Talk2Radar: Talking to mmWave Radars via Smartphone Speaker

Kaiyan Cui[1,2], Leming Shen[2], Yuanqing Zheng[2], Fu Xiao[1], Jinsong Han[3]
[1]Nanjing University of Posts and Telecommunications, Nanjing, China
[2]The Hong Kong Polytechnic University, Hong Kong, China
[3]Zhejiang University, Hangzhou, China

Abstract*—Integrated Sensing and Communication (ISAC) is gaining a tremendous amount of attention from both academia and industry. Recent work has brought communication capability to sensing-oriented mmWave radars, enabling more innovative applications. These solutions, however, either require hardware modifications or suffer from limited data rates. This paper presents *Talk2Radar*, which builds a faster communication channel between smartphone speakers and mmWave radars, without any hardware modification to either commodity smartphones or off-the-shelf radars. In *Talk2Radar*, a smartphone speaker sends messages by playing carefully designed sounds. A mmWave radar acting as a data receiver captures the emitted sounds by detecting the sound-induced smartphone vibrations, and then decodes the messages. *Talk2Radar* characterizes smartphone speakers for speaker-to-mmWave radar communication and addresses a series of technical challenges, including modulation and demodulation of extremely weak sound-induced vibrations, multi-speaker concurrent communication and human motion suppression. We implement and evaluate *Talk2Radar* in various practical settings. Experimental results show that *Talk2Radar* can achieve a data rate of up to 400bps with an average BER of less than 5%, outperforming the state-of-the-art by approximately 33×.

I. Introduction

Millimeter wave (mmWave) radars are widely deployed in robots, smart home appliances, vehicles, and road infrastructure to sense surrounding objects. Compared to other sensing technologies, mmWave radars offer unprecedented sensing resolution and can work in challenging weather conditions, enabling various innovative applications, including object localization and tracking [1, 2], human activity recognition [3, 4] and micro-vibration detection [5–7].

To explore more possibilities, recent work aims to integrate additional communication capability to mmWave radars which are primarily designed for wireless sensing. Such an ISAC system efficiently reuses wireless spectrum as well as hardware components for sensing and communication [8]. As illustrated in Fig. 1, this integration holds immense potential to expand the application scenarios of mmWave radars in smart homes, smart factories and smart traffic environments. With ISAC, a mmWave radar can receive messages while sensing its surroundings. Moreover, each message inherently carries the context information of the sensed object (*e.g.*, location, trajectory, velocity, *etc*), thereby enhancing support for location-based services, human-machine interactions, roadside-to-vehicle communication, and others. For example, upon receiving messages from a user's smartphone, an ISAC-

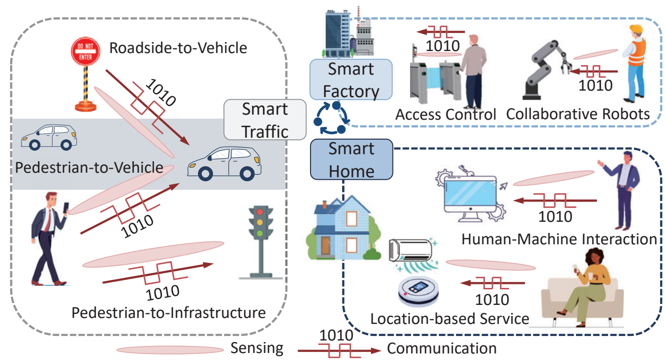


Fig. 1. New applications enabled by ISAC mmWave radars.

enabled mmWave radar can leverage sensed information for personalized interactions with more context information.

Previous works attempt to incorporate communication capability into these sensing-oriented radars by introducing extra communication modules or designing dual-functional radio waveforms [9–12]. For example, Sidense [12] exploits radio side lobes for sensing while utilizing the main lobe for communication. Other work introduces reconfigurable intelligent surfaces to communicate with mmWave radars [13–15]. For example, ROS [15] designs passive mmWave tags with different layouts as road signs, thereby sending roadside information to vehicles with mmWave radars. These solutions, however, require hardware modifications or specialized mmWave tags. mmRipple [16] communicates with mmWave radars by modulating smartphone vibra-motors to reflect incoming mmWave signals with tiny propagation variations. Yet, the peak data rate, limited by the narrow bandwidth and mechanical inertia of vibra-motors, can only be tens of bps.

In this paper, we aim to empower mmWave radars with communication capability and boost the data rate without any hardware modification. To this end, we present *Talk2Radar*, which builds a faster communication channel between smartphones and mmWave radars through a speaker that operates in a higher and wider frequency band [17–20]. In particular, a smartphone sends messages by playing modulated sounds with on-board speakers, while a mmWave radar senses the sound-induced vibrations and decodes the modulated messages. As summarized in Table I, *Talk2Radar* provides a speaker-to-mmWave radar communication channel with several unique advantages. Compared to acoustic communication, *Talk2Radar* utilizes a mmWave radar as a receiver to separate multiple sound sources, which can better support multi-speaker scenar-

TABLE I
COMPARISON WITH RELATED WORKS.

Type	Tx-Rx	No Modif.	Multi-Object Comm.	Higher Data Rate	ISAC
Acoustic	Speaker to Microphone [24]	✓	×	✓	×
Vibration	DC motor to Microphone [25]	×	✓	✓	×
	Vibra-motor to mmWave radar [16]	✓	✓	×	✓
	Speaker to mmWave radar	✓	✓	✓	✓

ios, especially in noisy environments. ISAC radars can also provide the context information about a sound source. In contrast to vibration-based communication, *Talk2Radar* achieves much higher data rates without any hardware modification.

Turing the above basic idea of *Talk2Radar* into practical systems is challenging. First, it is challenging to reliably modulate and send messages with extremely weak sound-induced vibrations. To address this issue, *Talk2Radar* adopts chirp spread spectrum (CSS) modulation and plays wideband audio chirps with different initial frequencies to send messages. The energy concentration effect of CSS demodulation [21–23] benefits the detection and decoding of weak vibrations.

Second, the cross-modal communication between smartphone speakers and mmWave radars entails unique challenges. i) Synchronization issues and hardware heterogeneity cause unwanted offsets and decoding errors. Therefore, we design a special preamble with a known initial frequency as a reference for offset correction. ii) The sound-induced vibrations are weak and suffer from severe frequency-selective fading. Consequently, the chirp signal recovered by mmWave radar exhibits uneven energy distribution and becomes susceptible to interference and noise. To address this issue, we propose a shape-aware audio chirp refinement method, which utilizes a wavelet synchrosqueezed transform (WSST) to reconstruct clean and energy-uniform audio chirps for demodulation.

To further improve the practicality, we address a series of challenges, such as multi-speaker concurrent communication by leveraging spatial and signal diversities. Moreover, *Talk2Radar* suppresses the impact of human motion to improve the system robustness.

We implement *Talk2Radar* and conduct comprehensive experiments under different conditions. *Talk2Radar* achieves $400bps$ data rate at $1m$ communication range and $87.5bps$ at $3m$, with a mean BER of $< 5\%$. The contributions of this paper can be summarized as follows.

- *Talk2Radar* builds a faster communication channel between smartphones and mmWave radars, outperforming the state-of-the-art by approximately 33× in data rate.
- *Talk2Radar* addresses the practical challenges involved in cross-modal communication between smartphones and mmWave radars, including offset correction, modulation and demodulation of the sound-induced vibrations. We also develop novel algorithms to support concurrent communication of multiple transmitters and mitigate the impact of human motion.
- We implement *Talk2Radar* using commodity devices, and conduct comprehensive experiments. We report lessons learned in the development of *Talk2Radar*.

II. SOUND-INDUCED VIBRATION

Talk2Radar measures the sound-induced vibrations to build the communication channel. Therefore, we conduct several experiments to characterize the sound-induced vibrations.

A. Sound-induced Smartphone Vibration

Fig. 2(a) shows the structure of an electro-dynamic speaker widely used in smartphones. Its maximum displacement of the diaphragm $\eta(f)$ at frequency f can be represented as [26, 27]:

$$\eta(f) = \frac{e}{2\pi f_0 B l Q_{ES}} |\gamma(f)| \tag{1}$$

where e is the voltage at the speaker, B is the magnetic field, l is the length of voice coil, and Q_{ES} is the electrical damping. f_0 represents the resonance frequency of the speaker, and $\gamma(f)$ is a dimensionless frequency response function given by:

$$\gamma(f) = 1/[1 - (\frac{f}{f_0})^2 + j\frac{1}{Q_{TS}} \cdot \frac{f}{f_0}] \tag{2}$$

where Q_{TS} represents the total damping effect, including the electrical damping Q_{ES} and the mechanical damping Q_{MS}. Fig. 2(b) plots the normalized displacement with different values of Q_{TS}. Due to the damping, the displacement of diaphragm $\eta(f)$ is non-linear over sound frequency and approximately reaches the peak at the resonance frequency f_0. Generally, the Q_{TS} of micro-speakers is greater than 0.4 with resonance frequencies f_0 between 500Hz and 1000Hz [28].

As the vibration follows a typical harmonic motion, the sound-induced time-varying vibration displacement $\delta(t)$ can be represented as $\delta(t) = \eta(f)cos(2\pi ft)$. When we use a mmWave radar to detect the smartphone emitting sounds in the range bin r and the angle bin θ, the sound-induced vibration will influence the range between the smartphone and radar, leading to the phase changes of the received mmWave signal. The phase measurement $\phi_{r,\theta}(t)$ of the reflected signal from the smartphone $S_{r,\theta}(t)$ can be represented as:

$$\phi_{r,\theta}(t) = 4\pi R(t)/\lambda = 4\pi[R_0 + \underbrace{\eta(f)cos(2\pi ft)}_{\text{sound-induced vibration } \delta(t)}]/\lambda \tag{3}$$

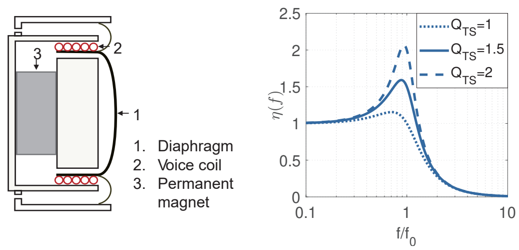
where $R(t)$ is the range between smartphone and radar, *i.e.*, $R(t)$=R_0+$\delta(t)$. R_0 is the initial range and λ is the wavelength.

From Eq. 3, we can see that the phase measurement $\phi_{r,\theta}(t)$ can capture sound-induced smartphone vibrations $\delta(t)$. Note that we assume the vibrating direction of the smartphone aligns with the sensing direction of the mmWave radar. The impact of misalignment will be evaluated and discussed in §VI.

B. Characterizing the Sound-induced Smartphone Vibration

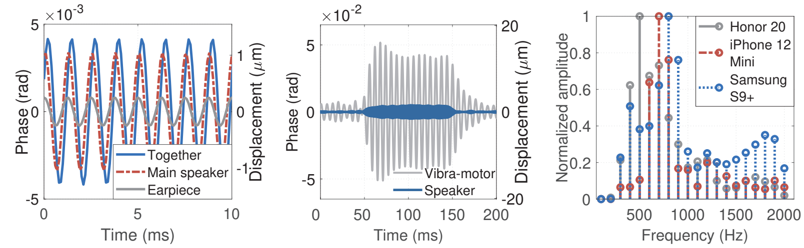
We use a TI AWR1642 mmWave radar to capture the sound-induced vibrations when the smartphone speaker is playing single tones. Each tested smartphone is fixed on a tripod at a distance of 50cm from the radar.

Speakers in a smartphone. A smartphone typically has two speakers, *i.e.*, an earpiece at the top and a main speaker at the bottom [29]. Fig. 3(a) plots the smartphone vibrations



(a) Speaker structure (b) Displacement-frequency

Fig. 2. Illustration of speaker structure.



(a) The type of speaker (b) Vibration amplitude (c) Frequency response

Fig. 3. The characteristics of sound-induced smartphone vibrations.

recovered by mmWave radar (§IV-B) where speakers are playing a single tone at 800Hz. We see that mmWave radar can capture smartphone vibrations generated by speakers within a frequency error of 3Hz. The vibration effect is the strongest when both speakers play the same sound simultaneously. The earpiece produces a smaller vibration effect due to its smaller size and lower power output. By default, we set the two speakers to work simultaneously for better vibration effect.

Vibration amplitude. The vibration generated by speakers is extremely weak, making it difficult to precisely measure its vibration amplitude. To compare the vibration amplitude induced by a smartphone speaker and a vibra-motor, we control a smartphone (Samsung 9+) to play a single tone of 800Hz using a speaker and vibrate at 150Hz by triggering a vibra-motor. As shown in Fig. 3(b), we see that the peak-to-peak vibration amplitude of the speaker is approximately $3\mu m$, much weaker than that triggered by the vibra-motor.

Frequency response. Fig. 3(c) plots the normalized sound-induced vibration amplitudes measured by the mmWave radar with different sound frequencies. We observe that the frequency response is not flat, with the strongest vibrations detected near the resonant frequency between 500Hz and 800Hz. This is because the vibration displacement of the speaker varies nonlinearly with frequency, causing the sound-induced vibrations to be heavily frequency selective. Note that high-frequency vibrations ($>$2000Hz) are much weaker because of attenuation and absorption [30–32].

Based on these observations, we see that although the extremely weak sound-induced vibrations and non-flat frequency responses pose significant challenges, mmWave radars can potentially capture such weak vibrations generated by smartphone speakers. Compared to the smartphone vibra-motors, smartphone speakers operate in a higher and wider frequency band, thus offering the possibility of increasing the data rate between mmWave radars and smartphones.

III. Transmitter Design

Fig. 4 shows the workflow of *Talk2Radar*. In this section, we focus on the transmitter design.

In *Talk2Radar*, a smartphone transmits messages using modulated sound waves. Effective communication relies on the careful modulation of sound waves. The following two characteristics make the design of *Talk2Radar* unique and challenging: i) *Talk2Radar* is a cross-modal communication from speakers to mmWave radars through sound-induced vibrations; ii) The sound-induced vibrations are extremely weak and suffer from severe frequency-selective fading.

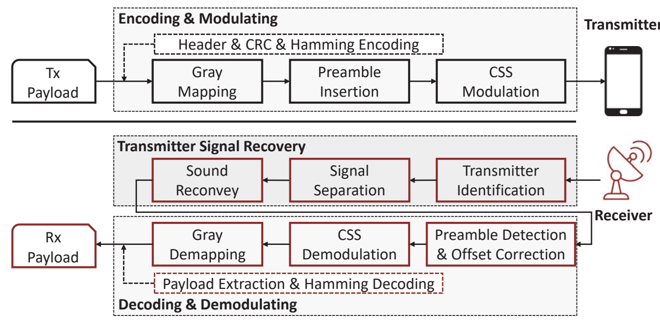


Fig. 4. Workflow of *Talk2Radar*.

CSS modulation. To overcome these challenges, we adopt the Chirp Spread Spectrum (CSS) modulation, which can effectively concentrate energy to improve the target signal strength [24, 33–35] and mitigate the impact of frequency selective fading. In particular, CSS uses a wideband linear frequency modulated signal (*i.e.*, chirp) to encode data. In CSS modulation, a *base chirp* sweeps from f_{min} to f_{max} within chirp duration T_c and can be denoted as $C(t) = \cos[2\pi(f_{min} + \frac{\mu}{2}t)t]$, where μ denotes the chirp sweep rate and $f_{max} = f_{min} + \mu T_c$. CSS changes the initial frequency of the base chirp to modulate different symbols. Hence, the signal of a symbol $x(t, f_i')$ is:

$$x(t, f_i') = \cos[2\pi(f_{min} + f_i' + \frac{\mu}{2}t)t] \tag{4}$$

where f_i' is the frequency shift of the symbol $x(t, f_i')$ and its start frequency f_i is equal to $f_{min} + f_i'$.

Transmission capacity. The spreading factor (SF) is an important parameter to control the chirp shape and data rate. We have 2^{SF} initial frequencies for encoding SF bits data. The modulation bit rate R_b can be expressed as:

$$R_b = SF \times \frac{BW}{2^{SF}} \tag{5}$$

where BW is the chirp bandwidth. Based on the frequency response in Fig. 3(c), we empirically set the chirp frequency varying from 300Hz to 1900Hz (BW=1600Hz) to produce better sound-induced vibrations. Table II illustrates the corresponding chirp duration (symbol duration) and bit rates at different SFs. The maximum bit rate can be up to $600bps$.

Note that as SF increases, the chirp duration becomes larger, and the SNR of chirp signal gets better, implying a longer communication range. On the other hand, the bit rate decreases with the increase of SF. In practice, we can adjust the value of SF to balance the communication range and data transmission rate according to application requirements.

TABLE II
CSS MODULATION PROPERTIES (BW = 1600HZ)

SF	3	4	5	6	7
T_c(ms)	5	10	20	40	80
R_b(bps)	600	400	250	150	87.5

Preamble. To synchronize the transmitter and the receiver, we add a preamble to each packet. In *Talk2Radar*, a basic preamble is a base chirp followed by a down chirp (sweeping from f_{max} to f_{min}) with the same SF as the subsequent payload. This up-and-down chirp has a better self-correlation [36, 37] and can indicate the SF value of the corresponding transmitter. To increase the detection probability of the preamble, the length of the preamble is configurable, with a default length of three up-and-down chirps.

Coding scheme and error correction. We can also leverage several coding techniques to ensure reliable packet transmission. We use Gray code for mapping payload bits to symbols to reduce bit errors caused by adjacent misidentification. In addition, forward error correction codes can be added to correct bit errors.

IV. RECEIVER DESIGN

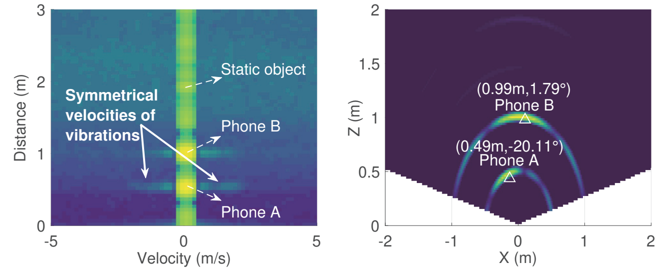
In this section, we present the technical detail of *Talk2Radar* receiver. After capturing the reflected signals from the surroundings, *Talk2Radar* receiver (mmWave radar) first identifies the target transmitters, and then separates out the reflected signals from each transmitter (§IV-A). Next, for each transmitter, *Talk2Radar* extracts the sound-induced vibrations to recover the transmitting sound waves (§IV-B). After that, the sound waves are demodulated for decoding (§IV-C, §IV-D).

In the following, we conduct a feasibility study where a mmWave radar is used to capture messages from two smartphones, *i.e.*, Phone A at $(0.5m, -20°)$ and Phone B at $(1m, 0°)$. With the empirical results, we go through each key component at the receiver side and describe our design.

A. Transmitter Identification and Signal Separation

Talk2Radar first performs the Range-FFT to resolve objects in range and then conducts the Doppler-FFT to separate objects in velocity [38]. Fig. 5(a) shows the obtained Range-Doppler spectrogram, where the bright spots indicate higher probabilities of the existence of objects. Unlike other objects, we observe that the transmitters produce both positive and negative velocities in the range of $[-2m/s,\ 2m/s]$ due to reciprocating vibrations. Therefore, we search all range bins where candidate objects (bright spots) are detected, and then locate the transmitters' range bins by finding the symmetrical velocity pattern within the target velocity range [5].

After obtaining the transmitters' range bins, we perform Angle-FFT on the received signals from multiple antennas to get the location of each transmitter as shown in Fig. 5(b). After that, *Talk2Radar* utilizes the spatial resolution of mmWave radar to separate transmitters at different ranges and angles, and further acquires the reflected signals $S_{r,\theta}(t)$ of each



(a) Range-Doppler spectrogram (b) Transmitter localization

Fig. 5. Illustration of transmitter identification and localization.

transmitter located at the range bin r and angle bin θ, which can be represented as:

$$
\begin{aligned}
& s^{[n]}(t) = \alpha^{[n]} \exp\left[j4\pi(f_c + Kt)R^{[n]}(t)/c\right], n \in N \\
\xrightarrow[\text{in range bin } r]{\text{Range-FFT}} \; & S_r^{[n]}(t) = \alpha^{[n]} \exp\left[j4\pi f_c R_r^{[n]}(t)/c\right], n \in N \\
\xrightarrow[\text{in angle bin } \theta]{\text{Angle-FFT}} \; & S_{r,\theta}(t) = \alpha \exp\left[j4\pi R_{r,\theta}(t)/\lambda\right]
\end{aligned}
\tag{6}
$$

where α is the path loss. f_c, λ and K are the starting frequency, wavelength, and the slope of the FMCW signal, respectively. N is the number of receiver antennas and $R_{r,\theta}(t)$ is the distance between the target smartphone and mmWave radar.

B. Sound Recovery

This step is to extract phase measurements of each transmitter to recover the modulated sound waves.

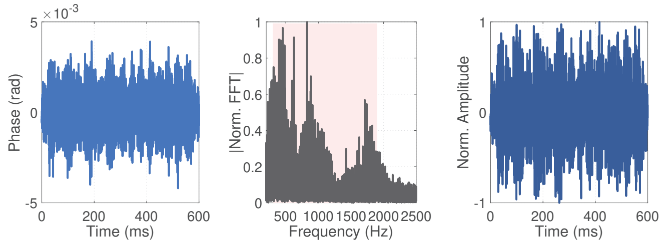
Extracting phase measurement. Besides the target transmitter, *i.e.*, the vibrating smartphone, other static objects also reflect signals that may fall in the same range bin and angle bin as the target. Thus, the received signal $\vec{S}_{r,\theta}$ from the target smartphone location (r, θ) is a superposition of the target smartphone signal $\vec{S}_{\text{phone}}$ and static interference $\vec{S}_0$.

$$
\vec{S}_{r,\theta} = \underbrace{\alpha \exp[j4\pi f_c R_{r,\theta}(t)/c]}_{\vec{S}_{\text{phone}}} + \underbrace{\sum_i \alpha_0^{[m]} \exp[j4\pi f_c R_{r,\theta}^{[m]}/c]}_{\vec{S}_0} \tag{7}
$$

where $\alpha_0^{[m]}$ and $R_{r,\theta}^{[m]}$ represent the signal strength and propagation distance of the m-th background reflection, respectively. To extract the phase measurements from the target component $\vec{S}_{\text{phone}}$, we first remove the static component $\vec{S}_0$ through circle fitting [39]. Phase drifts and noises caused by hardware imperfections are mitigated by detrending the phase fluctuations within each frame [30, 32]. Fig. 6(a) shows the extracted phase measurements of Phone A in Fig. 5(b).

Recovering sound waves. As the transmitter plays the modulated sounds, its phase measurements exhibit the sound-induced vibrations in the pre-configured frequency band. As shown in Fig. 6(b), the phase measurements of Phone A show a higher frequency response from 300Hz to 1900Hz. Due to the hardware characteristics of the speakers, the frequency response is non-flat with the highest response around the resonance frequency of 800Hz.

Based on this observation, we apply a band-pass filter (BPF) on the phase measurements to extract the target sound-induced vibrations and recover the corresponding sound waves, while filtering out noise. The lower and upper stopping frequencies



(a) Raw phase (b) FFT on phase (c) Recovered sound

Fig. 6. Illustration of sound recovery.

of the BPF filter are determined by the modulation frequency band on the transmitter side. Fig. 6(c) shows the recovered sound waves of Phone A. Unfortunately, the recovered sounds have poor intelligibility because of weak vibrations and non-flat frequency response. As such, we apply the short-time Fourier transform (STFT) on the recovered sound waves to display the time-localized frequency information. After that, we can see in the upper panel of Fig. 7(a) where the recovered data packet starts with a preamble including three up-and-down chirps followed by multiple audio chirps.

C. Demodulation and Decoding

Next, we demodulate and decode the recovered sounds.

Preamble detection. We first locate the preamble embedded at the start of each packet. As the preamble is pre-defined, we calculate the correlation between the recovered sound waves and a template, and then find the correlation peak for preamble detection. Fig. 7(a) shows the edges of preambles detected from the recovered sound waves of Phone A in Fig. 6(c).

Symbol demodulation. After locating the preamble, we can determine the starting point of the payload. Then we demodulate each symbol by measuring its initial frequency. Based on Eq. 3 and Eq. 4, the extracted phase measurements of the i-th audio chirp symbol can be expressed as:

$$\phi(t, f_i') = 4\pi[R_0 + \eta(f_{min} + f_i' + \mu t) \cdot x(t, f_i')]/\lambda \quad (8)$$

After applying a band pass filter and eliminating constant values, the recovered audio chirp for the i-th symbol is:

$$y(t, f_i') = \eta(f) \cdot x(t, f_i') \quad (9)$$

where $\eta(f)$ denotes the vibration amplitude of the smartphone speaker at frequency f and $f = f_{min} + f_i' + \mu t$.

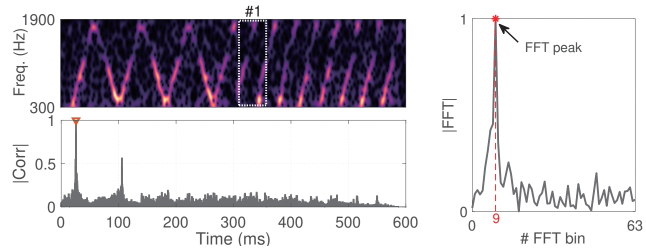
We demodulate the received audio chirps in two steps. First, we de-spread the received symbol $y(t, f_i')$ by multiplying it with the base chirp $C(t)$ based on $\cos A \cos B = \frac{1}{2}(\cos(A+B)+\cos(A-B))$. By filtering the high-frequency part $\cos(A+B)$, we obtain the de-chirped signal of the i-th symbol:

$$y(t, f_i')_\downarrow = \frac{1}{2}\eta(f) \cdot \cos(2\pi f_i' t) \quad (10)$$

We observe that the de-chirped signal $y(t, f_i')_\downarrow$ is a single tone with the frequency of f_i'. Second, we apply FFT on the de-chirped signal, *i.e.*, $Z(f) = |\text{FFT}(y(t, f_i')_\downarrow)|$ to find f_i':

$$f_i' = \arg\max_{f_m} \|Z(f_m)\| \quad (11)$$

where $f_m = m \cdot \frac{BW}{2^{SF}}$ denotes the frequency shift of the m-th FFT bin ($m = 0, 1, 2, ..., 2^{SF} - 1$). Hence, we can determine f_i' by finding the frequency bin with the maximum energy to demodulate the recovered audio chirp.



(a) Preamble detection (b) Demodulation result

Fig. 7. CSS demodulation.

Fig. 7(b) shows the demodulation results of the recovered symbol #1 in Fig. 7(a). The FFT peak is located at FFT bin #9, indicating its initial frequency and data. After gray demapping, we can finally obtain the messages from the transmitter. Note that the demodulation concentrates the power of each chirp symbol to a specific FFT bin, so that the audio chirp recovered from tiny vibrations can still be detected and decoded.

D. Demodulation Enhancement

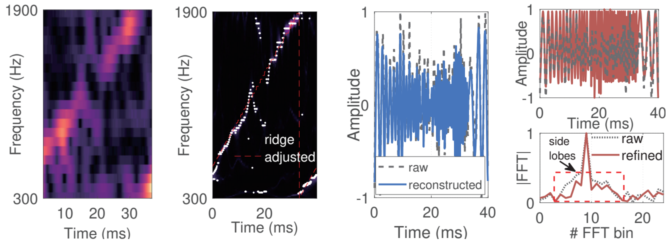
Adopting CSS modulation in cross-modal communication between smartphone speakers and mmWave radars entails some unique challenges.

Offset correction. In practice, the frequency of the recovered packet is affected by synchronization issues, hardware imperfections and differences in cross-modal communication, resulting in some unwanted frequency offsets. To address this issue, our basic idea is to estimate the frequency offset from the known preamble. Since a preamble is composed of several base up-and-down chirps, the FFT peak of each chirp is located at the FFT bin #0 after demodulation. Hence, we can determine the frequency offset Δf by searching possible frequency offsets that can maximize the magnitude of the FFT bin #0 after compensating Δf on the recovered signals [33]. Based on our observation that the frequency offset is within $\pm$5Hz, we use binary search to find it in this range.

Shape-aware audio chirp refinement. Due to the inherent non-flat frequency response of speaker, the recovered audio chirp by mmWave radar is a multi-component signal that comprises frequency-modulated and amplitude-modulated components. Fig. 8(a) shows a recovered audio chirp in the T-F domain. We can see that its amplitude and frequency vary with time, resulting in signals at lower vibration amplitudes being easily overwhelmed by noise. Hence, we propose a shape-aware audio chirp refinement method with four steps:

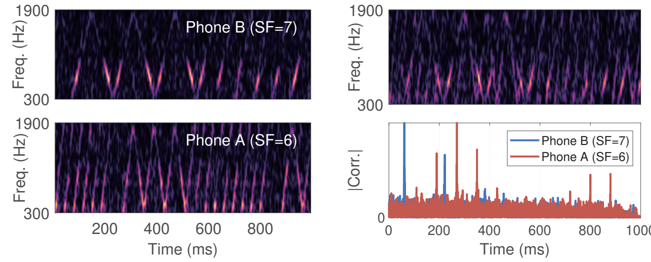
1) Sharpening the audio chirp with WSST: Wavelet synchrosqueezed transform (WSST) is a T-F analysis method, which employs a wavelet transform to decompose the signal into different frequency bands and then squeezes them, improving the T-F localization. Therefore, we utilize WSST based on the analytic Morlet wavelet to transform the raw audio chirp and generate the sharper audio chirp in Fig. 8(b). We can see that WSST sharpens the T-F representation of raw audio chirp along the frequency direction, making it easier to isolate the target audio chirp from noise.

2) Identifying the audio chirp based on chirp shape: This step aims to locate the region in the T-F plane that corresponds



(a) Audio chirp (b) WSST (c) Construction (d) Refinement

Fig. 8. Denoising and refinement.



(a) Separated signals (b) Packet collision

Fig. 9. Concurrent transmission based on location diversity.

to the target audio chirp. Specifically, we first identify the ridge (local maxima) in the magnitudes of the WSST spectrogram in Fig. 8(b). However, the ridge may jump as the region of the highest energy in the T-F plane changes between the target signal and noise. To address this issue, our basic idea is to adjust the ridge in the T-F plane until it resembles a chirp with maximum energy. After adjustment, we can accurately locate the target chirp signal amid the noise and other components.

3) Reconstructing the audio chirp: WSST is invertible as time information is preserved. Hence, we leverage inverse WSST (IWSST) to extract target audio chirp from localized regions in the T-F plane. Fig. 8(c) shows the reconstructed signals, exhibiting lower noise levels compared to the raw signal, especially at lower amplitudes.

4) Compensating for amplitude fluctuations: In Fig. 8(c), we observe that the amplitude of the reconstructed chirp signal still fluctuates with time due to the non-flat frequency response of the speaker, which may cause some unwanted side-lobes around the FFT peak after demodulation. To further mitigate the impact of amplitude fluctuations, we first obtain the envelope of the target signal, and then equalize and compensate amplitude fluctuations. Fig. 8(d) shows the refined chirp signal and its demodulation result. After compensation, we observe that the amplitude of the refined signal is more flattened compared to the raw signal. The side-lobes are further dampened. As a result, these four steps can effectively mitigate the impact of noise and amplitude fluctuations, generating more stable and accurate demodulation outcomes.

V. Improving Robustness and Practicality

A. Enabling Concurrent Communication

At the receiver side, *Talk2Radar* can support concurrent transmissions of multiple smartphones by exploiting spatial diversity and orthogonal parameter configuration.

Spatial diversity. The high spatial resolution of mmWave radars allows *Talk2Radar* to separate multiple transmitters at different locations (§IV-A). Fig. 9(a) plots the recovered sounds from two transmitters in Fig. 5(b). The audio chirps of Phone B with a larger SF present a lower rate of frequency change over time. Note that the concurrent communication based on location diversity is limited by the sensing resolution of radar. When the spacing between two transmitters is less than the resolution limits, it is challenging to separate them.

Transmission parameter diversity. *Talk2Radar* adopts CSS modulation, which allows multiple signals with different SFs to coexist in the same frequency band without significant interference. Hence, the concurrent communication capability of *Talk2Radar* can be improved by assigning different SFs to different transmitters. Fig. 9(b) presents the recovered sound waves from two smartphones whose spacing is below the resolution limit. We can see that the recovered signals consist of mixed packets and cannot be separated using location diversity. To decode these two conflicting packets, we first locate their respective preambles using the corresponding preamble templates. As shown in the lower panel of Fig. 9(b), we observe that the preamble detection result for each SF has a correlation peak, implying that there is a preamble with its corresponding SF. By doing so, two preambles belonging to different transmitters are detected. Then, we decode the payload for each transmitter using different demodulation windows associating with the SF (§IV-C and §IV-D).

Talk2Radar allows multiple transmitters at different locations or different SFs to transmit signals simultaneously, fine-grained range and angular resolution are first used to separate objects, and then the difference in SF is further employed for collision decoding. We believe these two dimensions are sufficient for our target scenarios, as the distance between users is usually larger than the limit of spatial resolution.

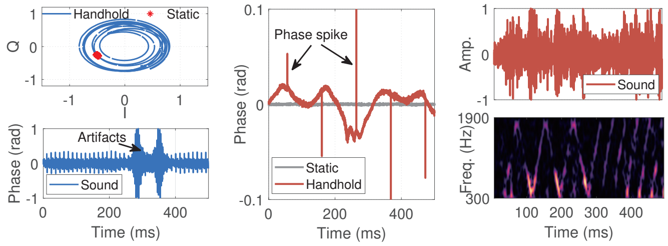
B. Suppressing the Impact of Human Motions

In practice, a smartphone is often carried by a user in hand. Hand motion or other micro-motion from the human body will interfere with the recovered target sound waves. To investigate its impact, we ask a user to hold a smartphone and move back and forth during communication. Fig. 10(a) shows the extracted reflected signal from the smartphone on the I/Q plane and the recovered sound wave. Compared to a static object fixed on a tripod, the signal trajectory of the smartphone in motion exhibits helical curves as both the amplitude and phase of the reflected signal change with the hand motion, resulting in evident artifacts on the recovered sound wave.

To mitigate this impact, our key idea is to split the received signal into multiple short segments as the speed of human motion v can be assumed to be constant within a short period (*e.g.*, 102.4ms in our setting). Hence, we first extract the phase measurements of every short segment by removing the static component (§IV-B). The extracted phase measurement is:

$$\phi(t, f_t') = \frac{4\pi}{\lambda}[R_0 + \underbrace{\eta(f)\cdot x(t, f_i')}_{\text{sound-induced vibration}} + \underbrace{vt}_{\text{motion}}] \tag{12}$$

We notice that the impact of motion can be removed by taking the first-order derivative of phase measurement $\phi'(t)$,



(a) Raw signal (b) Phase diff. (c) After suppression

Fig. 10. Motion suppression.

as vt associated with motion can be transformed into a DC component v. In practice, we calculate the phase difference of two consecutive measurements to approximate the first-order derivative of the phase. Fig. 10(b) plots the obtained phase difference and we observe that segment-based motion suppression results in a phase spike between two successive segments. To address these issues, we first locate the phase spikes and replace them with the linear interpolation of neighbouring. DC component and lower frequency fluctuation associated with human motion are then removed through a highpass filter with a cutoff frequency of 200Hz. Finally, we reconstruct the phase measurement from the refined phase difference and recover the target sound waves. Compared with the raw sound waves in Fig. 10(a), the proposed motion suppression method can effectively mitigate the impact of human artifacts and produce a clearer sound wave in Fig. 10(c).

VI. Evaluation

A. Implementation and Methodology

Hardware and software. We implement *Talk2Radar* on commodity devices. On the transmitter side, we test 6 smartphones from various vendors. The receiver in our prototype is a TI AWR1642 mmWave radar board, operating in the 77GHz$\sim$81GHz frequency band. Data processing algorithms are implemented using Matlab and are executed on a computer with an Intel Xeon E5-2620 v4 2.10GHz CPU.

Experiment setup. Fig. 11 shows our experiment setting. By default, the smartphone is placed on a tripod playing the modulated audio chirps at 80% of the maximum volume. The bandwidth of transmitting sound waves is 1600Hz (300$\sim$1900Hz) and SF is 6. On the receiver side, we let one Tx antenna of mmWave radar send the FMCW signals with the bandwidth of 3GHz and four Rx antennas receive the reflected signals. In our setting, the range resolution is about $5cm$, and the angular resolution is about $28.65°$. We set the chirp duration to $100\mu s$, resulting in a phase sampling rate of 10KHz. The communication range r, communication angle θ, and smartphone orientation β are also defined in Fig. 11.

Metrics. We evaluate the performance in terms of bit error rate (BER). The transmission data from senders are recorded as ground truth. A BER of <10% is considered acceptable as error correction codes can recover those errors [40].

B. Overall Performance

Transmission data rate. We first evaluate the data rate of *Talk2Radar*. A smartphone (Samsung S9+) is placed in

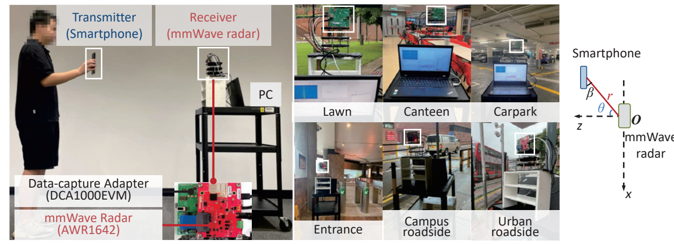
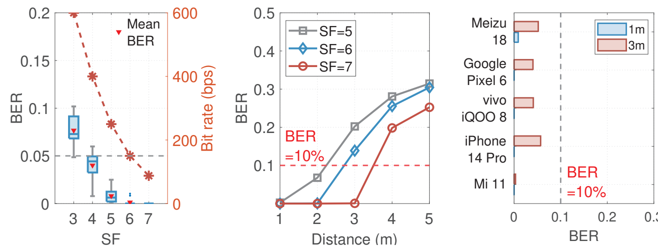


Fig. 11. Experiment setup of *Talk2Radar*.



(a) Data rate (b) Comm. range (c) Generalization

Fig. 12. Overall performance of *Talk2Radar*.

front of the radar at $1m$ and transmits the modulated sounds at different SFs. Fig. 12(a) plots the BER and data rate. As the SF increases, we can see that BER and bit rate gradually decrease. This is because a larger SF results in a longer chirp duration and higher SNR (Table II). The bit rate can reach $600bps$ with a mean BER of 7.67% at an SF of 3. When the SF is greater than 3, the bit rate is $\leq 400bps$ and the mean BER becomes less than 5%. Compared with the state-of-the-art vibra-motor based method ($12.12bps$) [16], *Talk2Radar* achieves $33\times$ improvement in data rate. Moreover, it can adjust the data rate according to different requirements.

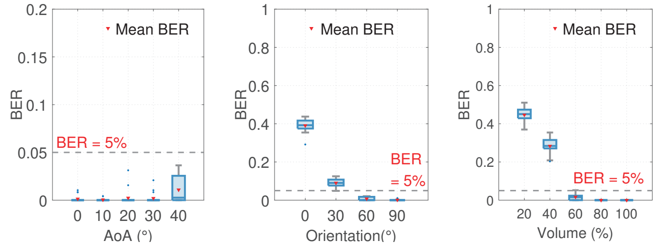
Communication range. We place the smartphone at different positions to evaluate the communication range. As expected, an increase in communication range leads to a higher BER as shown in Fig. 12(b). In addition, larger SFs achieve longer communication ranges. The communication range can be $3m$ with a BER of $<1\%$ when SF is 7. We believe such a communication capability can support our target scenarios, as *Talk2Radar* aims to allow users to send commands or messages to mmWave radars within their sights. The enhancement of communication range and data rate is worthy of further exploration.

Evaluation with different smartphones. We repeat the experiments (SF=7) using different types of smartphones as shown in Fig. 12(c). We observed that the BER of *Talk2Radar* is less than 1% at $1m$ communication distance and 5.7% at $3m$ across different devices, which is below the BER threshold of 10%. The results indicate that *Talk2Radar* can be readily deployed in different smartphones.

C. Controlled Experiment

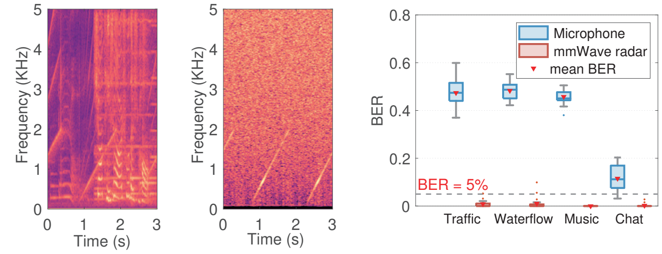
We place a smartphone (Samsung S9+) directly in front of the radar at $1m$ and perform controlled experiments to further evaluate our system under various conditions. By default, the smartphone is fixed on a tripod and oriented towards the radar.

Communication angle. In practical scenarios, the communication angle θ between the smartphone and the mmWave



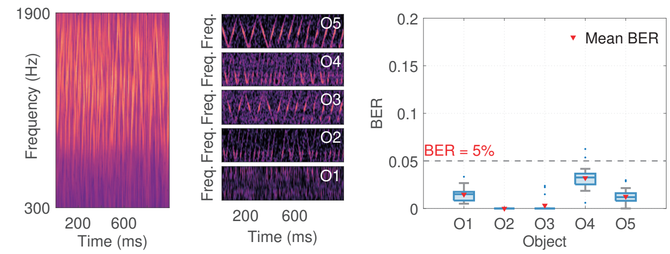
(a) Comm. angle (b) Phone orientation (c) Speaker volume

Fig. 13. Result of controlled experiments in the lab environment.



(a) Microphone (b) Radar (c) BER under ambient noises.

Fig. 14. Noisy environment.



(a) Microphone (b) Radar (c) BER under multi-speaker comm.

Fig. 15. Multi-speaker communication.

radar can vary. To evaluate its impact, we conduct experiments by changing the communication angle within a range of $0°$ to $40°$. Our experiment results from Fig. 13(a) demonstrate that BER remains consistently below 5% within the radar's field of view ($\pm 40°$). However, as the smartphone approaches the boundary of this field, the BER gradually increases. Hence, within the sensing area of the mmWave radar, we can achieve better communication performance.

Smartphone orientation. This experiment changes the smartphone orientation related to the radar (β in Fig. 11) and evaluates its impact. When β is 90°, the phone screen directly faces to the mmWave radar. In Fig. 13(b), we see that the BER decreases as the orientation angle increases. This is because when the smartphone screen faces the radar without misalignment, the radar can capture stronger vibrations with larger RCS and vibration displacement, resulting in a lower BER. Hence, we suggest users maintain the smartphone orientation within a range of 30° to 90° for better performance.

Speaker volume. A higher speaker volume increases diaphragm displacement and enhances smartphone vibration effects. Hence, we repeat the experiments at different speaker volumes. In Fig. 13(c), we see the speaker volume significantly influences the BER. The BER is below 5% when the volume is greater than 40% of the maximum volume.

Since a smartphone absorbs sound-induced vibrations at higher frequency ($>$ 2KHz) [30, 32], *Talk2Radar* utilizes audible sounds with better vibration effects for communication. As the speaker volume increases, communication performance improves. In practice, we empirically set the speaker volume within 60-80% of the phone's maximum volume, which balances the performance and user experience. In this case, the transmitted audible sounds are usually muffled by ambient noises. In the future, we may explore a covert communication, such as randomly shifting the phase of each frequency to transform the emitted audio chirps into imperceptible white noises [41].

D. Noisy Environment

To compare mmWave radars with microphones, we put them together to capture the same sound from a smartphone speaker at a distance of $1m$. The smartphone plays an audio chirp from 200Hz to 2000Hz, and two loudspeakers play music noise at a distance of $50cm$. As shown in Fig. 14(a), the received signal by the microphone is heavily contaminated by noise and harmonics caused by non-linear components, making it challenging to extract the target signal. Conversely, in Fig 14(b), the mmWave radar can accurately separate the target signal from noise, as the sound is detected directly from the source of vibrations.

We further evaluate the communication performance under various noise conditions. When a smartphone is playing modulated audio chirps with SF of 6 at distance of $1m$ (70 dB-SPL), two loudspeakers are placed at $50cm$ playing four ambient noises including traffic, waterflow, music and chatting with corresponding typical SPL levels (60-80dB). We observe the background noise degrades the performance of the microphone based benchmark and its mean BER exceeds 10% even in typical chatting conditions. In contrast, the BER of recovered signal by the mmWave radar ($<$2%) is less affected.

E. Multiple Transmitters

This experiment evaluates the concurrent communication performance of *Talk2Radar*. We place five smartphones in an area of $2m \times 1m$ in front of the mmWave radar. For performance comparison, a microphone is placed next to the mmWave radar. With five smartphones playing modulated sounds simultaneously, we observe that the microphone receives the mixed signals (Fig. 15(a)), while the mmWave radar can separate them and extract the signals for each transmitter (Fig. 15(b)). After decoding, *Talk2Radar* achieves a mean BER of less than 5% for all five transmitters.

TABLE III
TRANSMISSION PARAMETER DIVERSITY.

Range diff. (cm)	0	5	10	15	20
BER of Phone 1 /% (SF=6)	4.67	3.97	0.83	0.25	0.13
BER of Phone 2 /% (SF=7)	0.5	0.32	0.11	0	0

Besides the location diversity, *Talk2Radar* can resolve multiple transmitters using transmission parameter diversity. To evaluate it, we place two smartphones separated by a varied distance. The two phones are approximately $1m$ away from the

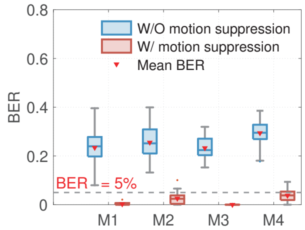


Fig. 16. Motion suppression

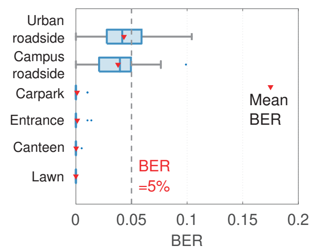


Fig. 17. Practical scenarios

radar and assigned with different SFs. As shown in Table III, we can see that *Talk2Radar* can still separate two objects (BER $<5\%$) even if they are very close in range ($\leq 5cm$ of sensing resolution). Hence, the transmission parameter can be used as an additional dimension to enhance the concurrency.

Although *Talk2Radar* designs a many-to-one messaging, it is a one-way uplink communication from smartphone speakers to a mmWave radar. Fortunately, out-of-band downlink channels (*e.g.*, light, sound, and visual cues) can be used to acknowledge smartphone users (*e.g.*, blinking light).

F. Motion Artifacts

To evaluate *Talk2Radar* under human movement, we invite four volunteers (two females and two males) to carry the smartphone and conduct experiments in four motion scenarios: (M1) sit: sitting while holding the smartphone; (M2) sit + arm wiggles: sitting while holding the smartphone with arm wiggles; (M3) stand: standing while holding the smartphone; and (M4) stand + body wiggles: standing while holding the smartphone with body wiggles. In Fig. 16, we can see that our proposed motion suppression method can effectively mitigate the impact of human motion with a BER of $<5\%$.

G. Robustness in Practical Scenarios

To evaluate the robustness of *Talk2Radar*, we conduct experiments in six practical scenarios as shown in Fig. 11. In each scenario, the distance between the smartphone and the mmWave radar is maintained between $1m$ to $1.5m$, and 640 symbols with SF of 6 are collected. In Fig. 17, we can see that *Talk2Radar* has good performance with BER below 5% across all tested scenarios with strong multipath effects. That is because *Talk2Radar* can filter out interference signals and accurately identify the target transmitters.

VII. Related Work

Micro-vibration detection with mmWave radars. Benefiting from the high-resolution, mmWave radars are used for micro-vibration sensing. mmVib [5] uses mmWave radars to detect micrometer-level industrial vibrations. GWaltz [6] can measure sub-mm-level 2D orbits of rotating machinery. Multi-Vib [7] achieves multi-point vibration monitoring by placing physical markers. Unlike these works that aim to sense vibrations, *Talk2Radar* modulates and demodulates vibrations for the communication purpose.

Sound sensing with mmWave radars. mmWave radar has also emerged as a potential radio microphone as it can separate and recover multiple sound sources [31]. WaveEar [42] captures the near-throat skin vibrations to recover the human voice. mmMIC [43] extracts the multi-modal features from lip motion and vocal-cords vibrations to improve speech recognition performance. AmbiEar [44] exploits the surrounding objects as ears and implements mmWave-based voice recognition in non-line-of-sight scenarios. RadioMic [31] constructs various types of sounds from both active sources (*e.g.*, human throat) and passive sources (*e.g.*, paper bag). Moreover, mmWave radars can eavesdrop remotely the sounds from smartphone loudspeakers and headsets [30, 32, 45–48]. *Talk2Radar* builds on prior sensing techniques and empowers mmWave radars with communication capability at much higher data rates.

Vibration-based communication. Vibration is a common modality of data communication. Ripple [49] utilizes vibra-motors to modulate messages, which are then decoded using contact based IMU sensors. Bleep [50] leverages motor vibrations to enable UAVs to communicate with each other. MotorBeat [25] connects small appliances to a smart speaker by capturing vibrations from direct current motors. Recent works [51, 52] use mmWave radars to sense water surface vibrations caused by underwater sonars enabling cross-media communication. mmRipple [16] modulates vibrations of smartphone vibra-motors to send data to mmWave radars. The narrow bandwidth of vibra-motors, however, limits its data rate. *Talk2Radar* develops a new speaker-to-mmWave radar communication channel with much higher data rates.

VIII. Conclusion

This paper presents *Talk2Radar*, which builds a speaker-to-mmWave radar communication channel with sound-induced vibrations. Without any hardware modification, *Talk2Radar* empowers commodity mmWave radars with communication capability. Talk2Radar substantially improves data rates compared with the state-of-the-art vibration based methods. We overcome several practical challenges involved in encoding data with extremely weak and frequency-selective sound-induced vibrations, supporting multi-speaker concurrent communication, and suppressing hand and body motion.

Acknowledgment

This work is supported in part by the Hong Kong GRF under grant 15206123, National Science Fund for Distinguished Young Scholars of China under grant No. 62125203, The Key Program of the National Natural Science Foundation of China under grant No. 61932013, National Natural Science Foundation of China under grant 62372400, Research Institute of Cyberspace Governance in Zhejiang University, "Pioneer" and "Leading Goose" R&D Program of Zhejiang under grant No. 2023C01033, Natural Science Research Start-up Foundation of Recruiting Talents of Nanjing University of Posts and Telecommunications (Grant No. NY223171). Yuanqing Zheng and Fu Xiao are the corresponding authors.

REFERENCES

[1] X. Shuai, Y. Shen, Y. Tang, S. Shi, L. Ji, and G. Xing, "millieye: A lightweight mmwave radar and camera fusion system for robust object detection," in *ACM/IEEE IoTDI*, 2021.

[2] H. Ding, Z. Chen, C. Zhao, F. Wang, G. Wang, W. Xi, and J. Zhao, "MI-Mesh: 3d human mesh construction by fusing image and millimeter wave," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2023.

[3] X. Yang, J. Liu, Y. Chen, X. Guo, and Y. Xie, "Mu-id: Multi-user identification through gaits using millimeter wave radios," in *IEEE INFOCOM*, 2020.

[4] P. S. Santhalingam, A. A. Hosain, D. Zhang, P. Pathak, H. Rangwala, and R. Kushalnagar, "mmASL: Environment-independent asl gesture recognition using 60 ghz millimeter-wave signals," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2020.

[5] J. Guo, Y. He, C. Jiang, M. Jin, S. Li, J. Zhang, R. Xi, and Y. Liu, "Measuring micrometer-level vibrations with mmwave radar," *IEEE Transactions on Mobile Computing*, 2023.

[6] J. Guo, M. Jin, Y. He, W. Wang, and Y. Liu, "Dancing waltz with ghosts: Measuring sub-mm-level 2d rotor orbit with a single mmwave radar," in *ACM IPSN*, 2021.

[7] Y. Yang, H. Xu, Q. Chen, J. Cao, and Y. Wang, "Multi-Vib: Precise multi-point vibration monitoring using mmwave radar," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2023.

[8] Y. He, J. Liu, M. Li, G. Yu, J. Han, and K. Ren, "SenCom: Integrated sensing and communication with practical wifi," in *ACM MobiCom*, 2023.

[9] Q. Zhang, H. Sun, X. Gao, X. Wang, and Z. Feng, "Time-division isac enabled connected automated vehicles cooperation algorithm design and performance evaluation," *IEEE Journal on Selected Areas in Communications*, 2022.

[10] X. Chen, Z. Feng, Z. Wei, P. Zhang, and X. Yuan, "Code-division ofdm joint communication and sensing system for 6g machine-type communication," *IEEE Internet of Things Journal*, 2021.

[11] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, "Toward dual-functional radar-communication systems: Optimal waveform design," *IEEE Transactions on Signal Processing*, 2018.

[12] Q. Yang, H. Wu, Q. Huang, J. Zhang, H. Chen, W. Li, X. Tao, and Q. Zhang, "Side-lobe can know more: Towards simultaneous communication and sensing for mmwave," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2023.

[13] E. Soltanaghaei, A. Prabhakara, A. Balanuta, M. Anderson, J. M. Rabaey, S. Kumar, and A. Rowe, "Millimetro: mmwave retro-reflective tags for accurate, long range localization," in *ACM MobiCom*, 2021.

[14] M. H. Mazaheri, A. Chen, and O. Abari, "mmTag: A millimeter wave backscatter network," in *ACM SIGCOMM*, 2021.

[15] J. Nolan, K. Qian, and X. Zhang, "RoS: passive smart surface for roadside-to-vehicle communication," in *ACM SIGCOMM*, 2021.

[16] K. Cui, Q. Yang, Y. Zheng, and J. Han, "mmRipple: Communicating with mmwave radars through smartphone vibration," in *ACM IPSN*, 2023.

[17] T. Zheng, C. Cai, Z. Chen, and J. Luo, "Sound of motion: Real-time wrist tracking with a smart watch-phone pair," in *IEEE INFOCOM*, 2022.

[18] C. Cai, H. Pu, P. Wang, Z. Chen, and J. Luo, "We hear your pace: Passive acoustic localization of multiple walking persons," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2021.

[19] Y. He, W. Wang, L. Mottola, S. Li, Y. Sun, J. Li, H. Jing, T. Wang, and Y. Wang, "Acoustic localization system for precise drone landing," *IEEE Transactions on Mobile Computing*, 2023.

[20] Z. Wang, Y. Wang, M. Tian, and J. Shen, "HearFire: Indoor fire detection via inaudible acoustic sensing," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2023.

[21] X. Xia, Q. Chen, N. Hou, Y. Zheng, and M. Li, "XCopy: Boosting weak links for reliable lora communication," in *ACM MobiCom*, 2023.

[22] N. Hou, X. Xia, and Y. Zheng, "Don't miss weak packets: Boosting lora reception with antenna diversities," *ACM Transactions on Sensor Networks*, 2023.

[23] Q. Yang and Y. Zheng, "AquaHelper: Underwater sos transmission and detection in swimming pools," in *ACM Sensys*, 2023.

[24] H. Lee, T. H. Kim, J. W. Choi, and S. Choi, "Chirp signal-based aerial acoustic communication for smart devices," in *IEEE INFOCOM*, 2015.

[25] W. Wang, J. Li, Y. He, X. Guo, and Y. Liu, "Motorbeat: Acoustic communication for home appliances via variable pulse width modulation," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2022.

[26] A. Izzo, L. Ausiello, C. Clemente, and J. J. Soraghan, "Loudspeaker analysis: A radar based approach," *IEEE Sensors Journal*, 2019.

[27] L. L. Beranek and T. Mellow, *Acoustics: sound fields and transducers*. Academic Press, 2012.

[28] "Speaker and receiver." Goertek. [Online]. Available: https://www.goertek.com/en/content/details20_877.html

[29] C. Cai, R. Zheng, and J. Luo, "Ubiquitous acoustic sensing on commodity iot devices: A survey," *IEEE Communications Surveys & Tutorials*, 2022.

[30] C. Wang, F. Lin, T. Liu, K. Zheng, Z. Wang, Z. Li, M.-C. Huang, W. Xu, and K. Ren, "mmEve: Eavesdropping on smartphone's earpiece via cots mmwave device," in *ACM MobiCom*, 2022.

[31] M. Z. Ozturk, C. Wu, B. Wang, and K. Liu, "RadioMIC: Sound sensing via mmwave signals," *arXiv preprint arXiv:2108.03164*, 2021.

[32] S. Basak and M. Gowda, "mmspy: Spying phone calls using mmwave radars," in *IEEE S&P*, 2022.

[33] X. Xia, N. Hou, Y. Zheng, and T. Gu, "PCube: Scaling lora concurrent transmissions with reception diversities," in *ACM MobiCom*, 2021.

[34] H. Yang, Z. Sun, H. Liu, X. Xia, Y. Zhang, T. Gu, G. Hancke, and W. Xu, "Chirpkey: a chirp-level information-based key generation scheme for lora networks via perturbed compressed sensing," in *IEEE INFOCOM*, 2023.

[35] Z. Sun, H. Yang, K. Liu, Z. Yin, Z. Li, and W. Xu, "Recent advances in lora: A comprehensive survey," *ACM Transactions on Sensor Networks*, 2022.

[36] R. Nandakumar, K. K. Chintalapudi, V. Padmanabhan, and R. Venkatesan, "Dhwani: Secure Peer-to-Peer Acoustic NFC," *ACM SIGCOMM*, 2013.

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

[38] C. Iovescu and S. Rao, "The fundamentals of millimeter wave sensors," *Texas Instruments*, 2017.

[39] J. Ma, Z. Chang, F. Zhang, J. Xiong, B. Jin, and D. Zhang, "Mobi2Sense: enabling wireless sensing under device motions," in *ACM MobiCom*, 2022.

[40] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[41] A. Wang, J. E. Sunshine, and S. Gollakota, "Contactless infant monitoring using white noise," in *ACM MobiCom*, 2019.

[42] C. Xu, Z. Li, H. Zhang, A. S. Rathore, H. Li, C. Song, K. Wang, and W. Xu, "Waveear: Exploring a mmwave-based noise-resistant speech sensing for voice-user interface," in *ACM MobiSys*, 2019.

[43] L. Fan, L. Xie, X. Lu, Y. Li, C. Wang, and S. Lu, "mmMIC: Multi-modal speech recognition based on mmwave radar," in *IEEE INFOCOM*, 2023.

[44] J. Zhang, Y. Zhou, R. Xi, S. Li, J. Guo, and Y. He, "AmbiEar: Mmwave based voice recognition in nlos scenarios," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2022.

[45] P. Hu, Y. Ma, P. S. Santhalingam, P. H. Pathak, and X. Cheng, "Milliear: Millimeter-wave acoustic eavesdropping with unconstrained vocabulary," in *IEEE INFOCOM*, 2022.

[46] C. Wang, F. Lin, T. Liu, Z. Liu, Y. Shen, Z. Ba, L. Lu, W. Xu, and K. Ren, "mmphone: Acoustic eavesdropping on loudspeakers via mmwave-characterized piezoelectric effect," in *IEEE INFOCOM*, 2022.

[47] Y. Feng, K. Zhang, C. Wang, L. Xie, J. Ning, and S. Chen, "mmEavesdropper: Signal augmentation-based directional eavesdropping with mmwave radar," in *IEEE INFOCOM*, 2023.

[48] P. Hu, W. Li, R. Spolaor, and X. Cheng, "mmEcho: A mmwave-based acoustic eavesdropping method," in *IEEE INFOCOM*, 2023.

[49] N. Roy, M. Gowda, and R. R. Choudhury, "Ripple: Communicating through physical vibration," in *USENIX NSDI*, 2015.

[50] A. Bannis, H. Y. Noh, and P. Zhang, "Bleep: motor-enabled audio side-channel for constrained uavs," in *ACM MobiCom*, 2020.

[51] M. R. Romero, R. M. Narayanan, E. H. Lenzing, D. C. Brown, and K. L. Greenert, "Wireless underwater-to-air communications via water surface modulation and radar detection," in *Radar Sensor Technology XXIV*, 2020.

[52] F. Tonolini and F. Adib, "Networking across boundaries: enabling wireless communication through the water-air interface," in *ACM SIGCOMM*, 2018.