SCALAR: Self-Calibrated Acoustic Ranging for Distributed Mobile Devices

Lei Wang, Haoran Wan, Ting Zhao, Ke Sun, Shuyu Shi, Haipeng Dai, *Member, IEEE*, Guihai Chen, *Senior Member, IEEE*, Haodong Liu, and Wei Wang, *Member, IEEE*

***Abstract*—Acoustic ranging has been viewed as a promising Human-Computer Interaction (HCI) technology in many scenarios, such as Augmented Reality (AR)/Virtual Reality (VR) and smart appliances. Most ranging systems with distributed devices undergo an extra calibration process to remove the timing errors. However, the calibration process needs user intervention. Furthermore, it should assume that the clock drifts are linear and stable, which is disabled within tens of minutes. In this paper, we introduce a self-calibrated acoustic ranging system that achieves sub-millimeter accuracy on distributed asynchronous devices. Based on our theoretical timing model, we precisely cancel both the system delay and the nonlinear clock drift with carefully designed Orthogonal Frequency-Division Multiplexing (OFDM) ranging signals. Our synchronization scheme achieves a timing accuracy of 1.9 microseconds, which allows us to build large-scale virtual acoustic arrays. Based on such a calibration scheme, our localization system achieves a ranging error of 0.39 mm within three meters in real-world experiments.**

***Index Terms*—Acoustic sensing, distributed ranging.**

I. Introduction

With the rapid development of mobile devices, various device-based tracking applications have been enabled in many scenarios, including 3-D user interaction [1], [2], [3], [4], [5], [6], motion tracking for AR/VR [7], [8], [9], [10], and drone navigation [11]. Due to the six-order-of-magnitude difference between the speed of sound and the speed of light, acoustic tracking systems achieve a higher tracking precision than radio-based systems on resource limited mobile devices. Sub-millimeter tracking accuracy has been demonstrated on commercial smartphones [7], while state-of-the-art radio-based tracking systems are limited to decimeter-level accuracy [12], [13], [14], [15], [16], [17]. Furthermore, with ubiquitous acoustic devices surrounding us, including voice assistants, televisions, and mobile phones, we can reuse these on-device speakers and microphones that are widely available to enable numerous distributed tracking applications in the future smart home, such as quantitative remote control and video gaming. Therefore, the distributed tracking system based on audio signal becomes an excellent candidate for smart home applications, as shown in Fig. 1.

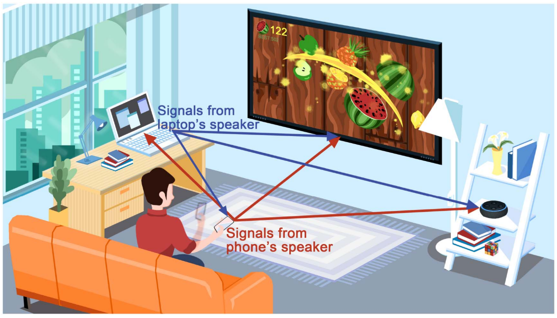


Fig. 1. A typical application scenario for distributed tracking. With the high-precision distributed ranging, it is possible to derive the highly accurate initial location of the mobile device and further enable tracking applications using multiple distributed devices.

While promising, there is a severe issue hindering the adoption of acoustic distributed tracking: *the mobile device's initial location*. Without the initial locations, the system can only track the relative movement of the device and the resulting trajectory could be severely distorted. Most acoustic tracking systems resort to an extra calibration process to infer the initial location. For example, users have to move the device along a given trajectory [18], [19] or to a known position [7], which needs user intervention and cannot be performed automatically. Furthermore, users need to re-calibrate the system within tens of minutes to ensure the accuracy of the location [7].

The main reason for such continuous calibration requirement is the timing errors caused by the unstable built-in reference clocks in distributed devices. There are two major sources of timing errors. First, applications cannot precisely control the underlying system delay in audio playback and recording. Second, there are clock drifts between the transmitting and the receiving audio devices. Such timing errors introduce a *dynamic* bias in distance measurements. Furthermore, most calibration

Manuscript received 19 April 2022; revised 7 November 2022; accepted 19 January 2023. Date of publication 1 February 2023; date of current version 8 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grants 62102006, 61872173, and 62272223. Recommended for acceptance by B. Liang. *(Corresponding author: Wei Wang.)*

Lei Wang is with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China, and also with Soochow University, Suzhou 215021, China (e-mail: wangl@smail.nju.edu.cn).

Haoran Wan, Ting Zhao, Shuyu Shi, Haipeng Dai, Guihai Chen, and Wei Wang are with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China (e-mail: wanhr@smail.nju.edu.cn; zhaoting@smail.nju.edu.cn; ssy@nju.edu.cn; haipengdai@nju.edu.cn; gchen@nju.edu.cn; ww@nju.edu.cn).

Ke Sun is with the Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093 USA (e-mail: kesun@eng.ucsd.edu).

Haodong Liu is with Hisilicon Co.Ltd., Shanghai 518129, China (e-mail: liuhaodong@hisilicon.com).

Digital Object Identifier 10.1109/TMC.2023.3241304

1536-1233 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.

algorithms assume that the clock drifts are linear and stable [7], [18], [19]. It takes a few seconds to estimate the slope of the drift but may lose synchronization within tens of minutes due to estimation errors or nonlinearity in the clock [7]. Therefore, synchronizing physically separated devices without modifying the hardware is the the key to high-precision distributed acoustic localization systems.

There are automatic calibration solutions that use radio signals [20] or round-trip sound signals [21], [22] to calibrate distributed devices on-the-fly. However, these solutions only provide coarse-grained calibration that incurs centimeter-level ranging errors [20], [21] and decimeter-level 3-D localization errors [22]. Compared with the *tracking* task, which requires an initial location and continuous monitoring of the movement, the *ranging* task that gives the absolute distance between devices is much harder. There is still a huge gap between the centimeter-level ranging accuracy and sub-millimeter level tracking accuracy.

In this paper, we introduce the SCALAR system, a Self-Calibrated Acoustic Ranging system that achieves sub-millimeter ranging accuracy on distributed commercial devices. We prove that SCALAR perfectly cancels timing errors caused by both the system delay and the nonlinear clock drift so that we can directly measure the Time-of-Flight (ToF) between the speaker and the microphone. In real-world experiments, we show that our calibration scheme achieves timing accuracy of $1.9\,\mu s$ (microseconds), which converts to a ranging error of $0.39\,mm$ on commercial mobile devices running Android, iOS, and Linux. In our implementations, SCALAR only uses high-level audio APIs provided by the operating system, without precise timing control requirement on either the audio device or the wireless transceiver. Furthermore, SCALAR does not need user intervention and outputs the correct distance within 0.5 seconds after the system starts playing sound. This allows low duty cycle operation, e.g., only operate for 0.5 seconds in every ten seconds, to save energy when performing long-term monitoring tasks. With the self-calibration capability, we can build virtual transmit/receive acoustic arrays using distributed devices with synchronized phases of the acoustic signal. Such virtual arrays improve the coverage of the acoustic sensing system and provide robust localization even if some of the Line-of-Sight (LOS) paths to some of the devices are occluded. Furthermore, the application of such large-scale distributed acoustic systems is not limited to ranging and tracking. They could also be useful in device-free acoustic sensing [23], [24] as well as in sound-field reconstruction [25].

We face three key technical challenges when developing SCALAR. The first challenge is to cancel the clock offset at a precision that is less than 10% of audio sampling intervals. The widely supported $48\,kHz$ audio sampling rate has a sampling interval of $20.8,\mu s$, during which the sound travels by $7\,mm$ in distance. The sampling clock of two non-synchronized devices could be misaligned by a fraction of the sampling interval that leads to a millimeter-level distance bias. To achieve sub-millimeter accuracy, we use the phase of the same central frequency measured by different devices to cancel the clock offset. Our phase-based operation can perfectly cancel dynamical sub-sample misalignment caused by clock drifts. The second challenge is to remove the range ambiguity of phase-based measurements. As the phase measurements have a limited range of $-\pi$ to π, we get the same phase when the device is moved by a full wavelength. To remove such ambiguity, we design an OFDM ranging signal that can measure both the coarse-grained cross-correlation estimation and the fine-grained phase estimation. By combining the measurements with different resolutions, we can resolve the range ambiguity with a bandwidth of just 4 kHz. The third challenge is to identify and precisely measure the ToF over multiple sound paths. To address this challenge, we use the excellent auto-correlation property of our OFDM signal to separate different sound paths so that we are able to remove the impacts of objects that are more than 30 cm from the LOS path and capture the ToF of four paths with a single measurement for every two pairs of microphone and speaker. Our experimental results show that SCALAR achieves a 1-D ranging error of 0.39 mm within a distance of three meters. Furthermore, it can perform one-shot measurement that directly returns the correct distance within 500 ms after cold starts. With the new capability, we show that SCALAR is able to use multiple distributed devices to localize a device with a error of 1.07 mm and 1.86 mm in the 2-D and 3-D space, respectively.

II. Model for Acoustic Ranging

In this section, we first analyze sources of timing errors in acoustic ranging systems. We then show that these errors can be canceled when two devices transmit and receive at the *same time* and at the *same central frequency*.

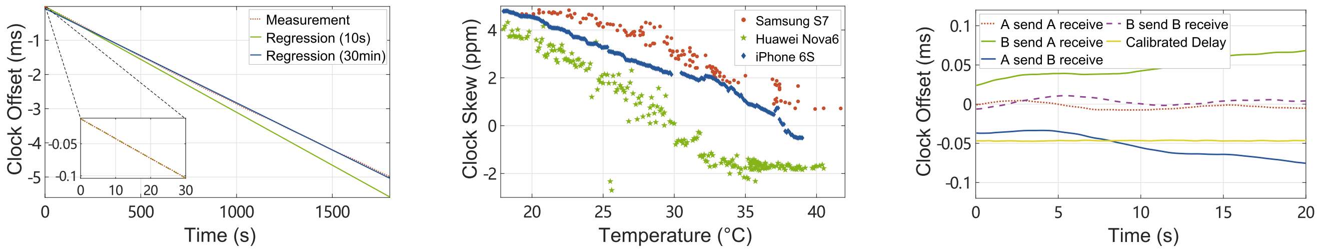
A. Timing Error Analysis

Timing precision is the key to acoustic ranging. There are two major sources of timing error when measuring the ToF of sound waves traveling from one device to another device.

The first one is the system delay [21], [26], [27], [28]. When playing audio, there is a system delay between the time that application puts the digital audio sample into the playback buffer and the time that the audio sample is really played out by the speaker. Similarly, there is a system delay when recording audio from the microphone. Such system delays include both the software delay of the audio drivers and the hardware delay of amplifiers and filters in the audio system. The system delay leads to a random offset that may change every time when the application starts playback/recording. The second source of timing error is the sampling clock offset. Audio systems on mobile devices use local oscillators to generate the sampling clock. Due to hardware imperfections and temperature changes, the frequency generated by the local oscillator may have an error of up to ±50 ppm (part-per-million) [29]. Using different sampling clocks, the digital samples taken by different devices are misaligned, and the sampling offset between the two digital sequences keeps changing because of the accumulated clock drifts.

Traditional acoustic ranging systems perform calibration every time that the audio system starts playback/recording to correct the timing error [7], [18], [30]. To estimate the system

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.



(a) Long-term linear regression errors

(b) Clock skew variations for different pairs of mobile devices

(c) Short-term nonlinear clock offsets between two mobile devices (A and B)

Fig. 2. Clock offset measurement results on commercial devices.

delay, the user has to put devices at known positions [7] or move them along a specific trajectory [18], [30] after the system starts. Then, the clock skew is estimated based on the linear clock offset model. Given that the time on the clock of device A is t_A, the linear clock offset model assumes the clock on device B is

$$t_B = t_A - pt - b, \tag{1}$$

where b is the initial clock offset, p is the relative clock skew (slope of the clock drift) between the two local clocks, and t is the wall clock time. Traditional calibration schemes require the two devices to remain static for a few seconds so that the clock skew p can be estimated effectively. We can then compensate the clock offset using (1), by setting the initial clock offset b according to the system delay estimation.

There are several unsolved issues in the traditional calibration process. First, it requires user intervention every time that the system restarts. Second, the linear model is not accurate enough to keep long-term synchronization. Fig. 2(a) shows the sampling offset between two static mobiles measured by the phase offset of Continuous Wave (CW) at a frequency of 18 kHz with a sampling rate of 48 kHz. The clock skew is estimated with a ten-second time window using linear regression. While the short-term sampling offset is very close to the linear model in (1) as shown in the enlarged part in Fig. 2(a), the extrapolated clock offset could be as large as 0.6 ms after 30 minutes, which leads to a ranging error of 206 mm when the speed of sound is $c = 343$ m/s. Third, the sampling clock offset could be non-linear and the clock skew p may have long-term and short-term variations. In the long term, the clock offset may deviate from the linear module so that it cannot be corrected by linear regression. In the sample shown in Fig. 2(a), even if we perform the linear regression over a 30-minute period, the residual regression error is still as large as $60.5\,\mu s$, which leads to a ranging error of 20.8 mm. This is because the clock skew changes with the temperature so that it is not a constant in the long-term. Fig. 2(b) shows the clock skew measured on three pairs of smartphones, where the transmitter and receiver are of the same brand. We heat the receiving device to change the temperature of the device and measure the temperature using a FLIR E5-XT Infrared camera [31]. All clock skews are less than ±5 ppm, much smaller than the standard requirement of ±50 ppm [29]. However, the clock skew significantly changes with the temperature so that linear calibrations are no longer valid when the temperature of the device changes. In the short term, Fig. 2(c) shows the clock offset between two iPhone 6S, measured within ten seconds after device B starts playback/recording. We observe that clock offset in the recorded signals of device B has perceivable short-term nonlinearity as it is no longer a straight line. We suspect such nonlinear clock offset comes from the initialization of Phase-Locked Loop (PLL). Such nonlinearity prevents accurate estimation of the clock skew within a few seconds after the audio system starts.

B. Synchronization Model

SCALAR uses the following timing model to calibrate distributed audio systems. Given a device A, we denote the time indicated by its own clock as

$$t_A = t - q_A(t), \tag{2}$$

where t is the standard time (e.g., given by GPS) and $q_A(t)$ is the function of clock offset for device A when compared with the standard time. We *do not* assume that the offset function $q_A(t)$ is linear as the linear model in (1). Instead, we make the following assumption in our timing model.

Assumption: Both the system delay and the clock offset of the audio systems are quasi-static so that we can approximate them as constants within a short time period, e.g., 20 ms, for the sound traveling from the transmitter to the receiver.

The above assumption is valid for most commercial audio systems. First, as discussed in the previous section, the system delay should be quasi-static after the playback/recording starts so that there are no distortion in the continuous audio signal. Second, the clock skew in commercial systems are so small that the clock offset changes by a negligible amount during a short time period. For example, with a clock skew of 5 ppm as shown in Fig. 2(b), the clock offset changes by $20\,\text{ms} \times 5 \times 10^{-6} = 0.1\,\mu s$ within 20 ms. The sound wave only travels by a negligible distance of 0.03 mm within $0.1\,\mu s$. Thus, our assumption is valid for most commercial mobile devices. The propagation delay of 20 ms is sufficient for most indoor applications, as the sound travels by a distance of 6.8 meters within 20 ms. Therefore, based on our assumption, we can approximate the time of device A as a constant $t_A = t - q_A(t_0)$ during a *single* transmission that starts

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.

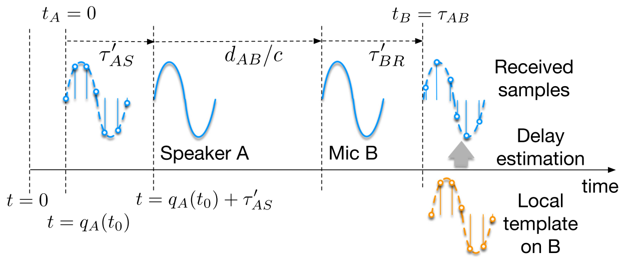
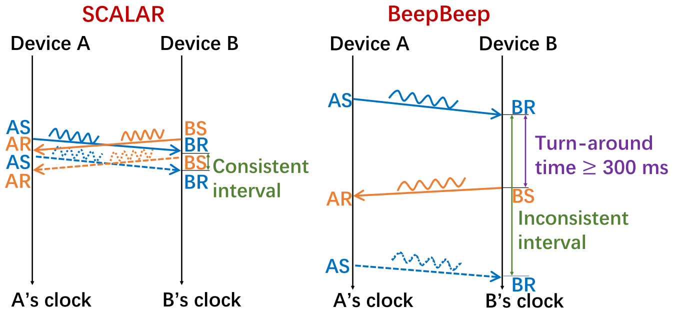


Fig. 3. Delay for sound transmission and receiving.



(a) SCALAR model

(b) BeepBeep model

Fig. 4. Comparison between SCALAR and BeepBeep. (a) SCALAR model (b) BeepBeep model.

at t_0. Similarly, we can use $t_B = t - q_B(t_0)$ to approximate the time at the receiving device B.

We use a simple two-device system where both device A and B can transmit and receive sound signals to demonstrate our model. The playback and recording delays in our timing model are shown in Fig. 3. At local time $t_A = 0$, device A starts transmitting the ranging signal by aligning the start of the ranging signal to the first playback sample in the digital audio sequence. Due to the clock offset, device A actually aligns the first sample at a standard time $t = t_A + q_A(t_0) = q_A(t_0)$. After the system delay τ'_{AS} for audio transmission on device A, the start of the signal reaches the speaker at time $t = q_A(t_0) + \tau'_{AS}$. The sound wave travels from speaker A to microphone B with a ToF of d_{AB}/c, where c is the speed of sound and d_{AB} is the distance between A and B. After a system delay of τ'_{BR}, for recording on device B, the sound signal is received by device B as digital samples. While the samples could be misaligned, e.g., the start of the received ranging signal is not aligned to any sampling point on device B as shown in Fig. 3, B can still use the delay measurement scheme described in later sections to get a precise estimation of ToF. On the standard clock, the time that B receives the starting sample is $t = q_A(t_0) + \tau'_{AS} + d_{AB}/c + \tau'_{BR}$. However, as the *measured* ToF τ_{AB} is based on the clock on device B with reading of $t_B = t - q_B(t_0)$, we have

$$\tau_{AB} = q_A(t_0) + \tau'_{AS} + d_{AB}/c + \tau'_{BR} - q_B(t_0). \tag{3}$$

Based on our assumption, both τ'_{AS} and $q_A(t_0)$ are constant in this short period, we can combine them to define the transmission delay as $\tau_{AS} = q_A(t_0) + \tau'_{AS}$ and drop the variable t_0 in our later discussions. We also define the receiving delay as $\tau_{BR} = \tau'_{BR} - q_B(t_0)$. Thus, the measured delay is given by

$$\tau_{AB} = \tau_{AS} + d_{AB}/c + \tau_{BR}. \tag{4}$$

Both the transmission and receiving delay τ_{AS} and τ_{BR} are unknown and slowly changing with time. However, when the two devices can transmit and receive at the same time, we can measure four delays on device A and B at the same time. We denote these delays as τ_{AA}, τ_{AB}, τ_{BA}, and τ_{BB}, using the subscripts to indicate the transmitter and the receiver. For example, τ_{AA} is the delay of the sound sent by A and measured also by device A itself. We show in the following theorem that all the unknown transmission/receiving delays can be canceled.

Theorem 1. For two devices that can transmit and receive at the same time, we can calculate the distance between the two devices using four delay measurements as

$$d_{AA} + d_{BB} - d_{AB} - d_{BA} = c(\tau_{AA} + \tau_{BB} - \tau_{BA} - \tau_{AB}). \tag{5}$$

Proof. We observe that when device B measures the delay of the ranging signal transmitted by device A, we have

$$\tau_{AB} = \tau_{AS} + d_{AB}/c + \tau_{BR}. \tag{6}$$

Similarly, we also have

$$\tau_{AA} = \tau_{AS} + d_{AA}/c + \tau_{AR}, \tag{7}$$

$$\tau_{BA} = \tau_{BS} + d_{BA}/c + \tau_{AR}, \tag{8}$$

$$\tau_{BB} = \tau_{BS} + d_{BB}/c + \tau_{BR}. \tag{9}$$

Using (6)–(9), we can cancel unknown delays by

$$\begin{aligned} &\tau_{AA} + \tau_{BB} - \tau_{BA} - \tau_{AB} \\ &= (d_{AA} + d_{BB} - d_{BA} - d_{AB})/c, \end{aligned}$$

which directly leads to (5).

The key innovation of our model is that it considers sub-sample clock drifts. While (5) is similar to previous round-trip based calibration schemes, such as BeepBeep [21], Theorem 1 remains valid for delays that are fractions of the sampling interval. The resolution of existing schemes are limited by the sampling interval, which is 20.8 μs at 48 kHz (around 7 mm in distance). In comparison, SCALAR works for sub-sample drifts so that we can calibrate the phase of the sound signal, which leads to sub-millimeter accuracy. Moreover, as shown in Fig. 4, there are two fundamental differences between SCALAR and the round-trip calibration used in BeepBeep. First, the unstable interval of the I/O operation (introduced in Section II-A) in the turn-around process may cause a system delay of up to 10 ms [27], leading to calibration failure. To this end, MotionBeep [27] proposes to keep the audio processing alive all the time. However, this method may result in increased power consumption. Instead, SCALAR just requires both devices to transmit signals at the same time and calculate the four delay measurements subsequently. According to Theorem 1, the system delay can be canceled out perfectly by performing addition operations. We leave the detailed description on how to enable the simultaneous transmission and reception with two devices in Section IV-A.

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.

Second, the round-trip calibration assumes that the clock-drift is negligible during the turn-around time. However, such drifts may lead to errors of more than 1 mm as the turn-around time could be as high as 300 ms for low-end devices to detect the pulse and send back the response. In contrast, the four delay measurements in the SCALAR model are derived within the same frame period (i.e., 20 ms for the frame size of 960 samples). Thus, the clock-drift within 20 ms is negligible compared to that in the round-trip calibration (i.e., BeepBeep and MotionBeep).

C. Ranging Model

Theorem 1 shows that we can use four delay measurements to cancel the unknown transmission/receiving delays. However, in the derived distance of $d_{AA} + d_{BB} - d_{AB} - d_{BA}$, there are still four unknown distances to be solved. Existing works assume the distance between the speaker and the microphone of a given device, i.e., d_{AA} or d_{BB}, is fixed and can be measured in advance [21]. Furthermore, when devices are on a straight line, we can also assume that $d_{AB} = d_{BA}$. With these three additional constraints, it is possible to solve the distance between device A and B as

$$d_{AB} = \frac{d_{AA} + d_{BB} - c(\tau_{AA} + \tau_{BB} - \tau_{BA} - \tau_{AB})}{2}.$$

Our derivation in Theorem 1 leads to two key insights that help understanding more general scenarios. First, we do not assume that the transmission and receiving delay $\tau_{AS}, \tau_{AR}, \tau_{BS}$, and τ_{BR} are related to each other. This means our cancellation scheme still works even if the speaker and the microphone on device A are separated. Second, when multiple speakers are transmitting at the same time, each microphone can independently measure its delay to all of the speakers. This provides us a system of linear equations that could be used for deriving multiple constraints on distances.

Consider a distributed acoustic system that has m transmitters and n receivers. In this case, we can rewrite (6) as

$$\tau_{ij} = \tau_{iS} + d_{ij}/c + \tau_{jR}, \tag{10}$$

where τ_{ij} is the delay of speaker i measured at microphone j, τ_{iS} and τ_{jR} are the unknown transmission/receiving delay for speaker i and microphone$j$, and d_{ij} is the distance between speaker i and microphone j. With $m \times n$ delay measurements, we will have a linear system

$$\begin{bmatrix} 1 & 0 & \cdots & 1 & 0 & \cdots \\ 1 & 0 & \cdots & 0 & 1 & \cdots \\ \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 & \cdots & 1 \\ 0 & 1 & \cdots & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \tau_{1S} \\ \tau_{2S} \\ \vdots \\ \tau_{1R} \\ \vdots \\ \tau_{nR} \end{bmatrix} = \begin{bmatrix} \tau_{11} - d_{11}/c \\ \tau_{12} - d_{12}/c \\ \vdots \\ \tau_{mn} - d_{mn}/c \end{bmatrix}, \tag{11}$$

which is a matrix multiplication of an $mn \times (m+n)$ matrix $\mathbf{A}$ with an $m+n$ vector $\mathbf{D}$: $\mathbf{AD} = \mathbf{R}$. Each row k in $\mathbf{A}$ represents a transmitter/receiver pair i and j, with elements of zero expect $a_{ki} = 1$ and $a_{k(m+j)} = 1$ for speaker i and microphone j. The unknown delays of τ_{iS} and τ_{jR} are listed in the $m+n$ vector $\mathbf{D}$, with m transmit delays followed by n receiving delays. Finally, the mn elements in $\mathbf{R}$ are the linear combinations of delay measurements and unknown distance $\tau_{ij} - d_{ij}/c$.

When we have two transmitters and two receivers, our formulation reduces to the specific case studied in [21], [32]

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tau_{1S} \\ \tau_{2S} \\ \tau_{1R} \\ \tau_{2R} \end{bmatrix} = \begin{bmatrix} \tau_{11} - d_{11}/c \\ \tau_{12} - d_{12}/c \\ \tau_{21} - d_{21}/c \\ \tau_{22} - d_{22}/c \end{bmatrix}. \tag{12}$$

Note that in this case the 4×4 matrix $\mathbf{A}$ only has a rank of three. Therefore we can combine the four equations and cancel all unknown delays to get $4 - 3 = 1$ equation of τ_{ij} and d_{ij}. For the four variables of d_{ij} we need to assume that we know d_{11} and d_{22}, also we should assume $d_{12} = d_{21}$ to reduce the number of unknown parameters to one.

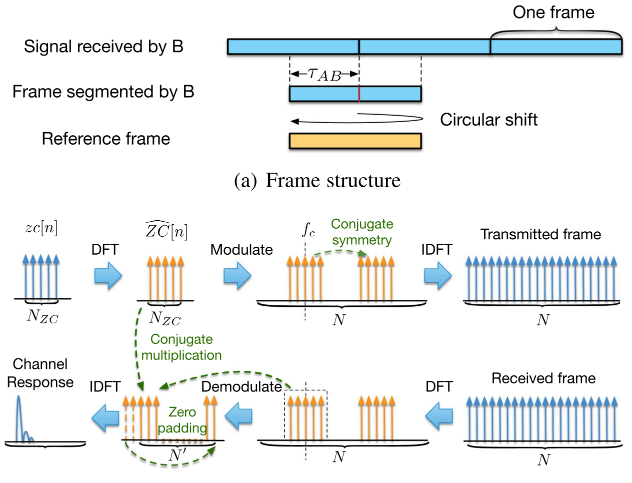
In the general case with m transmitters and n receivers while all of them can hear each other, we can perform $m \times n$ measurements at the same time. As matrix $\mathbf{A}$ is the transpose of the incidence matrix of a complete bipartite graph $K_{m,n}$, it only has a rank of $m + n - 1$. This gives us $mn - (m + n - 1) = (m - 1)(n - 1)$ independent equations of measurements and distances in the null space of $\mathbf{A}$, as it has mn rows. Since we have mn unknown distance values of d_{ij}, we need to know at least $m - n + 1$ constraints to solve for the rest unknown distances.

There are two types of constraints that can be used for solving the equations. The first type is some of precise distances that are known in advance. For example, the distance between the speaker and microphone on the same device is fixed once the device is manufactured. It is also possible to calibrate the distance between two separate but fixed devices, e.g., distance between a TV set and a voice assistant, after they are deployed. The second type of constraint comes from knowing that some speakers or microphones are driven by the same audio chip. In this case, these speakers and microphones are synchronized as different channels of the same audio stream. For example, if a microphone array has two microphones with identity of 1 to 2. Then, we would have a constraint for $\tau_{1R} = \tau_{2R}$. We can prove that these two types of constraints are interchangeable, i.e., knowing the distance between devices can be converted to knowing the synchronize relationship between devices (detailed proofs are omitted due to page limitations). Therefore, when the distance between some of speakers and microphones is known, we can build a virtual speaker or microphone array using distributed commercial devices. Such distributed system allows us to cover larger areas, as well as providing more measurements of the same target so that the target can be localized in a space with higher degrees of freedom.

III. Ranging Signal Design

The sub-sample calibration resolution of SCALAR calls for delay measurements at similar resolutions. Traditional

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.



(a) Frame structure

(b) OFDM modulation and demodulation (blue: time domain, orange: frequency domain)

Fig. 5. Structure of the ranging signal.

correlation-based systems cannot fulfill this requirement, as they measure the delay by counting the number of samples [21]. Therefore, we design an OFDM ranging signal that supports phase-based delay estimation, which can directly cancel the clock offset in phase measurements.

A. Signal Structure

We use a two-device system to demonstrate our signal design. Our ranging signal consists of periodical frames as shown in Fig. 5(a). We use a default frame size of 20 ms (i.e., 960 samples at 48 kHz) to achieve an unambiguous round-trip range of 3.4 meters which is sufficient for most applications. Note that we can increase the frame size to achieve larger unambiguous round-trip range.

When device B receives the signal, it first samples the signal using its own sampling clock and then segments the digital samples into frames consisting of 960 samples. Based on our assumption, the difference between the 20 ms frame duration measured by the clock of A and B is negligible. Therefore, the signal frame received by B is a circularly shifted version of the signal transmitted by A, as A repeats the frame with the same period of 20 ms. To ensure the signal received at the end of the frame experiences the same channel state as that received at the start of the frame, we need to further assume that the coherence time T_c is larger than 20 ms. The coherence time is the duration that the channel remains stable and can be estimated as $T_c \approx c/(4 f_c v)$, where f_c is the carrier frequency and v is the doppler speed [33]. The movement speed of the devices should be smaller than 22.6 cm/s, for a carrier frequency of $f_c = 19$ kHz to ensure $T_c \geq 20$ ms.

Our goal is to measure the amount of circular shift of the received signal at sub-sample resolution. Denote the digital sample of the frame received by B as $y[n], n = 0 \ldots, N-1$ with $N = 960$, and its Discrete Fourier Transform (DFT) as $Y[n]$. In the following discussions, we use the non-capitalized/capitalized symbols to denote signals in the time/frequency domain. We denote the frame transmitted by A as $x[n]$ and its DFT as $X[n]$. By the time shifting theorem [34], we have

$$Y[n] = e^{-j 2\pi n \tau f_s / N} X[n], \tag{13}$$

where $x[n]$ is circularly shifted by a delay of τ, where f_s is the sampling frequency and j is the imaginary unit. The time shifting theorem holds for delays of a fraction of sample, because we can treat the received signal as a continuous periodical signal. As the transmitted signal $x[n]$ is known, we can use the phase-shift in the frequency domain to measure the delay with sub-sample resolution. The measurement is based on the clock of device B, shown as τ_{AB} in Fig. 3.

Algorithm 1. Modulation Algorithm

```
Output: A modulated sequence $x[n]$ of length $N$
1: Generate $zc[n]$, $n = 0, \ldots, 2h_{zc}$; Perform $N_{zc}$-point
   DFT on $zc[n]$ to get $ZC[n]$;
2: Frequency domain rearrangement:
   $\widehat{ZC}[n] \Leftarrow ZC[n - h_{zc}], n = h_{zc}, \ldots, 2h_{zc}$,
   $\widehat{ZC}[n] \Leftarrow ZC[N_{zc} - 1 - n], n = 0, \ldots, h_{zc} - 1$;
3: OFDM modulation: $X[n] \Leftarrow 0, n = 0 \ldots N - 1$,
   $X[n + n_c - h_{zc}] \Leftarrow \widehat{ZC}[n], n = 0 \ldots 2h_{zc}$,
   $X[n] \Leftarrow X^*[N - n], n = N/2 + 1, \ldots, N - 1$;
4: Perform $N$-point IFFT on $X[n]$ to get $x[n]$;
```

We use the OFDM modulated Zadoff-Chu (ZC) sequence [35] to measure the phase-shift of the signal. We denote the ZC sequence as $zc[n]$ and its DFT as $ZC[n]$, both have length of N_{zc}. The ZC sequence has constant amplitude and optimal auto-correlation so that it is a good candidate for acoustic sensing tasks [36].

Modulation: Algorithm 1 shows our frequency domain modulation scheme. We first generate the DFT of the ZC sequence, $ZC[n]$, and rearrange it as $\widehat{ZC}[n]$ so that the DC (zero frequency) component is at the center of the sequence, i.e., $n = (N_{zc} - 1)/2$. To simplify our notations, we define $h_{zc} = (N_{zc} - 1)/2$. The rearranged $\widehat{ZC}[n]$ is copied to the frequency domain frame buffer $X[n]$, with the DC component of $\widehat{ZC}[n]$ aligned to the carrier frequency, as shown in Fig. 5(b). With a carrier frequency of f_c, a frame length of N, and a sampling frequency of f_s, the carrier frequency is at the frequency point $n_c = N f_c / f_s$ in the frequency domain. We copy the conjugate of $\widehat{ZC}[n]$ to the negative frequency part of $X[n]$ so that the resulting $X[n]$ satisfies the conjugate symmetry conditions for a real-valued signal [34]. The rest parts of the frequency domain frame buffer are set to zero. After an N-point IDFT on $X[n]$, we get the real-valued time signal $x[n]$. The time-domain signal is a periodical signal with a period of N and a bandwidth of $f_s N_{zc} / N$.

Demodulation: The receiver first segments the received signal into frames of length N and then performs an N-point DFT on each segment to get $Y[n]$. To demodulate, the receiver selects the N_{zc} frequency points centered at the carrier frequency f_c and then multiplies them with the conjugate of the ZC template,

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.

Algorithm 2. Demodulation Algorithm

```
Input: Received signal sequence $y[n]$
Output: Channel response sequence $cir[n]$ of length $N'$
        for each frame
1 Segment the received signal into frames with equal length
  of $N$;
2 foreach frame $y[n]$ of length $N$ do
3     Perform $N$-point FFT on $y[n]$ to get $Y[n]$;
4     Conjugate multiplication: $\widehat{CFR}[n] \Leftarrow \widehat{ZC}^*[n] \times$
        $Y[n + n_c - h_{zc}], n = 0, \ldots, 2h_{zc},$;
5     Zero Padding: $CFR[n] \Leftarrow 0, n = 0 \ldots N' - 1,$
        $CFR[n] = \widehat{CFR}[n + h_{zc}], n = 0, \ldots, h_{zc},$
        $CFR[N' - 1 - n] = \widehat{CFR}[n], n = 0, \ldots, h_{zc} - 1;$
6     Perform $N'$-point IFFT on $CFR[n]$ to get $cir[n]$;
7 end
```

$\widehat{ZC}^*[n]$, where * means conjugation. We denote the multiplication result as $\widehat{CFR}[n]$. Since conjugate multiplication in the frequency domain is equivalent to circular cross-convolution of two signals [34], the resulting time domain sequence will be the circular cross-correlation of the ZC sequence with the received data frame. As the auto-correlation of the ZC sequence is a perfect unit impulse function [35], we can get the complex-valued Channel Impulse Response (CIR) by converting $\widehat{CFR}[n]$ back to the time domain by IDFT. We use spectral zero padding on $\widehat{CFR}[n]$ by inserting zeros between the positive frequency and negative frequency to expand the sequence to $N' > N_{zc}$ points to improve the time resolution of the CIR. Our frequency domain modulation/demodulation scheme is equivalent to the scheme used in [36]. However, our scheme has lower computational costs due to efficient DFT/IDFT operations. With a frame length of $N = 3 \times 5 \times 128$, DFT can be calculated through a combination of radix-3, radix-5, and radix-2 Fast Fourier Transforms (FFT), which is supported by most mobile devices.

B. Phase Measurements on OFDM Signals

We use the following theorem to measure the delay of the ranging signal.

Theorem 2. With the modulation/demodulation scheme in Section III-A, if the transmitted signal is delayed by τ, the CIR will be a sinc function with the peak at $m = \tau f_s N'/N$ and the phase of the peak is given by $\varphi = -2\pi\tau f_c$.

Proof. In the frequency domain, the modulated signal $X[n]$ is non-zero only in the neighborhood of the central frequency, i.e., $n = n_c - h_{zc}, \ldots, n_c + h_{zc}$. Using the time shifting theorem, we have

$$\begin{aligned} Y[n + n_c - h_{zc}] &= e^{-j2\pi(n+n_c-h_{zc})\tau f_s/N} X[n + n_c - h_{zc}] \\ &= e^{-j2\pi(n+n_c-h_{zc}))\tau f_s/N} \widehat{ZC}[n]. \qquad (14) \end{aligned}$$

where n is belong to $[0, 2h_{zc}]$. By the property of the Zadoff-Chu sequence, we have $\widehat{ZC}[n] \times \widehat{ZC}^*[n] = 1, \forall n \in [0, 2h_{zc}]$. Therefore, we get

$$\begin{aligned} \widehat{CFR}[n] &= \widehat{ZC}^*[n] \times Y[n + n_c - h_{zc}] \\ &= e^{-j2\pi(n+n_c-h_{zc})\tau f_s/N}, \qquad (15) \end{aligned}$$

for $n \in [0, 2h_{zc}]$. After zero padding and rearranging frequency components (Line 5 in Algorithm 2), we have

$$\begin{aligned} CFR[n] &= R[n] \times e^{-j2\pi(n+n_c)\tau f_s/N} \\ &= R[n] \times e^{-j2\pi n_c \tau f_s/N} \times e^{-j2\pi n\tau f_s/N}, \qquad (16) \end{aligned}$$

where $R[n]$ is the circulate rectangular function centered around sample 0

$$R[n] = \begin{cases} 1 & 0 \le n \le h_{zc}, \\ 1 & N' - h_{zc} \le n \le N' - 1, \\ 0 & h_{zc} < n < N' - h_{zc}. \end{cases} \qquad (17)$$

We observe that the CFR is the product of three components. The first component $R[n]$ is a rectangular function that is a real-valued sinc function with peak at $n = 0$ in the time domain. The second component $e^{-j2\pi n_c \tau f_s/N}$ is a constant phase shift. Since we have $f_c = n_c f_s/N$, this phase shift is equivalent to $e^{-j2\pi\tau f_c}$. The third component $e^{-j2\pi n\tau f_s/N}$ is a phase offset that is linearly related to n. Based on the time shifting theorem, this phase offset is equivalent to a circular shift of $\tau f_s N'/N$ samples in the time domain. Therefore, the resulting time domain $cir[n]$ is a time shifted sinc function with a constant phase offset of $\varphi = -2\pi\tau f_c$.

Theorem 2 shows that delaying the signal by τ has two effects on the CIR. First, the peak of the CIR will be circularly shifted by an offset of $m = \tau f_s N'/N$ sample points. With the offset of the correlation peak, we can get the delay in terms of the number of sample points. For example, if we set $N' = N$, the offset estimation will have a coarse resolution of 7 mm at a sampling rate of 48 kHz. Second, the sinc function is real-valued with a positive peak so that the phase of the peak is equal to φ. As the wavelength of the sound wave is $\lambda_c = c/f_c$ and the delay $\tau = d/c$, we have $\varphi = -2\pi d/\lambda_c$. Therefore, when the path length d changes by a wavelength, the phase φ changes by 2π and measuring the phase leads to sub-millimeter ranging resolutions.

Both the circular shift offset m and the phase φ are linear functions of the delay τ. Therefore, we can use Theorem 1 to directly cancel the unknown system delay and clock offsets in both m and φ, as shown by the following corollaries.

Corollary III.1. Given two devices that have CIR peak offsets of m_{AA}, m_{AB}, m_{BA}, and m_{BB} for the corresponding LOS path, we have

$$\begin{aligned} & m_{AA} + m_{BB} - m_{AB} - m_{BA} \\ &\quad = -\frac{f_s N'(d_{AB} + d_{BA} - d_{AA} - d_{BB})}{cN} \mod N'. \end{aligned}$$

The goal of merging the offset of the LOS path is to get a coarse-grained estimation with an error smaller than the wavelength so that we can further use the phase information to get the fine-grained sub-wavelength resolution using Corollary III.2. To reduce the rounding errors when cancelling at the number of samples, we over-sample the CIR by setting the N' to be four times of the frame length, e.g., $N' = 4N$. This reduces the rounding errors to around 1.75 mm, which is sufficient for

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.

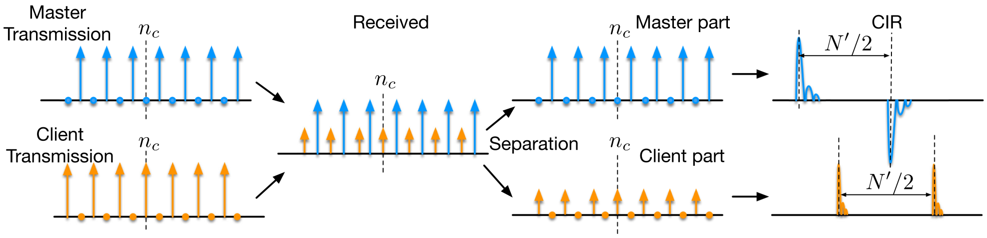


Fig. 6. OFDMA ranging signal design.

a coarse estimation at the wavelength level, which is around 18 mm.

Corollary III.2. Given two devices that have phase measurements of φ_{AA}, φ_{AB}, φ_{BA}, and φ_{BB} on the LOS path, we have

$$\begin{aligned}&\varphi_{AA} + \varphi_{BB} - \varphi_{AB} - \varphi_{BA}\\&\quad = 2\pi \frac{d_{AB} + d_{BA} - d_{AA} - d_{BB}}{\lambda_c} \mod 2\pi,\end{aligned}$$

where λ_c is the wavelength of the central frequency f_c.

In practice, our cancellation scheme has the capability to remove the *non-linear* and *dynamic* clock offsets as predicted by our theoretical timing model. Fig. 2(c) shows the delay measured by the phase of CIR for a pair of phones. The non-linear clock offsets incur a time deviation of more than 0.042 ms in both τ_{AB} and τ_{BA} in twenty seconds. We observe that our cancellation scheme successfully removes the non-linear offset to produce a constant delay, where the deviation is reduced by more than 160 times to 0.26 μs (0.09 mm in distance).

IV. System Design

In this section, we present how SCALAR coordinates multiple devices to transmit at the *same time* in the *same frequency* to enable precise cancellation of the clock offset.

A. OFDMA Ranging

Our Orthogonal Frequency-Division Multiple Access (OFDMA) ranging scheme allows multiple devices to send at the same central frequency so that each device can measure the offset m and phase φ of different transmitting sources using the *same* audio frame. We use a two-device system as an example, where one device acts as the master device that coordinates the calibration process. The other device, the client device, should demodulate the received audio signal and feedback the measured m and φ to the master through other communication channels, e.g., a Wi-Fi or a Bluetooth channel. Each feedback will be tagged with a timestamp so that the master can merge measurements from both devices taken within our delay constraint of 40 ms.

We allocate interleaved subcarriers to the master and the client as shown in Fig. 6. When transmitting, the master transmits only in odd OFDM subcarriers and sets even subcarriers to zero. Similarly, the client uses even subcarriers only. Thus, the receiver can separate the transmissions from the master and the client by gathering the odd/even subcarriers from a single frame. With a frame length of 960 and a sampling frequency of 48 kHz, the frequency interval between neighboring subcarriers is 50 Hz. Given a clock skew that leads to a frequency offset of less than five Hertz, the interference between neighboring subcarriers is small.

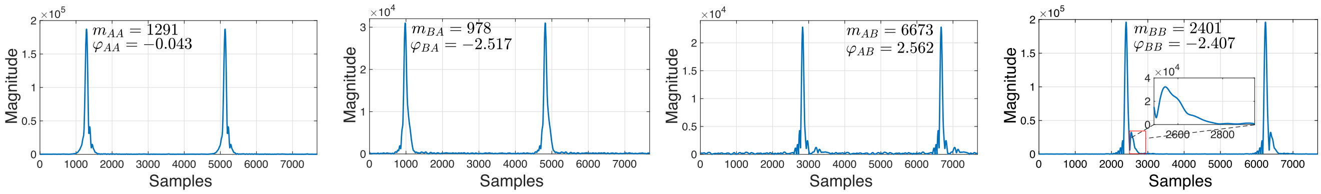
The key advantage of using interleaved subcarriers for different devices is that this scheme keeps the same ranging resolution as if all devices are using the full bandwidth. To understand this, we observe that the signal of the client device is the original full bandwidth signal (with both even and odd subcarriers) multiplied by a discrete impulse train

$$I[n] = \begin{cases} 1 & n \bmod 2 = 0, \\ 0 & n \bmod 2 = 1, \end{cases} \tag{18}$$

in the frequency domain. Multiplication with $I[n]$ in the frequency domain is equivalent to convolution with the IDFT of $I[n]$ in the time domain. As the N'-point IDFT of $I[n]$ has only two non-zero points at $n = 0$ and $n = N'/2$, both of which have a value of 1, the convolution leads to two identical copies of peaks that are separated by $N'/2$ samples. Therefore, using only the even-subcarriers, the correlation peaks of the CIR will have the same width as the full bandwidth signal, which gives the same range resolution as using the full bandwidth. However, this scheme reduces the unambiguous round-trip range of our system by half, e.g., to 3.4 meters when using a 20 ms frame length. Fig. 7 shows the CIR measurements of two devices, where A is the master device. We can get four offsets and four phases from CIR measurements of A and B and derive the distance using Corollaries III.1 and III.2.

The master uses odd subcarriers in the OFDM signal, where the impulse train $I[n]$ is shifted by one point in the frequency domain. This will lead to a phase shift in the time domain [34], which adds an extra phase shift of π to the second correlation peak in the CIR. To help the client identify the right peak, the master first sends in the full-bandwidth, i.e., using all the subcarriers. With full-bandwidth transmissions, the CIR of the master will only have a single correlation peak which corresponds to the first peak of the odd subcarriers. After determining the offset of the right peak, the client will notify the master so that the master switches to the odd subcarriers. In this way, both devices can determine which correlation peak is the first peak for the master by comparing the measured CIR before and after the switching action. For the client signal, we can use the phase of either the first or the second peak, since both peaks have the same phase for even subcarriers.

To summarize, our ranging scheme has the following steps:



(a) A → A (b) B → A (c) A → B (d) B → B

Fig. 7. The magnitude of CIR measurements for a pair of devices (with four combinations of sender → receiver).

1. The master device first sends in all subcarriers and notifies the client device using other wireless channels.

2. Both the master and the client record the master transmission and determine the position of the LOS path.

3. The client notifies the master by other wireless channels.

4. The client starts transmission in even subcarriers and the master only transmits in the odd subcarriers.

5. The client measures m_{AB}, m_{BB}, φ_{AB}, and φ_{BB}, then transmits the result and timestamps to the master.

6. The master merges the result from the client to derive the fine-grained distance based on Corollaries III.1 and III.2.

B. Discussions and Limitations

Our ranging scheme has the following limitations.

LOS Transmission: Our system is based on the assumption that the LOS path exists. When the LOS path is blocked, the phase measurements and peak offsets may have large errors. For example, when we put two devices at a distance of 60 cm, the ranging error is around 0.15 mm. However, when we placed a thick sweater between the devices, the ranging error suddenly increased to 31.42 cm. Our system can detect whether the LOS is blocked or not by comparing the energy of peaks. When the energy of the first correlation peak is no longer larger than other peaks, we determine that the LOS path is blocked and mark the ranging results as unreliable. In case that we have multiple speakers and microphones that form a virtual array, we have redundant measurements so that the target can be localized even if some of the LOS paths are blocked, see Section V.

Multipath Effects: When the sound travels through multiple paths to reach the microphone, the resulting CIR is the linear combination of circularly shifted *sinc* function, i.e., $sinc(x) = \sin(x)/x$, since both the modulation/demodulation operations and the channel are linear. Fig. 7 shows examples of real-world CIR measurements. The LOS path corresponds to the highest peak and other small peaks are paths reflected by nearby objects, e.g., the second small peak in Fig. 7(d). When the LOS exists, its peak can be easily separated from multipath and the phase of the LOS path in CIR is not affected by paths that are more than 30 cm away in our system. Thus, the distance measurement is robust to surrounding objects that are more than 30 cm to the LOS. In our real-world experiments, we found that the phase measurement φ is more robust to interference than the correlation offset m. This coincides with the observations in MilliSonic [7]. Thus, most of the measurement errors come from the correlation offset m. To mitigate such multipath interference, we propose to use regression algorithms based on correlation parameters, e.g., the width and symmetry of the peak, to further compensate for the multipath effect. However, as our current scheme can handle most multipath conditions, we leave the study of the regression algorithm as our future work.

More Than Two Acoustic Links: When there are more than two pairs of speakers and microphones, we can measure the distances pair-by-pair in a time division multiplex manner. Moreover, we can also use the OFMDA scheme to allow more than two speakers to transmit at the same time. For example, for four speakers, we can allocate one fourth of the subcarriers to each speaker, i.e., speaker with an id of k transmits at subcarriers that has $n \bmod 4 = k$. The receiving microphones can separate the four transmissions in the frequency domain and measure the CIR of four speakers within a single frame. In this case, the correlation peaks in the CIR is duplicated by four times and the unambiguous range is reduced to a quarter of the frame length.

V. Evaluations

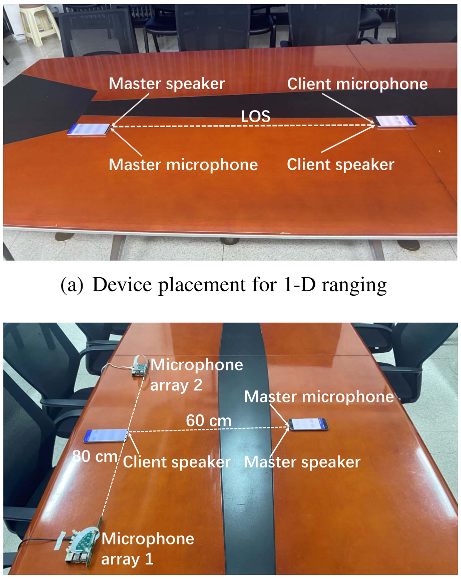
A. Implementation

We have implemented SCALAR on iOS, Android, and Linux systems. SCALAR operates at a central frequency of 19 kHz, occupies a bandwidth of 4 kHz (with $N_{zc} = 81$), and uses a frame length of 20 ms if not specified. The sound signals transmitted by SCALAR are inaudible to most users. We use the same set of parameters on all platforms so that devices with different operating systems are compatible with each other. For all three platforms, we develop stand-alone applications that perform the signal processing and the ranging algorithm in real-time. And for data virtualization and quick demo implementation, we also develop Python and MATLAB codes to receive the signal from different devices through networks and process in real-time. We use two smartphones (Samsung Galaxy S10 and S7), serving as the master and client devices, respectively, to evaluate the 1-D ranging performance, as shown in Fig. 8(a). We use multiple distributed devices, including two smartphones and two ReSpeaker 4-mic linear arrays mounted on Raspberry Pi 3B+, to evaluate the performance of our virtual array, as shown in Fig. 8(b).

B. Evaluation on 1-D Ranging

SCALAR achieves an average 1-D ranging error of less than 0.39 mm within a distance of 3 meters. Fig. 9(a) shows the ranging error of SCALAR when compared with a vernier with a

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.



(a) Device placement for 1-D ranging

(b) Device placement for 2-D localization

Fig. 8. Experimental setup.

resolution of 0.01 mm. Our ranging experiments are conducted in an indoor environment with a homogeneous temperature of $15 \sim 19\,^{\circ}$C. We moved one of the devices (Samsung S7) by a high-resolution moving platform with a step size of 0.5 mm for one centimeter in each set of experiment. Due to the limited range of the vernier, five sets of experiments were performed in the range of $30 \sim 50$ cm. The results show that the average ranging error is smaller than 0.1 mm when compared to the vernier. Fig. 9(b) shows the box plot of the measurement error when the devices are separated by a longer distance and the ground truth is obtained by a ruler. The unambiguous single-trip range of our default system setting is 6.86 meters (i.e., frame size of 1920) so that we set the maximum distance as 5 meters. The average errors are less than 0.13 mm within one meter and 0.39 mm within three meters, and the 75th percentile errors are 0.22 mm and 0.55 mm, respectively. Unfortunately, when the distance exceeds four meters, the accuracy drops significantly to millimeter-level due to the decrease in signal-to-noise ratio (SNR) and strong multipath effect. The average errors at distances of four and five meters are 0.99 mm and 3.05 mm, respectively. Note that although the excellent auto-correlation property of our OFDM signal has the advantage of identifying the LOS path, when the SNR decreases the peak value regarding the LOS path may reduce significantly, while that regarding the multipath increases, resulting in a large error in peak measurement. Additionally, there is a small number of outliers with an error close to half-wavelength beyond a distance of three meters, which is mainly due to the error of our coarse-grained correlation when the signal is weak at long distances.

SCALAR achieves an average ranging error of less than 0.2 mm for different types of devices at a distance of 60 cm. Fig. 10(a) shows the error of different types of mobile devices when placed at a distance of 60 cm, including iPhone 6S, Huawei Nova6, and Samsung S7. We observe that different types of devices have a similar performance, where the average errors are between 0.09 mm and 0.2 mm. SCALAR also works well when the two devices are from different vendors, e.g., using Huawei Nova6 as the sender and Samsung S7 as the receiver. Furthermore, based on the above setup, we measured the frequency response of SCALAR in the frequency range of $17-21\ kHz$ to explore the sensor sensitivity of each pair of devices. Fig. 11(a) shows that all device pairs have strong signal transmitting/receiving capability over the bandwidth occupied by the system. Compared with the other four device pairs, the Samsung-Samsung pair has a more consistent frequency response and achieves the best sensing performance in consequence, as indicated by Fig. 10(a).

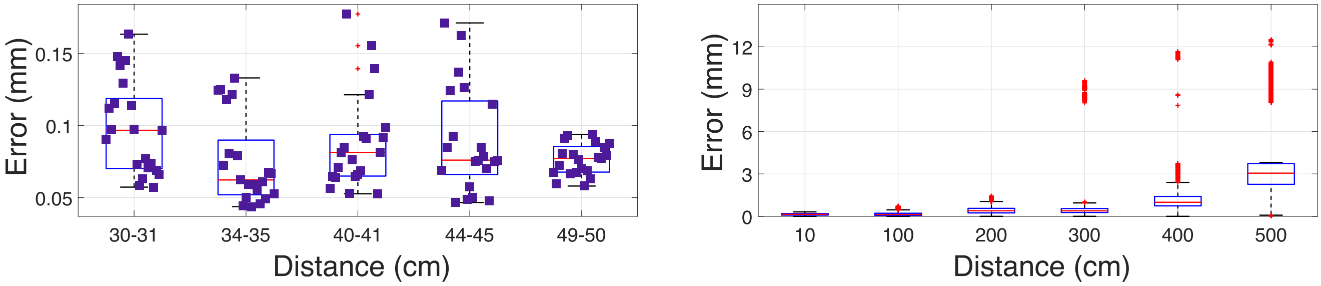
SCALAR is robust to noises and achieves an average distance error of less than 0.15 mm at a distance of 60 cm under different types of noises. Fig. 10(b) shows the performance when there are environmental noises around the devices. We observe that surrounding human speeches and musics only slightly increase the localization error from 0.097 mm to 0.114 mm and 0.105 mm, respectively. To explore the noise distribution of microphones on different devices, we placed the above devices in the office, where there will be speaking voices and sound of keystrokes, and recorded the sound in the frequency range from 0 to 24 kHz. Fig. 11(b) shows that the energy of the environmental noise is mainly below 4 kHz. In the frequency band occupied by our system, the noise is almost negligible, which is consistent with the result in Fig. 10(b).

SCALAR is robust to multipath inferences and achieves an average error of less than 0.25 mm at a distance of 60 cm under multipath conditions. Fig. 10(c) shows the performance under different multipath conditions. We place objects with a size of $17 \times 24 \times 5.5$ cm at a distance of 30 cm and 40 cm to the LOS path to introduce static multipaths. We also asked other users to walk around the devices at a distance of one meter to introduce dynamical multipaths. We observe that multipaths only slightly increase the average error to 0.24 mm, 0.13 mm, and 0.14 mm for static multipath at 30 cm, 40 cm, and dynamic multipath conditions, respectively.

SCALAR is robust against long-term clock drift. As introduced in the previous sections, SCALAR can cancel the clock drift to provide results that are stable in the long-term. We place two phones at a distance of 50 cm, where the distance is measured by a combination of ruler and vernier. We continuously run SCALAR for 30 minutes. The result shows that our system can keep a stable sub-millimeter 1-D ranging accuracy for more than 30 minutes, as shown in Fig. 13. The 75th percentile errors for the six time slots of five-minutes are 0.20 mm, 0.38 mm, 0.37 mm, 0.30 mm, 0.26 mm, and 0.27 mm. We observe that the errors of SCALAR do not exhibit an increasing trend as existing systems such as MilliSonic, whose accuracy will drop to around 2.5 mm in ten minutes [7].

SCALAR achieves a high ranging accuracy with up to ten devices. In this experiment, we set the number of devices, M, as 2, 4, 6, 8, 10, 12, to evaluate the 1-D ranging performance. Note that each of M devices occupies the full-bandwidth while

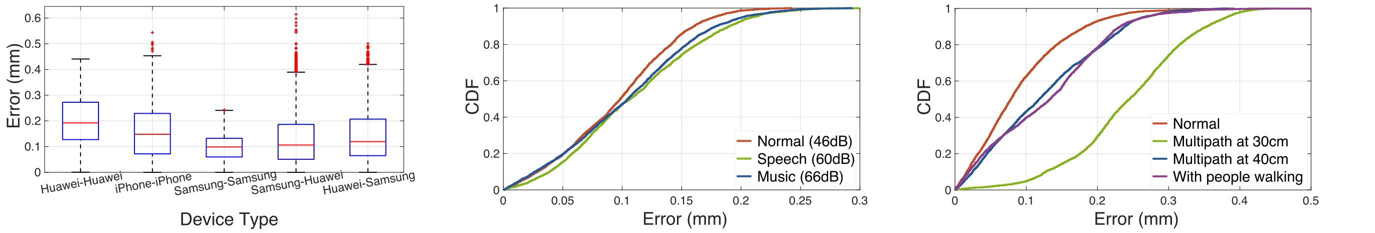
Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.



(a) Ranging error vs. verniers

(b) Ranging error distribution

Fig. 9. Ranging error distribution at different distances between devices.

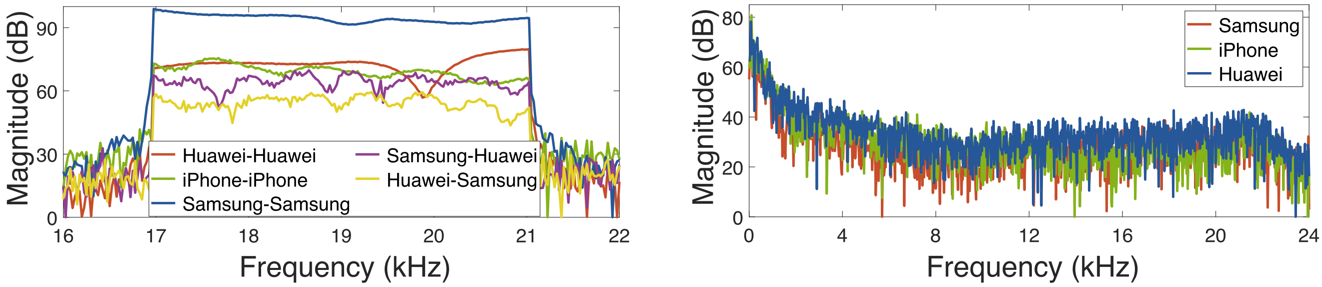


(a) Different types of devices

(b) Different noise levels

(c) Multipath environments

Fig. 10. Robustness of distance measurements.



(a) Frequency response of different devices

(b) Noise distribution of microphones on different types of devices

Fig. 11. Sensitivity of different types of devices.

assigned with different subcarriers according to the OFDMA scheme illustrated in Section IV-A. We placed the two target mobile devices at a distance of 50 cm with no strict placement requirements for the other involved devices. The results are shown in Fig. 12 , where we can observe that SCALAR keeps the high ranging accuracy with $M \leq 10$. However, when M is greater than 10, there will be a drastic decrease in the ranging accuracy. This is mainly because the correlation property is severely weakened when M is extremely large. Based on the above results, SCALAR can support a larger number of devices (i.e., $M \geq 12$) with the following two approaches. First, we can use one device with two transmitting channels and one receiving channel then we can use a large number of passive receiving devices to get the accurate localization result, according to our timing model. Second, we can use different frequency bands to double the transmitting devices without damaging the correlation property.

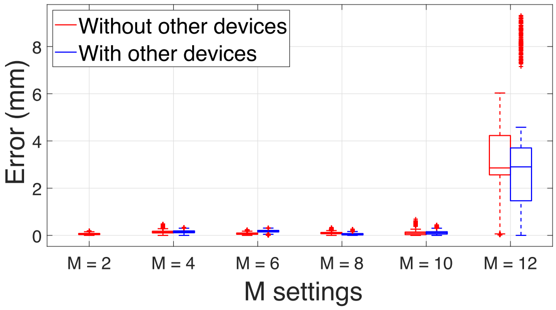


Fig. 12. 1D localization errors with different M settings.

C. Evaluation on 2-D Localization With Two Devices

For 2-D localization, we use a Samsung S7 mobile phone as the master device and a ReSpeaker 4-mic linear array on

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.

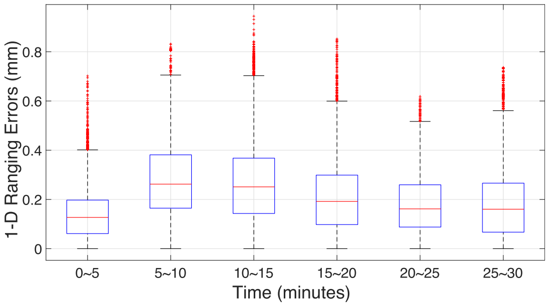
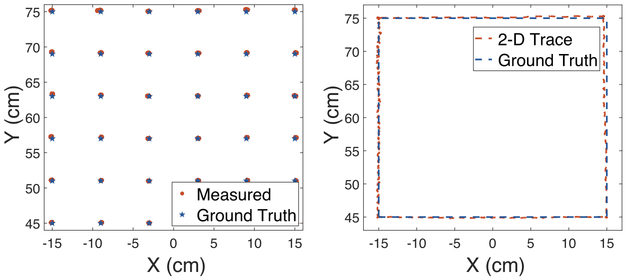


Fig. 13. Long-term clock drift of 1-D ranging.



(a) 2-D localization accuracy (b) Tracking example

Fig. 14. Performance of 2-D localization and tracking.

Raspberry Pi that runs Linux as the client device. We localize the master device using two microphones on the ReSpeaker that are separated by 15 cm.

SCALAR can localize targets in the 2-D plane with an average error of 1.71 mm. Fig. 14(a) shows the localization error in a 30 × 30 cm square with the center at 60 cm to the microphone array. The ground truth locations are marked on the table as a grid. We repeat the measurement for 100 times for each grid point and the measured results are shown as red points in the figure. Note that our localization is performed by directly placing the device on grid points without moving the device or using historical traces.

SCALAR can track moving objects in the 2-D plane with an error of 1.45 mm. Fig. 14(b) shows the trace of moving the mobile phone along a square trajectory with an average speed of 18 cm/s, which is close to hand movement speeds [24]. The average tracking error is 1.45 mm and the maximum error is 5.2 mm. This demonstrates that SCALAR also achieves an excellent tracking performance comparable to state-of-the-art, i.e., MilliSonic. However, benefiting from the advantage of distributed ranging, we believe SCALAR will act as an attractive alternative to the traditional tracking system which is based on the distance change for long-term tracking.

D. Evaluation on Distributed Virtual Array

To compare the performance of the distributed virtual array with the two-device solution, we deploy two ReSpeaker 4-mic linear arrays driven by two separate Raspberry Pis with a smartphone (the array smartphone) to form a virtual array as shown in Fig. 8(b). The two ReSpeaker arrays are separated by a distance of 80 cm. We transmit ranging signals from both the array smartphone and the target smartphone, while receiving from the two ReSpeaker arrays and the target smartphone. Therefore, our virtual array system has two transmitting speakers and nine receiving microphones on four separate devices. The microphones within the same ReSpeaker array are synchronized and we use SCALAR to synchronize the signal between separated devices. As discussed in Section II-C, our virtual array provides redundant measurements so that we can either use regression schemes on redundant measurements to improve localization accuracy or localize the target in a higher dimension.

SCALAR can reduce the localization error from 1.71 mm to 1.07 mm by using the virtual array solution. We compare our virtual array performance with the two-device deployment as in Section V-C. As shown in Fig. 15(a), when only using one microphone on each of the ReSpeaker arrays, the average localization error is reduced to 1.36 mm as the baseline (distance between microphones) of the 2-D localization array is increased. If we use all eight microphones on the two ReSpeaker arrays, the localization error can be further reduced to 1.07 mm by combining more measurements. Furthermore, when we block the LOS path of one of the ReSpeaker arrays, the virtual array will automatically switch to the remaining microphones. The virtual array is robust to such partial LOS path blockage and maintains an average localization error of 1.14 mm, which is worse than using all 8 microphones but better than the 2-mic scenario. This greatly improves the usability of the system, as the chance of full blockage of all the distributed devices is small.

SCALAR can locate static targets in 3-D space with an average error of 1.86 mm. We use the similar virtual array deployment as in the 2-D plane to further evaluate the performance of SCALAR for 3-D localization. The target phone is placed on a platform with adjustable heights and the ground truth of the location is measured through predefined grid points on the platform and the height of the platform. We perform the measurements on 27 grid points within a $15 \times 15 \times 15\ cm$ cube with 100 repetitions on each grid point. Fig. 15(b) shows the localization result distribution where the average error is 1.86 mm.

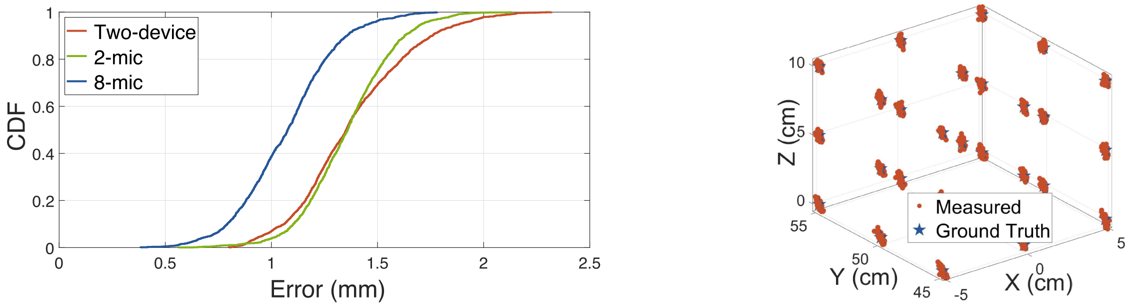
E. System Performance

TABLE I
SYSTEM PERFORMANCE ON DIFFERENT DEVICES

Device	Signal processing (ms)	Fusion (ms)	Battery life (h)
iPhone 6S	0.185±0.01	0.0014±0.0001	3.4
Samsung S7	8.654±1.995	0.0082±0.0034	4.7
Raspberry Pi 3B+	14.570±1.656	2.0423±0.6224	-

SCALAR can process the audio signal in real-time on commercial iOS, Android mobile phones and Raspberry Pi. Table I shows the time for SCALAR to process one audio frame with a duration of 20 ms on different devices. The signal processing delay includes the procedures of segmenting the raw signal,

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.



(a) 2-D localization performance of the virtual array

(b) 3-D localization error distribution

Fig. 15. Localization performance on distributed virtual array.

TABLE II
RECENT WORKS ON ACOUSTIC DEVICE RANGING

System	Setup	User intervention	Dimension	Accuracy	Long-term drift errors	Latency	Refresh rate	Range	Audible
BeepBeep [21]	Phone-Phone	N	1-D	2 cm	Not Given	50 ms	20 Hz	12 m	Y
Swordfight [49]	Phone-Phone	N	1-D	2 cm	Not Given	46 ms	12 Hz	3 m	Y
CAT [18]	Speaker-Phone	Y	3-D	9 mm	10 mm error in 10 min	40 ms	25 Hz	7 m	N
SoundTrak [1]	Speaker-Watch	Y	3-D	13 mm	4 mm error in 10 min	12 ms	86 Hz	20 cm	Y
Sonoloc [56]	Phone-Phone	Y	2-D	6 cm	Not Given	3.2∼48 s	< 0.3 Hz	17 m	Y
MilliSonic [7]	Mic-Phone	Y	1-D	0.7 mm[1]	2 mm error in 10 min	15∼30 ms	40 Hz	3 m	N
			3-D	2.6 mm		25∼40 ms			
SCALAR	**Speaker-Phone**	**N**	**1-D**	**0.39** mm	**Long-term stable**	**1∼14** ms	**50** Hz[2]	**3** m	**N**
			3-D	**1.86** mm		**26∼36** ms			

[1] MilliSonic reports this result within 1 m range and their median errors increase to 1.74 mm in the range of $1 \sim 2$ m.

[2] We can adjust the refresh rate using different frame size to satisfy the application's requirements.

demodulation, and extracting the peak offsets and phases. The fusion delay is the process for the master device to merge the measurements and output the distance. With the vDSP acceleration framework on iOS, SCALAR incurs negligible computational cost with a CPU load of less than 0.5% on iPhone 6S. In Android implementation, we use Java Native Interface (JNI) to process the audio signal in C language. It also meets the real-time processing requirements, but is an order of magnitude slower than the vDSP version. Even on resource extremely limited devices like Raspberry Pi, our algorithm can still process the data in real-time with Python.

SCALAR returns the accurate measurement within 411 ms in cold start. We repeated the cold start process for more than one hundred times on a pair of Samsung S7. On average, the audio signals take 2.94 frames (118 ms) to stabilize for reliable offset and phase measurements. Furthermore, it takes 293 ms for the client to notify the master to switch to the odd subcarriers through Wi-Fi. This gives an overall cold start latency of 411 ms.

SCALAR has a power consumption of less than 500 mW on the Android platform. We use Powertutor [37] to measure the power consumption on Samsung S7. The average power consumption for CPU and Audio are 115.6 mW and 384.3 mW when SCALAR is operating continuously. In addition, we measured the battery life during continuous execution of SCALAR on two mobile phones, i.e., Samsung S7 and iPhone 6S. During the experiments, the phone's screen was on. As shown in Table I, the battery lasts 4.7 h and 3.4 h on the Samsung S7 and iPhone 6S, respectively, which indicates that SCALAR consumes very little energy and is promising for practical deployment in the future.

VI. RELATED WORK

Recent works related to SCALAR are in the following four categories.

Device-Free Acoustic Ranging. Device-free acoustic ranging systems use sound signals reflected by a moving object to track the target. As signals are transmitted and received by the same device and these systems, they are synchronized by the hardware and no calibration is needed [23], [24], [38]. Synchronized device-free systems have been widely used for gesture recognition [23], [39], [40], vital sign monitoring [41], [42], [43], and localization [44], [45]. In addition to in-air acoustic propagation, synchronized systems can also measure the structure-borne sound to localize objects on solid surfaces with centimeter-level accuracy [36], [46], [47], [48]. As physical connections are required for synchronized operation, these systems cannot be deployed in a distributed way.

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.

Device-Based Acoustic Ranging. Device-based ranging systems localize a target device that is actively transmitting/receiving sound waves [1], [30], [49]. Most device-based ranging systems use the phase-based approach to improve tracking accuracy. CAT develops a distributed FMCW system to achieve a tracking accuracy of 5 mm [18]. Vernier uses an efficient phase-change estimation algorithm to track a moving device with an accuracy of 4 mm [19], [50]. MilliSonic provides sub-millimeter tracking accuracy by combining the correlation-based and phase-based ranging [7]. However, these acoustic ranging systems require a calibration process to remove the unknown clock offset [7], [18], [50], [51] or have to use synchronized sources to localize the target [1], [52], [53]. This additional calibration requirement limits their applications to tracking the relative movement instead of measuring the true distance between devices [11], [54]. The recent works on acoustic ranging are summarized in Table II.

Calibration Schemes. To synchronize distributed devices, Cricket uses Radio Frequency (RF) transmissions to calibrate the audio system [20]. BeepBeep measures round-trip delay to reduce the impact of system delay [21]. However, these calibration schemes do not consider the sampling clock drift. Thus, their accuracy is limited to the centimeter-level [21], [55], which is an order-of-magnitude larger than the millimeter-level tracking accuracy achieved by phase-based schemes. Phase-based calibration has been applied in RF-based ranging to achieve centimeter-level accuracy [32]. However, such systems are susceptible to multipath effects as they measure a single frequency at a time.

VII. Conclusion

In this paper, we developed SCALAR, a fine-grained calibration scheme for acoustic ranging systems on distributed mobile devices. Our solution uses existing low-cost hardware on mobile devices to achieve repeatable sub-millimeter level ranging accuracy. By building a virtual acoustic array with the synchronization capability provided by SCALAR, we can perform reliable 2-D/3-D localization with millimeter level accuracy.

References

[1] C. Zhang et al., “SoundTrak: Continuous 3D tracking of a finger using active acoustics,” in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2017, Art. no. 30.

[2] J. Wang, D. Vasisht, and D. Katabi, “RF-IDraw: Virtual touch screen in the air using RF signals,” in *Proc. ACM Conf. SIGCOMM*, 2014, pp. 235–246.

[3] Y. Zhuang, Y. Wang, Y. Yan, X. Xu, and Y. Shi, “ReflecTrack: Enabling 3D acoustic position tracking using commodity dual-microphone smartphones,” in *Proc. 34th Annu. ACM Symp. User Interface Softw. Technol.*, 2021, pp. 1050–1062.

[4] Y. Wang et al., “FaceOri: Tracking head position and orientation using ultrasonic ranging on earphones,” in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2022, Art. no. 290.

[5] N. Xiao, P. Yang, X.-Y. Li, Y. Zhang, Y. Yan, and H. Zhou, “MilliBack: Real-time plug-n-play millimeter level tracking using wireless backscattering,” in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2019, Art. no. 112.

[6] H. Zijun, Z. Lu, X. Wen, L. Guo, and J. Zhao, “Towards 3D centimeter-level passive gesture tracking with two WiFi links,” *IEEE Trans. Mobile Comput.*, early access, Oct. 29, 2021, doi: 10.1109/TMC.2021.3123694.

[7] A. Wang and S. Gollakota, “MilliSonic: Pushing the limits of acoustic motion tracking,” in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, Art. no. 18.

[8] W. Mao, W. Sun, M. Wang, and L. Qiu, “DeepRange: Acoustic ranging via deep learning,” in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2021, Art. no. 143.

[9] W. Mao, M. Wang, W. Sun, L. Qiu, S. Pradhan, and Y.-C. Chen, “RNN-based room scale hand motion tracking,” in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, 2019, Art. no. 38.

[10] H. Cheng and W. Lou, “Push the limit of device-free acoustic sensing on commercial mobile devices,” in *Proc. IEEE Conf. Comput. Commun.*, 2021, pp. 1–10.

[11] W. Mao, Z. Zhang, L. Qiu, J. He, Y. Cui, and S. Yun, “Indoor follow me drone,” in *Proc. 15th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2017, pp. 345–358.

[12] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, “SpotFi: Decimeter level localization using WiFi,” in *Proc. ACM Conf. Special Int. Group Data Commun.*, 2015, pp. 269–282.

[13] D. Vasisht, S. Kumar, and D. Katabi, “Decimeter-level localization with a single WiFi access point,” in *Proc. 13th Usenix Conf. Netw. Syst. Des. Implementation*, 2016, pp. 165–178.

[14] J. Xiong and K. Jamieson, “ArrayTrack: A fine-grained indoor location system,” in *Proc. 10th USENIX Conf. Netw. Syst. Des. Implementation*, 2013, pp. 71–84.

[15] R. Ayyalasomayajula, D. Vasisht, and D. Bharadia, “BLoc: CSI-based accurate localization for BLE tags,” in *Proc. 14th Int. Conf. Emerg. Netw. EXperiments Technol.*, 2018, pp. 126–138.

[16] Y. Xie, J. Xiong, M. Li, and K. Jamieson, “mD-track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking,” in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, 2019, Art. no. 8.

[17] H. Ding et al., “Trio: Utilizing tag interference for refined localization of passive RFID,” in *Proc. IEEE Conf. Comput. Commun.*, 2018, pp. 828–836.

[18] W. Mao, J. He, and L. Qiu, “CAT: High-precision acoustic motion tracking,” in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 69–81.

[19] Y. Liu et al., “Vernier: Accurate and fast acoustic motion tracking using mobile devices,” *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 754–764, Feb. 2021.

[20] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, “The cricket location-support system,” in *Proc. 6th Annu. Int. Conf. Mobile Comput. Netw.*, 2000, pp. 32–43.

[21] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, “BeepBeep: A high accuracy acoustic ranging system using COTS mobile devices,” in *Proc. 5th Int. Conf. Embedded Netw. Sensor Syst.*, 2007, pp. 1–14.

[22] J. Qiu, D. Chu, X. Meng, and T. Moscibroda, “On the feasibility of real-time phone-to-phone 3D localization,” in *Proc. 9th ACM Conf. Embedded Netw. Sensor Syst.*, 2011, pp. 190–203.

[23] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, “FingerIO: Using active sonar for fine-grained finger tracking,” in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2016, pp. 1515–1525.

[24] W. Wang, A. X. Liu, and K. Sun, “Device-free gesture tracking using acoustic signals,” in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 82–94.

[25] Z. Yang and R. R. Choudhury, “Personalizing head related transfer functions for earables,” in *Proc. ACM SIGCOMM Conf.*, 2021, pp. 137–150.

[26] C. Cai, R. Zheng, and J. Luo, “Ubiquitous acoustic sensing on commodity IoT devices: A survey,” *IEEE Commun. Surveys Tut.*, vol. 24, no. 1, pp. 432–454, First Quarter 2022.

[27] R. Jin, C. Cai, T. Deng, Q. Li, and R. Zheng, “MotionBeep: Enabling fitness game for collocated players with acoustic-enabled IoT devices,” *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10755–10765, Jul. 2021.

[28] C. Cai, M. Hu, X. Ma, K. Peng, and J. Liu, “Accurate ranging on acoustic-enabled IoT devices,” *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3164–3174, Apr. 2019.

[29] I. E. Commission, “Digital audio interface–Part1: General,” IEC 60958–1, 2014.

[30] S. Yun, Y.-C. Chen, and L. Qiu, “Turning a mobile device into a mouse in the air,” in *Proc. 13th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2015, pp. 15–29.

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.

[31] FLIR E5-XT infraed camera, 2015. [Online]. Available: https://www.flir.com/products/e5-xt/

[32] M. Maróti et al., “Radio interferometric geolocation,” in *Proc. 3rd Int. Conf. Embedded Netw. Sensor Syst.*, 2005, pp. 1–12.

[33] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[34] A. V. Oppenheim, A. S. Willsky, and S. Hamid, *Signals and Systems*. London, U.K.: Pearson, 1996.

[35] B. M. Popovic, “Generalized chirp-like polyphase sequences with optimum correlation properties,” *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1406–1409, Jul. 1992.

[36] K. Sun, T. Zhao, W. Wang, and L. Xie, “VSkin: Sensing touch gestures on surfaces of mobile devices using acoustic signals,” in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 591–605.

[37] L. Zhang et al., “Accurate online power estimation and automatic battery behavior based power model generation for smartphones,” in *Proc. IEEE/ACM/IFIP Int. Conf. Hardware/Softw. Codesign Syst. Synth.*, 2010, pp. 105–114.

[38] S. Yun, Y.-C. Chen, H. Zheng, L. Qiu, and W. Mao, “Strata: Fine-grained acoustic-based device-free tracking,” in *Proc. 15th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2017, pp. 15–28.

[39] W. Ruan, Q. Z. Sheng, L. Yang, T. Gu, P. Xu, and L. Shangguan, “AudioGest: Enabling fine-grained hand gesture detection by decoding echo signal,” in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 474–485.

[40] K. Sun, W. Wang, A. X. Liu, and H. Dai, “Depth aware finger tapping on virutal displays,” in *Proc. 16th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2018, pp. 283–295.

[41] R. Nandakumar, S. Gollakota, and N. Watson, “Contactless sleep apnea detection on smartphones,” in *Proc. 13th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2015, pp. 45–57.

[42] T. Wang, D. Zhang, Y. Zheng, T. Gu, X. Zhou, and B. Dorizzi, “C-FMCW based contactless respiration detection using acoustic signal,” in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2018, Art. no. 170.

[43] A. Wang, J. E. Sunshine, and S. Gollakota, “Contactless infant monitoring using white noise,” in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, 2019, Art. no. 52.

[44] B. Zhou, M. Elbadry, R. Gao, and F. Ye, “BatMapper: Acoustic sensing based indoor floor plan construction using smartphones,” in *Proc. 15th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2017, pp. 42–55.

[45] Y.-C. Tung and K. G. Shin, “EchoTag: Accurate infrastructure-free indoor location tagging with smartphones,” in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 525–536.

[46] Y. C. Tung and K. G. Shin, “Expansion of human-phone interface by sensing structure-borne sound propagation,” in *Proc. 14th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2016, pp. 277–289.

[47] J. Liu, C. Wang, Y. Chen, and N. Saxena, “VibWrite: Towards finger-input authentication on ubiquitous surfaces via physical vibration,” in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2017, pp. 73–87.

[48] J. Liu, Y. Chen, M. Gruteser, and Y. Wang, “VibSense: Sensing touches on ubiquitous surfaces through vibration,” in *Proc. IEEE 14th Annu. Int. Conf. Sens. Commun. Netw.*, 2017, pp. 1–9.

[49] Z. Zhang, D. Chu, X. Chen, and T. Moscibroda, “SwordFight: Enabling a new class of phone-to-phone action games on commodity phones,” in *Proc. 10th Int. Conf. Mobile Syst. Appl. Serv.*, 2012, pp. 1–14.

[50] Y. Zhang, J. Wang, W. Wang, Z. Wang, and Y. Liu, “Vernier: Accurate and fast acoustic motion tracking using mobile devices,” in *Proc. IEEE Conf. Comput. Commun.*, 2018, pp. 1709–1717.

[51] G. Cao et al., “EarphoneTrack: Involving earphones into the ecosystem of acoustic motion tracking,” in *Proc. 18th Conf. Embedded Netw. Sensor Syst.*, 2020, pp. 95–108.

[52] Q. Lin, Z. An, and L. Yang, “Rebooting ultrasonic positioning systems for ultrasound-incapable smart devices,” in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, 2019, Art. no. 2.

[53] J. Yang et al., “Detecting driver phone use leveraging car speakers,” in *Proc. 17th Annu. Int. Conf. Mobile Comput. Netw.*, 2011, pp. 97–108.

[54] H. Zhang, W. Du, P. Zhou, M. Li, and P. Mohapatra, “DopEnc: Acoustic-based encounter profiling using smartphones,” in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 294–307.

[55] P. Lazik, N. Rajagopal, B. Sinopoli, and A. Rowe, “Ultrasonic time synchronization and ranging on smartphones,” in *Proc. IEEE Real-Time Embedded Technol. Appl. Symp.*, 2015, pp. 108–118.

[56] V. Erdélyi, T.-K. Le, B. Bhattacharjee, P. Druschel, and N. Ono, “Sonoloc: Scalable positioning of commodity mobile devices,” in *Proc. 16th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2018, pp. 136–149.

Lei Wang received the PhD degree from the Department of Computer Science and Technology, Nanjing University, China, in April 2020. He has been invited as a joint PhD student by University of New South Wales, Australia, from October 2018 to October 2019. He was a postdoctoral assistant researcher with the School of Computer Science, Peking University, China, from September 2020 to November 2022. He is currently a distinguished associate professor with Soochow University, China. His research interests are in the areas of pervasive computing, wireless sensing.

Haoran Wan received the BS degree from the School of Information and Communication, University of Electronic Science and Technology of China. He is currently working toward the third year master’s degree with the Department of Computer Science and Technology, Nanjing University. His research interests are in the areas of mobile computing and wireless sensing.

Ting Zhao received the BE degree from the School of Electronic Science and Engineering and the MD degree from the Department of Computer Science and Technology both from Nanjing University, Nanjing, China, in 2018 and 2021, respectively. She is now a staff in the NetEase, Inc.

Ke Sun received the BS degree in computer science from the Nanjing University of Aeronautics and Astronautics, Jiangsu, China, in 2016. He is currently working toward the master’s degree with Nanjing University. His research interests are in the area of mobile computing.

Shuyu Shi received BE degree from the University of Science and Technology of China (USTC), in 2011, and the PhD degree from SOKENDAI, where she was with National Institute of Informatics and Department of Informatics, Japan. She is currently a research associate professor with the Department of Computer Science, Nanjing University, China. She was a research fellow with the Wireless And Networked Distributed Sensing (WANDS) System Group in Parallel and Distributed Computing Centre (PDCC), School of Computer Science and Engineering, Nanyang Technological University (NTU) from 2016 to 2018, Singapore. She was also a JSPS research fellow from April 2015 to October 2016. Her research interests focus on mobile and ubiquitous computing.

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.

Haipeng Dai (Member, IEEE) received the BS degree from the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2010, and the PhD degree from the Department of Computer Science and Technology, Nanjing University, Nanjing, China, in 2014. His research interests are mainly in the areas of data mining, Internet of Things, and mobile computing. He is an associate professor with the Department of Computer Science and Technology, Nanjing University. His research papers have been published in many prestigious conferences and journals such as ACM UbiComp, IEEE INFOCOM, VLDB, IEEE ICDE, ACM WWW, ACM SIGMETRICS, ACM MobiHoc, ACM MobiSys, IEEE ICNP, IEEE ICDCS, *IEEE/ACM Transactions on Networking*, *IEEE Journal on Selected Areas in Communications*, *IEEE Transactions on Parallel and Distributed Systems*, *IEEE Transactions on Mobile Computing*, *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Dependable and Secure Computing*, and *IEEE Transactions on Information Forensics and Security*. He is an ACM member. He serves/ed as poster chair of the IEEE ICNP'14, track chair of the ICCCN'19 and the ICPADS'21, TPC member of the ACM MobiHoc'20-21, IEEE INFOCOM'20-22, IJCAI'21-22, IEEE SC'22, IEEE ICDCS'20-21, IEEE ICNP'14, IEEE IWQoS'19-21, and IEEE IPDPS'20-22. He received Best Paper Award from IEEE ICNP'15, Best Paper Award Runner-up from IEEE SECON'18, and Best Paper Award Candidate from IEEE INFOCOM'17.

Guihai Chen (Senior Member, IEEE) received the BS degree from Nanjing University, in 1984, the ME degree from Southeast University, in 1987, and the PhD degree from the University of Hong Kong, in 1997. He is a professor of Nanjing University, China. He had been invited as a visiting professor by many universities including Kyushu Institute of Technology, Japan, in 1998, University of Queensland, Australia, in 2000, and Wayne State University, during 2001 to 2003. He has a wide range of research interests with focus on sensor network, peer-to-peer computing, high-performance computer architecture and combinatorics. He has published more than 200 peer-reviewed papers, and more than 120 of them are in well-archived international journals such as *IEEE Transactions on Parallel and Distributed Systems*, *Journal of Parallel and Distributed Computing*, *Wireless Network*, *The Computer Journal*, *International Journal of Foundations of Computer Science*, and *Performance Evaluation*, and also in well-known conference proceedings such as HPCA, MOBIHOC, INFOCOM, ICNP, ICPP, IPDPS, and ICDCS.

Haodong Liu received the bachelor's degree in electronic engineering from the University of Electronic Science and Technology of China, in 1999. He is currently an engineer in Hisilicon Co.Ltd. His research interests are in the area of mobile and ubiquitous computing.

Wei Wang (Member, IEEE) received the MS and PhD degrees from the ESE Department of Nanjing University and the ECE Department of National University of Singapore, in 2000 and 2008 respectively. He is currently an associate professor with the CS Department of Nanjing University. His research interests are in the area of wireless networks, including device-free sensing, cellular network measurements, and software defined radio systems.

Authorized licensed use limited to: Univ of Calif San Diego. Downloaded on January 09,2024 at 22:00:16 UTC from IEEE Xplore. Restrictions apply.