595 \log_m \left(1 + \frac{f}{700}\right),$$

where m represents the Mel frequency of human auditory sensing, and f is the frequency (in Hz) of the audio signal. If the distribution of the Mel scale is uniform, the gap between the actual frequencies becomes larger and larger. In the Mel frequency domain, the human auditory sensing ability is obviously lin-

early related, indicating that the two audio frequencies are in the Mel frequency domain. The audio feature parameter extracted according to the auditory sensing ability of the human ear has an excellent effect on distinguishing different key audios. The Mel spectrum can be obtained by preprocessing the signal, framing, windowing, and Fourier transform, passing through the Mel filter, and then performing logarithmic operations and discrete cosine transform (DCT).

3.5 Wireless Ranging

In HCI, it is often necessary to use a wireless signal to measure the distance between two objects or devices. The location algorithms for wireless sensor networks depend on a variety of distance measurement techniques. There are many factors that affect the accuracy of a location algorithm. Therefore, the selection of the infinite distance algorithm should be based on various applications, such as network structure, sensor density in the area, number of anchors, and geometry of the measurement area. However, the type of measurements and the corresponding accuracy fundamentally determine the accuracy of the location algorithm. Common wireless ranging technologies are time of arrival (ToA)/time of flight (ToF), time difference of arrival (TDOA), angle of arrival (AOA), phase difference, and RSSI attenuation models.

3.5.1 Time of Arrival (ToA)/Time of Flight (ToF)

ToA is the time at which a signal travels between the transmitter and the receiver.

Combining with the propagation speed v of the signal, the propagation distance $d = v \times ToA$ can be calculated. In acoustics, $v = 344$ m/s. This method requires precise synchronization between the transmitter and the receiver to avoid measurement errors.

Due to the large environmental impact, the ToF range selected based on signal strength is unstable. In practice, when the speed of signal propagation in the medium is known, the signal propagation time (also called the flight time) is often used to measure the signal distance. There are three common ways to measure ToF, which are time synchronous measurement, signal reflection, and waves velocity difference.

Time Synchronous Measurement. Assuming precise time synchronization between the sender and the receiver, when the sender sends a packet to the receiver, the receiver records the arrival time of the signal after receiving the message, and obtains the ToF by subtracting the sending time stamp from the receiving time stamp. The pass distance is derived as $d = c \times t$, where d is the distance from the sender to the receiver, c is the speed at which the signal travels, and t is the measured flight time.

Signal Reflection. Since it is difficult to directly use ToF measurements, it has been proposed to use signal reflection to calculate the flight time in order to avoid the time synchronization between the transceiver and the receiver. This is usually done in two ways. The first method is the direct reflection of the transmitted signal, which uses the range finder as a reflector. For example, to measure ToF using FMCW as the transmission signal, this method requires that the transceiver can operate in full duplex mode and the ranging object has a certain volume. The second method is to use two devices as the sender and the receiver respectively, where the sender transmits the signal, the receiver receives it and waits for a period of time to return the same wave, and finally the sender records the time of reply to calculate the distance. The formula for the same is as follows:

$$d = (v \times (t_1 - t_0 - \Delta t))/2,$$

where t_1 is the moment when the sender receives the reply from the receiver, t_0 is the time when the sender sends the signal and Δt is the time when the receiver returns the same wave. This method requires bidirectional communication between the sender and the receiver, which is divided into two parts—one-sided bidirectional ranging, which measures the distance from a single-sided round-trip communication, and two-sided bidirectional ranging, which reduces the one-sided bidirectional error. However, this method is still affected by the clock drift of both devices.

Waves Velocity Difference. This method uses the wave velocity difference between the two signals to measure the distance. The sender sends two different wireless signals at the same time, and then records the arrival time of the two signals at the receiver. Based on the different arrival time, the distance between the sender and the receiver can be calculated as follows:

$$d = \frac{v_r v_s (t_s - t_r)}{v_r - v_s}.$$

In the above formula, r and s are two types of signals, t_r and t_s are the time when the two signals reach the

receiving end respectively, and v_r and v_s are the corresponding speeds at which the two signals propagate respectively. The formula for calculating the distance can be simplified if the propagation speed of one signal is much smaller than that of the other. Assuming that the s-signal is much smaller than the r-signal, it can be simplified.

3.5.2 Time Difference of Arrival (TDOA)

In fact, TDOA is a modification of the ToA algorithm, which uses the time difference between the signals reflected by the object to be measured and the different receivers to analyze the distance difference.

Assuming that the time difference between the arrival of the two objects to be measured at different transmitting (receiving) sections is Δt, multiplying by the speed of signal transmission v, the travel distance difference Δd of the wireless signal to different base stations can be found. The TDOA algorithm is mainly divided into two cases: the first case is that to send data from the object to the receiver, the receiver receives the data, gets the time stamp, and then can calculate the distance difference between the two objects to be measured to each receiver. The other case is that all the transmitters send data to the object to be measured at the same time, and the time difference of the signal arrival is recorded by the object to be measured.

3.5.3 Angle of Arrival (AOA)

The core idea of AOA is to calculate the relative orientation of the receiving and transmitting nodes through a hardware device. It usually relies on multiple antenna arrays. In a multi-antenna array, for signals arriving at the antenna array at different angles, there will be a time difference between the individual antennas, which corresponds to the angle of arrival (AOA). There are three common methods to obtain the angle of arrival: the method using signal time delay estimation; the method using Multiple Signal Classification (MUSIC), and the method using beamforming (Beamforming).

Method Using Signal Delay. The time delay of the received signal of the array is determined, and the information of the angle of arrival is obtained by combining the propagation speed of the signal and the geometric distribution of the array. This method has more applications in acoustics, such as working[116, 117].

Method Using Multiple Signals. It is an algorithm based on subspace decomposition, which uses the orthogonality of the signal and noise subspaces to construct a spatial spectral function and estimate the parameters of the signal by spectral peak search. The basic idea is to decompose the covariance matrix of the output data of an arbitrary array into features, so as to obtain the signal subspace corresponding to the signal component and the noise subspace orthogonal to the signal component, and then to use the orthogonality of these two subspaces to estimate the parameters of the signal, such as the direction of incidence, polarization information and signal strength, and so on. An example is the Music method. This method has high directional accuracy, high resolution for the lateral direction of the signal within the antenna beam, is suitable for short data cases, and can be processed in real time using high-speed processing techniques, but when the wavelength is less than twice the high-frequency component of the array element spacing, the array element cannot receive the signal; and in radar systems, with the increasing requirements of anti-stealth and resolution of the target, the assumption of narrow-band signals is no longer in line with the actual situation.

Beamforming. This method uses antenna arrays to enhance signals in different directions, and detects signal strength information in different directions to determine the arrival angle. There are two common beam shaping methods, one is based on delay sum and the other is based on SRP (Steered-Response Power). After the arrival angle is obtained by the three methods mentioned above, the position of the point to be measured can be obtained by locating it with the theorem of triangle. For example, Gallo and Magone[118] suggested to associate the measured WiFi RSSI with the electronic compass data on a smartphone to calculate the angle of arrival of the smartphone signal and the movement of the smartphone.

3.5.4 Phase Difference

The received signal phase method uses the carrier phase (or phase difference) to estimate the distance[119]. This method is also known as the phase of arrival (POA)[120]. Assume that all transmitters emit a sinusoidal signal with a frequency of F and a phase offset of zero. In order to determine the phase of the signal received at the target point, the signal sent from each transmitter to the receiver requires a limited transmission delay. As shown in Fig.10, transmitter stations B to E are placed in a specific location in a fic-

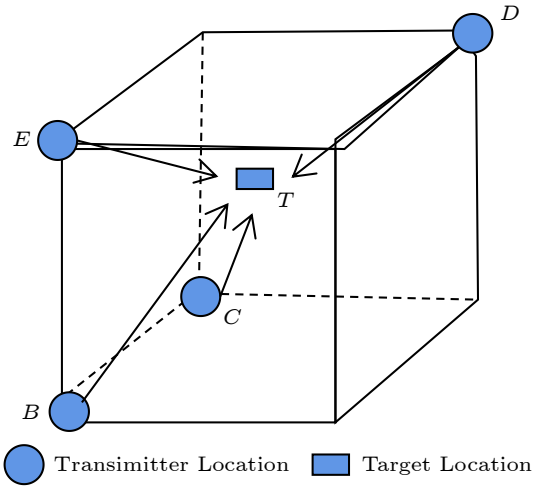


Fig.10. Positioning based on the signal phase.

tional cubic building and T is the target location. The delay is expressed as a fraction of the wavelength of the signal as $S_i(t) = \sin(2\pi ft + \phi_i)$, where $I_{\phi_i} = (2\pi f D_i)/c$, $i \in (B, C, D, E)$ and c is the velocity of light. As long as the wavelength of the transmitted signal is greater than the diagonal of the cubic building, i.e., $0 \leqslant \phi_i \leqslant 2\pi$, the distance can be estimated as $D_i = c\phi_i/(2\pi f)$.

For indoor positioning systems, the signal phase method can be used in combination with ToA/TDOA or the RSS method to fine-tune positioning. However, the receiving signal phase method needs to overcome the ambiguity of the carrier phase measurement. It requires LOS signaling; otherwise it will cause more errors in the indoor environment.

3.5.5 *RSSI Attenuation Model*

This ranging method is mainly used in WiFi. Path loss refers to the loss of radio signals in transmission. Generally, RSSI is affected by four factors: transmission power, path attenuation, reception gain and system processing gain. Therefore, the RSSI can be expressed as:

$$RSSI = Tx_{\text{Power}} + Path_{\text{Loss}} + Rx_{\text{Gain}} + System_{\text{Gain}},$$

where $Tx_{\text{Power}}, Path_{\text{Loss}}, Rx_{\text{Gain}}$, and $System_{\text{Gain}}$ represent the four influencing factors of transmission power, path attenuation, reception gain, and system processing gain, respectively.

Due to the multipath effect, signals from different propagation paths have different delays and energy decay. Intuitively, the farther the distance, the lower the signal strength (it can also be seen in the free-propagation-based path consumption model in theory). The distance measurement using the signal strength is based on a free-space propagation path consumption model[①], which is used to predict the strength of the received signal in a distance-of-sight environment without any obstacle between the receiver and the transmitter.

For the propagation of wireless signals in free space, the path consumption has the same receiving and transmitting distance, which will cause power change loss in the distance of 100 m to 1 000 m. In this scenario, the formula for calculating the received power in logarithmic terms is:

$$10\lg(P_r) = 10\lg(c_0 P_t) - 10n\lg(r) = A - 10n\lg(r),$$

where P_r and P_t are the receiving power and the transmitting power of the wireless signal respectively, r is the distance between transceivers, c_0 is a constant related to antenna parameters and signal frequency, and n is the propagation factor whose value depends on the environment in which the wireless signal is propagated. Since the transmission power is known, $A - 10n\lg(r)$ can be viewed as the power to transmit 1 m long-time received signal. From the above formula, the values of constants A and n can be obtained, which determine the relationship between the received signal strength and the signal transmission distance.

As shown in Fig.11, if the signal propagation factor n is fixed, the intensity of the wireless signal decreases rapidly when it travels near the field, and slowly and linearly when it travels long distances. When A is fixed, the smaller the attenuation of the signal in the propagation process, the longer the signal can travel[①]. The RSSI measurement uses the theoretical or empirical loss of the signal propagation model, calculates the distance between the receiver and the receiver through the path distance formula, and calculates the signal loss during the transmission.

To sum up, if we know n and A received when the wireless transceiver nodes are together for 1 m, we can calculate the distance. Generally, these two values are empirical, which are closely related to the used hardware nodes and the environment in which wireless signals are propagated. Therefore, before ranging, these two empirical values must be calibrat-

[①]Wikipedia contributors. Log-distance path loss model—Wikipedia, the free encyclopedia, 2022. https://en.wikipedia.org/wiki/Log-distance_path_loss_model, Aug. 2022.

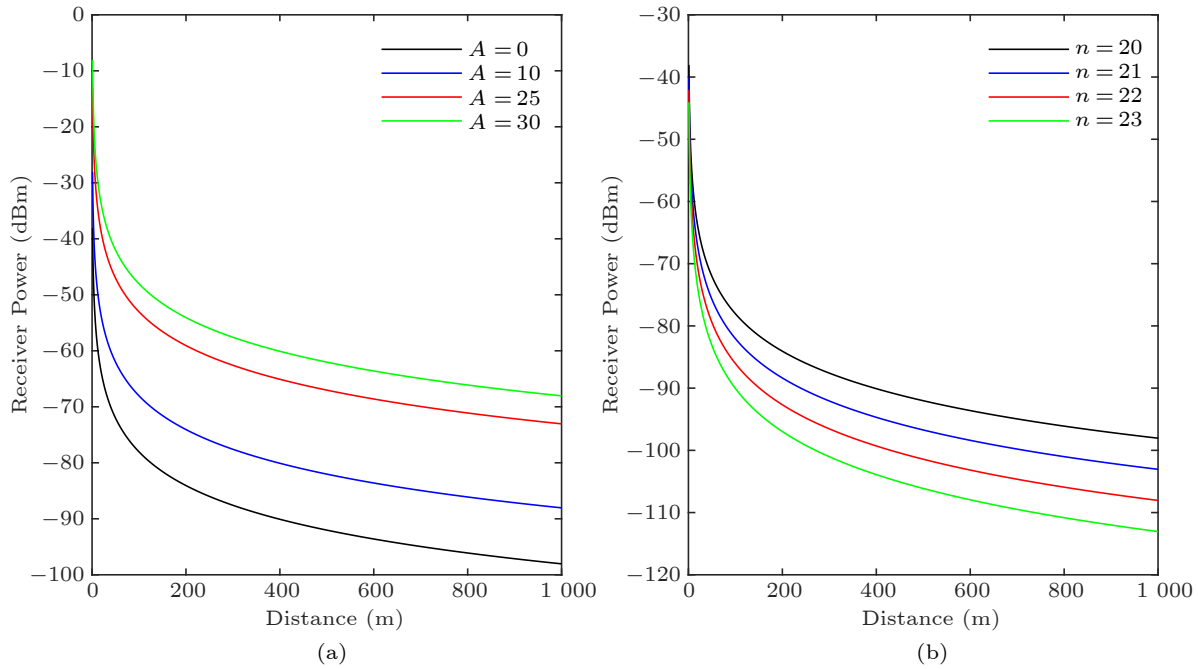


Fig.11. Attenuation curves of electromagnetic wave energy with distance in free space. (a) Fixed A. (b) Fixed n.

ed well in the application environment. The accuracy of the calibration is directly related to the accuracy of ranging and positioning. Therefore, measuring distances using RSSI can encounter the instability of RSSI, which can be smoothened by designing filters, e.g., the mean filter and the weighted filter.

3.6 Wireless Tracking

Visually, tracking can achieve the goal of tracking a target by constantly calculating its location. However, in practice, we can achieve the effect of tracking without calculating the specific location. As long as the initial position of the target is known, tracking and positioning of the target can be achieved continuously. The tracking technology can use the signal characteristics of frequency, phase, etc. of the object to be measured to know the change in its spatial position, so as to achieve the purpose of tracking. In target tracking, often the object to be measured is tracked based on the change in distance. In the field of acoustics and WiFi, three tracking methods are commonly used: Doppler tracking, FMCW (Frequency Modulated Continuous Wave) tracking, and signal phase tracking.

3.6.1 Doppler Tracking

Based on the Doppler effect, the velocity of motion of the receiver can be calculated as follows:

$$v = \frac{f - f_0}{f_0} c = \frac{\Delta f}{f_0} c,$$

where f_0 denotes the frequency at which the source emits sound and c denotes the speed of sound. Δf denotes the frequency change produced by the Doppler effect. In the time window, the frequency domain information of the signal can be obtained by SSFT (short time Fourier transform), which can then be used to obtain the frequency of the signal. If the original signal is fixed, then it can be obtained as Δf.

The variation in the distance of the receiver's motion can be obtained by integrating the velocity with time, that is $d = \int_0^T v \mathrm{d}t$, where T represents the time from the beginning to the end of the receiver motion. If the initial location of the recipient is known, the final location of the target can be calculated to track the target.

3.6.2 FMCW Tracking

The basic principle of FMCW (Frequency Modulated Continuous Wave) is to emit a frequency wave continuously whose frequency increases or decreases periodically. The frequency of FMCW periodically increases (from $f_{\min}$ to $f_{\max}$) or decreases (from $f_{\max}$ down to $f_{\min}$). FMCW tracking can be conducted in two ways. The first way is to measure ToF by using the frequency offset from the mixing of the reflected and transmitted signals, and then to measure the dis-

tance between the source and the reflected object. The second way is to measure the change in distance by the frequency difference between the received and the transmitted signals.

If the first method is used for distance measurement, an equipment is required to eliminate the emitted strong signal when receiving the reflected signal, so as to avoid its interference to the latter. The distance measurement by using the second method requires precise clock synchronization between the transmitting and receiving devices. These two methods are difficult to implement in actual scenarios. We can avoid the precise time synchronization and analyze the change in distance by using the FMCW signal received by the receiving device and virtually sending the signal to track the movement.

3.6.3 *Signal Phase Tracking*

Phase location tracking is a common method in the positioning and tracking of the Internet of Things. Especially in recent years, a series of phase-based positioning and tracking methods have emerged in many fields of research. The phase-based tracking methods will be improved continuously, and its basic principle is to measure the phase change of a signal. Assume that the fixed frequency signal sent by the signal source is $R(t) = A\cos(2\pi ft)$, the signal propagates through path p, and the length of the propagation path changes with time to $d_p(t)$. Then, the received sound signal via path p can be expressed as:

$$R_p(t) = A_p(t)\cos\left(2\pi ft - \frac{2\pi f d_p(t)}{c} - \theta_p\right),$$

where $A_p(t)$ is the amplitude of the received signal, $2\pi f d_p(t)/c$ is the phase offset caused due to the propagation, and c is the sound speed. θ_p is the phase offset caused by hardware delay, half-wave loss due to the reflection, etc., which can be considered constant and does not change with time. If the phase information can be obtained from the received signal $R_p(t)$, the change of the propagation path length $d_p(t)$ can be obtained, and the receiver's motion path can be tracked.

If multiple sound sources with different frequencies are used to send sound waves at different locations, the spatial position of the device can be calculated based on the distance between the device and different sound sources over a period of time when the starting location is known, so as to achieve high-precision positioning and tracking.

4 Application

4.1 Behavior Recognition and Tracking

The framework for Subsection 4.1 is shown in Fig.12.

4.1.1 *Daily Activity*

Human action recognition is one of the important research contents in intelligent application. Daily behavior detection consists of activity recognition and tracking. Activity recognition refers to the classification and recognition of human behaviors according to certain algorithms, such as the deep learning methods by measuring certain signal data generated by human while performing various actions. Tracking is more about tracking the activity process and trajectory through physical methods. By accurately identifying and tracking the human behavior, the quality of

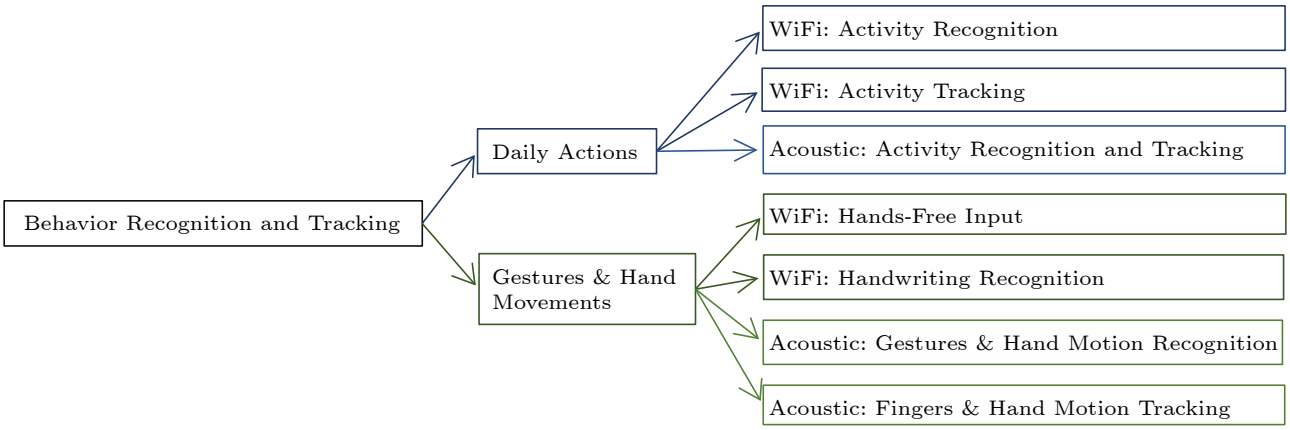


Fig.12. Behavior recognition and tracking structure diagram.

human-computer interaction can be improved and the scope of the intelligent application can be expanded. It is the future development trend of intelligent life, which will have great application prospects and economic value for the research of smart home, intelligent teaching, and medical assistance. According to Pedretti and Early[116], daily behaviors mainly include eating, bathing, dressing, going to the bathroom, walking and such other behaviors. Nowadays, there is a growing trend towards intelligent home, in which WiFi and acoustic play a crucial role, which makes many applications to realize human behavior perception emerge. The action recognition and tracking based on WiFi and acoustic sensing are introduced below respectively.

1) *Activity Recognition*

In the last decade, activity recognition was one of the hottest research topics in wireless sensing. Focusing on large human movements, activity recognition also facilitates exciting applications, such as motion detection, fall detection, and multi-activity classification where multiple activities are combined for detection. Different activities identify different targets with different emphasis. Next are three common types of applications.

Some applications aim to detect the presence of moving objects. Traditional methods, like [43], use a predetermined threshold to distinguish such motion-induced burst signals. However, since wireless signals are sensitive to the environment and environmental changes, they require different calibrations for different environments. WiDetect[42] uses statistical theory to model the signal state in a scattering-rich environment. For example, WiBorder[44] analyzes CSI conjugate multiplication in detail and uses a sensing boundary determination method. Therefore, it can accurately determine whether a person has entered into a given place.

Some applications aim to detect specific actions of everyday activities. Unlike their predecessors, these applications need a distinction that clearly distinguishes between specific actions and others. A typical example of such an application is the fall detection. Falling is a major threat to the life and health of elderly people. Timely detection and rescue of falling can greatly reduce the consequences of falling. Han *et al.*[9] proposed WiFall, which was the first to propose a WiFi-based fall detection mechanism. In this system, the Local Outlier Factor (LOF) algorithm is used to detect the outlier data in the CSI stream and the one-class Support Vector Machine (SVM) is used to identify the descent action. After that, more and more researchers began to pay attention to this problem. For example, RT-Fall[10], FallDeFi[12] and DeFall[13] have made use of more powerful data processing methods and achieved better detection performance. More recently, researchers[14–17] have proposed new solutions to solve the three major challenges of fall detection: user-related, environment-dependent, and motion interference. For example, since the AOA (angle of arrival) reflected by the human body does not depend on the environment and the subject, AFall[14] proposes to use the Dynamic-Music algorithm[56] to model the relationship between human falls and changes in the AOA of WiFi signals reflected by the human body. Therefore, when the environment changes slightly, the performance of falling can remain stable.

There are also applications that detect multiple activities in daily life. This type of applications is related to the category of activities. In general, the extracted features are different for different scenarios with different activity contents. TW-See[48] is a human activity recognition method based on wall-penetrating passive CSI. The method tracks six human activities: walking, sitting, standing, falling, arm swinging, and boxing. The method uses two key techniques for tracking. First, the inverse Robust PCA (OR-PCA) method is applied to obtain the correlation between human activities and CSI value changes, and then a normalized variance sliding window method is applied to segment the OR-PCA waveform of human actions. On the other hand, Wi-Motion[45] divides a human motion into macroscopic and microscopic motions. It uses a human motion detection method based on CSI amplitude and phase information, which minimizes the random phase offset and uses different signal processing methods to obtain clean datasets.

In many previous studies, machine learning models were used to classify signals by directly comparing them. For example, WiSee[46] and CARM[41] convert the wireless signal into the frequency domain, then get environment-independent DFS (Doppler Frequency Shift) (representing motion speed information), and finally use the model to classify. However, more and more recent studies have been done to improve the robustness of the environment. EI[38] and DeepMV[39] combine DNN with domain discriminator to extract the general characteristics of ac-

tivities. Alternatively, in-depth learning models are also combined to identify activities. For example, DeepSense[37] proposes a HAR method based on CSI and in-depth learning. The proposed activities (walking, standing, lying, running and empty running) were classified by using a long-term recursive convolution network of self-encoders with an accuracy of 97.4%. Chen *et al.*[40] proposed a CSI macro activity recognition method based on deep learning. This method divides macro-activities into six categories: fall, walk, sit, lie down, stand up and run, and uses two-way long and short memories (ABLSTM) to learn the representative characteristics of the original CSI from two directions.

2) *Activity Tracking*

In WiFi, both the large motion and the small motion tracking take place. The large motion tracking mainly includes the calculation of steps, human position tracking, the number of people derived from human position calculation and other activity tracking applications. The tracking of minor movements mainly involves lip reading. Location-aware studies focus on the coordinates of the target people, population counts, and the number of people in a specific area.

In the application of step counting, the corresponding signals from legs or feet may be covered by stronger signals reflected from the trunk. Therefore, it is difficult to track steps directly. WiStep[55] identifies walking modes through time-frequency analysis, separates the leg-induced signals from the trunk-induced signals using wavelet transform, and calculates steps using a peak detection method. WiStep's step count accuracy in the two-dimensional space is higher than 87.6%.

Since WiFi sensing enables non-inductive localization and tracking, human localization has attracted a lot of researchers' attention in the field of WiFi human sensing. In the application of human position tracking, many studies have proposed the use of CSI for indoor localization and activity recognition (see Subsection 4.3.2 for details). As human location has attracted more and more attention, WiFi wireless sensing has also emerged into applications derived from human locations. Examples include crowd counts that focus on the exact number of people in a particular place. Since CSI provides more fine-grained channel information (i.e., amplitude and phase information) through multiple subcarriers, it can be known from the work [35] that the number of subjects has different impact on the amplitudes of CSI data on different subcarriers. More people can induce higher CSI variance through WiFi links. An integrated crowd management system was proposed in [35]. Unlike previous studies, the proposed system uses existing WiFi traffic and uses robust semi-supervised learning to estimate population density, population count, walking speed and direction. The method can be easily extended to new environments. The human intrusion described in Subsection 4.4.3 is also a derived application of human localization.

In addition, human micro-motion tracking is also used. For example, the Smokey[47] system uses CSI to detect smoking by monitoring different smoke-related actions, such as holding, lifting, sucking, dropping, inhaling and exhaling. It is also evaluated in an indoor environment with multiple users and achieves good performance. Human tiny motion tracking based on CSI can be extended to sense lip movements. WiHear[33] uses a directional antenna to send WiFi signals in the direction of the target user's face and recognizes 14 syllables and 32 words with an accuracy of 91%. WiTalk[34] obtains mouth movement information by measuring DFS based on the scene of lip reading while making a phone call, and uses DTW to classify the 12 syllables. The relative position between the phone and the mouth is constant where the user is or which direction he/she is facing. The accuracy of WiTalk is higher than 82.5%.

Compared with acoustic, the current research on WiFi in daily behavior recognition is more in-depth and extensive. But the recognition and tracking of daily behavior based on sound wave sensing also involves some aspects. Early work SoundSense[73] is the first general-purpose sound sensing system specifically designed for mobile phones by modeling sound events on mobile phones and monitoring people's daily activities, such as walking, driving, riding in cars, and so on, demonstrating its ability to identify meaningful sound events that occur in the user's daily life. The extracted acoustic features are the processed phase, signal strength, and frequency and bandwidth, which are then classified into specific categories using decision trees and Markov models. The feature extraction relies on human knowledge and experience. BodyScope[74] uses a wearable activity recognizer based on commercial headphones, extracts MFCC of the captured sound and then uses SVM to classify and monitor oral movements, such as eating, laughing, talking, and so on. In EarSense[75], it was found that the actions of the teeth, i.e., tapping and sliding,

create vibrations in the jaw and skull, and these lineups are strong enough to propagate to the face and create a vibrational signal at the earphones. Six different tooth movements can be detected in real time by analyzing the two headphone signals. HearFit[76] turns smart speakers into active sonar, designs a fitness detection method based on Doppler frequency shift, and uses short-term energy to segment fitness actions as fitness quality guidance, which can assist to improve fitness and prevent injuries. It can detect 10 movements with and without dumbbells and give accurate statistics in various environments. In a recent study, Liang and Thomaz[77] proposed an audio-based activity recognition framework that can leverage millions of embedded features from public online video sound clips. Based on the combination of oversampling and deep learning methods, this framework does not require further feature processing or outlier filtering like previous work. Fifteen daily activities of 14 participants achieved an average recognition accuracy of 64.2% and 83.6% in Top-1 and Top-3, respectively. The work[78] performs unsupervised tracking of daily household activities through acoustic sensing, a system that uses captured sound to identify shifts in typical activities without the need for activity tags. It relies on sound embedding, through pre-trained models and a new dimensionality reduction algorithm, and applies dynamic time warping for pattern matching. The accuracy reported in this paper reached a precision of 0.99 and a recall of 0.95.

4.1.2 Gestures and Hand Movements

Gestures and hand movements are the most important ways for human interaction. From the technical route, gesture recognition is mainly based on machine learning or deep learning after extracting the wireless signal characteristics. Such applications are data-driven and focus on identifying the contents of hand movements. For finger-hand motion tracking, physical modeling is the key to its motion process tracking, and the focus also includes the content of the action at each moment in the process itself, which provides more flexible functions to support a variety of human-computer interaction programs. The following first describes the practical applications of WiFi-based and sonic-based sensing in the areas of gestures and hand movements. Gesture/hand movement recognition in WiFi work is mainly divided into two parts: hands-free input and handwriting recognition.

1) *WiFi: Hands-Free Input*

The gesture recognition technology is an important means of human-computer interaction, which is crucial to the development of human-computer computing. Traditional gesture recognition methods include computer vision, infrared recognition and dedicated sensors. However, computer vision methods are easily limited by light conditions, and infrared and sensor recognition are complex to deploy, inconvenient to carry, and expensive also. Therefore, they cannot well meet the application scene of smart home. As WiFi-based motion sensing is gradually maturing, WiFi-based hand motion recognition has attracted widespread attention.

In the application of recognizing freely changing gestures, the wireless signal changes are much smaller than other behavior recognition signals as the hand and finger gestures are smaller than the body movements. Such small signal changes can easily be masked by strong signal changes caused by other parts of the person's body or moving objects. Extracting these changes is not a simple task. Therefore, researchers often limit the interaction range to a small space to amplify the hand and finger signals. WiG[20] is an early work on WiFi gesture perception, which extracts four features from CSI amplitude to train an SVM classifier to distinguish four common gestures (i.e., right, left, push and pull). The authors of WiKey[18] observed that each keystroke has a unique CSI pattern due to different gestures. Therefore, the keystroke recognition accuracy rate reaches 93.47%. However, the system is very sensitive to the relative position changes between the user and the device. Then, Mudra[19] implements accurate detection of finger-level gesture signals independent of the position. Mudra performs this with the help of interference cancellation techniques based on the differences in received signals between antennas at different positions. In the past two years, researchers[21–29] also proposed various solutions to the challenges of environment, user or motion interference in gesture recognition in the WiFi field. For example, WiHF[23] improves the seam carving algorithm to extract motion change patterns in real time to provide a solution for motion changes. [25] establishes a WiFi frequency theoretical model to demonstrate that the commonly-used motion velocity and motion direction features are position-dependent, and address the environment dependence by extracting two position-independent features.

2) *WiFi: Handwriting Recognition*

Existing work on handwriting recognition mainly uses wireless tracking in wireless signals. WiDraw[30] uses an over-the-air handwriting recognition method. This method utilizes the AOA of the wireless signal at the receiver and has an average accuracy of 91% for multiple writing.

There are some problems in gesture recognition. The most important requirement is to distinguish gestures accurately, where a major challenge is that hand-induced signals are much lower than those in other parts of the body. Existing systems often deploy sensors near their hands. However, application scenarios are limited. With this in mind, researchers can use the directional antenna and beamforming technology to enlarge the hand space like mmASL[117], or to explore more robust signal models, such as CSI-quotient[32]. The second challenge is to extract distinguishable features. Since gesture modeling is relatively difficult, we expect to use more data-driven models in future.

3) *Acoustic: Gesture and Hand Motion Recognition Based on Acoustic*

Since sound wave sensing is more fine-grained than WiFi and other sensing methods, it is more suitable for more complex and accurate fine-grained action recognition and tracking, such as gesture recognition and hand motion tracking. It has been widely investigated and applied. The application and related work are introduced below in detail from the two aspects of gesture (mainly finger motion) recognition and hand motion tracking. Human gestures and hand movements are the main ways for interaction. The frequency shift caused by the Doppler effect is the most common and direct method in the recognition technology based on sound wave sensing. Consider a person writing in air next to a smart device, such as a mobile phone as an example. The speaker of the mobile phone emits modulated ultrasonic waves. Due to the Doppler frequency shifting effect caused by the human writing, the frequency of the sound wave received by the mobile phone microphone will change. Such systems usually include steps, such as data acquisition and preprocessing, short-time Fourier transform, feature extraction, and classification and recognition. In addition to the Doppler frequency offset, there are other methods based on FMCW and CIR combining the distance between the target and the sound source. DopLink[79], AirLink[121], etc. are through the connection or interaction between device and device, relying on the user to hold the device and wave it to complete the pairing, information transfer and action recognition between devices. Compared with relatively distant and coarse-grained hand motion recognition, tighter and finer-grained finger gesture motion recognition is increasingly important in human-computer interaction. The Dolphin system designed by Yang *et al.*[80] uses the built-in speaker and microphone to transmit and receive continuous 21 kHz ultrasonic signals, extract the frequency-domain features related to the Doppler effect, and use machine learning models to achieve up to the recognition of 17 volley gestures? SoundWrite[122] and SoundWrite II[83] describe handwritten features using amplitude spectral density and some other acoustic features, such as MFCC, and use *K*NN to match the captured features with labeled features in the database. Using the ZC sequence, a periodic pulse signal, as the acoustic signal for sensing gestures, VSkin[81] enables touch gesture sensing on all surfaces of the mobile device, not just the touch screen area, by measuring the amplitude and phase, which use structure-borne sound (that is, the sound that travels through the structure of the device) and air-borne sound (that is, the sound that travels through air) to sense finger taps and movements, enabling fine-grained gestures on the back of the mobile devices based on induced acoustic signals. Wang *et al.*[123] proposed a dynamic speed warping (DSW) algorithm, based on the observation that the gesture type is determined by the trajectory of the hand component rather than the movement speed, by dynamically scaling the velocity distribution and tracking the movement distance of the trajectory. It can match gesture signals from different domains with ten-fold speed differences, achieving 97% accuracy using only one training sample for each gesture type from four trained users.

Regarding some recent research advances, the gesture recognition pursues more fine-grained, higher accuracy, smaller training set size, and more goals. With the development of techniques, such as the transfer learning[124], few-shot learning[125], and generative adversarial networks[126], these techniques have also been applied to the field of gesture recognition. In [127] a transfer learning based convolutional neural network was used for gesture recognition, whose accuracy is better than those of the existing work on sign language digits and Thomas Moeslund's gesture recognition datasets. In [128], EMG was used for the recognition of learning gesture of small samples. Al-

though it is different from sound-based sensing, its method is easy to reuse on the spectrogram obtained by sound-wave sensing. In [129], a generative adversarial network GAN was applied, and a scene transfer network was developed, which not only uses the real samples of the scene, but also uses real samples from another available scene to generate virtual samples to train and test a small sample dataset on the mmWave-based data and test platform. Although the carriers and methods are different, the above methods also have some inspirations for gesture recognition based on sound wave sensing. Furthermore, Ultragesture[85] is based on the channel impulse response (CIR). CIR measurements can provide a resolution of 7 mm, which is sufficient to identify slight finger movements. It encapsulates CIR measurements into images with better accuracy than Doppler-based schemes, and it can run on commercial speakers and microphones that have already existed in most mobile devices without requiring any hardware modification. RobuCIR[86] uses a frequency-hopping mechanism to avoid signal interference by mitigating frequency-selective fading. This high-accuracy CIR work can recognize 15 gestures. AMT[130] defines a new concept of primary echo to better represent the target motion by using multiple speaker-microphone pairs, which perform multi-point localization of actions, detect primary echoes and filter out secondary echoes, eliminate target bulge multipath effects instead of assuming them as particles, improve tracking accuracy, and achieve multi-target tracking at the centimeter level. Aimed at the challenge of adaptively responding to expected movements instead of unexpected ones in real-time tracking movements systems of gesture recognition, Amaging[87] gives an independent sensing dimension of acoustic two-dimensional hand forming images. Amaging has multiplicative expansion of sensing capabilities and two-dimensional parallel hand shape and gesture-trajectory recognition, and its hand shape imaging performance and immunity to mobile interference have been verified through experiments and simulations.

4) *Acoustic: Finger and Hand Motion Tracking Based on Acoustic*

The next issue is about the finger and hand motion tracking. Physical modeling is the key to tracking its motion process. Motion tracking provides more flexible functions to support various human-computer interaction programs. The system designed by Yun *et al.*[131] transforms signals into an inaudible frequency band at different frequencies, and uses Doppler shift to estimate the speed and distance of hand movement. CAT[132] analyzes FMCW of the acoustic signal and converts the time difference mapping into frequency shifts for further improving the tracking accuracy without requiring any precise synchronization. EchoTrack[133] measures the distance from the hand to the speaker array embedded in a smart phone via the chirp's Time of Flight (ToF). A unique triangular geometry is generated from the speaker array and hand to localize the gesture. Doppler shift compensation and trajectory correction are used to improve the trajectory accuracy. LLAP[84] uses a commercial smartphone to achieve motion tracking at the millimeter level by measuring the phase change in the sound signal caused by the gesture movement, and converting the phase change into the distance of movement. FingerIO[134] also uses commercial smartphones only. Using the orthogonal frequency division multiplexing in wireless frequency division multiplexing, the OFDM technology enables finer-grained finger tracking, and finally prototypes a smartwatch-shaped finger I/O device to demonstrate that it can extend the interaction space to $0.5\ \mathrm{m} \times 0.25\ \mathrm{m}$ on both sides of the device area of $0.25\ \mathrm{m}^2$. It works well even when fingers are completely occluded. Strata[135] estimates the channel impulse effect (CIR) in order to explicitly account for multipath propagation, and to select well-behaved channels and extract the phase change in the selected channel signal to accurately estimate the distance change of the finger, and uses a new optimization framework to estimate the absolute distance of the finger according to the change in CIR. The core work of Vernier[82] is calculated with a small signal window phase transition, whose number of local maxima corresponds to the number of cycles of the phase transition, removes complex frequency analysis and long windows of signal accumulation, and significantly reduces tracking delay and overhead. The evaluated results show that its tracking error is less than 4 mm, and the speed is also faster. The phase-based approach is improved by a factor of 3. Lu *et al.*[136] designed a tracking system for a conventional computer without a touch screen, emitting inaudible acoustic signals from the two speakers of the laptop, and then analyzed the energy of the acoustic signal received by the microphone features and Doppler shifts to track the hand motion trajectories. For more complex indoor environments, it is sometimes difficult for acoustic-based methods to achieve accurate

motion tracking due to the multipath fading and limited sampling rates of the mobile devices. PAMT[137] defines a new parameter, namely the multipath effect ratio (MER), to represent the effect of the multipath fading on the received signals at different frequencies, and develops a new multipath effect mitigation technique based on MER and the phase-based acoustic motion tracking method PAMT, by using multiple speakers to calculate the phase change in the acoustic signal and track the corresponding moving distance. The measurement errors of one-dimensional and two-dimensional scenes on an Android smartphone are less than 2 mm and 4 mm, respectively.

4.2 Health Caring

The framework of Subsection 4.2 is shown in Fig.13.

Health concerns mainly include attention to human physiological indicators and detection of driving behaviors. Human physiological indicators include hundreds of items, such as heart beat, breathing, and blood pressure, which can reflect the health of the human body. Traditional methods, such as camera-based methods (e.g., distance PPG[138]) and sensor-based methods (e.g., geophone[139, 140]), can accurately track vital signs. However, these methods require either bright lighting or complex installation and maintenance. In contrast, new sensing methods based on wireless signals have become more attractive due to their low cost, no contact and easy deployment. Next, we will introduce the related work and applications of WiFi-based sensing and sound-based sensing for relative health, driving behavior monitor, and other medical aspects.

4.2.1 *Health Related*

In the monitoring of life signs of human physiological indicators, their accurate monitoring is conducive to timely understanding of physical conditions, especially for elderly people. Since the detection of breathing and heartbeat in wireless sensing is the detection of subtle chest movements, the work based on WiFi sensing mainly includes the followings.

Respiratory monitoring systems usually use peak detection to calculate respiratory rates based on repetitive patterns of chest movements. Wang *et al.*[49] designed a Fresnel zone based model for monitoring human breathing without training, which is robust to different locations and directions. Zeng *et al.*[32] extended the monitoring range to 8 meters using the CSI quotient model. In the last few years, there has been some work on breath detection on smartphones using Nexmon firmware such as WiPhone[50], MoBreath[51]. On the other hand, with the emergence of the COVID-19 disease, there are some jobs that are also focusing on it. For example, Wi-COVID[141] monitors COVID-19 patient respiration rate (RR) via WiFi and tracks RR for medical providers. Due to the high rate of transmission of the COVID-19 disease, healthcare systems around the world are potentially under-resourced to help large numbers of patients at once; non-critical patients are suggested to self-isolate at home. Wi-COVID offers novel, rapid and safe solution to detect and rapidly report patient symptoms to healthcare providers.

Wu *et al.*[5] extended the breath detection from sleep to standing position for still body detection. WiSleep[4] is the first CSI-based method to monitor the respiratory rate of a person during sleep. Liu *et al.*[6] proposed a system to detect vital signs and pos-

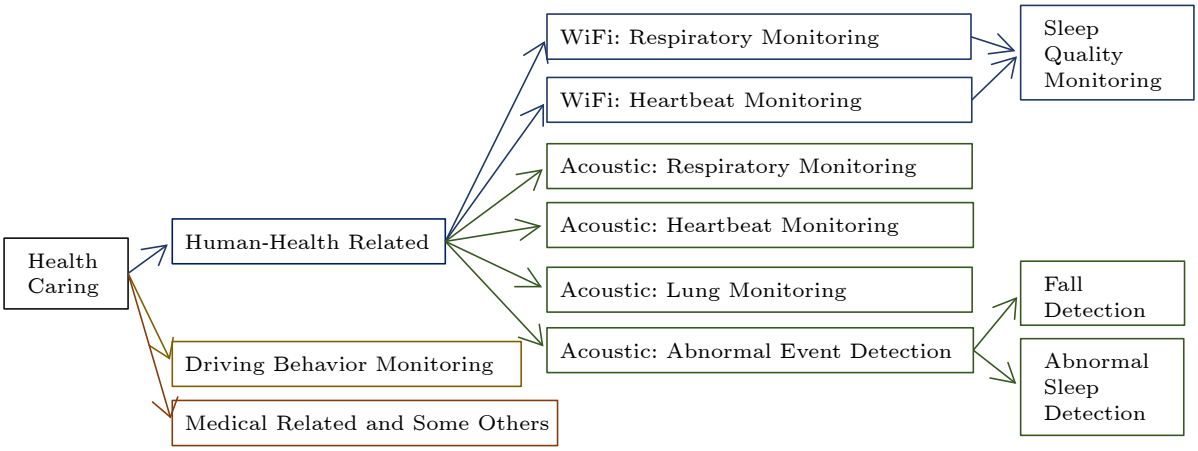


Fig.13. Health caring structure diagram.

ture during sleep by tracking CSI fluctuations caused by small human movements, and to detect the respiratory rate of one or two people in bed. Similarly, PhaseBeat[8] uses DWT for exercise separation and simultaneously monitors breathing and heartbeat with accuracy of 0.5 bpm and 1 bpm, respectively.

Health-related applications based on acoustic mainly rely on the detection of physiological conditions, such as the capture of human breathing, heartbeat, and so on, or some event monitoring, such as the fall detection, abnormal sleep behaviors, and so on, as well as medical-related assistance control to assist body recovery, etc. Larson *et al.*[88–91, 142] captured physiological signals, such as human respiration, heartbeat, vital capacity, chest wall motion, and some other non-voice sounds, such as swallowing during eating. Based on these physiological signals, Nahdakumar *et al.*[142] turned a cell phone into an active sonar system that emits modulated sound signals and detects breathing-induced minute movements of the chest and abdomen from reflexes, and developed the identification of various types of sleeps from sonar reflexes algorithms for apnea events, including obstructive apneas, central apneas, and hypopneas. In [89] inaudible acoustic signals were used to accurately monitor heartbeats, by using only common commercially available microphones and speakers, through transmission sound signals and their reflections on the human body to identify the heart beat rate and heartbeat rhythm, and generate an acoustic electrocardiogram (ACG). Based on this, many subsequent studies also emphasized on capturing heartbeat features by using machine learning techniques to classify the required research. The work[91], continuous multi-person respiration tracking using an acoustic-based COTS device, employing a two-stage algorithm to separate and recombine respiration signals from multiple paths in a short period of time, can distinguish the respiration of at least four subjects within a distance of three meters. SpiroSmart[88] uses a built-in microphone for spirometry. SpiroSonic[90] measures the human chest wall motion through acoustic sensing and interprets it as an index of the lung function based on clinically validated correlations.

Audio equipment can also identify events and perform event detection through inaudible sound signals, such as fall detection[143]. Modulated ultrasonic waves are emitted through speakers, reflected signals are recorded by microphones to detect Doppler shifts caused by events, and features are extracted from the spectrogram to represent fall patterns that are distinguishable from normal activities. Afterwards, SVD (singular value) and K-means algorithms are used to reduce the data feature dimensions and cluster data. The final detection and recognition are accomplished by different methods, such as the Hidden Markov Model for training classification, or modeling each fall through the Gaussian mixture model. In addition, the speakers and microphones on smart devices, such as smartphones, can also collect specific, but distinguishable, body movements and sound signals that accompany each sleep stage of a person. Using the data, it is possible to construct the sleep and wake state, and daily sleep quality. A model that can detect abnormal sleep behaviors (see [142]), and some more detailed sleep events, including snoring, coughing, rolling over, getting up, etc. (see [144]). The recent work Apnoea-Pi[92] presents an open-source surface acoustic wave (SAW) platform to monitor and recognize apnoea in patients. The authors[92] argued that Thin-film SAW devices outperform standard and off-the-shelf capacitive electronic sensors in response and accuracy for human respiration tracking purposes. Combined with embedded electronics, it provides a suitable platform for human respiratory monitoring and sleep disorder identification.

4.2.2 Driving Behavior Monitor

Driver fatigue is a major cause of road accidents, which lead to serious injury, and even to death. The existing work on fatigue detection mainly focuses on the visual and electroencephalogram (EEG)-based detection methods. However, vision-based methods suffer from visual blocking or visual distortion problems, while EEG-based systems are intrusive that inherently bring uncomfortable driving sensations, which may further worsen the driver fatigue. In addition, these systems are expensive to install.

On the contrary, the WiFi signal has the advantage of being non-invasive and device-free. For specific work, WiDrive[70] discusses about the detection method of dangerous driving behaviors. In the driving scenario, the driver is fixed on the seat and the interior environment is stable. Therefore, WiDrive does not need to consider the robustness of any environmental change. Peng and Jia[71] observed that WiFind is a device-free system, which can detect the fatigue degree through two modes: breathing mode and movement mode through WiFi signal. The accuracy

of this method is 89.6%.

The driving behavior monitoring based on acoustic sensing includes the monitoring of the driver's behaviors, such as making and receiving calls[145], inattention[146, 147], and identifying different driving behaviors through multiple classifiers. Concentration behaviors are detected 50% earlier and alerted in advance in order to ensure people's driving safety. Furthermore, as introduced above, the breathing pattern is one of the key indicators of health status. Breathlistener[148] uses audio devices in smartphones to estimate fine-grained breathing waveforms in driving environments by extracting the energy of acoustic signals spectral density feature (ESD) and further to design a GAN-based architecture to generate fine-grained breathing waveforms. Related experiments show that it can capture the breathing patterns of the driving environment in the driving environment. The resulting breathing pattern can be further used to keep track of the driver's current driving situation. Additionally, in the medical field, UbiEar[149] was designed, which is a smartphone-based acoustic event sensing and notification system for the hearing-impaired to achieve location-independent acoustic event recognition.

4.3 Localization and Navigation

The common ranging methods are introduced in Subsection 3.5. ToF, TDOA, angle of arrival (AOA) measurements, phase difference, and RSSI attenuation models are used to calculate the distance between the object and the device or between the sender and the receiver. With the distance information, we can gain more information, such as the location of the object. In wireless sensing, the positioning of devices or objects is also one of the important issues. The four most commonly used methods are trilateral positioning, triangular positioning, hyperbolic positioning, and wireless fingerprint positioning. Next, we will describe them separately.

4.3.1 Methods

1) *Trilateral Positioning.* The method is to make a circle at three distances from the measured object, and the intersection of the three circles is the coordinates of the measured object's position. Suppose we want to locate an object in the room. It is in the reference point of the three known locations. The distances from the three reference points to the object to be measured are obtained by the method as described in Subsection 3.5. In Fig.10, assuming the location node as $D(x, y)$, the known coordinates of E, B, C are $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ respectively. The distances from them to D are d_1, d_2 and d_3 respectively. The location of D can be obtained by either of the following equations:

$$\begin{cases} (x-x_1)^2 + (y-y_1)^2 = d_1^2, \\ (x-x_2)^2 + (y-y_2)^2 = d_2^2, \\ (x-x_3)^2 + (y-y_3)^2 = d_3^2. \end{cases}$$

In practice, however, it is not common for three circles to intersect at a single point, which may result in an inaccurate location of the target to be measured using the trilateral positioning. However, in practice, if there are many APs at the known locations, then we can select the optimal base station by using the optimal quadratic method, and then use the weighted method, the centroid method and so on to approximate the coordinates of the measured object.

2) *Triangular Positioning.* The Triangular positioning algorithm is a common wireless positioning algorithm. Its core idea is to sense the arrival angles of the signals sent by other devices through hardware devices, usually relying on multiple antenna arrays. In a multi-antenna array, for signals arriving at the antenna array at different angles, there remains a time difference between any two antennas which is denoted by the angle of arrival. The angle is calculated using AOA, and then the location of the unknown node is calculated using geometric methods.

3) *Hyperbolic Positioning.* Trilateral positioning requires knowing the distance between the object to be measured and multiple APs before it can be used. However, we can actually locate the object to be measured using a hyperbolic positioning algorithm that uses the TDOA algorithm. Assuming that the time difference between the objects to be measured arriving at different send (receive) periods is Δt, multiplying the speed of signal transmission v, we can find the distance difference between the wireless signals to different base stations Δd. If we know the mileage difference between the two sending (receiving) ends of the object to be measured, then on this basis, we can use geometry knowledge (that is, hyperbola) to solve the location of the target.

4) *Wireless Fingerprint Positioning.* This is a common method in the field of WiFi location[58–62, 69]. The WiFi fingerprint positioning uses RSSI or CSI signals in the WiFi signal to collect each sample point

in the space for training, forming a fingerprint database. If a certain location point needs to be located, its signal characteristics are collected and compared with the fingerprint database. The training sample point in the database closest to the signal characteristics of the location point is used as the positioning result for changing the location point. For example, during the offline training phase, deep fingerprints are generated from all the weights obtained by in-depth learning in the DeepFi[61] system. In the online positioning phase, the system uses a probability method based on radial basis functions to estimate the targeted location.

4.3.2 Indoor Localization

Today, the popularity of smartphones and a range of WiFi terminals further promote the rapid development of wireless base stations. WiFi is ubiquitous because of its wide distribution of hot spots, low access conditions and high flexibility, which make related indoor positioning technology widely usable in public safety, industry, medical treatment and other fields. As one of the research directions of the WiFi-based indoor positioning technology, the related research of WiFi indoor positioning is quite mature, and many achievements have also been made in related fields.

In RSSI-based location applications, a typical passive location infrastructure includes AP and monitoring points (MPs). MPs detect WiFi signal changes from AP to passive locators. For example, Alkandari *et al.*[63] used one AP and one MP to estimate the speed of movement in indoor environments. Similar to that in [63], one MP was added to the work in [64] to experiment for providing better performance. In other studies, Oguntala *et al.*[65] used a ranging method of passive RFID reception signal strength to locate people. Particle filter algorithms analyze and compute RSSI to obtain targeted locations in indoor environments.

Since the results obtained based on RSSI are susceptible to multipath effects, FIMD[66] uses CSI temporal stability and frequency diversity in order to reduce the multipath effects in the above signal propagation. A false alarm filter and a data fusion scheme are also used to improve the detection accuracy. FILA[67] uses a trilateral measurement method and CSI to mitigate the multipath effect at the receiver. The Widar series[57, 58] rely on the relationship between the motion speed and associated DFS to track users in 2D coordinates through two orthogonal WiFi links. WiDE[68] is a WiFi-distance estimation based group profiling system using LightGBM to learn powerful hidden features automatically. WiDE can automatically learn powerful hidden features from the proposed features for between-user distance estimation, and infer group membership with the estimated distance in a three-floor campus building and a shopping center.

Since each localization method has certain disadvantages, some systems combine wireless fingerprint localization methods to improve its accuracy. Abdel-Nasser *et al.*[60] proposed MonoPHY, a device-free localization system based on wireless LAN that uses a single wireless stream. The CSI data at each location is modeled as a Gaussian mixture and stored in the fingerprint. Some studies combine the information dimension features of multiple positioning methods to improve the accuracy. Xie *et al.*[59] introduced the mD-Track, a device-free WiFi tracking system. The system can jointly fuse information from as many signal dimensions as possible, such as AOA, ToF, Doppler shift, and so on, to overcome the resolution limitation of each dimension. In addition, some studies have established theoretical models. Qian *et al.*[57] introduced Widar series, a WiFi-based passive tracking system, where the moving speed (speed and direction) and position of the user are estimated at the decimeter level.

The research on WiFi-based indoor positioning technology is relatively mature, and some achievements have also been made based on sound wave sensing. For the specific research on sound waves using ToA and TDoA for ranging, some studies, such as the work in [150] used the Doppler shift of the acoustic signal for direction finding. In addition, Zhang *et al.*[151] developed the SwordFight system using sound sensors on mobile devices. Liu *et al.*[152] also proposed the use of acoustic ranging techniques to constrain the positional relationship between devices, thereby eliminating the large error problem in localization. The Centaur localization system framework proposed by Nandakumar *et al.*[153] uses sound ranging and localization for Bayesian inference. The algorithm is designed to make sound ranging more robust in non-line-of-sight situations, and to make sound-only ranging devices to participate in sonolocation. Tarzia *et al.*[154] proposed an Acoustic Background Spectrum (ABS) ambient sound fingerprint, determined by measuring the current room fingerprint and then se-

lecting the "closest" fingerprint from the database, adding ABS improves the localization accuracy of WiFi-only rooms from 30% to 69%. The technology can be used without WiFi. GuoGuo[93] localizes the target by measuring the ToA of the acoustic signal. This work could increase the position updating rate by providing adequate coverage by the then-advanced signal processing techniques and by increasing the transmission speed of the acoustic signal through a symbol-interleaved signal structure, which can be averaged in normal environments to a localization accuracy of 0.25 m; EchoTag[155] actively generates acoustic features by transmitting the sound signal through the mobile phone speaker and sensing its reflection with the mobile phone microphone. Compared with the widely used passive sensing, this active sensing provides more fine-grained control over the collected signatures. Since the sensed signal is controlled by the EchoTag, it can be intentionally chosen to enrich the sensed signature and remove noise from useless reflections. Swadloon[156] uses the smartphone's acoustic direction finding in combination with inertial sensors for fine-grained indoor localization to track the smartphone's displacement relative to the acoustic orientation with a resolution of less than 1 mm. Orientation is then obtained by combining the velocity from the displacement with that from the inertial sensors. Pradhan *et al.*[95] developed a smartphone-based indoor space mapping system that allows ordinary users to quickly map indoor spaces by simply walking around by carrying their mobile phones. The system accurately measures the distance to the nearby reflectors, estimates the user's trajectory, and pairs different reflectors that the user encounters during walking to automatically build contours. Its experimental results, that the median errors are 1.5 cm for a single wall and 6 cm for multiple walls, show that the median error of 30 cm and a 90-percentile error of 1 m for the entire system outperform the previous best-performing BatMapper[94]. The constructed indoor profile can also be used to predict the wireless RSS. In some recent work related to localization, CARACAL[96] is a low-cost, custom-designed hardware and software system that can extract and locate weak acoustic signals and apply them to gunshot location, prey location, animal call location, etc. The system is open source and can be customized to suit a variety of wildlife research applications. In [97], a method (Structures Containing Unknown Empty Areas, SUEA) is proposed to identify the shape, size and position of the hollow area in the unknown area by activating the active AE (acoustic emission) source and using the collected AE arrival signals. Then, the unknown AE source is located by combining the identified void. This method can provide a more accurate solution for the AE source location of complex structures including unknown void areas such as tunnels, bridges, railways and caves in practical engineering. [98] uses the active acoustic wave method to locate cracks in water supply pipes, and proposes an active detection and location method based on low-frequency acoustic wave propagation in water pipes to detect and locate leaks in water supply systems. To overcome the major difficulty of statistical processing of time delays associated with multiple sound paths in reverb environments, two approaches are used in this paper: 1) the classical signal decomposition technique (Prony's method) and 2) a clustering pre-processing approach called Spectral Mean-Shift Clustering.

4.4 Privacy and Security

The framework of Subsection 4.4 is shown in Fig.14.

With the smart devices and mobile devices becoming important parts of our daily lives, they are often used to store important information, including personal identity and other sensitive data. They bring convenience, but become threats also to potential security and privacy issues. As the most common built-in device in mobile phones, microphones and speakers

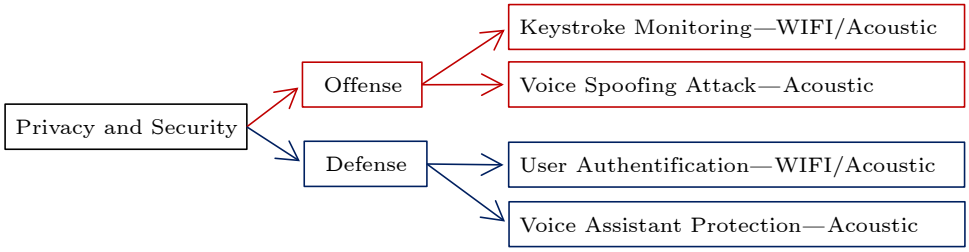


Fig.14. Privacy and security structure diagram.

can be used to potentially disrupt privacy attacks or enhance user security. This subsection addresses the privacy and security issues of WiFi-based and sonic-based sensings. It introduces the work related to privacy security of new sensings from two perspectives: offense and defense. It mainly involves keystroke spying attacks, voice spoofing attacks, user authentication and voice assistant protection.

4.4.1 *Keystroke Sniffing Attack*

Among all kinds of sound-based attacks, keystroke monitoring is one of the most common forms at present. Since finger tracking systems can locate the coordinates of a finger on a two-dimensional plane, researchers can reconstruct keyboards based on the likelihood of keyboard strokes. As a result, keystroke tracking exposes privacy issues about keystroke eavesdropping.

Based on WiFi sensing, Ali *et al.*[18] proposed a keystroke recognition scheme based on WiFi signal CSI, namely WiKey. They believed that when the hand is typing on the keyboard, WiKey detects the keys that are pressed as the target user's hands and fingers form a unique shape and direction that results in a unique pattern in the received CSI. They used threshold-based segmentation, DWT feature extraction, and the *K*NN classifier to identify 37 keystroke types, achieving 97.5% keystroke detection rate and 96.4% one-key accuracy. However, the system is very sensitive to relative position changes between users and devices.

Based on acoustic, the pink keys on the physical keyboard are used to monitor the physical keyboard with sound wave signals. Based on the physical location relationship, the attack mainly uses TDOA to process the signal obtained in the time domain, according to the acoustic characteristics, and uses the mobile phone to monitor the keyboard to type. According to the short-term energy value, the audio information is used to check the records and endpoints. After that, the key audio in the separated audio information segment is mainly extracted from the audio feature parameters on the frequency domain.

Recently active research has been extended to signal processing for keystroke monitoring. In addition to frequency domain features, Ubik[99] improves the accuracy of keystrokes on solid surfaces and enables fingerprinting of acoustic differences due to multipath fading. Zhu *et al.*[157] disclosed a context-independent keystroke listening attack based on keystroke sound wave spillover, using three cooperative mobile phones to locate possible keystroke areas according to the TDoA localization principle described above. But it requires three mobile phones that cooperate with each other to limit its application in real scenarios. Liu *et al.*[158] used a 192 KHz mobile phone with a high sampling rate to obtain a millimeter-level sound wave ranging function. Only one mobile phone is needed to monitor the keystroking. The mobile phone is placed after the keyboard and the two microphones are connected in parallel to the long side of the keyboard. TDoA is measured at this moment, and the two microphones of the mobile phone are used to monitor the keyboard tapping. There is a difference in the time when the key audio signal reaches the two microphones. At this moment, the keys on the same half of the hyperbola can be aggregated. Liu *et al.*[159] combined accelerometer sensors and keystroke acoustic waves to improve the recognition performance. The recent work of FANG[160] demonstrated a position-independent keystroke monitoring, using TDoA and acoustic signatures, to determine the relative position of the keyboard to the phone and the possible keystroke area, achieving 93% accuracy. And the work[109] shows the possibility of reconstructing the acoustic side channel attack of a PIN (called PinDrop) by analyzing the acoustic signature of individual keys on a PIN plate. Its attack on ATM is at a distance of 2 meters and can recover up to 57% of 4-digit PINs and 39% of 5-digit PINs in three attempts.

4.4.2 *Voice Spoofing Attack*

Voice spoofing attacks are aimed at applications based on sound wave sensing. In this subsection, we introduce the relevant contents of current voice spoofing attacks. Existing voice spoofing is mainly aimed at recognition systems, including voice assistants. Although there are differences in attack methods, the goal is to generate a voice signal, which can be recognized by the speech recognition system as the voice signal of the attacked user, and then controls the speech recognition system and performs specific malicious operations. At present, the existing main voice spoofing attacks can be divided into identifiable attacks, such as the imitation attacks, replay attacks, speech synthesis attacks and speech conversion attacks according to whether the generated or used signals can be understood or perceived by users, and

those that are not easily identifiable, such as the adversarial example based attacks and silent attacks. The voice signal used by the identified attack can be heard and understood by the attacked user. When the user hears the voice signal that does not belong to him/her, he/she can know that his/her voice recognition system has been attacked. Among them, the imitation attack is the simplest type of attacks against the speech recognition system. The attacker controls the smart device to perform related operations by imitating the voice of the victim[161]. This kind of attacks is easy to be detected. Due to the limited degree of imitation, it is a threat to the device. Replay attacks are also limited, which refer to the possibility that the attacker may obtain the target user's speech samples through eavesdropping, recording, etc. and replay them through speakers[162]. Due to their easy detection, their actual impact is limited. But there are also more advanced and dangerous basis of attacks. Based on this idea, there are speech synthesis methods, in which the attacker synthesizes the target user's speech from text input and plays it through the speaker[100] and speech conversion methods. In such a case, the attacker converts a piece of voice input into the target user's voice and plays it through the speaker[163] enabling the attacker to generate fake sentences, and the similarity with the target user's voice is determined by using synthesis and transformation techniques.

The above-mentioned attacks that can be identified, such as the imitation attacks, replay attacks, speech synthesis attacks and speech conversion attacks, have limited attack effects when the attacker is present or under surveillance. Recent studies have shown that the recognition model of a deep neural network is likely to classify some samples wrongly with little disturbance to other samples. When such small disturbance samples are artificially generated, adversarial sample attacks can be carried out on the network. These adversarial samples might sound like ordinary speech content, like "Excuse me" to a human, but the tiny perturbations contained in it might make the whole speech be recognized as "Do it". After a major breakthrough in the generative adversarial network (GAN), researchers, such as Carlini *et al.*[164], studied the use of the GAN's generator to generate malicious attack signals doped with tiny noises, so that they retain enough acoustics features, but it is difficult for humans to understand them. CommanderSong[103] embeds a given voice command into a randomly selected song that sounds perfectly normal to the user, but it is recognized by the speech system as a specific voice command. Carlini and Wagner[165] combined the contents of [164] and [103] to make it possible to generate an audio very similar to the waveform from any given voice command waveform, and the newly generated audio can be recognized by the speech recognition system for voice commands. In addition to the attacks based on adversarial samples, there are also some silent attacks, such as DolphinAttack[102]. DolphinAttack[102] proposes a new and effective attacking method for speech recognition systems, which modulates any audible voice command into the ultrasonic frequency band, and uses the non-communication of the microphone circuit. This "dolphin sound attack" can silently inject voice commands into the microphone circuit, and then the voice signal will be demodulated and recovered, so as to be recognized by the voice assistant, and finally control the smart device to perform corresponding operations, including silent activation of Siri, and initiate calls on the iPhone, etc. Kasmi and Esteves[166] achieved a silent attack on speech recognition systems by injecting voice commands modulated in electromagnetic signals into smart devices with earphones or power cords. The recent work[167] has identified two new acoustic features to improve the performance of spoofing attacks. The first feature consists of two cepstrum coefficients and a LogSpec feature extracted from the linear prediction (LP) residual signal. The second feature is the subband ratio feature of harmonic noise, which can reflect the difference of the interactive motion of the sound tract and glottic airflow between the real speech and the spoofed speech.

4.4.3 *User Authentication*

Authentication refers to the confirmation of the user's identity through certain technology. In real life, identity authentication plays a role in protecting personal privacy and property security. However, the authentication technology is still a huge challenge. On the one hand, human beings can recognize people, and they are familiar with not only by their facial characteristics, but also by voice and even gait behavioral characteristics. On the other hand, humans may need a token to identify strangers. Similar challenges arise: how can computers identify legitimate users while rejecting deceivers? As the first level of defense for smart and mobile devices, user authentication can be

divided into three modes/factors. 1) What do you know (password, etc.)? 2) What do you have (keys, PIN cards, specific devices, etc.)? 3) Who are you (fingerprint, iris, voiceprint, etc.)? The first two modes/factors are obviously more vulnerable to attackers when they are single. As long as they have the corresponding content, they can be attacked and cannot be distinguished if they are used by the users themselves. Therefore, the third mode, especially biometric-based solutions, such as fingerprints, faces, iris, gait, sound, and so on, is becoming the key to user authentication. The identification process requires different types of sensors and supporting schemes: camera, inertial measurement unit IMU, WiFi, RFID, and microphone. The methods based on camera, user voice, and IMU may have intrusive behaviors or privacy issues. The RFID-based and WiFi-based methods require the deployment of sensors or devices. These problems more or less limit their applications to real-world scenarios. In contrast, WiFi-based and sound-wave based authentications have attracted considerable attention from researchers due to their low-cost nature and widespread deployment of speakers and microphones in mobile devices.

The WiFi-based authentication is divided into two main categories, one is to use the unique gait to identify a specific user, and the other is to use the uniqueness of life characteristics to identify specific users. Firstly, the gait recognition provides a silent authentication solution when the target user passes through a specific sensing area. WiWho[53] uses gait characteristics to distinguish different people by perceiving the human movement gait. It can identify up to six people at a time. When the number of total identifiers changes from two to six, the corresponding recognition rate for WiWho becomes 92%–80%. Unlike WiWho, WiFi-ID[54] does not extract specific gait features, but directly analyzes the entire walking behavior. WiFi-ID selects a 20 Hz–80 Hz frequency band for analyzing and transforming WiFi signal (CSI) into a time-frequency combined domain by continuous Fourier transform. The recognition rate is 93%–77% for two to six people. Secondly, the user's life characteristics are unique. FingerPass[72] collects WiFi signals of continuous finger movements, extracts gestures, movements and user characteristics, identifies finger movements through LSTM models, and identifies the executor. The recognition accuracy of this method is 91.4%.

In recent years, the identity authentication process based on sound wave perception has used a variety of different technologies and activities, ranging from individual breathing differences, ear canal echo differences, acoustic face echoes, oral teeth occlusion, lip reading, and acoustic signal detection gait. Researches on acoustic signal detection signatures and various two-factor authentications are carried out, which are both practical and highly accurate. BreathPrint[101] uses GFCC (gammatone frequency cepstral coefficients) to extract human breath sounds from the human breath sounds that may be heard at three levels: sniffing, normal breathing, and deep breathing by extracting human acoustic features. EchoPrint[168] actively emits a barely audible sound signal from the earpiece speaker, combined with the facial landmark position. Since the echo features depend on the 3D facial geometry, it is not easily attacked by 2D visual images or video spoofing attacks. BiLock[169] uses human tooth occlusion, i.e., tooth clicks, for identity authentication, and achieves an average false rejection rate of less than 5% and an average false acceptance rate of 0.95% in actual experimental evaluation. It has advantages in terms of robustness to noise and security against replay attacks. LipPass[105] is based on lip reading, where the Doppler curve of the acoustic signal changes from the smartphone's built-in audio device caused by the user's speaking lips. According to the unique lip motion pattern that exists in each individual, effective features are extracted from Doppler contours with deep learning. EarEcho[104] uses ear canal echo for wearable authentication, which is 95.16% accurate with one-time authentication. The accuracy rate of 97.57% and continuous authentication reflects that the unique physical and geometric features of the human ear canal can be used in identity authentication. AcousticID[170] uses commercial off-the-shelf equipment to generate acoustic signals, and analyzes how much each part of the human body responds to acoustic signals during walking. Puller effect proves the feasibility of gait recognition, and then extracts fine-grained gait features that can distinguish different people from both macroscopic and microscopic dimensions. ASSV[171] is a device-free online handwritten signature verification system that provides paper-based handwritten signature authentication. It uses a novel chord-based method to estimate phase-dependent changes induced by minute movements. Then, based on the estimation, frequency domain features are extracted by the discrete cosine transform (DCT). In addition, a deep convolutional

neural network (CNN) model with a distance matrix is designed to verify the signature. SilentSign[106] utilizes the speaker in the smart device to emit sound, and the microphone receives the reflected frequency-shifted sound waves to measure the distance change in the pen tip during signing. EarGate[107] observes gait-based recognition from walking-induced sounds, utilizes an in-ear microphone to detect the user's gait from within the ear canal through the occlusion effect of the headset, and achieves a balanced accuracy (BAC) of 97.26%.

In addition to the methods mentioned above, with the popularity of mobile devices in recent years, two-factor authentication (2FA) has received more and more attention. The three modes mentioned above can verify a person's identity. If three factors namely the secret information, personal items and physiological characteristics, need to be provided at the same time during authentication, then it is called the two-factor authentication. The bank card is the most common two-factor authentication, that is, users must provide both the bank card and the password to get cash. At present, password and mobile phone verification code have become the most common two-factor authentication scheme. But short messages, SIM cards and ID cards are all at risk of forgery. Acoustic-based two-factor authentications include Home Alone[172] that uses active notification sounds generated by the user's smartphone to measure proximity to browsers. Listening Watch[173] uses human speech as a sound factor to detect proximity to smartwatches and browsers, both of which are authenticated by a second factor with a randomly selected acoustic signal. The recent Proximity-Echo[108] utilizes the proximity of the user's registered mobile phone and the logged-in device as a second-factor authentication without user interaction or pre-built device fingerprints. It derives location features from alternating beep signals from two devices and senses echoes with microphones, and compares the extracted signatures for proximity detection. Given the received beep signal, the system designs a period selection scheme to accurately identify two sound segments: the chirp period which is the sound segment propagating directly from the speaker to the microphone, and the echo period which is the sound segment reflected back by surrounding objects. In two pieces of related work in the last year, TeethPass[110] uses earplugs to collect bite sounds in the binaural canal to achieve authentication, extracting unique characteristics from three aspects: bone structure, bite position, and bite sound. Based on an incremental learning based Siamese network, the classifier has an accuracy of 96.8% and can resist nearly 99% spoofing attacks through a large number of experiments. ToothSonic[111] uses toothmarks induced by users performing tooth gestures for audibility authentication. It designs a representative tooth gesture that produces an effective sound wave carrying acoustic fingerprint information. Toothprint is caught via a user's private teeth-ear channel, which modulates and encrypts sound waves and is resistant to spoofing attacks. The related work involved is shown in Table 1.

4.4.4 *Voice Assistant Protection*

The voice assistant protection problem is also aimed at the sensing method based on sound waves. Regarding the protection problem of voice assistants, the research mainly focuses in two directions: attack detection by using deception to attack its own acoustic defects and the difference in living body information between humans and speakers. Acoustic flaws in spoofing attacks may exist in both hardware and software. With the development of more advanced microphones and speakers on hardware, speech synthesis and conversion tools, recordings that do not distinguish between living bodies, such as users and spoofing attackers only, are likely to be difficult to achieve.

Another research direction is to find the living difference between human and speaker vocalizations. Even if the acoustic defect is very weak, the way people and speakers vocalize is completely different, even the user's own human vocalization and its recording, by distinguishing the way the mouth moves. Vibration with diaphragms is also achievable. Feng *et al.*[174] proposed the use of wearable devices, such as glasses, to measure the human body conduction of sound. In addition, some studies have explored speakers or other properties unique to humans. Chen *et al.*[175] proposed to use the magnetometer in the mobile phone to detect the magnetic field generated by the speaker, and Wang *et al.*[176] proposed to identify living users by detecting the noise produced by humans in exhaling while speaking. In addition, the recent review[177] has conducted a comprehensive study and summary on the countermeasures of voice assistant against various attacks, aiming at the problem that various attacks and independent defense in the literature often lack a systematic perspective, which makes it difficult for designers to correctly identify,

Table 1. Comparison of Acoustic-Based User Authentication Work

Work	Year	Type	Evaluation	Technology and Content
BreathPrint[101]	2017	Passive	ACC: 94%	Sniffing, normal breathing and deep breathing; extracting the acoustic feature GFCC as a voiceprint distinction
EchoPrint[168]	2018	Active	ACC: 98.05%	Acoustic face echo & visual face recognition
BiLock[169]	2018	Passive	FRR: 5%; FAR: 0.95%	Teeth bite
LipPass[105]	2018	Active	ACC: 93.1%	Lip reading
EarEcho[104]	2019	Passive	ACC: 97.57%	Ear canal echo
AcousticID[170]	2019	Active	ACC: 96.6%	Gait-based acoustic signal
ASSV[171]	2019	Active	AUC: 98.7%; EER: 5.5%	Phase emits an acoustic signal to estimate phase-dependent changes caused by tiny movements, DCT frequency-domain signatures
SilentSign[106]	2020	Active	AUC: 98.2%; EER: 1.25%	Phase sends an acoustic signal to measure changes in pen tip distance
EarGate[107]	2021	Passive	FAR: 3.23%; EER: 2.25%	In-ear microphone detects gait from the ear canal
Proximity-Echo[108]	2021	Active	EER: 4.3%	Two-factor authentication (2FA): the two devices emit a beep signal alternately and the location features are derived through the microphone inductive echoes, and the extracted signatures are compared
TeethPass[110]	2022	Passive	ACC: 96.8%	Use the sound wave effect produced by the tooth structure
ToothSonic[111]	2022	Passive	ACC: 95%	The main working principle is similar to the above one tooth gesture

understand and mitigate the security threats against voice assistant.

5 Limitations and Open Issues

Although researchers have conducted extensive studies in WiFi and acoustic sensing, there still exist some limitations in existing work and open issues to be explored in the future.

5.1 Limitations

Hardware Restriction. Although CSI outperforms RSSI in sensing granularity and precision, it can only be extracted from specific NICs (such as Intel 5300, Atheros series, AX210, and so on) at present. Moreover, the CSI sampling rate depends on the selected WiFi working mode, which indicates that CSI data can easily be affected by the parameters of the communication equipment. The hardware restriction has affected the development of WiFi sensing applications, which requires joint efforts of chip manufacturers to facilitate the acquisition of CSI from commercial hardware. Although acoustic sensors are commonly equipped in smart devices, some model-based motion tracking applications require multiple pairs of the microphone and speaker which are not supported by existing commercial devices.

Sensing Robustness. Due to the well-known multipath effect, both WiFi and acoustic signals are relatively sensitive to external environments, such as the layout of different objects and random human movements. They vary with different environments and may exhibit different patterns even for the same kind of gestures or activities. Consequently, a WiFi or acoustic sensing system is easily interfered by external factors such as locations, layouts, and other objects' presence or movements. This makes a well-designed sensing system fail to work in diverse real-world scenarios and hinders the deployment of such systems in real world.

Domain Shift Problem. Most existing sensing systems follow a machine learning approach which relies on collecting massive data dependent of specific tasks in the model training stage. However, considering the complexity of practical usage scenarios, the trained model probably fails to work in the environments or settings different from those in the training stage. This is the so-called domain shift problem which weakens sensing scalability to various scenarios and adds the overhead of retraining models. Although some domain adaptation and generalization techniques have been proposed by recent work, they are application-specific and lack generalization ability to different applications.

5.2 Open Issues

Cross-Modal Sensing. With the increasing richness of sensing modalities, it is worth trying to combine them and explore cross-modal learning techniques to improve the performance of WiFi and acoustic sensing systems. Person-in-WiFi[178] tries body segmentation and posture estimation based on commercial WiFi hardware with the aid of annotations of RGB videos. WiSIA[179] is a multifunctional

system which simultaneously accomplishes low-cost WiFi imaging, multi-target segmentation, and fine-grained contrast enhancement. Through these examples, we envision that multimodal learning can be utilized to enhance the sensing capability of wireless sensing systems in more applications.

Integrated Sensing and Communication (ISAC). Recently, ISAC is a hotspot in the community of wireless sensing and communication which aims at optimizing communication and sensing capabilities simultaneously of radio frequency signals[180–182]. To achieve this goal, there are some enabling techniques including the design of transmit waveform, the modeling of propagation environment, signal processing methods, etc. Nevertheless, there exist multiple challenges to realize ISAC including diverse sensing applications, varied performance requirements, and massive data to be processed in real time. There are many open issues for ISAC research such as collaborative sensing among multiple devices and sensing-assisted communication[182].

6 Conclusions

In this paper, we conducted a comprehensive survey of WiFi and acoustic sensing in terms of the principles, technologies, and applications, in order to enable readers to obtain a good understanding of this promising area. Specifically, we introduced the basic principles, fundamental techniques, and significant applications of wireless sensing based on WiFi and acoustic signals. We also discussed several key limitations of existing research work and put forward open issues remaining to deal with in the future. According to the survey, we held the viewpoint that wireless sensing based on WiFi, acoustic, or other kinds of signals opens up a new paradigm of intelligent sensing and shows promising potential in touch-free human-computer interaction applications. However, before making it come true, researchers need to overcome several critical limitations such as sensing granularity, robustness, and cross-domain problems.

We expect that our survey attracts more people to pay attention to this area and acts as a valuable guide for them.

References

[1] Bahl P, Padmanabhan V N. RADAR: An in-building RF-based user location and tracking system. In *Proc. the 19th Annual Joint Conference of the IEEE Computer and Communications Societies*, Mar. 2000. DOI: 10.1109/INFCOM.2000.832252.

[2] Halperin D, Hu W J, Sheth A, Wetherall D. Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM Computer Communication Review*, 2011, 41(1): 53. DOI: 10.1145/1925861.1925870.

[3] Liu J, Wang Y, Chen Y Y, Yang J, Cheng J. Tracking vital signs during sleep leveraging off-the-shelf WiFi. In *Proc. the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Jun. 2015, pp.267–276. DOI: 10.1145/2746285.2746303.

[4] Liu X F, Cao J N, Tang S J, Wen J Q. Wi-Sleep: Contactless sleep monitoring via WiFi signals. In *Proc. the 2014 IEEE Real-Time Systems Symposium*, Dec. 2014, pp.346–355. DOI: 10.1109/RTSS.2014.30.

[5] Wu C S, Yang Z, Zhou Z M, Liu X F, Liu Y H, Cao J N. Non-invasive detection of moving and stationary human with WiFi. *IEEE Journal on Selected Areas in Communications*, 2015, 33(11): 2329–2342. DOI: 10.1109/JSAC.2015.2430294.

[6] Liu J, Chen Y Y, Wang Y, Chen X, Cheng J, Yang J. Monitoring vital signs and postures during sleep using WiFi signals. *IEEE Internet of Things Journal*, 2018, 5(3): 2071–2084. DOI: 10.1109/JIOT.2018.2822818.

[7] Wang X Y, Yang C, Mao S W. PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices. In *Proc. the 37th IEEE International Conference on Distributed Computing Systems*, Jun. 2017, pp.1230–1239. DOI: 10.1109/ICDCS.2017.206.

[8] Wang X Y, Yang C, Mao S W. On CSI-based vital sign monitoring using commodity WiFi. *ACM Trans. Computing for Healthcare*, 2020, 1(3): Article No. 12. DOI: 10.1145/3377165.

[9] Wang Y X, Wu K S, Ni L M. WiFall: Device-free fall detection by wireless networks. *IEEE Trans. Mobile Computing*, 2017, 16(2): 581–594. DOI: 10.1109/TMC.2016.2557792.

[10] Wang H, Zhang D Q, Wang Y S, Ma J Y, Wang Y X, Li S J. RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices. *IEEE Trans. Mobile Computing*, 2017, 16(2): 511–526. DOI: 10.1109/TMC.2016.2557795.

[11] Zhang L, Wang Z R, Yang L. Commercial Wi-Fi based fall detection with environment influence mitigation. In *Proc. the 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, Jun. 2019. DOI: 10.1109/SAHCN.2019.8824989.

[12] Palipana S, Rojas D, Agrawal P, Pesch D. FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018, 1(4): Article No. 155. DOI: 10.1145/3161183.

[13] Hu Y Q, Zhang F, Wu C S, Wang B B, Liu K J R. DeFall: Environment-independent passive fall detection using WiFi. *IEEE Internet of Things Journal*, 2022, 9(11): 8515–8530. DOI: 10.1109/JIOT.2021.3116136.

[14] Chen S, Yang W, Yang X, Geng Y Y, Xin B Z, Huang L S. AFall: Wi-Fi-based device-free fall detection system using spatial angle of arrival. *IEEE Trans. Mobile Computing*, 2022. DOI: 10.1109/TMC.2022.3157666.

[15] Ji S J, Xie Y X, Li M. SiFall: Practical online fall detection with RF sensing. arXiv: 2301.03773, 2023. https://arxiv.org/abs/2301.03773, Jan. 2023.

[16] Yang Z, Zhang Y, Zhang Q. Rethinking fall detection with Wi-Fi. *IEEE Trans. Mobile Computing*, 2022. DOI: 10.1109/TMC.2022.3188779.

[17] Wang Y C, Yang S, Li F, Wu Y, Wang Y. FallViewer: A fine-grained indoor fall detection system with ubiquitous Wi-Fi devices. *IEEE Internet of Things Journal*, 2021, 8(15): 12455–12466. DOI: 10.1109/JIOT.2021.3063531.

[18] Ali K, Liu A X, Wang W, Shahzad M. Keystroke recognition using WiFi signals. In *Proc. the 21st Annual International Conference on Mobile Computing and Networking*, Sept. 2015, pp.90–102. DOI: 10.1145/2789168.2790109.

[19] Ouyang Z, Srinivasan K. Mudra: User-friendly fine-grained gesture recognition using WiFi signals. In *Proc. the 12th International Conference on Emerging Networking EXperiments and Technologies*, Dec. 2016, pp.83–96. DOI: 10.1145/2999572.2999582.

[20] He W F, Wu K S, Zou Y P, Ming Z. WiG: WiFi-based gesture recognition system. In *Proc. the 24th International Conference on Computer Communication and Networks (ICCCN)*, Aug. 2015. DOI: 10.1109/ICCCN.2015.7288485.

[21] Zhang Y, Zheng Y, Qian K, Zhang G D, Liu Y H, Wu C S, Yang Z. Widar3.0: Zero-effort cross-domain gesture recognition with Wi-Fi. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2022, 44(11): 8671–8688. DOI: 10.1109/TPAMI.2021.3105387.

[22] Venkatnarayan R H, Mahmood S, Shahzad M. WiFi based multi-user gesture recognition. *IEEE Trans. Mobile Computing*, 2021, 20(3): 1242–1256. DOI: 10.1109/TMC.2019.2954891.

[23] Li C N, Liu M N, Cao Z C. WiHF: Gesture and user recognition with WiFi. *IEEE Trans. Mobile Computing*, 2022, 21(2): 757–768. DOI: 10.1109/TMC.2020.3009561.

[24] Tan S, Yang J, Chen Y Y. Enabling fine-grained finger gesture recognition on commodity WiFi devices. *IEEE Trans. Mobile Computing*, 2022, 21(8): 2789–2802. DOI: 10.1109/TMC.2020.3045635.

[25] Niu K, Zhang F S, Wang X Z, Lv Q, Luo H T, Zhang D Q. Understanding WiFi signal frequency features for position-independent gesture sensing. *IEEE Trans. Mobile Computing*, 2021, 21(11): 4156–4171. DOI: 10.1109/TMC.2021.3063135.

[26] Xiao R, Liu J W, Han J S, Ren K. OneFi: One-shot recognition for unseen gesture via COTS WiFi. In *Proc. the 19th ACM Conference on Embedded Networked Sensor Systems*, Nov. 2021, pp.206–219. DOI: 10.1145/3485730.3485936.

[27] Gao R Y, Li W W, Xie Y X, Yi E Z, Wang L Y, Wu D, Zhang D Q. Towards robust gesture recognition by characterizing the sensing quality of WiFi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2022, 6(1): Article No. 11. DOI: 10.1145/3517241.

[28] Zhou Y X, Chen H X, Huang C Y, Zhang Q. WiADv: Practical and robust adversarial attack against WiFi-based gesture recognition system. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2022, 6(2): Article No. 92. DOI: 10.1145/3534618.

[29] Wu D, Gao R Y, Zeng Y W, Liu J Y, Wang L Y, Gu T, Zhang D Q. FingerDraw: Sub-wavelength level finger motion tracking with WiFi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2020, 4(1): Article No. 31. DOI: 10.1145/3380981.

[30] Sun L, Sen S, Koutsonikolas D, Kim K H. WiDraw: Enabling hands-free drawing in the air on commodity WiFi devices. In *Proc. the 21st Annual International Conference on Mobile Computing and Networking*, Sept. 2015, pp.77–89. DOI: 10.1145/2789168.2790129.

[31] Hernandez S M, Bulut E. Performing WiFi sensing with off-the-shelf smartphones. In *Proc. the 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Mar. 2020. DOI: 10.1109/PerComWorkshops48775.2020.9156194.

[32] Zeng Y W, Wu D, Xiong J, Yi E Z, Gao R Y, Zhang D Q. FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019, 3(3): Article No. 121. DOI: 10.1145/3351279.

[33] Wang G H, Zou Y P, Zhou Z M, Wu K S, Ni L M. We can hear you with Wi-Fi! In *Proc. the 20th ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*, Sept. 2014, pp.593–604. DOI: 10.1145/2639108.2639112.

[34] Du C L, Yuan X Q, Lou W J, Hou Y T. Context-free fine-grained motion sensing using WiFi. In *Proc. the 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, Jun. 2018. DOI: 10.1109/SAHCN.2018.8397118.

[35] Guo X N, Liu B, Shi C, Liu H B, Chen Y Y, Chuah M C. WiFi-enabled smart human dynamics monitoring. In *Proc. the 15th ACM Conference on Embedded Network Sensor Systems*, Nov. 2017, Article No. 16. DOI: 10.1145/3131672.3131692.

[36] Xi W, Zhao J Z, Li X Y, Zhao K, Tang S J, Liu X, Jiang Z P. Electronic frog eye: Counting crowd using WiFi. In *Proc. the 2014 IEEE Conference on Computer Communications*, Apr. 27–May 2, 2014, pp.361–369. DOI: 10.1109/INFOCOM.2014.6847958.

[37] Zou H, Zhou Y X, Yang J F, Jiang H, Xie L H, Spanos

C J. DeepSense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network. In *Proc. the 2018 IEEE International Conference on Communications (ICC 2018)*, May 2018. DOI: 10.1109/ICC.2018.8422895.

[38] Jiang W J, Miao C L, Ma F L, Yao S C, Wang Y Q, Yuan Y, Xue H F, Song C, Ma X, Koutsonikolas D, Xu W Y, Su L. Towards environment independent device free human activity recognition. In *Proc. the 24th Annual International Conference on Mobile Computing and Networking*, Oct. 2018, pp.289–304. DOI: 10.1145/3241539.3241548.

[39] Xue H F, Jiang W J, Miao C L, Ma F L, Wang S Y, Yuan Y, Yao S C, Zhang A D, Su L. DeepMV: Multi-view deep learning for device-free human activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2020, 4(1): Article No. 34. DOI: 10.1145/3380980.

[40] Chen Z H, Zhang L, Jiang C Y, Cao Z G, Cui W. WiFi CSI based passive human activity recognition using attention based BLSTM. *IEEE Trans. Mobile Computing*, 2018, 18(11): 2714–2724. DOI: 10.1109/TMC.2018.2878233.

[41] Wang W, Liu A X, Shahzad M, Ling K, Lu S L. Understanding and modeling of WiFi signal based human activity recognition. In *Proc. the 21st Annual International Conference on Mobile Computing and Networking*, Sept. 2015, pp.65–76. DOI: 10.1145/2789168.2790093.

[42] Zhang F, Wu C S, Wang B B, Lai H Q, Han Y, Liu K J R. WiDetect: Robust motion detection with a statistical electromagnetic model. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019, 3(3): Article No. 122. DOI: 10.1145/3351280.

[43] Wang W, Liu A X, Shahzad M, Ling K, Lu S L. Understanding and modeling of WiFi signal based human activity recognition. In *Proc. the 21st Annual International Conference on Mobile Computing and Networking*, Sept. 2015, pp.65–76. DOI: 10.1145/2789168.2790093.

[44] Li S J, Liu Z P, Zhang Y, Lv Q, Niu X P, Wang L Y, Zhang D Q. WiBorder: Precise Wi-Fi based boundary sensing via through-wall discrimination. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2020, 4(3): Article No. 89. DOI: 10.1145/3411834.

[45] Li H J, He X, Chen X K, Fang Y Y, Fang Q. Wi-Motion: A robust human activity recognition using WiFi signals. *IEEE Access*, 2019, 7: 153287–153299. DOI: 10.1109/ACCESS.2019.2948102.

[46] Pu Q, Gupta S, Gollakota S, Patel S. Whole-home gesture recognition using wireless signals. In *Proc. the 19th Annual International Conference on Mobile Computing & Networking*, Sept. 2013, pp.27–38. DOI: 10.1145/2500423.2500436.

[47] Zheng X L, Wang J L, Shangguan L F, Zhou Z M, Liu Y H. Smokey: Ubiquitous smoking detection with commercial WiFi infrastructures. In *Proc. the 35th Annual IEEE International Conference on Computer Communications*, Apr. 2016. DOI: 10.1109/INFOCOM.2016.7524399.

[48] Wu X H, Chu Z B, Yang P L, Xiang C C, Zheng X, Huang W C. TW-See: Human activity recognition through the wall with commodity Wi-Fi devices. *IEEE Trans. Vehicular Technology*, 2019, 68(1): 306–319. DOI: 10.1109/TVT.2018.2878754.

[49] Wang H, Zhang D Q, Ma J Y, Wang Y S, Wang Y X, Wu D, Gu T, Xie B. Human respiration detection with commodity WiFi devices: Do user location and body orientation matter? In *Proc. the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Sept. 2016, pp.25–36. DOI: 10.1145/2971648.2971744.

[50] Liu J Y, Zeng Y W, Gu T, Wang L Y, Zhang D Q. WiPhone: Smartphone-based respiration monitoring using ambient reflected WiFi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2021, 5(1): Article No. 23. DOI: 10.1145/3448092.

[51] Yin Y Q, Yang X, Xiong J, Lee S I, Chen P P, Niu Q. Ubiquitous smartphone-based respiration sensing with Wi-Fi signal. *IEEE Internet of Things Journal*, 2022, 9(2): 1479–1490. DOI: 10.1109/JIOT.2021.3088338.

[52] Korany B, Karanam C R, Cai H, Mostofi Y. XModal-ID: Using WiFi for through-wall person identification from candidate video footage. In *Proc. the 25th Annual International Conference on Mobile Computing and Networking*, Aug. 2019, Article No. 36. DOI: 10.1145/3300061.3345437.

[53] Zeng Y Z, Pathak P H, Mohapatra P. WiWho: WiFi-based person identification in smart spaces. In *Proc. the 15th ACM/IEEE International Conference on Information Processing in Sensor Networks*, Apr. 2016. DOI: 10.1109/IPSN.2016.7460727.

[54] Zhang J, Wei B, Hu W, Kanhere S S. WiFi-ID: Human identification using WiFi signal. In *Proc. the 2016 International Conference on Distributed Computing in Sensor Systems*, May 2016, pp.75–82. DOI: 10.1109/DCOSS.2016.30.

[55] Xu Y, Yang W, Wang J X, Zhou X, Li H, Huang L S. WiStep: Device-free step counting with WiFi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018, 1(4): Article No. 172. DOI: 10.1145/3161415.

[56] Li X, Li S J, Zhang D Q, Xiong J, Wang Y S, Mei H. Dynamic-MUSIC: Accurate device-free indoor localization. In *Proc. the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Sept. 2016, pp.196–207. DOI: 10.1145/2971648.2971665.

[57] Qian K, Wu C S, Yang Z, Liu Y H, Jamieson K. Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi. In *Proc. the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Jul. 2017, Article No. 6. DOI: 10.1145/3084041.

3084067.

[58] Qian K, Wu C S, Zhang Y, Zhang G D, Yang Z, Liu Y H. Widar2.0: Passive human tracking with a single Wi-Fi link. In *Proc. the 16th Annual International Conference on Mobile Systems, Applications, and Services*, Jun. 2018, pp.350–361. DOI: 10.1145/3210240.3210314.

[59] Xie Y X, Xiong J, Li M, Jamieson K. mD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking. In *Proc. the 25th Annual International Conference on Mobile Computing and Networking*, Aug. 2019, Article No. 8. DOI: 10.1145/3300061.3300133.

[60] Abdel-Nasser H, Samir R, Sabek I, Youssef M. MonoPHY: Mono-stream-based device-free WLAN localization via physical layer information. In *Proc. the 2013 IEEE Wireless Communications and Networking Conference*, Apr. 2013, pp.4546–4551. DOI: 10.1109/WCNC.2013.6555311.

[61] Wang X Y, Gao L J, Mao S W, Pandey S. DeepFi: Deep learning for indoor fingerprinting using channel state information. In *Proc. the 2015 IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2015, pp.1666–1671. DOI: 10.1109/WCNC.2015.7127718.

[62] Li H, Chen X, Wang J, Wu D, Liu X. DAFI: WiFi-based device-free indoor localization via domain adaptation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2021, 5(4): Article No. 167. DOI: 10.1145/3494954.

[63] Alkandari M, Basu D, Hasan S F. A Wi-Fi based passive technique for speed estimation in indoor environments. In *Proc. the 2nd Workshop on Recent Trends in Telecommunications Research*, Feb. 2017. DOI: 10.1109/RTTR.2017.7887877.

[64] Basu D, Hasan S F. Assessing device-free passive localization with a single access point. In *Proc. the 14th International Conference on Dependable, Autonomic and Secure Computing, 14th International Conference on Pervasive Intelligence and Computing, 2nd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress*, Aug. 2016, pp.493–496. DOI: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.96.

[65] Oguntala G, Obeidat H, Al Khambashi M, Elmegri F, Abd-Alhameed R A, Yuxiang T, Noras J. Design framework for unobtrusive patient location recognition using passive RFID and particle filtering. In *Proc. the 2017 Internet Technologies and Applications*, Sept. 2017, pp.212–217. DOI: 10.1109/ITECHA.2017.8101941.

[66] Xiao J, Wu K S, Yi Y W, Wang L, Ni L M. FIMD: Fine-grained device-free motion detection. In *Proc. the 18th IEEE International Conference on Parallel and Distributed Systems*, Dec. 2012, pp.229–235. DOI: 10.1109/ICPADS.2012.40.

[67] Wu K S, Xiao J, Yi Y W, Gao M, Ni L M. FILA: Fine-grained indoor localization. In *Proc. the 2012 IEEE INFOCOM*, Mar. 2012, pp.2210–2218. DOI: 10.1109/INFCOM.2012.6195606.

[68] Jiang G Y, Li M L, Liu X J, Liu W P, Jia Y F, Jiang H B, Lei J L, Xiao F, Zhang K. WiDE: WiFi distance based group profiling via machine learning. *IEEE Trans. Mobile Computing*, 2023, 22(1): 607–620. DOI: 10.1109/TMC.2021.3073848.

[69] Chen X, Li H, Zhou C Y, Liu X, Wu D, Dudek G. Fidora: Robust WiFi-based indoor localization via unsupervised domain adaptation. *IEEE Internet of Things Journal*, 2022, 9(12): 9872–9888. DOI: 10.1109/JIOT.2022.3163391.

[70] Bai Y H, Wang Z J, Zheng K Y, Wang X R, Wang J M. WiDrive: Adaptive WiFi-based recognition of driver activity for real-time and safe takeover. In *Proc. the 39th IEEE International Conference on Distributed Computing Systems*, Jul. 2019, pp.901–911. DOI: 10.1109/ICDCS.2019.00094.

[71] Peng H J, Jia W J. WiFind: Driver fatigue detection with fine-grained Wi-Fi signal features. In *Proc. the 2017 IEEE Global Communications Conference*, Dec. 2017. DOI: 10.1109/GLOCOM.2017.8253925.

[72] Kong H, Lu L, Yu J D, Chen Y Y, Kong L H, Li M L. FingerPass: Finger gesture-based continuous user authentication for smart homes using commodity WiFi. In *Proc. the 20th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Jul. 2019, pp.201–210. DOI: 10.1145/3323679.3326518.

[73] Lu H, Pan W, Lane N D, Choudhury T, Campbell A T. SoundSense: Scalable sound sensing for people-centric applications on mobile phones. In *Proc. the 7th International Conference on Mobile Systems, Applications, and Services*, Jun. 2009, pp.165–178. DOI: 10.1145/1555816.1555834.

[74] Yatani K, Truong K N. BodyScope: A wearable acoustic sensor for activity recognition. In *Proc. the 2012 ACM Conference on Ubiquitous Computing*, Sept. 2012, pp.341–350. DOI: 10.1145/2370216.2370269.

[75] Prakash J, Yang Z J, Wei Y L, Hassanieh H, Choudhury R R. EarSense: Earphones as a teeth activity sensor. In *Proc. the 26th Annual International Conference on Mobile Computing and Networking*, Apr. 2020, Article No. 40. DOI: 10.1145/3372224.3419197.

[76] Xie Y D, Li F, Wu Y, Wang Y. HearFit: Fitness monitoring on smart speakers via active acoustic sensing. In *Proc. the 2021 IEEE Conference on Computer Communications*, May 2021. DOI: 10.1109/INFOCOM42981.2021.9488811.

[77] Liang D W, Thomaz E. Audio-based activities of daily living (ADL) recognition with large-scale acoustic embeddings from online videos. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019, 3(1): Article No. 17. DOI: 10.1145/3314404.

[78] Nicolaou P, Efstratiou C. Tracking daily routines of elderly users through acoustic sensing: An unsupervised learning approach. In *Proc. the 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom*

Workshops), Mar. 2022, pp.391–396. DOI: 10.1109/PerComWorkshops53856.2022.9767404.

[79] Aumi M T I, Gupta S, Goel M, Larson E, Patel S. DopLink: Using the Doppler effect for multi-device interaction. In *Proc. the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Sept. 2013, pp.583–586. DOI: 10.1145/2493432.2493515.

[80] Yang Q F, Tang H, Zhao X B, Li Y, Zhang S F. Dolphin: Ultrasonic-based gesture recognition on smartphone platform. In *Proc. the 17th IEEE International Conference on Computational Science and Engineering*, Dec. 2014, pp.1461–1468. DOI: 10.1109/CSE.2014.273.

[81] Sun K, Zhao T, Wang W, Xie L. VSkin: Sensing touch gestures on surfaces of mobile devices using acoustic signals. In *Proc. the 24th Annual International Conference on Mobile Computing and Networking*, Oct. 2018, pp.591–605. DOI: 10.1145/3241539.3241568.

[82] Zhang Y T, Wang J L, Wang W Y, Wang Z, Liu Y H. Vernier: Accurate and fast acoustic motion tracking using mobile devices. In *Proc. the 2018 IEEE Conference on Computer Communications*, Apr. 2018, pp.1709–1717. DOI: 10.1109/INFOCOM.2018.8486365.

[83] Luo G, Chen M S, Li P, Zhang M T, Yang P L. SoundWrite II: Ambient acoustic sensing for noise tolerant device-free gesture recognition. In *Proc. the 23rd IEEE International Conference on Parallel and Distributed Systems (ICPADS)*, Dec. 2017, pp.121–126. DOI: 10.1109/ICPADS.2017.00027.

[84] Wang W, Liu A X, Sun K. Device-free gesture tracking using acoustic signals. In *Proc. the 22nd Annual International Conference on Mobile Computing and Networking*, Oct. 2016, pp.82–94. DOI: 10.1145/2973750.2973764.

[85] Ling K, Dai H P, Liu Y T, Liu A X, Wang W, Gu Q. UltraGesture: Fine-grained gesture sensing and recognition. *IEEE Trans. Mobile Computing*, 2020, 21(7): 2620–2636. DOI: 10.1109/TMC.2020.3037241.

[86] Wang Y W, Shen J X, Zheng Y Q. Push the limit of acoustic gesture recognition. *IEEE Trans. Mobile Computing*, 2020, 21(5): 1798–1811. DOI: 10.1109/TMC.2020.3032278.

[87] Wang P H, Jiang R B, Liu C. Amaging: Acoustic hand imaging for self-adaptive gesture recognition. In *Proc. the 2022 IEEE Conference on Computer Communications*, May 2022, pp.80–89. DOI: 10.1109/INFOCOM48880.2022.9796906.

[88] Larson E C, Goel M, Boriello G, Heltshe S, Rosenfeld M, Patel S N. SpiroSmart: Using a microphone to measure lung function on a mobile phone. In *Proc. the 2012 ACM Conference on Ubiquitous Computing*, Sept. 2012, pp.280–289. DOI: 10.1145/2370216.2370261.

[89] Qian K, Wu C S, Xiao F, Zheng Y, Zhang Y, Yang Z, Liu Y H. Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart devices. In *Proc. the 2018 IEEE Conference on Computer Communications*, Apr. 2018, pp.1574–1582. DOI: 10.1109/INFOCOM.2018.8485978.

[90] Song X Z, Yang B Y, Yang G, Chen R R, Forno E, Chen W, Gao W. SpiroSonic: Monitoring human lung function via acoustic sensing on commodity smartphones. In *Proc. the 26th Annual International Conference on Mobile Computing and Networking*, Apr. 2020, Article No. 52. DOI: 10.1145/3372224.3419209.

[91] Wan H R, Shi S Y, Cao W Y, Wang W, Chen G H. RespTracker: Multi-user room-scale respiration tracking with commercial acoustic devices. In *Proc. the 2021 IEEE Conference on Computer Communications*, May 2021. DOI: 10.1109/INFOCOM42981.2021.9488881.

[92] Vernon J, Canyelles-Pericas P, Torun H, Binns R, Ng W P, Fu Y Q. Apnoea-Pi: Sleep disorder monitoring with open-source electronics and acoustics. In *Proc. the 26th International Conference on Automation and Computing (ICAC)*, Sept. 2021. DOI: 10.23919/ICAC50006.2021.9594073.

[93] Liu K K, Liu X X, Li X L. Guoguo: Enabling fine-grained indoor localization via smartphone. In *Proc. the 11th Annual International Conference on Mobile Systems, Applications, and Services*, Jun. 2013, pp.235–248. DOI: 10.1145/2462456.2464450.

[94] Zhou B, Elbadry M, Gao R P, Ye F. BatMapper: Acoustic sensing based indoor floor plan construction using smartphones. In *Proc. the 15th Annual International Conference on Mobile Systems, Applications, and Services*, Jun. 2017, pp.42–55. DOI: 10.1145/3081333.3081363.

[95] Pradhan S, Baig G, Mao W G, Qiu L L, Chen G H, Yang B. Smartphone-based acoustic indoor space mapping. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018, 2(2): Article No. 75. DOI: 10.1145/3214278.

[96] Wijers M, Loveridge A, Macdonald D W, Markham A. CARACAL: A versatile passive acoustic monitoring tool for wildlife research and conservation. *Bioacoustics*, 2021, 30(1): 41–57. DOI: 10.1080/09524622.2019.1685408.

[97] Dong L J, Tao Q, Hu Q C, Deng S J, Chen Y C, Luo Q M, Zhang X H. Acoustic emission source location method and experimental verification for structures containing unknown empty areas. *International Journal of Mining Science and Technology*, 2022, 32(3): 487–497. DOI: 10.1016/j.ijmst.2022.01.002.

[98] Kafle M D, Fong S, Narasimhan S. Active acoustic leak detection and localization in a plastic pipe using time delay estimation. *Applied Acoustics*, 2022, 187: 108482. DOI: 10.1016/j.apacoust.2021.108482.

[99] Wang J J, Zhao K C, Zhang X Y, Peng C Y. Ubiquitous keyboard for small mobile devices: Harnessing multipath fading for fine-grained keystroke localization. In *Proc. the 12th Annual International Conference on Mobile Systems, Applications, and Services*, Jun. 2014, pp.14–27. DOI: 10.1145/2594368.2594384.

[100] Alegre F, Vipperla R, Evans N, Fauve B. On the vulnerability of automatic speaker recognition to spoofing attacks with artificial signals. In *Proc. the 20th European Signal Processing Conference (EUSIPCO)*, Aug. 2012,

pp.36–40.

[101] Chauhan J, Hu Y N, Seneviratne S, Misra A, Seneviratne A, Lee Y. BreathPrint: Breathing acoustics-based user authentication. In *Proc. the 15th Annual International Conference on Mobile Systems, Applications, and Services*, Jun. 2017, pp.278–291. DOI: 10.1145/3081333.3081355.

[102] Zhang G M, Yan C, Ji X Y, Zhang T C, Zhang T M, Xu W Y. DolphinAttack: Inaudible voice commands. In *Proc. the 2017 ACM SIGSAC Conference on Computer and Communications Security*, Oct. 2017, pp.103–117. DOI: 10.1145/3133956.3134052.

[103] Yuan X J, Chen Y X, Zhao Y, Long Y H, Liu X K, Chen K, Zhang S Z, Huang H Q, Wang X F, Gunter C A. CommanderSong: A systematic approach for practical adversarial voice recognition. arXiv: 1801.08535, 2018. https://arxiv.org/abs/1801.08535, February 2023.

[104] Gao Y, Wang W, Phoha V V, Sun W, Jin Z P. EarEcho: Using ear canal echo for wearable authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019, 3(3): Article No. 81. DOI: 10.1145/3351239.

[105] Lu L, Yu J D, Chen Y Y, Liu H B, Zhu Y M, Liu Y F, Li M L. LipPass: Lip reading-based user authentication on smartphones leveraging acoustic signals. In *Proc. the 2018 IEEE Conference on Computer Communications*, Apr. 2018, pp.1466–1474. DOI: 10.1109/INFOCOM.2018.8486283.

[106] Chen M Q, Lin J W, Zou Y P, Ruby R, Wu K S. SilentSign: Device-free handwritten signature verification through acoustic sensing. In *Proc. the 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Mar. 2020. DOI: 10.1109/PerCom45495.2020.9127372.

[107] Ferlini A, Ma D, Harle R, Mascolo C. EarGate: Gait-based user identification with in-ear microphones. In *Proc. the 27th Annual International Conference on Mobile Computing and Networking*, Oct. 2021, pp.337–349. DOI: 10.1145/3447993.3483240.

[108] Ren Y Z, Wen P, Liu H B, Zheng Z R, Chen Y Y, Huang P C, Li H W. Proximity-Echo: Secure two factor authentication using active sound sensing. In *Proc. the 2021 IEEE Conference on Computer Communications*, May 2021. DOI: 10.1109/INFOCOM42981.2021.9488866.

[109] Balagani K, Cardaioli M, Cecconello S, Conti M, Tsudik G. We can hear your PIN drop: An acoustic side-channel attack on ATM PIN pads. In *Proc. the 27th European Symposium on Research in Computer Security*, Sept. 2022, pp.633–652. DOI: 10.1007/978-3-031-17140-6_31.

[110] Xie Y D, Li F, Wu Y, Chen H J, Zhao Z Y, Wang Y. TeethPass: Dental occlusion-based user authentication via in-ear acoustic sensing. In *Proc. the 2022 IEEE Conference on Computer Communications*, May 2022, pp.1789–1798. DOI: 10.1109/INFOCOM48880.2022.9796951.

[111] Wang Z, Ren Y L, Chen Y Y, Yang J. ToothSonic: Earable authentication via acoustic toothprint. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2022, 6(2): Article No. 78. DOI: 10.1145/3534606.

[112] Ma Y S, Zhou G, Wang S Q. WiFi sensing with channel state information: A survey. *ACM Computing Surveys*, 2019, 52(3): Article No. 46. DOI: 10.1145/3310194.

[113] Xie Y X, Li Z J, Li M. Precise power delay profiling with commodity WiFi. In *Proc. the 21st Annual International Conference on Mobile Computing and Networking*, Sept. 2015, pp.53–64. DOI: 10.1145/2789168.2790124.

[114] Jiang Z P, Luan T H, Ren X C, Lv D T, Hao H, Wang J, Zhao K, Xi W, Xu Y S, Li R. Eliminating the barriers: Demystifying Wi-Fi baseband design and introducing the PicoScenes Wi-Fi sensing platform. *IEEE Internet of Things Journal*, 2022, 9(6): 4476–4496. DOI: 10.1109/JIOT.2021.3104666.

[115] Zheng F, Zhang G L, Song Z J. Comparison of different implementations of MFCC. *Journal of Computer Science and Technology*, 2001, 16(6): 582–589. DOI: 10.1007/BF02943243.

[116] Pedretti L W, Early M B. Occupational Therapy: Practice Skills for Physical Dysfunction. Mosby London, 2001.

[117] Santhalingam P S, Hosain A A, Zhang D, Pathak P, Rangwala H, Kushalnagar R. mmASL: Environment-independent ASL gesture recognition using 60 GHz millimeter-wave signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2020, 4(1): Article No. 26. DOI: 10.1145/3381010.

[118] Gallo P, Mangione S. RSS-eye: Human-assisted indoor localization without radio maps. In *Proc. the 2015 IEEE International Conference on Communications*, Jun. 2015, pp.1553–1558. DOI: 10.1109/ICC.2015.7248545.

[119] Liu H, Darabi H, Banerjee P, Liu J. Survey of wireless indoor positioning techniques and systems. *IEEE Trans. Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2007, 37(6): 1067–1080. DOI: 10.1109/TSMCC.2007.905750.

[120] Pahlavan K, Li X R, Makela J P. Indoor geolocation science and technology. *IEEE Communications Magazine*, 2002, 40(2): 112–118. DOI: 10.1109/35.983917.

[121] Chen K Y, Ashbrook D, Goel M, Lee S H, Patel S. AirLink: Sharing files between multiple devices using in-air gestures. In *Proc. the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Sept. 2014, pp.565–569. DOI: 10.1145/2632048.2632090.

[122] Zhang M T, Yang P L, Tian C, Shi L, Tang S J, Xiao F. SoundWrite: Text input on surfaces through mobile acoustic sensing. In *Proc. the 1st International Workshop on Experiences with the Design and Implementation of Smart Objects*, Sept. 2015, pp.13–17. DOI: 10.1145/2797044.2797045.

[123] Wang X, Sun K, Zhao T, Wang W, Gu Q. Dynamic speed warping: Similarity-based one-shot learning for de-

vice-free gesture signals. In *Proc. the 2020 IEEE Conference on Computer Communications*, Jul. 2020, pp.556–565. DOI: 10.1109/INFOCOM41043.2020.9155491.

[124] Weiss K, Khoshgoftaar T M, Wang D D. A survey of transfer learning. *Journal of Big Data*, 2016, 3(1): Article No. 9. DOI: 10.1186/s40537-016-0043-6.

[125] Wang Y Q, Yao Q M, Kwok J T, Ni L M. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys*, 2021, 53(3): Article No. 63. DOI: 10.1145/3386252.

[126] Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, 2019, 58: 101552. DOI: 10.1016/j.media.2019.101552.

[127] Ozcan T, Basturk A. Transfer learning-based convolutional neural networks with heuristic optimization for hand gesture recognition. *Neural Computing and Applications*, 2019, 31(12): 8955–8970. DOI: 10.1007/s00521-019-04427-y.

[128] Rahimian E, Zabihi S, Asif A, Farina D, Atashzar S F, Mohammadi A. FS-HGR: Few-shot learning for hand gesture recognition via electromyography. *IEEE Trans. Neural Systems and Rehabilitation Engineering*, 2021, 29: 1004–1015. DOI: 10.1109/TNSRE.2021.3077413.

[129] Wang J, Zhang L, Wang C C, Ma X R, Gao Q H, Lin B. Device-free human gesture recognition with generative adversarial networks. *IEEE Internet of Things Journal*, 2020, 7(8): 7678–7688. DOI: 10.1109/JIOT.2020.2988291.

[130] Liu C, Wang P H, Jiang R B, Zhu Y M. AMT: Acoustic multi-target tracking with smartphone MIMO system. In *Proc. the 2021 IEEE Conference on Computer Communications*, May 2021. DOI: 10.1109/INFOCOM42981.2021.9488768.

[131] Yun S K, Chen Y C, Qiu L L. Turning a mobile device into a mouse in the air. In *Proc. the 13th Annual International Conference on Mobile Systems, Applications, and Services*, May 2015, pp.15–29. DOI: 10.1145/2742647.2742662.

[132] Mao W G, He J, Qiu L L. CAT: High-precision acoustic motion tracking. In *Proc. the 22nd Annual International Conference on Mobile Computing and Networking*, Oct. 2016, pp.69–81. DOI: 10.1145/2973750.2973755.

[133] Chen H J, Li F, Wang Y. EchoTrack: Acoustic device-free hand tracking on smart phones. In *Proc. the 2017 IEEE Conference on Computer Communications*, May 2017. DOI: 10.1109/INFOCOM.2017.8057101.

[134] Nandakumar R, Iyer V, Tan D, Gollakota S. FingerIO: Using active sonar for fine-grained finger tracking. In *Proc. the 2016 CHI Conference on Human Factors in Computing Systems*, May 2016, pp.1515–1525. DOI: 10.1145/2858036.2858580.

[135] Yun S K, Chen Y C, Zheng H H, Qiu L L, Mao W G. Strata: Fine-grained acoustic-based device-free tracking. In *Proc. the 15th Annual International Conference on Mobile Systems, Applications, and Services*, Jun. 2017, pp.15–28. DOI: 10.1145/3081333.3081356.

[136] Lu L, Liu J, Yu J D, Chen Y Y, Zhu Y M, Kong L H, Li M L. Enable traditional laptops with virtual writing capability leveraging acoustic signals. *The Computer Journal*, 2021, 64(12): 1814–1831. DOI: 10.1093/comjnl/bxz153.

[137] Liu Y, Zhang W X, Yang Y, Fang W D, Qin F, Dai X W. PAMT: Phase-based acoustic motion tracking in multipath fading environments. In *Proc. the 2019 IEEE Conference on Computer Communications*, Apr. 29–May 2, 2019, pp.2386–2394. DOI: 10.1109/INFOCOM.2019.8737366.

[138] Kumar M, Veeraraghavan A, Sabharwal A. DistancePPG: Robust non-contact vital signs monitoring using a camera. *Biomedical Optics Express*, 2015, 6(5): 1565–1588. DOI: 10.1364/BOE.6.001565.

[139] Jia Z H, Bonde A, Li S G, Xu C R, Wang J X, Zhang Y Y, Howard R E, Zhang P. Monitoring a person's heart rate and respiratory rate on a shared bed using geophones. In *Proc. the 15th ACM Conference on Embedded Network Sensor Systems (SenSys 2017)*, Nov. 2017, Article No. 6. DOI: 10.1145/3131672.3131679.

[140] Jia Z H, Alaziz M, Chi X, Howard R E, Zhang Y Y, Zhang P, Trappe W, Sivasubramaniam A, An N. HB-phone: A bed-mounted geophone-based heartbeat monitoring system. In *Proc. the 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, Apr. 2016. DOI: 10.1109/IPSN.2016.7460676.

[141] Li F Y, Valero M, Shahriar H, Khan R A, Ahamed S I. Wi-COVID: A COVID-19 symptom detection and patient monitoring framework using WiFi. *Smart Health*, 2021, 19: 100147. DOI: 10.1016/j.smhl.2020.100147.

[142] Nandakumar R, Gollakota S, Watson N. Contactless sleep apnea detection on smartphones. In *Proc. the 13th Annual International Conference on Mobile Systems, Applications, and Services*, May 2015, pp.45–57. DOI: 10.1145/2742647.2742674.

[143] Li Y, Zeng Z L, Popescu M, Ho K C. Acoustic fall detection using a circular microphone array. In *Proc. the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Aug. 31–Sept. 4, 2010, pp.2242–2245. DOI: 10.1109/IEMBS.2010.5627368.

[144] Ren Y Z, Wang C, Yang J, Chen Y Y. Fine-grained sleep monitoring: Hearing your breathing with smartphones. In *Proc. the 2015 IEEE Conference on Computer Communications (INFOCOM)*, Apr. 26–May 1, 2015, pp.1194–1202. DOI: 10.1109/INFOCOM.2015.7218494.

[145] Yang J, Sidhom S, Chandrasekaran G, Vu T, Liu H B, Cecan N, Chen Y Y, Gruteser M, Martin R P. Detecting driver phone use leveraging car speakers. In *Proc. the 17th Annual International Conference on Mobile Computing and Networking*, Sept. 2011, pp.97–108. DOI: 10.1145/2030613.2030625.

[146] Xu X Y, Gao H, Yu J D, Chen Y Y, Zhu Y M, Xue G T, Li M L. ER: Early recognition of inattentive driving leveraging audio devices on smartphones. In *Proc. the*

2017 IEEE Conference on Computer Communications, May 2017. DOI: 10.1109/INFOCOM.2017.8057022.

[147] Xu X Y, Yu J D, Chen Y Y, Zhu Y M, Qian S Y, Li M L. Leveraging audio signals for early recognition of inattentive driving with smartphones. *IEEE Trans. Mobile Computing*, 2018, 17(7): 1553–1567. DOI: 10.1109/TMC.2017.2772253.

[148] Xu X Y, Yu J D, Chen Y Y, Zhu Y M, Kong L H, Li M L. BreathListener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals. In *Proc. the 17th Annual International Conference on Mobile Systems, Applications, and Services*, Jun. 2019, pp.54–66. DOI: 10.1145/3307334.3326074.

[149] Liu S C, Zhou Z M, Du J Z, Shangguan L F, Han J, Wang X. UbiEar: Bringing location-independent sound awareness to the hard-of-hearing people with smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2017, 1(2): Article No. 17. DOI: 10.1145/3090082.

[150] Nishimura Y, Imai N, Yoshihara K. A proposal on direction estimation between devices using acoustic waves. In *Proc. the 8th International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, Dec. 2011, pp.25–36. DOI: 10.1007/978-3-642-30973-1_3.

[151] Zhang Z B, Chu D, Chen X M, Moscibroda T. SwordFight: Enabling a new class of phone-to-phone action games on commodity phones. In *Proc. the 10th International Conference on Mobile Systems, Applications, and Services*, Jun. 2012, pp.1–14. DOI: 10.1145/2307636.2307638.

[152] Liu H B, Gan Y, Yang J, Sidhom S, Wang Y, Chen Y Y, Ye F. Push the limit of WiFi based localization for smartphones. In *Proc. the 18th Annual International Conference on Mobile Computing and Networking*, Aug. 2012, pp.305–316. DOI: 10.1145/2348543.2348581.

[153] Nandakumar R, Chintalapudi K K, Padmanabhan V N. Centaur: Locating devices in an office environment. In *Proc. the 18th Annual International Conference on Mobile Computing and Networking*, Aug. 2012, pp.281–292. DOI: 10.1145/2348543.2348579.

[154] Tarzia S P, Dinda P A, Dick R P, Memik G. Indoor localization without infrastructure using the acoustic background spectrum. In *Proc. the 9th International Conference on Mobile Systems, Applications, and Services*, Jun. 2011, pp.155–168. DOI: 10.1145/1999995.2000011.

[155] Tung Y C, Shin K G. EchoTag: Accurate infrastructure-free indoor location tagging with smartphones. In *Proc. the 21st Annual International Conference on Mobile Computing and Networking*, Sept. 2015, pp.525–536. DOI: 10.1145/2789168.2790102.

[156] Huang W C, Xiong Y, Li X Y, Lin H, Mao X F, Yang P L, Liu Y H. Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones. In *Proc. the 2014 IEEE Conference on Computer Communications*, Apr. 27–May 2, 2014, pp.370–378. DOI: 10.1109/INFOCOM.2014.6847959.

[157] Zhu T, Ma Q, Zhang S F, Liu Y H. Context-free attacks using keyboard acoustic emanations. In *Proc. the 2014 ACM SIGSAC Conference on Computer and Communications Security*, Nov. 2014, pp.453–464. DOI: 10.1145/2660267.2660296.

[158] Liu J, Wang Y, Kar G, Chen Y Y, Yang J, Gruteser M. Snooping keystrokes with mm-level audio ranging on a single phone. In *Proc. the 21st Annual International Conference on Mobile Computing and Networking*, Sept. 2015, pp.142–154. DOI: 10.1145/2789168.2790122.

[159] Liu X Y, Zhou Z, Diao W R, Li Z, Zhang K H. When good becomes evil: Keystroke inference with smartwatch. In *Proc. the 22nd ACM SIGSAC Conference on Computer and Communications Security*, Oct. 2015, pp.1273–1285. DOI: 10.1145/2810103.2813668.

[160] Fang Y Y, Zhao Z W, Wang Z, Min G Y, Cao Y, Huang H J, Yin H. Eavesdrop with PoKeMon: Position free keystroke monitoring using acoustic data. *Future Generation Computer Systems*, 2018, 87: 704–711. DOI: 10.1016/j.future.2017.10.039.

[161] Wu Z Z, Evans N, Kinnunen T, Yamagishi J, Alegre F, Li H Z. Spoofing and countermeasures for speaker verification: A survey. *Speech Communication*, 2015, 66: 130–153. DOI: 10.1016/j.specom.2014.10.005.

[162] Wu Z Z, Gao S, Cling E S, Li H Z. A study on replay attack and anti-spoofing for text-dependent speaker verification. In *Proc. the 2014 Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Dec. 2014. DOI: 10.1109/APSIPA.2014.7041636.

[163] Wu Z Z, Li H Z. Voice conversion and spoofing attack on speaker verification systems. In *Proc. the 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Oct. 29–Nov. 1, 2013. DOI: 10.1109/APSIPA.2013.6694344.

[164] Carlini N, Mishra P, Vaidya T, Zhang Y K, Sherr M, Shields C, Wagner D, Zhou W C. Hidden voice commands. In *Proc. the 25th USENIX Security Symposium (USENIX Security 16)*, Aug. 2016, pp.513–530.

[165] Carlini N, Wagner D. Audio adversarial examples: Targeted attacks on speech-to-text. In *Proc. the 2018 IEEE Security and Privacy Workshops (SPW)*, May 2018. DOI: 10.1109/SPW.2018.00009.

[166] Kasmi C, Esteves J L. IEMI threats for information security: Remote command injection on modern smartphones. *IEEE Trans. Electromagnetic Compatibility*, 2015, 57(6): 1752–1755. DOI: 10.1109/TEMC.2015.2463089.

[167] Wei L Q, Long Y H, Wei H R, Li Y J. New acoustic features for synthetic and replay spoofing attack detection. *Symmetry*, 2022, 14(2): Article No. 274. DOI: 10.3390/sym14020274.

[168] Zhou B, Lohokare J, Gao R P, Ye F. EchoPrint: Two-factor authentication using acoustics and vision on smartphones. In *Proc. the 24th Annual International*

Conference on Mobile Computing and Networking, Oct. 2018, pp.321–336. DOI: 10.1145/3241539.3241575.

[169] Zou Y P, Zhao M, Zhou Z M, Lin J W, Li M, Wu K S. BiLock: User authentication via dental occlusion biometrics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018, 2(3): Article No. 152. DOI: 10.1145/3264962.

[170] Xu W, Yu Z W, Wang Z, Guo B, Han Q. AcousticID: Gait-based human identification using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019, 3(3): Article No. 115. DOI: 10.1145/3351273.

[171] Ding F, Wang D, Zhang Q, Zhao R. ASSV: Handwritten signature verification using acoustic signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019, 3(3): Article No. 80. DOI: 10.1145/3351238.

[172] Shrestha P, Shrestha B, Saxena N. Home alone: The insider threat of unattended wearables and a defense using audio proximity. In *Proc. the 2018 IEEE Conference on Communications and Network Security (CNS)*, May 30–Jun 1, 2018. DOI: 10.1109/CNS.2018.8433216.

[173] Shrestha P, Saxena N. Listening watch: Wearable two-factor authentication using speech signals resilient to near-far attacks. In *Proc. the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, Jun. 2018, pp.99–110. DOI: 10.1145/3212480.3212501.

[174] Feng H, Fawaz K, Shin K G. Continuous authentication for voice assistants. In *Proc. the 23rd Annual International Conference on Mobile Computing and Networking*, Oct. 2017, pp.343–355. DOI: 10.1145/3117811.3117823.

[175] Chen S, Ren K, Piao S X, Wang C, Wang Q, Weng J, Su L, Mohaisen A. You can hear but you cannot steal: Defending against voice impersonation attacks on smartphones. In *Proc. the 37th IEEE International Conference on Distributed Computing Systems (ICDCS)*, Jun. 2017, pp.183–195. DOI: 10.1109/ICDCS.2017.133.

[176] Wang Q, Lin X, Zhou M, Chen Y J, Wang C, Li Q, Luo X Y. VoicePop: A pop noise based anti-spoofing system for voice authentication on smartphones. In *Proc. the 2019 IEEE Conference on Computer Communications*, Apr. 29–May 2, 2019, pp.2062–2070. DOI: 10.1109/INFOCOM.2019.8737422.

[177] Yan C, Ji X Y, Wang K, Jiang Q H, Jin Z Z, Xu W Y. A survey on voice assistant security: Attacks and countermeasures. *ACM Computing Surveys*, 2022, 55(4): Article No. 84. DOI: 10.1145/3527153.

[178] Wang F, Zhou S P, Panev S, Han J S, Huang D. Person-in-WiFi: Fine-grained person perception using WiFi. In *Proc. the 2019 IEEE/CVF International Conference on Computer Vision*, Oct. 27–Nov. 2, 2019, pp.5451–5460. DOI: 10.1109/ICCV.2019.00555.

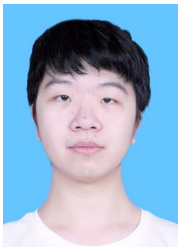
[179] Li C N, Liu Z, Yao Y G, Cao Z C, Zhang M, Liu Y H. Wi-Fi see it all: Generative adversarial network-augmented versatile Wi-Fi imaging. In *Proc. the 18th Conference on Embedded Networked Sensor Systems*, Nov. 2020, pp.436–448. DOI: 10.1145/3384419.3430725.

[180] Yang Q, Wu H X, Huang Q Y, Zhang J, Chen H, Li W C, Tao X F, Zhang Q. Side-lobe can know more: Towards simultaneous communication and sensing for mmWave. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2023, 6(4): Article No. 191. DOI: 10.1145/3569498.

[181] Huang Q Y, Luo Z Q, Zhang J, Wang W, Zhang Q. LoRadar: Enabling concurrent radar sensing and LoRa communication. *IEEE Trans. Mobile Computing*, 2022, 21(6): 2045–2057. DOI: 10.1109/TMC.2020.3035797.

[182] Wang J, Varshney N, Gentile C, Blandino S, Chuang J, Golmie N. Integrated sensing and communication: Enabling techniques, applications, tools and data sets, standardization, and future directions. *IEEE Internet of Things Journal*, 2022, 9(23): 23416–23440. DOI: 10.1109/JIOT.2022.3190845.

Jia-Ling Huang received her B.S. degree in computer science and technology from Chongqing Normal University, Chongqing, in 2021. She is currently pursuing her M.S. degree in the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen. Her major research interests include mobile and ubiquitous computing, and wireless networks.



Yun-Shu Wang received his B.S. degree in computer science and software engineering from Shenzhen University, Shenzhen, in 2022. He is currently pursuing his M.S. degree in the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen. His major research interests are ubiquitous computing and wireless sensing.

Yong-Pan Zou received his Ph.D. degree in computer science and engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2017. He is currently an associate professor in the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen. His main research interests include intelligent sensing, ubiquitous computing, and HCI.

Kai-Shun Wu received his Ph.D. degree in computer science and engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2011. He is currently a distinguished professor in the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen. His main research interests include Internet of Things, wireless networks, and HCI.

Lionel Ming-shuan Ni is the chair professor in the Data Science and Analytics Thrust at the Hong Kong University of Science and Technology (Guangzhou) and chair professor of computer science and engineering at the Hong Kong University of Science and Technology, Hong Kong, since 2019. He is a life fellow of IEEE. Dr. Ni has chaired over 30 professional conferences and has received eight awards for authoring outstanding papers.