Boosting the Sensing Granularity of Acoustic Signals by Exploiting Hardware Non-linearity

Xiangru Chen*[†], Dong Li*[§], Yiran Chen[†], Jie Xiong[§]

[†]Duke University, [§]University of Massachusetts Amherst
[†]{xc186, yiran.chen}@duke.edu, [§]{dli, jxiong}@cs.umass.edu

ABSTRACT

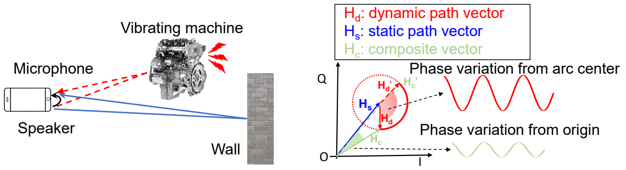
Acoustic sensing is a new sensing modality that senses the contexts of human targets and our surroundings using acoustic signals. It becomes a hot topic in both academia and industry owing to its finer sensing granularity and the wide availability of microphone and speaker on commodity devices. While prior studies focused on addressing well-known challenges such as increasing the limited sensing range and enabling multi-target sensing, we propose a novel scheme to leverage the non-linearity distortion of microphones to further boost the sensing granularity. Specifically, we observe the existence of the non-linear signal generated by the direct path signal and target reflection signal. We mathematically show that the non-linear chirp signal amplifies the phase variations and this property can be utilized to improve the granularity of acoustic sensing. Experiment results show that, by properly leveraging the hardware non-linearity, the amplitude estimation error for sub-millimeter-level vibration can be reduced from 0.137 *mm* to 0.029 *mm*.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**;

KEYWORDS

Sensing granularity, hardware non-linearity, acoustic sensing, higher-order non-linearity utilization


Vibrating machine
Microphone
Speaker
Wall
H_d: dynamic path vector
H_s: static path vector
H_c: composite vector
Q
Phase variation from arc center
H_d'
H_c'
H_s
H_d
Phase variation from origin
H_c
O
I

(a) The sensing scenario.

(b) The I/Q trajectory.

Figure 1: The illustration of phase variations caused by the target movement, e.g., machine vibration.

ACM Reference Format:
Xiangru Chen*[†], Dong Li*[§], Yiran Chen[†], Jie Xiong[§]. 2022. Boosting the Sensing Granularity of Acoustic Signals by Exploiting Hardware Non-linearity. In *The 21st ACM Workshop on Hot Topics in Networks (HotNets '22), November 14–15, 2022, Austin, TX, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3563766.3564091

1 INTRODUCTION

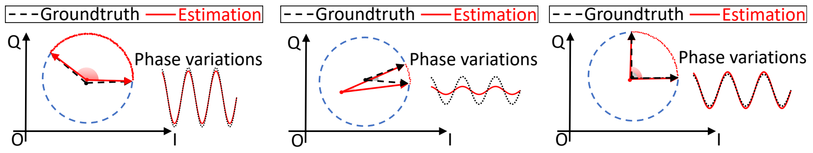
Acoustic sensing, as a new sensing modality, has attracted a tremendous amount of attention in recent years. The research community has devoted a lot of efforts to pushing the boundaries of acoustic sensing such as increasing the sensing range [10, 14, 15], and improving the sensing capability from a single target to multiple targets [9, 22, 32]. In this paper, for the first time, we demonstrate the possibility of boosting the sensing granularity by exploiting the non-linearity on commodity devices such as smartphones.

To achieve fine-grained sensing, the received signal is visualized on the complex plane [10, 20, 26, 29]. As shown in Figure 1a, the static path vector is the resultant of the direct path from the speaker to microphone and the reflections from the static objects (e.g., a wall), while the dynamic path vector is the reflection from the moving target (e.g., a vibration machine). For a subtle movement such as vibration, the dynamic path vector rotates with respect to the static path vector as shown in Figure 1b.

Phase variations are extracted from the I/Q trajectory to derive the fine-grained movement information such as displacement [26]. As shown in Figure 1b, due to the existence of static path, the phase variations directly extracted from the coordinate origin are inaccurate. To address this issue,

*Both authors contributed equally to the paper.

HotNets '22, November 14–15, 2022, Austin, TX, USA
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9899-2/22/11…$15.00
https://doi.org/10.1145/3563766.3564091



(a) Original signal for large movement. **(b) Original signal for small movement.** **(c) Intermodulated signal for small movement.**

Figure 2: Phase variations are accurate for (a) large movements but inaccurate for (b) small movements when estimated by the original signal. (c) They can be amplified and become more accurate for small movements when estimated by the intermodulated signal.

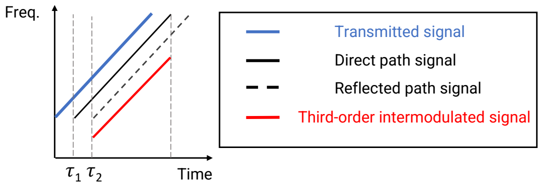


Figure 3: The illustration of the third-order intermodulated signal generated by the direct path signal and target reflection signal for the chirp-based acoustic signal.

prior studies [10, 12, 20] propose to estimate the position of the arc center by circle-fitting the I/Q trajectory and then extract accurate phase variations from the arc center. A larger phase variation leads to a better sensing performance [7]. From Figure 2a and Figure 2b, we can observe that, a longer I/Q trajectory can provide more accurate estimation of the arc center, and accordingly, more accurate phase extraction.

The length of the I/Q trajectory (i.e., the arc) reduces when the target movement distance decreases, resulting in limited sensing granularity for prior studies [10, 12, 20, 26, 29]. Based on the relationship between the amount of phase variation $\Delta\phi$ and the target displacement Δd, i.e., $\Delta\phi = \frac{4\pi f \Delta d}{c}$ [10, 27], the length of the I/Q trajectory is proportional to the frequency of the transmitted signal f given the same amount of displacement. Therefore, one naïve way to improve the sensing granularity is to increase the frequency of the transmitted signal. However, the constrained sampling rate on commodity devices, i.e., 48 kHz, only supports a maximum frequency of 24 kHz according to Nyquist sampling theorem, which is not sufficient for accurately sensing sub-millimeter-level movement. For example, if the target moves at a displacement of 0.1 mm, the induced phase variation is only 5°, which is too small to be accurately measured due to noise.

To break the limit of the constrained sampling rate, we propose a novel solution to boost the sensing granularity by exploiting the non-linearity of commodity devices. The non-linearity generally exists in the components of speakers and microphones on commodity devices such as amplifier and diaphragm [17, 30]. It introduces the intermodulation distortion at the received signal, which creates additional signals at high-order intermodulation frequencies (i.e., the sum and difference of the original frequencies) [28]. Prior studies have shown the feasibility of enabling new applications in security and communication domains by exploiting the second-order intermodulation [17, 30]. For example, they play two high-frequency tones (e.g., 40 kHz and 50 kHz) that humans cannot hear using ultrasound speakers. The two high-frequency sounds can create a low-frequency intermodulated signal (i.e., 10 kHz) due to the hardware non-linearity.

Different from prior studies [17, 30], we obtain an interesting observation when sensing with chirp acoustic signals, i.e., the third-order intermodulation can significantly boost the sensing granularity. More specifically, the received signal at the microphone is the superimposition of the direct path from the speaker and the reflection paths from the surrounding objects. As shown in Figure 3, due to various signal propagation delays, the frequencies of direct path signal and target reflection signal are different. This provides the prerequisite for creating the third-order and higher-order intermodulated signals on hardware. Through both mathematical analysis and experiment verification, we find that the generated intermodulated signals can result in phase variations that are multiple times larger than those at the original signal. As shown in Figure 2b and Figure 2c, for the same amount of movement, the induced phase variations at the intermodulated signal are much larger than those at the original signal. Therefore, the intermodulated signal can obtain more accurate arc center estimate, and thus, more accurate phase variation for sensing. To verify the observation, we perform experiments on ReSpeaker platform and five different brands of smartphones. We summarize our preliminary findings below:

- The intermodulated signal generally exists on the tested devices. However, the amplification of phase variations varies across devices.
- Besides the 3rd-order intermodulated signal, the 5th-order and 7th-order intermodulated signals can also be observed. Although a higher-order intermodulated signal can result in larger phase variations, its strength is weaker. The intermodulated signal with a suitable order should be selected for good sensing performance.
- We implement the vibration measurement prototype on both ReSpeaker platform and smartphones. Experiments show that we can achieve accurate vibration amplitude measurement on all the tested devices. The accuracy of sub-millimeter-level vibration sensing is improved from 0.137 mm to 0.029 mm.

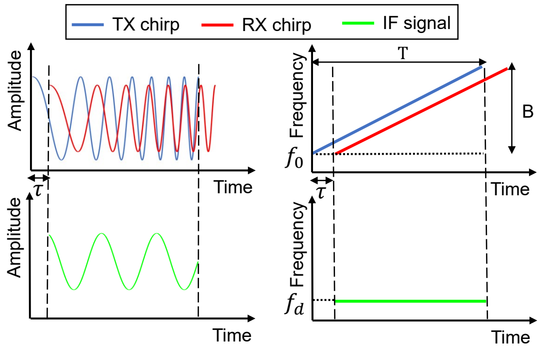


Figure 4: The transmitted signal (TX), received signal (RX) and intermediate frequency signal (IF).

- We analyze the factors impacting the sensing performance for intermodulated signal, including distance, target material, and device diversity. We find that the sensing performance is significantly impacted by the intensity of the input signal at the microphone.

2 PRELIMINARIES

2.1 Chirp-based Acoustic Sensing

Chirp signal is widely adopted in acoustic sensing. As shown in Figure 4, it is a frequency-modulated sine wave, whose frequency sweeps linearly over time. The chirp signal transmitted by the speaker can be represented as

$$S^T(t) = \cos\left(2\pi(f_0 t + \frac{B}{2T}t^2)\right), \tag{1}$$

where f_0 is the beginning frequency, T is the chirp duration, and B is the bandwidth. The received signal reflected from the target is a delayed version of the transmitted signal, which can be denoted as

$$S^R(t) = \alpha \cos\left(2\pi(f_0(t-\tau) + \frac{B}{2T}(t-\tau)^2)\right), \tag{2}$$

where α is an amplitude attenuation factor, and τ is the time-of-flight (ToF) in the air.

Then we multiply the received signal (RX) with the transmitted signal (TX) to derive the intermediate frequency (IF) signal that contains the phase information for subtle movement. After applying a low-pass filter, the IF signal becomes

$$\begin{aligned} S^{IF}(t) &= \frac{1}{2}\alpha \cos(2\pi \frac{B}{T} t\tau + 2\pi f_0 \tau - \frac{\pi B}{T}\tau^2) \\ &= \frac{1}{2}\alpha \cos(2\pi f_d t + \phi_d), \end{aligned} \tag{3}$$

where $f_d = \frac{B}{T}\tau$ is the beat frequency computed by the frequency difference between the transmitted signal and received signal. $\phi_d = 2\pi f_0 \tau - \frac{\pi B}{T}\tau^2 \approx 2\pi f_0 \tau$ is the initial phase. The approximation is based on the fact that $2\pi f_0 \tau$ is usually two orders of magnitude larger than $\frac{\pi B}{T}\tau^2$ due to the very small value of τ. Suppose that the distance between the device and target is d, the ToF τ can be computed as the round-trip distance divided by the signal speed in air c, i.e., $\frac{2d}{c}$. Therefore, the initial phase can be further denoted as $\phi_d = \frac{4\pi f_0 d}{c}$. If the target displacement is Δd, the phase variations caused by the target movement can be computed as $\Delta\phi = \frac{4\pi f_0 (d+\Delta d)}{c} - \frac{4\pi f_0 d}{c} = \frac{4\pi f_0 \Delta d}{c}$.

2.2 Non-linearity on Commodity Devices

Prior studies [17, 30] have shown that the components in a microphone can cause non-linear distortion at the received signal. Specifically, if the received signal reflected from the target is S_{in}, the output signal S_{out} after passing through the microphone can be represented as

$$\begin{aligned} S_{out} &= \sum_{i=1}^{\infty} k_i S_{in}^i \\ &= k_1 S_{in} + k_2 S_{in}^2 + k_3 S_{in}^3 + k_4 S_{in}^4 + ..., \end{aligned} \tag{4}$$

where k_i is the non-linear coefficient for the i_{th} term. The higher-order term is weaker as the order of distortion increases. Except for the first-order term, all of the remaining terms, i.e., the second-order term, the third-order term, etc., are non-linear distortions. In this paper, we mainly exploit the third-order term to boost the sensing granularity.

3 MATHEMATICAL DERIVATION

The received signal at the microphone is a superimposition of multiple paths, including the direct path and reflection paths from the surrounding objects. We observe that the third-order and even higher-order intermodulation between the direct path and reflection path from the target can be exploited to boost the sensing granularity. Next we mathematically prove this observation by taking the third-order intermodulated signal as the example.

According to Equation (2), the direct path signal $S_1(t)$ can be represented as

$$S_1(t) = \alpha_1 \cos\left(2\pi(f_0(t-\tau_1) + \frac{B}{2T}(t-\tau_1)^2)\right). \tag{5}$$

To remove the delay caused by the operating system, we align the received signal with the direct path signal [29]. Thus, Equation (5) can be further simplified as

$$S_1(t') = \alpha_1 \cos(2\pi f_1 t'), \tag{6}$$

where $t' = t - \tau_1$ denotes the alignment operation, and we simplify $S_1(t')$ as a single-frequency signal whose frequency is $f_1 = f_0 + \frac{B}{T}t'$. Similarly, we can denote the reflection path signal from the target $S_2(t')$ as

$$S_2(t') = \alpha_2 \cos(2\pi f_2(t' - \tau_2')). \tag{7}$$

where $\tau_2' = \tau_2 - \tau_1$ is the ToF of the reflection path signal after the alignment, and f_2 equals to $f_0 + \frac{B}{2T}(t' - \tau_2')$. For simplicity, we denote $S_1(t')$ as S_1 and $S_2(t')$ as S_2 hereafter. By substituting $S_1 + S_2$ into the third-order term of Equation (4), we can obtain

$$\begin{aligned} &k_3(S_1+S_2)^3 \\ =&k_3S_1^3 + 3k_3S_1^2S_2 + 3k_3S_1S_2^2 + k_3S_2^3. \end{aligned} \tag{8}$$

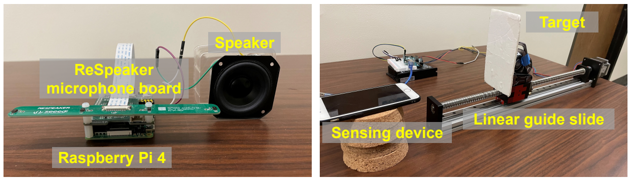
According to the power-reduction formula, $k_3S_1^3$ can be expanded as the sum of two terms, i.e., $\frac{3}{4}k_3\alpha_1^3\cos(2\pi f_1 t')$ and $\frac{1}{4}k_3\alpha_1^3\cos(2\pi \cdot 3f_1 \cdot t')$. The first term is the attenuated version of the direct path signal, and the second term will be filtered out since its frequency is much higher than the cut-off frequency of the microphone filter. The expansion of $k_3S_2^3$ is the same as $k_3S_1^3$. Therefore, no extra frequency components are introduced for both $k_3S_1^3$ and $k_3S_2^3$.

In the following, we analyze the remaining terms, i.e., $3k_3S_1^2S_2$ and $3k_3S_1S_2^2$. Using the power-reduction formula and product-to-sum identity, $3k_3S_1^2S_2$ can be expanded as the sum of three terms, i.e., $\frac{3}{2}k_3\alpha_1^2\alpha_2\cos(2\pi f_2(t'-\tau_2'))$, $\frac{3}{4}k_3\alpha_1^2\alpha_2\cos(2\pi\cdot(2f_1+f_2)\cdot t' - 2\pi f_2\tau_2')$, and $\frac{3}{4}k_3\alpha_1^2\alpha_2\cos(2\pi\cdot(2f_1-f_2)\cdot t' + 2\pi f_2\tau_2')$. The first term is the attenuated version of the target reflection, and the second term will be filtered out since it lies outside the microphone’s cut-off frequency. It is worth noting that the third term is the newly-introduced component at frequency $2f_1 - f_2$, which is kept. We denote the third term as $S^R_{2f_1-f_2}$ hereafter. Similarly, $3k_3S_1S_2^2$ also introduces a new component at frequency $2f_2 - f_1$.

Now we derive the intermediate frequency signals of the two newly-introduced components. Since the received signal is aligned with the direct path signal, we multiply the aligned signal with a delayed version of the transmitted signal whose ToF equals to that of the direct path signal $S^{DT} = \cos\left(2\pi(f_0(t-\tau_1) + \frac{B}{2T}(t-\tau_1)^2)\right)$. Using the product-to-sum identity and a low-pass filter, we can obtain

$$\begin{aligned} &S^{IF}_{2f_1-f_2} + S^{IF}_{2f_2-f_1} \\ =&S^{DT}\cdot S^R_{2f_1-f_2} + S^{DT}\cdot S^R_{2f_2-f_1} \\ =&\frac{3}{8}k_3\alpha_1^2\alpha_2\cos(2\pi f_d' t + \phi_d') + \\ &\frac{3}{8}k_3\alpha_1\alpha_2^2\cos(2\pi\cdot 2f_d'\cdot t + 2\phi_d'), \end{aligned} \tag{9}$$

where $f_d' = \frac{B}{T}(\tau_2 - \tau_1)$ and $\phi_d' = 2\pi f_0(\tau_2 - \tau_1)$. The first term $S^{IF}_{2f_1-f_2}$ is the attenuated version of the original intermediate frequency signal for the reflection path. Compared with the original intermediate frequency signal, the resulted intermodulated signal for the second term $S^{IF}_{2f_2-f_1}$ has two properties: (i) The beat frequency is twice larger than that of the original beat frequency; (ii) The phase variations are also twice larger than those of the original phase variations. Larger phase variations can provide better estimation of the arc center in the I/Q plane, and thus, yield more accurate phase extraction. We name the second term $S^{IF}_{2f_2-f_1}$ as the 3rd-order intermediate frequency (IF) signal hereafter. Similarly, we can obtain the 5th-order IF signal and 7th-order IF signal as $S^{IF}_{3f_2-2f_1}$ and $S^{IF}_{4f_2-3f_1}$, respectively.



(a) ReSpeaker platform. **(b) Sensing scenario.**

Figure 5: Experiment setup.

In a multipath-prevalent environment, besides the target reflection, there are reflections from other surrounding objects, resulting in a large number of potential intermodulated signals. However, most of the intermodulated signals are too weak to be detected. The strength of the intermodulated signal is proportional to the strength of the two signals that generate it. The intermodulated signals from two reflected signals are extremely weak and can be neglected. We only need to consider the intermodulated signal when one of the two signals is the strong direct path signal. Note that even if the strong direct path signal is involved, the intermodulated signal is still very weak when the other signal is reflected from an object at a distance. Therefore, only a limited number of intermodulated signals are strong enough to be detected.

4 IMPLEMENTATION

We implement our proposed system on the ReSpeaker platform and five different brands of smartphones. The received acoustic signals are analyzed in MATLAB using a laptop. To sense the fine-grained movement, we first estimate the target range bin corresponding to the original signal. Then we identify the target range bin corresponding to the non-linear IF signal by multiplying the beat frequency by a factor of $\frac{o+1}{2}$, where o is the intermodulation order. For example, we multiply 2 for the 3rd-order IF signal. At last, we extract the phase variations of the non-linear IF signal and utilize it to sense the fine-grained target movement.

Sensing Devices. We evaluate the performance of our proposed idea using a ReSpeaker 4-mic Linear Array board [19]. The microphone board and a general-purpose speaker are connected with Raspberry Pi 4 that controls the transmission and reception of acoustic signals, as shown in Figure 5a. Note that only one microphone is used for sensing. To verify the generalizability, we also conduct experiments using five smartphones, including iPhone 12, iPhone 6, Pixel 4, Samsung Galaxy S9+ and Sony Xperia G3423.

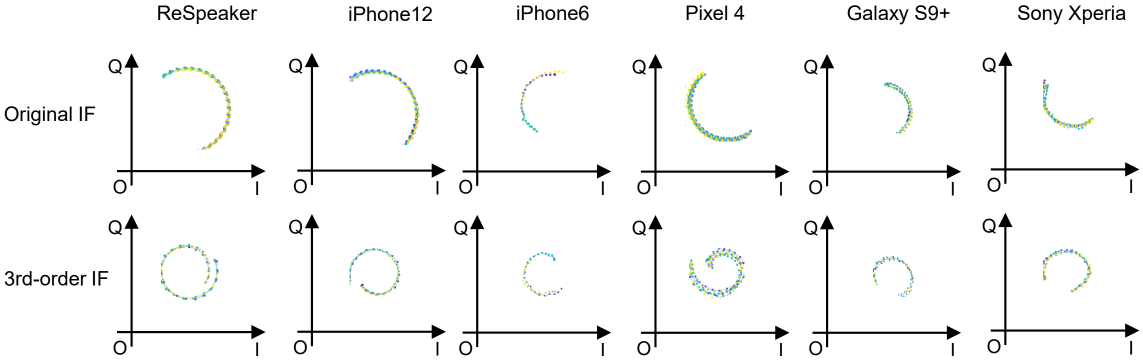


Figure 6: The illustration of IQ signals computed from the original IF signal and 3rd-order IF signal for different devices.

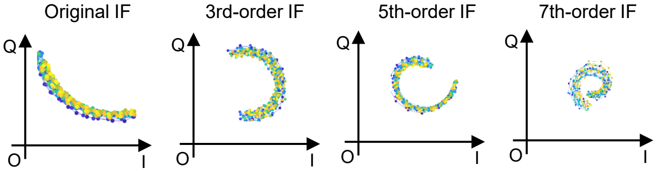


Figure 7: The comparison of IQ signals computed from the original IF signal and high-order IF signals.

Sensing Signals. The chirp signal adopted for the ReSpeaker microphone platform sweeps from 20 *kHz* to 22 *kHz*. Due to the poor high-frequency response [21], we configure the frequency of the chirp signal on smartphones from 16 *kHz* to 18 *kHz*. The duration of the chirp signal is set to 40 *ms*, and the chirp signal is sampled at 48 *kHz*.

Vibration System. To mimic subtle movement, we place the target on a linear slide that has a precision of 0.05 *mm* as shown in Figure 5b. Unless otherwise specified, we adopt the hand-sized cardboard as the target, and the distance between the device and target is set to 0.2 *m*.

5 EVALUATION

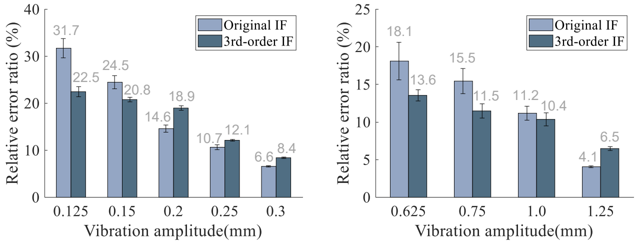
This section evaluates the performance of our proposed idea. We estimate the amplitude of each vibration, i.e., the maximum displacement that the target moves. We measure 15 vibrations for each trial and perform 10 trials for each setup.

5.1 Device Generalizability

We configure the slide to vibrate the target with an amplitude of 2.5 *mm*. Figure 6 displays the extracted IQ signals computed from both the original IF signal and 3rd-order IF signal. We can observe that the amounts of non-linear distortions vary across devices.

5.2 Higher-order IF Signal

Figure 7 compares IQ signals computed from the original IF signal and non-linear IF signals for the ReSpeaker platform.



(a) ReSpeaker board. **(b) iPhone 12.**

Figure 8: The overall performance comparison.

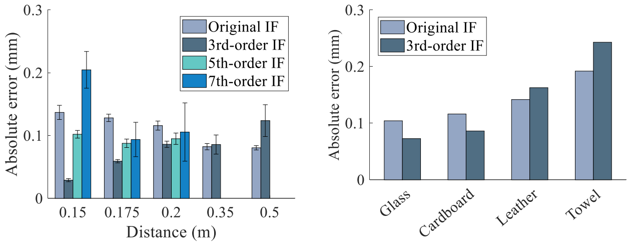
We can observe that the strength of the non-linear IF signal decreases as the order increases, which is consistent with our analysis in Sec. 3. Despite the non-linear IF signal is much weaker, we can still identify the amplified phase variations caused by the subtle movement.

5.3 Overall Performance

We conduct experiments to compare the performance of the original IF signal and 3rd-order IF signal. We adopt the relative error ratio as the evaluation metric, which is defined as the ratio of the absolute amplitude error to the vibration amplitude. Figure 8a and Figure 8b show the results for the ReSpeaker platform and iPhone 12, respectively. We can observe that the 3rd-order IF signal outperforms the original IF signal when the vibration amplitude decreases, indicating the effectiveness of boosting the sensing granularity. The amplitude estimation error for iPhone 12 is larger than that for the ReSpeaker platform due to larger hardware noise.

5.4 Impacting Factors

Impact of Distance. To evaluate how our proposed system works at various distances, we vary the distance between the sensing device (i.e., iPhone 12) and moving target (i.e., cardboard) from 0.15 *m* to 0.5 *m*. The vibration amplitude of



(a) Distance. **(b) Target material.**

Figure 9: The impacting factors.

the moving target is set to 0.75 *mm*. We adopt the absolute amplitude error as the evaluation metric here.

From Figure 9a, we can observe that, when the target is close to the sensing device, i.e., 0.15 *m* and 0.2 *m*, the 3rd-order IF signal performs better than the original IF signal. At 0.15 *m*, the median absolute amplitude error is reduced from 0.137 *mm* to 0.029 *mm*. However, as the distance increases, the original IF signal outperforms the 3rd-order IF signal. This is because when the distance between the device and target increases, the strength of the 3rd-order IF signal decreases much faster than that of the original IF signal.

We also plot the absolute amplitude errors for the 5th-order IF signal and 7th-order IF signal in Figure 9a. We can observe that, when the target is close to the device, both of them can accurately estimate the amplitude. However, as the distance increases, the high-order IF signal is too weak to be detected. Another observation is that a higher-order IF signal (e.g., 7th-order) actually results in a poorer performance because the intermodulated signal is too weak to be utilized for sensing.

Impact of Reflection Material. We evaluate the performance of our proposed system when objects made of different materials are used as the target. The distance between the device and target is set to 0.15 *m*, and the vibration amplitude is set to 0.75 *mm*. Figure 9b plots the absolute amplitude errors for the original IF signal and 3rd-order IF signal. We can observe that the performance for both signals varies across reflection materials. The reason is that the strength of the reflected signal varies across materials. For example, the signal reflected by the glass is stronger than that reflected by the towel due to the smooth surface. The strengths of the IF signals are both proportional to that of the reflected signal, resulting in performance variation across materials.

6 RELATED WORK

Recent years have witnessed an increasing interest in employing acoustic signals for human and environment sensing [4, 6, 8, 15, 17, 23, 24, 31]. Compared to other sensing modalities such as WiFi sensing, acoustic sensing can achieve a finer granularity owing to the low speed of acoustic signal in the air. Chirp signal is widely adopted for acoustic sensing due to its excellent performance against multipath and noise [5]. Existing studies focus on improving the performance of acoustic sensing in three directions, i.e., longer sensing range [10, 13–15], finer sensing granularity [12, 16, 29] and simultaneous multi-target sensing [9]. Our work falls in the second direction and is the first one to exploit the microphone non-linearity to boost the sensing granularity.

Intermodulated signal is produced by the non-linearity distortion from microphones, which has been adopted by prior studies to enable applications in localization [1, 11], communication [2, 17] and security [18, 30]. Instead, the work exploits the intermodulated signal for sensing purposes.

7 DISCUSSION

Limited sensing range. Due to weak target reflection, the sensing range for acoustic signals is limited [10, 15]. The strength of our adopted intermodulated signal is even weaker than that of the target reflection signal, which further reduces the sensing range of our proposed system. One potential solution is to increase the strength of the target reflection through spatial beamforming. In this paper, we trade off sensing range for higher sensing granularity. When we care more about the sensing granularity than the sensing range, the proposed method can be adopted.

Audible sensing signal. Speakers and microphones are primarily optimized for human voices and musics whose frequencies are usually below 4 *kHz* [29]. Therefore, commodity smartphones have good frequency responses in the audible frequency ranges but have poor frequency responses in the inaudible frequency ranges [21]. This work adopts the audible chirp (i.e., 16 *kHz* − 18 *kHz*) as the sensing signal. One potential solution to alleviate the audible noise is to mask the audible sensing signal with white noise [3, 8, 25].

8 CONCLUSION

This paper leverages intermodulated chirp signal generated by the direct path and target reflection due to the non-linear distortion of microphones to boost the granularity of acoustic sensing. We mathematically show the feasibility and further conduct benchmark experiments on six different devices to demonstrate the effectiveness of the proposed idea. We believe this work pushes the granularity boundary of acoustic sensing and can benefit a large range of real-life applications.

ACKNOWLEDGMENTS

This work was partially supported by National Institutes of Health (NIH) under Award Number 5R01MH122371-03 and National Science Foundation (NSF) under Award Number 2112562 and 2140247. We appreciate the valuable comments and feedback from our shepherd and anonymous reviewers.

REFERENCES

[1] Zhenlin An, Qiongzheng Lin, Lei Yang, and Yi Guo. 2020. Revitalizing ultrasonic positioning systems for ultrasound-incapable smart devices. *IEEE Transactions on Mobile Computing* 20, 5 (2020), 2007–2024.

[2] Yang Bai, Jian Liu, Li Lu, Yilin Yang, Yingying Chen, and Jiadi Yu. 2020. BatComm: enabling inaudible acoustic communication with high-throughput for mobile devices. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 205–217.

[3] Chao Cai, Zhe Chen, Henglin Pu, Liyuan Ye, Menglan Hu, and Jun Luo. 2020. AcuTe: Acoustic thermometer empowered by a single smartphone. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 28–41.

[4] Chao Cai, Rong Zheng, Jun Li, Linwei Zhu, Henglin Pu, and Menglan Hu. 2019. Asynchronous acoustic localization and tracking for mobile targets. *IEEE Internet of Things Journal* 7, 2 (2019), 830–845.

[5] Chao Cai, Rong Zheng, and Jun Luo. 2022. Ubiquitous acoustic sensing on commodity iot devices: A survey. *IEEE Communications Surveys & Tutorials* 24, 1 (2022), 432–454.

[6] Zhihui Gao, Ang Li, Dong Li, Jialin Liu, Jie Xiong, Yu Wang, Bing Li, and Yiran Chen. 2022. MOM: Microphone based 3D Orientation Measurement. In *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 132–144.

[7] Chengkun Jiang, Junchen Guo, Yuan He, Meng Jin, Shuai Li, and Yunhao Liu. 2020. mmVib: micrometer-level vibration measurement with mmwave radar. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–13.

[8] Dong Li, Shirui Cao, Sunghoon Ivan Lee, and Jie Xiong. 2022. Experience: practical problems for acoustic sensing. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*. 381–390.

[9] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2020. FM-track: pushing the limits of contactless multi-target tracking using acoustic signals. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 150–163.

[10] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2022. LASense: Pushing the Limits of Fine-grained Activity Sensing Using Acoustic Signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–27.

[11] Qiongzheng Lin, Zhenlin An, and Lei Yang. 2019. Rebooting ultrasonic positioning systems for ultrasound-incapable smart devices. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.

[12] Jialin Liu, Dong Li, Lei Wang, and Jie Xiong. 2021. BlinkListener: " Listen" to Your Eye Blink Using Your Smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–27.

[13] Jialin Liu, Dong Li, Lei Wang, Fusang Zhang, and Jie Xiong. 2022. Enabling Contact-free Acoustic Sensing under Device Motion. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–27.

[14] Wenguang Mao, Wei Sun, Mei Wang, and Lili Qiu. 2020. DeepRange: Acoustic Ranging via Deep Learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–23.

[15] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. 2019. RNN-Based Room Scale Hand Motion Tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*. ACM, 38.

[16] Kun Qian, Chenshu Wu, Fu Xiao, Yue Zheng, Yi Zhang, Zheng Yang, and Yunhao Liu. 2018. Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart devices. In *IEEE INFOCOM 2018-IEEE conference on computer communications*. IEEE, 1574–1582.

[17] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. 2017. Backdoor: Making microphones hear inaudible sounds. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. 2–14.

[18] Nirupam Roy, Sheng Shen, Haitham Hassanieh, and Romit Roy Choudhury. 2018. Inaudible Voice Commands: The {Long-Range} Attack and Defense. In *15th USENIX Symposium on Networked Systems Design and Implementation (NSDI 18)*. 547–560.

[19] Seeed. 2022. ReSpeaker 4-Mic Linear Array kit. https://wiki.seeedstudio.com/ReSpeaker_4-Mic_Linear_Array_Kit_for_Raspberry_Pi/. (2022).

[20] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. 2020. SpiroSonic: monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.

[21] Yu-Chih Tung, Duc Bui, and Kang G Shin. 2018. Cross-platform support for rapid development of mobile acoustic sensing applications. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 455–467.

[22] Haoran Wan, Shuyu Shi, Wenyu Cao, Wei Wang, and Guihai Chen. 2021. RespTracker: Multi-user Room-scale Respiration Tracking with Commercial Acoustic Devices. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. IEEE, 1–10.

[23] Haoran Wan, Lei Wang, Ting Zhao, Ke Sun, Shuyu Shi, Haipeng Dai, Guihai Chen, Haodong Liu, and Wei Wang. 2022. VECTOR: Velocity Based Temperature-field Monitoring with Distributed Acoustic Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–28.

[24] Anran Wang and Shyamnath Gollakota. 2019. Millisonic: Pushing the limits of acoustic motion tracking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.

[25] Anran Wang, Jacob E Sunshine, and Shyamnath Gollakota. 2019. Contactless infant monitoring using white noise. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.

[26] Wei Wang, Alex X Liu, and Ke Sun. 2016. Device-free gesture tracking using acoustic signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 82–94.

[27] Wei Wang, Lei Xie, and Xun Wang. 2017. Tremor detection using smartphone-based acoustic sensing. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 309–312.

[28] Wikipedia. 2022. Intermodulation distortion. https://en.wikipedia.org/wiki/Intermodulation. (2022).

[29] Fusang Zhang, Zhi Wang, Beihong Jin, Jie Xiong, and Daqing Zhang. 2020. Your Smart Speaker Can" Hear" Your Heartbeat! *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–24.

[30] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2017. Dolphinattack: Inaudible voice commands. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 103–117.

[31] Qian Zhang, Dong Wang, Run Zhao, Yinggang Yu, and Junjie Shen. 2021. Sensing to hear: Speech enhancement for mobile devices using acoustic signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–30.

[32] Ningzhi Zhu, Huangxun Chen, and Zhice Yang. 2021. Fine-grained Multi-user Device-Free Gesture Tracking on Today's Smart Speakers. In *2021 IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*. IEEE, 99–107.