RESEARCH ARTICLE Open Access

Biomimetic acoustic perception via chip-scale dual-soliton microcombs

Teng Tan[1,2†], Xin-Yue He[1†], Bing Chang[1†], Xu-Han Guo[3†], Heng Zhou[1,2†], Yong Geng[1], Yu Wu[1], Yu-Pei Liang[1], Ze-Ping Wang[1], Yong-Jun Huang[1], Ying-Zhan Yan[4], Si-Qin Ge[5], Yi-Kai Su[3], Chee Wei Wong[6] and Bai-Cheng Yao[1,2*]

Abstract

Acoustic perception is a fairly basic but extraordinary feature in nature, relying on multidimensional signal processing for detection, localization, and recognition. Replicating this capability in compact artificial systems, however, remains a formidable challenge due to limitations in scalability, sensitivity, and integration. Here, imitating the auditory system of insects, we introduce an opto-acoustic perception paradigm using fully-stabilized dual-soliton microcombs. By integrating digitally stabilized on-chip dual-microcombs, silicon optoelectronics and bionic fiber-microphone arrays on a single platform, we achieve parallelized interrogation of over 100 sensors. Leveraging the low-noise, multi-channel coherence of fully-stabilized soliton microcombs, this synergy enables ultra-sensitive detection of 29.3 nPa/Hz$^{1/2}$, sub centimeter precise localization, real-time tracking and identification for versatile acoustic targets. Bridging silicon photonics, optical fiber sensing and intelligent signal processing in a chiplet microsystem, our scheme delivers out-of-lab deployable capability on autonomous robotics. This work not only deepens the understanding of frequency comb science, but also establishes a concept of dual-comb-driven sensor networks as a scalable foundation for next-generation opto-acoustic intelligence.

1 Introduction

Listening is one of the most crucial capabilities for gathering information from surroundings [1]. Detection of acoustic waves is essential in a variety of applications, including but not limited to medical diagnostics [2, 3], sonar [4], navigation [5, 6], molecular tracing [7, 8], geoscience [9], and versatile industrial processes [10]. Analogous to hearing structures in animals such as insects [11, 12], the arrangement of multiple acoustic sensors in arrays facilitates the orientation and location tracking of acoustic signals [13–16]. In optics, this configuration has led to the evolution of techniques such as distributed acoustic sensing and acoustic radar [17–19]. Among kinds of opto-acoustic sensors, the fiber optical microphone (FOM) stands out due to its unique advantages such as ultrahigh sensitivity, microscale size, passive operation, broadband response, immunity against electromagnetic interference, and easy networking capabilities [20, 21]. Over recent years, the utilization of multiple

†Teng Tan, Xin-Yue He, Bing Chang, Xu-Han Guo and Heng Zhou have authors contributed equally.

*Correspondence:
Bai-Cheng Yao
yaobaicheng@uestc.edu.cn
[1] Key Laboratory of Optical Fiber Sensing and Communications (Education Ministry of China), University of Electronic Science and Technology of China, Chengdu 611731, China
[2] Engineering Center of Integrated Optoelectronic & Radio Meta-Chips, University of Electronic Science and Technology of China, Chengdu 611731, China
[3] State Key Laboratory of Advanced Optical Communication Systems and Networks, Shanghai Jiao Tong University, Shanghai 200240, China
[4] Information Science Research Institute, China Electronics Technology Group Corporation, Beijing 100042, China
[5] Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
[6] Fang Lu Mesoscopic Optics and Quantum Electronics Laboratory, University of California, Los Angeles, CA 90095, USA

FOMs has been successfully demonstrated in acoustic source localization [22, 23]. Traditional methods to drive and analyze an FOM array require the use of multiple independent lasers, amplified spontaneous emission light sources, and intricate optical filters. These methodologies have inherent limitations: for instance, the incongruity of optical frequencies restricts the performance variables, like sensitivity and signal-to-noise ratio (SNR) of the FOMs; additionally, the complex and bulk architecture impedes the capacity, simplicity, and integration of the entire system.

Integrated Kerr soliton microcomb is a promising light source for breaking the above bottlenecks. It offers multi-frequency output with high repetition and coherence at a chip-scale, for precision metrology and co-driving multiple photonic devices [24, 25]. Over the past two decades, the blooming development of microcomb technology has facilitated significant advances in wide applications [26], ranging from optical communication [27, 28], computation [29, 30], ranging [31, 32] to microwave control [33–35] and physical or biochemical sensing [36, 37]. In addition, frequency synthesis [38] based on dual-comb further provides a means to convert optical frequencies to radio frequencies. This technology allows for the illustration of broadband spectra while avoiding the need for large optical spectrometers with moving parts and limited optical resolution [39, 40]. It holds great promise in delivering high time–frequency resolution for parallel heterodyne measurements while reducing setup complexity. Utilizing dual-microcombs, recent advancements have demonstrated notable advantages in high-capacity communication, spectroscopy, and signal processing. However, the concept of microcomb-enhanced three-dimensional sensor networking remains largely unexplored.

Here we report a biomimetic optic-acoustic perception system based on coherently integrated dual-microcombs. In this design, the comb generators, wavelength division multiplexers and photodetectors are incorporated all on chips for seamless functionality. A pair of on-chip Kerr soliton microcombs generated in two silicon nitride microrings with a 4.1 MHz repetition difference are fully-stabilized via a compact ultra-stable optical reference based-on optical frequency division effect, culminating in a synchronized dual-comb interferometer. Uniquely, both the generation and stabilization of the dual soliton microcombs are digitally controlled in a compact field-programmable gate array (FPGA). After full stabilization, one comb performs as the probing light, and each of its comb line drives an individual FOM. In each FOM, a variety of acoustic response membranes are designed and prepared by mimicking the infraspinatus of different insects, which have excellent response characteristics covering wide acoustic bandwidth. Concurrently, the other comb serves as the local reference, facilitating parallel heterodyne measurements in the radio frequency domain.

This approach leads to advancements at both the device and system levels: First, the incorporation of integrated photonics into the fiber optic sensing network significantly reduces operational complexity and system size. The complementary metal–oxide–semiconductor (CMOS) compatible signal excitation and processing unit is minimized to a centimeter-level size, meanwhile the entire plug-and-play acoustic detection & recognition system evolves into a compact module. Second, the paradigm illustrates unprecedented high performances for dual-soliton microcomb stabilization. In each comb line, a remarkable optical linewidth down to 17 mHz is obtained, meanwhile stability of the dual comb beating reaches 8 μHz @ 1 s. Third, such high coherence of stabilized comb frequencies boosts the measurement accuracy of every biomimetic FOM, achieving an unprecedented minimum detectable pressure (MDP) down to 29.3 nPa/Hz$^{1/2}$, and keeping below nPa/Hz$^{1/2}$ in the broadband from 50 Hz to 20 kHz. Fourth, thanks to the full stabilization of every channel, we experimentally verify that the microcomb scheme can provision for over 100 optical frequency channels, thereby paving the way to simultaneously drive a huge number of acoustic sensors for high-precision stereoscopic and multi-target acoustic localization and tracking, with single centimeter accuracy. Furthermore, the entire system demonstrates flexible deployment and application capabilities out of laboratory, and suggests unique capability for high-precision sound recognition.

2 Results

Figure 1a illustrates the idea of our biomimetic acoustic mapper based on on-chip dual-microcomb. In nature, creatures like crickets and other insects utilize auditory perception to locate acoustic targets, fulfilling instinctive needs such as predator evasion, foraging, and mating. Throughout this process, a comprehensive biological information pathway is established. The brain acts as the central hub, connecting various subgenual organs located on the insect's legs through a neural network. These organs form an array that detects and localizes acoustic signals. Similarly, we mimick the auditory architecture of insects by using an on-chip dual comb as the optical source, which connects and empowers multiple miniature optical microphones with optical fibers. This forms an array that can perceive and locate sound information. As a result, it enables various functions such as acoustic target detection, eavesdrop, and incident warning.

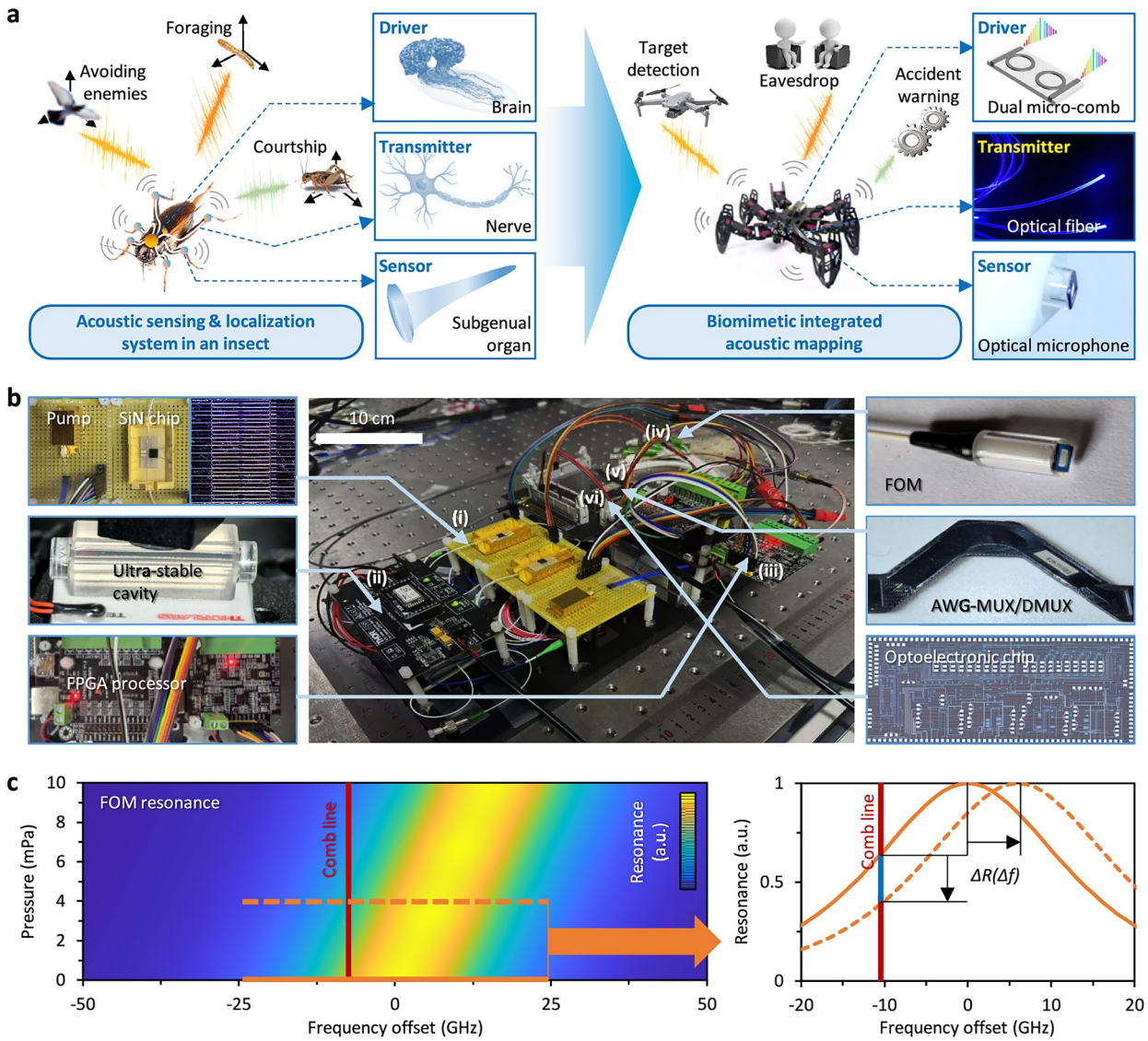

Fig. 1 Conceptual design of the dual-microcomb based biomimetic acoustic perception. **a,** Creatures like insects utilize auditory perception to locate acoustic targets, their auditory receptor network is driven by brain. Similarly, in our design, optical acoustic sensor array is driven by dual microcombs. **b,** Picture of the whole system, key optical and electronic components are integrated in a compact module, here we mark the devices: (i) the dual soliton microcomb generator, (ii) an ultra-stable vacuum F-P microcavity (USC) for comb stabilization, (iii) an FPGA-based electronic processor, (iv) FOMs fixed on fiber ends, (v) a pair of 108 channels arrayed waveguide gratings (AWGs) for frequency multiplexing and demultiplexing, (vi) a silicon chiplet containing filters, couplers and photodetectors. **c,** Calculated response of a comb line, which is reflected by an FOM. Here the orange curves show an example that acoustic pressure increases from 0 to 4 mPa

Figure 1b presents images of our system and its internal components. The complete microsystem is packaged into a portable and reliable $30\times20\times10$ cm^3 plug-and-play device, which incorporates photonic chips, connectors, electronic elements, and an FPGA module all in one. Here we also mark the key components within the portable device: (i) the dual soliton microcomb generator, (ii) an ultra-stable vacuum F-P microcavity for comb stabilization (packaged in the cooling box), (iii) an FPGA-based electronic processor, (iv) FOMs fixed on fiber ends, (v) a pair of 108 channels arrayed waveguide gratings (AWGs) for frequency multiplexing and demultiplexing, (vi) a silicon chiplet containing filters, couplers and photodetectors. Besides, auxiliary laser chips are hidden under the yellow card. Detailed device fabrication and characterization are provided in Fig. S1-1 and Supplementary Note S3.

Compared to other comb strategies like mode-locked lasers and EO combs, the on-chip soliton microcomb approach offers balanced performance and unique technical advantages [41]. For instance, our on-chip soliton microcomb can provide over 700 comb lines with 25 GHz spectral spacing, allowing for parallel driving of FOMs without spectral aliasing. Additionally, it features a compact footprint and eliminates the need for high-speed electrical drivers. Further discussions are available in Supplementary Note S4. In Fig. 1c, we present the acoustic sensing response of an FOM driven by a comb line. When acoustic pressure is applied to the biomimetic film of the FOM, the change in the F-P cavity length (L) alters the resonant frequency (f). Consequently, the reflected comb power is modulated by the reflection. Quantitatively, in the 1550 nm band, simulated results indicate that the sensitivity of resonant frequency to acoustic pressure reaches 1.506 GHz/mPa, allowing the power modulation for a comb line to achieve 11.2% per mPa in the quasi-linear region. During acoustic detection, the accuracy depends on the stability of the comb line. Detailed acoustic sensing mechanism and performance of the FOMs is shown in the Supplementary Note S2.

Prior to acoustic detection, we delve into the operation of our fully-stabilized dual microcombs and their associated parameter settings. Using compact optoelectronic feedback loops, we lock both the pump frequency (f_0), and the comb repetition rates ($f_{rep,1}$ and $f_{rep,2}$). Figure 2a schematically illustrates our stabilization strategy. First, we stabilize the Comb#1 via two-point locking scheme (the pump line and the 20th comb line), by using an ultra-stable vacuum cavity. Next, we stabilize $f_{rep,2}$ by the locking the repetition frequency difference (Δf_{rep}) between Comb #1 and Comb #2, via beating heterodyne. In this process, no expensive radio frequency references or modulators is in-need, and this ensures compactness of the whole system. We demonstrate the dual microcomb stabilization setup in Fig. S1-2. More technical details are also shown in the Methods and Supplementary Note S3. Leveraging this implementation, both carrier-envelope-offsets and repetition rates of the two soliton microcombs are well stabilized.

Figure 2b presents the optical spectra of Comb #1 and Comb #2, showcasing the optical band from 1545 to 1575 nm, which includes 150 lines in each comb. Both microcombs utilize the same pump wavelength of 1550 nm, with the residual pump filtered out to avoid DC interference. Through precise temperature control, the repetition rate of Comb #1 is maintained at 25.0031 GHz, while Comb #2 is set at 25.0072 GHz, indicating a Δf_{rep} of 4.1 MHz, significantly higher than acoustic frequencies in the kilohertz range. This separation effectively prevents frequency aliasing during acoustic detection. To meet the requirements of multiple sensor probes, we select 108 lines (Line #1 to Line #108 on the red side of the pump) from each comb using an integrated AWG. In the selected ≈ 2.7 THz band (highlighted in grey), the minimum power of the 108 comb lines approaches − 20 dBm, sufficient for driving the FOMs directly without further line-by-line amplification. As shown in Fig. S1-2, both soliton combs exhibit a sech^2-shaped envelope in the spectrum, confirming their single soliton state.

In Fig. 2c and d, we present the measured performance of comb#1, serving as the probe comb. Typically, the uncertainty in each comb line arises from inherent pump frequency drifts and noise superposition due to the optical frequency division effect both [42]. Pump frequency drifts stem from f_{ceo} noise, whereas noise superposition is dependent on repetition rate noise. In a free-running state, repetition rate noise predominantly contributes to the uncertainty in a comb line far away from the pump. Specifically, at 1 Hz offset, frequency noises of Line#1, #10 and #100 are 4.7×10^{11} Hz^2, 3.3×10^{13} Hz^2, and 3.8×10^{15} Hz^2, respectively. Upon full stabilization, both pump noise and repetition rate noise are significantly suppressed. Consequently, at 1 Hz offset, for Line#1, #10, and #100, the frequency noises decrease to 2.7×10^{6} Hz^2, 5.1×10^{6} Hz^2, and 5.7×10^{6} Hz^2, respectively. More importantly, the stabilization operation effectively suppresses frequency noise within the 20 Hz to 20 kHz band, which is crucial for achieving high accuracy in acoustic sensing. The effectiveness of stabilization is further demonstrated through Allan deviation measurements. Prior to locking, the 1-s optical uncertainty of Comb#1 (Line#100) is on the order of 10^{-8}, whereas after locking, it is reduced to the 10^{-12} level. The Comb#1 serves as the probe comb, and its stabilization enables remarkable suppression of optical frequency noise in the acoustic band. This feature is particularly beneficial for high-order comb lines, ensuring high precision in acoustic detection. Figure 2e and f depict the measured frequency noises of Comb#2, before and after full stabilization, respectively. The output of Comb#2 is utilized as the local signal. It is evident that Comb#2 exhibits similar stability compared to Comb#1. Notably, the full stabilization of Comb#2 is also essential for dual-comb-based sensing, as the final signal demodulation relies on analyzing the dual-comb heterodyne beat signal. In Fig. 2g and h, we make a summary. For Comb#1, during free-running operation, the minimum instantaneous linewidth increases from 1.2 Hz for Line#1 to 44 kHz for Line#100. However, after stabilization, the minimum instantaneous linewidths for Lines #1, #10, and #100 decrease to 0.27 Hz, 2.9 Hz, and 16.3 Hz, respectively. Similarly, for Comb#2, the minimum instantaneous linewidth suppression ratio reaches

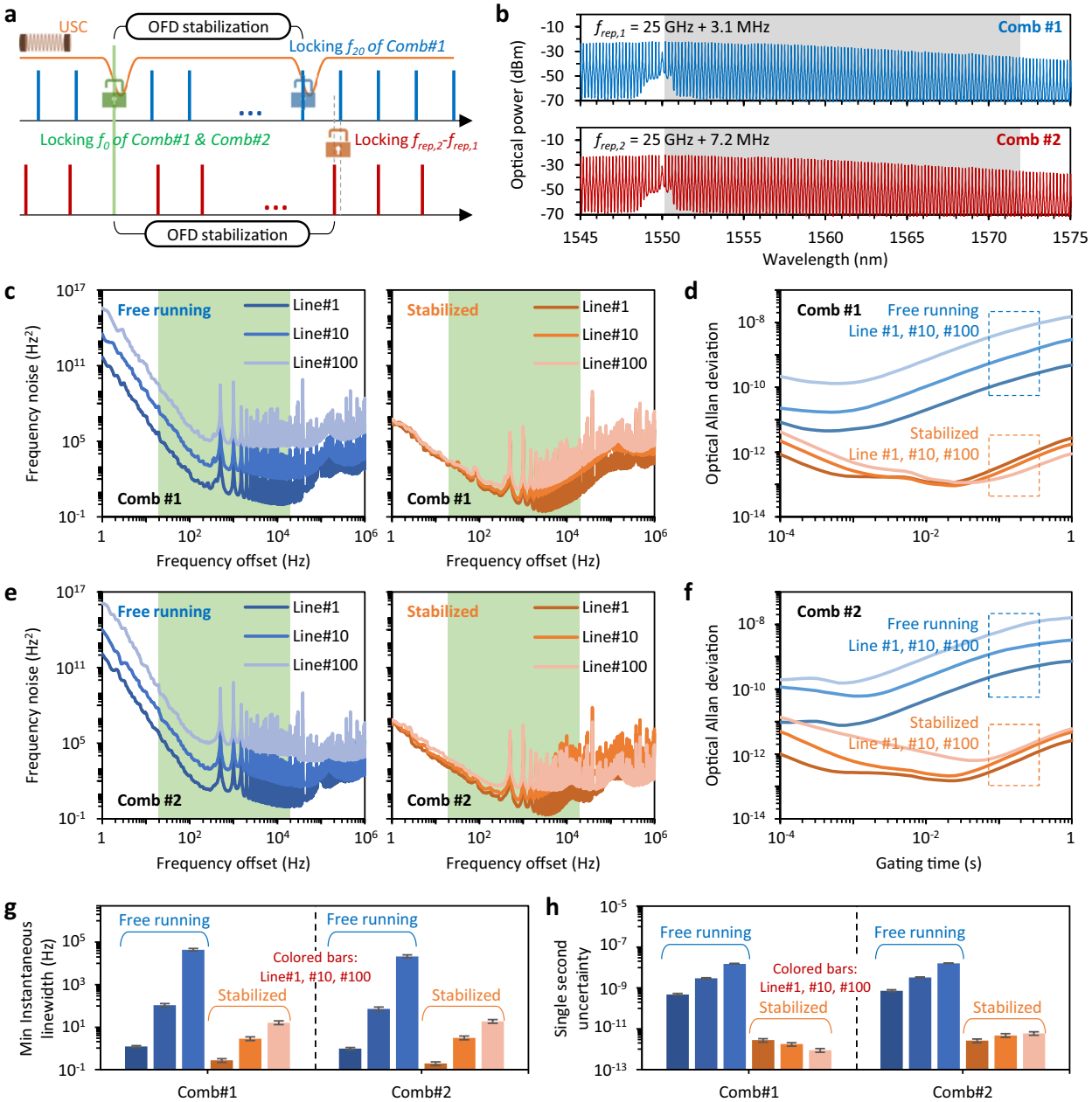

Fig. 2 Fully-stabilized dual microcombs. **a,** Scheme of the full stabilization based on optical frequency division (OFD). **b,** Optical spectra of the dual-microcombs (1545 nm to 1575 nm). Blue curve: Comb#1 (repetition 25.0031 GHz), red curve: Comb#2 (repetition 25.0072 GHz). In each microcomb, we select 108 comb lines to drive up to 108 FOMs after DMUX (grey shadow). **c,** Measured optical frequency noises of Comb#1. Left: free-running, right: fully-stabilized. **d,** Measured Allan deviation of Comb#1. **e,** Measured optical frequency noises of Comb#2. Left: free-running, right: fully-stabilized. **f,** Measured Allan deviation of Comb#2. In **(c-f)**, Blue curves: free-running, orange curves: fully-stabilized. **g-h,** Summary that the stabilization operation remarkably suppresses the instantaneous linewidth and 1-s uncertainty

10^4. Additionally, according to Allan deviation measurements, both Comb#1 and Comb#2 exhibit a significant improvement via full stabilization. On average, the 1 s stability is improved over 3 orders.

In Fig. 3, we illustrate that biomimetic design is first integrated into fiber optical microphone (FOM) probes, to enhance acoustic sensitivity and achieve distinctive responses. In nature, the micro-nano structures found in insect auditory receptors allow them to perceive weak and varied frequency characteristics of sound more effectively. By mimicking the subgenual film structures of Mecopoda elongata, Conocephalus gladiatus, and

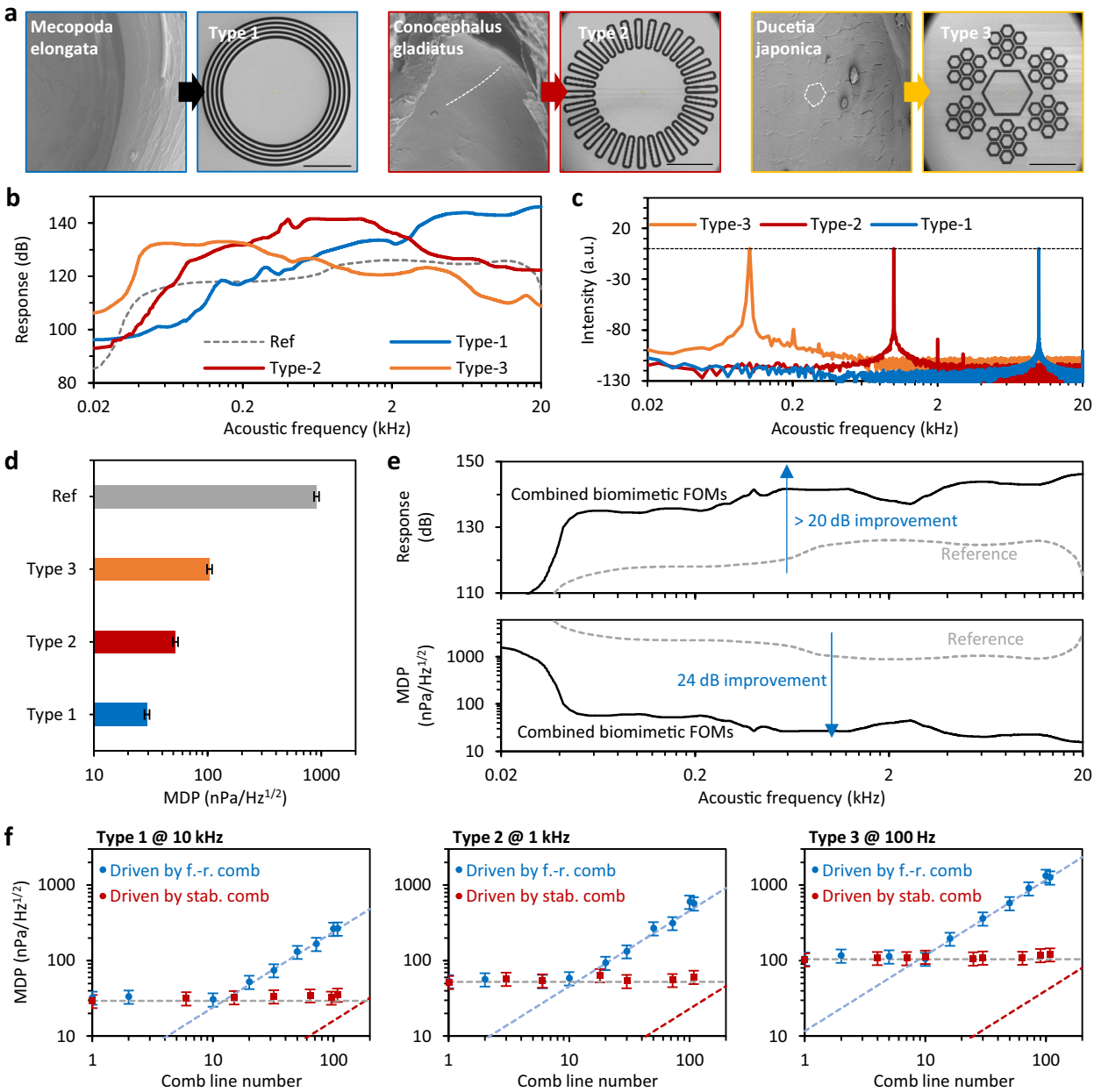

Fig. 3 Performances of the FOMs based on biomimetic acoustic films. **a,** Scanning electron microscope images of the biological prototypes alongside our biomimetic structures. Scale bar: 500 μm. **b,** Measured response spectra of FOMs with biomimetic acoustic films of types 1, 2, and 3. Varied biomimetic structures on film enables different response characteristics. **c,** SNR spectra when detecting sinusoidal acoustic signals. **d,** Measured MDP. Utilizing a fully-locked comb driven FOM, the minimum MDP is down to 29.3 nPa/Hz$^{1/2}$. Error bars: differences in repeated measurements. **e,** Measured response and MDP spectra of the coherently combined FOMs. Solid curves: combined FOM, dashed curves: reference. **f,** MDPs of channels from #1 to #108. Here blue dots (red dots) show the results that the FOMs are driven by free-running microcombs (fully stabilized microcombs), dashed lines show calculated limitations. Error bars: uncertainties in repeated measurements

Ducetia japonica, we fabricated three types of acoustic films using Si_3N_4 material. Figure 3a provides scanning electron microscope images of the biological prototypes alongside our biomimetic structures. Specifically, type 1 features multiple concentric circles; type 2 contains periodic radial stripes; and type 3 comprises a composite honeycomb structure. For each type, the central area of the film remains flat and smooth to ensure high-quality optical reflection. We show related simulations in Supplementary Note S2.

To further enhance sensitivity, a 20 nm thick Au layer is coated on the inner surface of the films and the fiber-end in our F-P cavity based FOMs. This modification improves the optical *Q* factor of the F-P cavity, enabling better acoustic pressure – optical drifting transform. More detailed discussions about the relationship between the *Q* factor and sensitivity can be found in Supplementary Note S4. Figure 3b presents the measured response spectra when using FOMs with biomimetic acoustic films of types 1, 2, and 3. The grey dashed curve indicates the response of a high-end commercial acoustic detector (B&K 4955), serving as a reference. Our three types of FOMs exhibit distinct response characteristics in the acoustic frequency domain: type 1 excels in the high-frequency region, type 2 in the mid-frequency range, and type 3 in the low-frequency domain. Their maximum responses reach 146.1 dB at 20 kHz, 141.4 dB at 800 Hz, and 132.4 dB at 60 Hz. Using these three types of probes simultaneously within one system enables high response across the entire acoustic band.

In Fig. 3c, we use fully stabilized comb lines to drive the FOMs based on film Types 1, 2, and 3, demonstrating their measured signal-to-noise ratios (SNRs). During the experiment, sinusoidal acoustic signals with a fixed acoustic pressure of $P_A = 37$ mPa and frequencies of approximately 10 kHz, 1 kHz, and 100 Hz are employed. The measured SNRs are 118 dB, 114 dB, and 108 dB, respectively, with a resolution bandwidth (BW) of 2 Hz. Here all the SNR numbers are measured within an acquisition time 20 ms. We note that achieving such high SNRs depends not only on the FOMs' high response performance but also on the low noise of the comb source. Subsequently, as represented in Fig. 3d, the minimum detectable pressure (MDP) for the three FOM types is calculated as 29.3 nPa/Hz$^{1/2}$@10 kHz, 52.2 nPa/Hz$^{1/2}$@1 kHz and 104 nPa/Hz$^{1/2}$@100 Hz respectively, using the transformation equation MDP = [$P_A{}^2$/(BW*SNR)]$^{1/2}$ [20, 43]. We find that these MDP values reach into the tens of nPa/Hz$^{1/2}$.

Since various biomimetic FOMs exhibit unique characteristic frequencies, it's thrilling to discover that one can utilize them simultaneously to achieve unprecedented high sensitivity across a broadband. In Fig. 3e, we examine the response and MDP spectra of the combined FOMs. The measured results indicate a response exceeding 130 dB within the range of 42 Hz to 20 kHz, suggesting an MDP below 100 nPa/Hz$^{1/2}$ in the band from 50 Hz to 20 kHz. This advancement allows for an acoustic detection system to perceive richer information with high resolution, which is crucial for applications such as precise sound recognition.

In Fig. 3f, we present the calculated and measured MDPs of our FOMs, which are driven by comb lines ranging from line#1 to line#108. By utilizing both a free-running comb and a fully-stabilized comb, we show the measured MDPs of FOMs in types 1, 2, and 3, in the panels from left to right. When using a free-running comb, due to the accumulation of optical frequency division noises, the MDPs of FOMs driven by comb teeth far from the central frequency gradually degrade. For instance, for the 100th FOM in types 1, 2, and 3, their MDPs are 267 nPa/Hz$^{1/2}$, 573 nPa/Hz$^{1/2}$, and 1261 nPa/Hz$^{1/2}$, respectively. On the other hand, when using a fully-stabilized comb, the uncertainties in all channels from #1 to #108 are below the MDPs determined by the FOMs. Specifically, for the 100th FOM in types 1, 2, and 3, their MDPs remain at 35.5 nPa/Hz$^{1/2}$, 61.2 nPa/Hz$^{1/2}$, and 121.3 nPa/Hz$^{1/2}$, respectively. Therefore, we conclude that the stability of a free-running comb is insufficient for driving 108 FOMs in parallel, whereas our fully stabilized comb exhibits excessive performance, effectively guaranteeing the performance of all sensing channels. Further discussions and comparisons are provided in Supplementary Note S4.

The on-chip dual comb based parallel optic-acoustic perception shows the capability to detect and localize outdoor acoustic targets such as drones and motors, which are important targets in both military and civilian applications, but usually have strong visual concealment, and often are hard to measure with traditional active techniques such as radar or optical ranging techniques. Furthermore, the synchronized operation of 108 FOMs driven by a single dual-microcomb source allows our system to be utilized in bionic robot clusters, significantly enhancing the networking capabilities of acoustic sensors. On each biomimetic robot, we employ 18 FOMs (including 3 types of biomimetic microphones on every foot) on a biomimetic hexapod robot, whose radius under maintenance condition is ≈ 25 cm. This parameter enables a maximum framing rate of 680 Hz. Figure 4b shows the detailed FOM distribution in one case, their 3D coordinates are shown in Table 1. For example, Fig. 4a demonstrates the experimental scenario in that we use this tool to localize a small unmanned aerial vehicle (UAV, DJI Air 3). We characterize the acoustic features of the UAV in Fig. S1-3.

Typically, the UAV shows multiple characteristic frequencies. In practice, for further promoting the SNR in detection, one can conveniently use an electrical filter or a computer recognition algorithm during the signal processing. Subsequently, we determine the delay difference of these detected acoustic traces across all pairwise FOMs by employing the equation $R_{i,j}(\tau) = \int_L \mathrm{Pod}_i(t)\,\mathrm{Pod}_j(t+\tau)dt$ [44]. Here L signifies temporal length of the sampled trace and $i \neq j$. The maximum value of $R_{i,j}(\tau)$ identifies the delay difference between Pod_i and

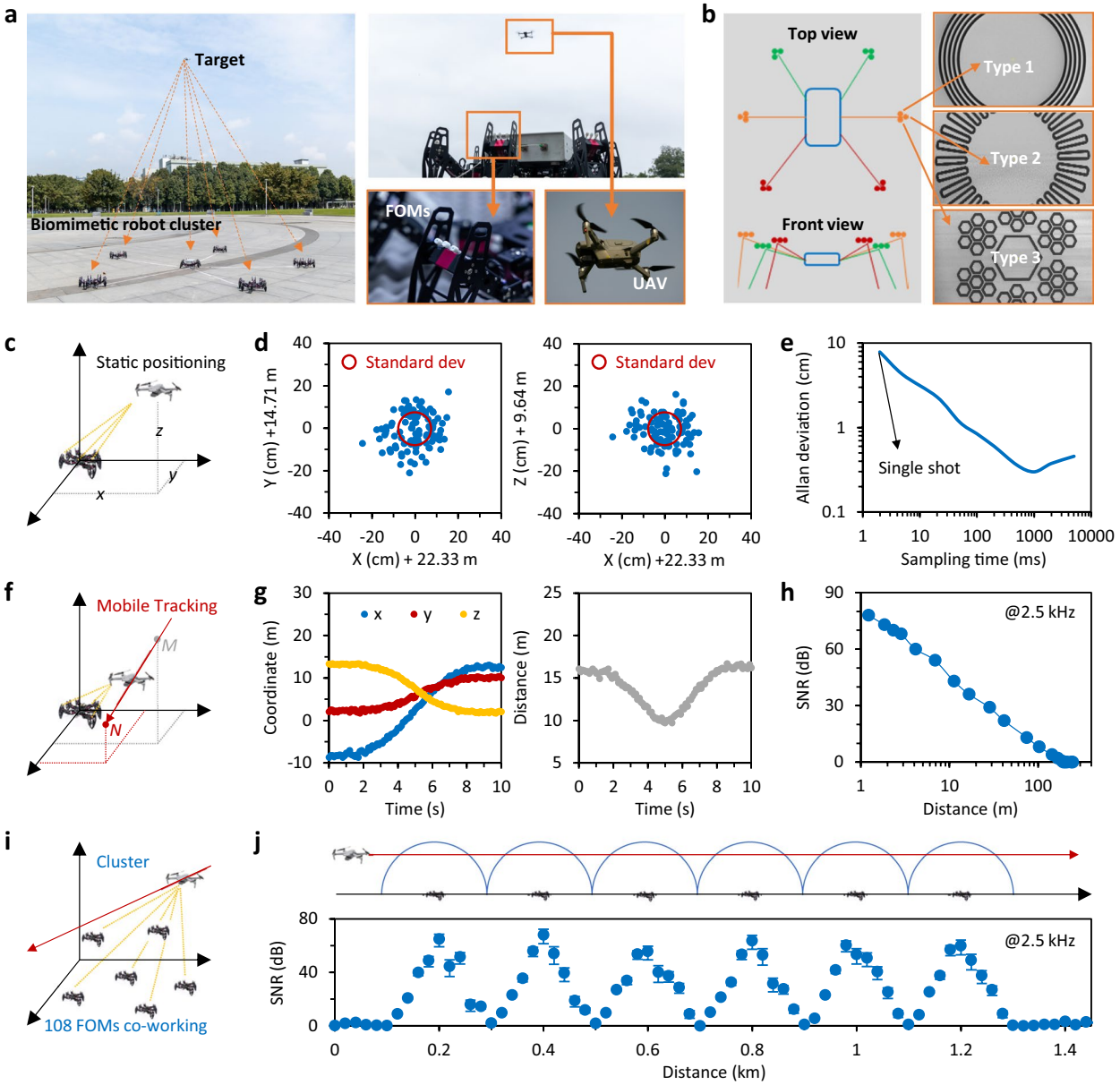

Fig. 4 Tracing a UAV out of the lab using a biomimetic hexapod robot equipped with the acoustic perception microsystem. **a,** Picture of the UAV detection scenario. **b,** Coordinates of the 18 FOMs distributed on 6 pods of the robot. **c-d,** Schematic diagram and measured results when the UAV is statically hovering. Blue dots show the measured coordinates, red circles mark the standard deviation. **e,** Allan deviation of the sensor-to-target distance. When the averaging time is 1 s, σ_d approaches 0.3 cm. **f-g,** Schematic diagram and measured results when the UAV moves from point *M* to point *N*. Here blue, red and yellow dots show the measured coordinates, while grey dots show the measured sensor-to-target distance. **h,** Distance versus SNR, suggesting a maximum UAV detectable distance over 194 m. **i-j,** schematic diagram and measured SNR when deploying 6 biomimetic robots as a cluster. The synchronized networking of over 100 FOMs enables remarkably extended sensing range

Pod$_j$ denoted by $\tau_{i,j}$. Then, by solving the matrix equations $\tau_{i,j}=|D_i\text{-}D_j|/v_A$, wherein $D_i=[(x_i\text{-}x)^2+(y_i\text{-}y)^2+(z_i\text{-}z)^2]^{1/2}$ is the distance from the target to Pod$_i$ and v_A is speed of sound, we can determine the spatial location of the UAV target.

Figure 4c shows the case that we localize the static hovering UAV, whose spatial coordinate is (x, y, z) = (22.33 m, 14.71 m, 9.64 m), therefore the sensor-to-target distance $d=(x^2+y^2+z^2)^{1/2}=28.424$ m. By repeatedly measuring its location 100 times, we record

Table 1 Coordinates of each FOMs on a biomimetic hexapod robot

FOM location	x (cm)	y (cm)	z (cm)
Pod 1	5	10	0.5
Pod 2	− 5	10	0.5
Pod 3	10	0	5
Pod 4	− 10	0	5
Pod 5	7	− 8	3.5
Pod 6	− 7	− 8	3.5

the measured points in the top-view and side-view maps in Fig. 4d. In statistics, standard deviation of the positioning is $\sigma_x = 7.87$ cm, $\sigma_y = 7.71$ cm, and $\sigma_z = 7.18$ cm, respectively. When the UAV hovers stably, such localizing errors can be reduced via continuous measurement.

Figure 4e shows the Allan deviation of d, here we set a framing rate of 500 Hz. For single shot measurement, σ_d is 7.86 cm; and when the averaging time is 1 s, σ_d reaches 0.3 cm. This number is already smaller than the size of the target. In Fig. 4f, we show the case that our on-chip dual comb based parallel optic-acoustic mapper can trace the dynamic movement of the UAV. When the UAV linearly flies from point M to point N, we can localize the UAV in real time. Originally, spatial coordinate of the UAV is M (− 8.71 m, 2.09 m, 13.42 m), while the terminal is N (12.51 m, 10.12 m, 2.09 m). By using our acoustic localizer, we record the 3-dimensional coordinate changes, as the blue, red, and yellow dots show (Fig. 4g). During this flight, we verify that the sensor-to-target distance d decreases from 16.1 m to 9.8 m, and then increases back to 16.2 m. Finally, we test the maximum measurement range for this UAV in Fig. 4h. In approximation, measured SNR decrement is 0.7 dB/m. When increasing d from 1.22 m to 265 m, we find that the SNR in our FOMs approaches 0 dB when $d > 194$ m. The on-line out-field test is shown in Supplementary Movie 1. In addition, the maximum acoustic localization range can be further extended by employing multiple robots in a distributed configuration, as illustrated in Fig. 4i-j. By positioning 6 robots (108 FOMs inside) at intervals of 200 m, our system is capable of tracking UAV within distances beyond 1.2 km. This operation is flexible and convenient, when compared to conventional sensor networking technology, our solution offers greater integration and cost advantages.

3 Discussions

In addition to localizing a single target, our system leverages dual-comb-based coherent demodulation for multiple sensors, enabling the simultaneous detection and positioning of more than one acoustic target. When these targets exhibit distinct characteristic frequencies, their separation can be straightforwardly accomplished through electronic filtering subsequent to a fast Fourier transform (FFT). Conversely, for targets whose acoustic frequencies overlap, alternative methods, such as neural network-based recognition algorithms [45], may be a good choice. The experimental demonstration of this capability is presented in Fig. 5. As illustrated schematically in Fig. 5a, we employ the dual-comb-based FOM array working in a complex acoustic environment. Here, acoustic emissions from a UAV, human speech, and a vehicle are detected concurrently, with each source being localized individually.

In the top panel of Fig. 5b, we present the recorded temporal trace of the composite acoustic wave, encapsulating both the frequency and localization information of the three targets. Subsequent FFT processing yields the acoustic spectrum (bottom panel). Utilizing electronic filters, we isolate the characteristic frequencies associated with each target. Specifically, the frequency bands for the UAV, human speech, and the vehicle are identified as 500 Hz ~ 750 Hz, 250 Hz ~ 400 Hz, and 1100 Hz ~ 1700 Hz, respectively. Post-filtering, we assess their target-to-sensor correlations and compute the variances in distance.

Figure 5c displays the pinpointed locations of the three targets, utilizing the sensor array depicted in Fig. 4. Distinct targets, each with unique coordinates, are vividly represented in three-dimensional space, with distances from the targets to the sensor array's center denoted by D_1 to D_3. When detecting fast moving target, Doppler effect should be considered, related discussion is provided in Supplementary Note S4. Figure 5d illustrates the outcomes of repeated measurements. For the UAV, the root mean square error in D_1 is measured at ± 6.8 cm; for human speech, D_2 at ± 5.6 cm; and for the vehicle, D_3 at ± 13.7 cm. Given that different acoustic sources produce sounds in varying volumes, these discrepancies are deemed acceptable.

Furthermore, in addition to the localization of multiple targets, Fig. 5e demonstrates that our acoustic mapping system can instantly recognize multiple targets with distinct acoustic characteristics, since sound data collection is continuous. When the environment is quiet, the system identifies the absence of sound with 100% accuracy. For a single target, the recognition accuracy exceeds 97.2% (blue columns). When two targets are detected, the accuracy is above 93.3% (red columns). In scenarios where human speech, UAVs, and vehicles are present simultaneously, the recognition accuracy is 91.5% (yellow column). In addition to distinguishing targets with distinct frequencies, our system efficiently achieves acoustic

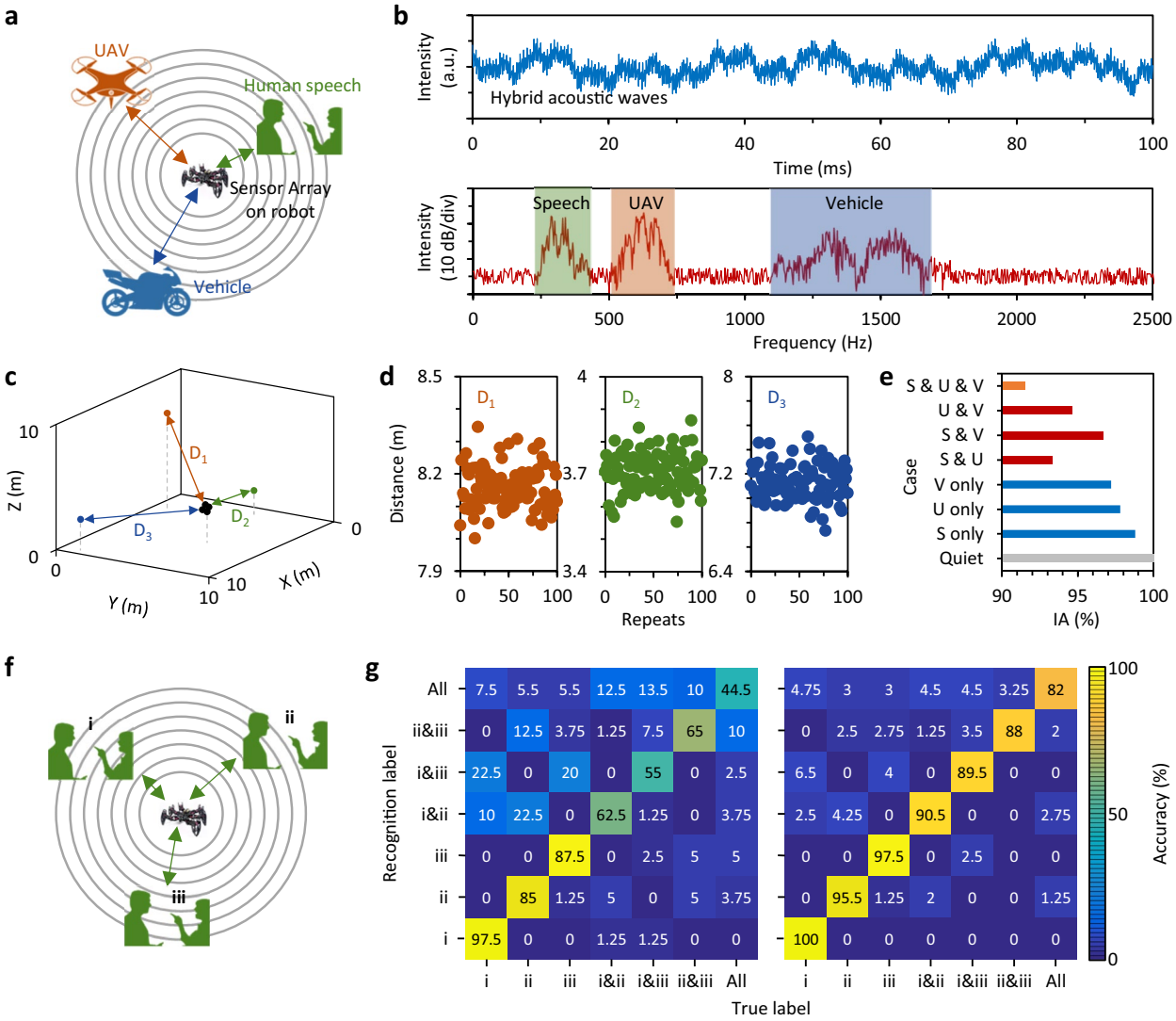

Fig. 5 Simultaneous multi-target localization and recognition. **a,** Schematic diagram shows the capability to detect and localize multiple targets. **b,** Measured temporal trace (top panel) and spectrum after filtering (bottom panel) of the hybrid acoustic wave, which contains information of 3 targets. **c,** Measured localizations of the 3 targets. **d,** Repeated measurements, here D_1, D_2, and D_3 can be individually demodulated. **e,** Recognition of different targets. IA: identification accuracy, S: speech, U: UAV, V: vehicle. **f,** A scenario demonstrating the use of the acoustic perception microsystem to identify mixed human speaking signals. **g,** Confusion matrices illustrating the sound recognition performance. Left: Using one FOM type; Right: Using all three FOM types in combination

recognition when multiple targets possess characteristic frequencies within the same range. This is facilitated by the CAM+ +based convolutional neural network algorithm implemented on our FPGA, integrated within the system. As an illustration, Fig. 5f depicts a scenario where the biomimetic robot can identify voices from three individuals (i, ii, and iii). Figure 5g presents the confusion matrices obtained from our measurements. Initially, using only one FOM type (anyone of the 3 types), the highest recognition accuracy for simultaneously speaking cases "i & ii & iii" was 44.5%. However, when all 3 FOM types are combinedly utilized, the recognition accuracy for the mixed signal "i & ii & iii" increases to 82%. Further information regarding the CNN-based sound recognition is provided in Supplementary Note S4.

In this work, we have developed a biomimetic plug-and-play acoustic perception microsystem by combining microcomb photonics, integrated optoelectronics and fiber sensing technology. Advanced optical synchronization enables instantaneous linewidth of a comb line below 0.2 Hz. Then, the use of coherent wavelength division multiplexing allows the on-chip dual-comb light

source to independently drive 108 miniature fiber optical microphones in parallel. This methodology not only facilitates listening with ultrahigh sensitivity at the level of tens nPa/Hz$^{1/2}$, but also allows for high-precision acoustic localization in three-dimensional space at the centimeter level. Furthermore, the dual-comb heterodyne-based signal collection enables simultaneous signal processing through a straightforward FPGA, enabling in-hardware acoustic recognition for distinct targets. Moreover, thanks to unique compactness and robustness, it suggests a unique advantage of flexible outdoor layout. This study provides an interdisciplinary concept, illustrating a physical paradigm in which sensing information can be precisely collected, discretely gathered and centrally processed via bionics. Practically, this system demonstrates a distinct ability to accurately and dynamically track visually concealed targets, and identify their acoustic characteristics effectively. In the future, the capacity of microcomb-based sensor networking could be further enhanced, suggesting that this approach may create a new path for various applications, including information interception, safety monitoring, and military perception with large scale and anti-reconnaissance capabilities.

4 Methods

4.1 Mechanism of the coherently parallel optic-acoustic detection and localization

The 3-dimensional acoustic positioning relies on solving the propagation paths of sound waves, based on measuring the arrival time differences. When using the sensor array system (with N sensors) to measure the three-dimensional coordinates of an acoustic target in the open air, we solve $N(N\text{-}1)/2$ acoustic path cross-correlations in an integrated processor. Based on numerical optimization methods, iterative algorithms are usually used to approximate the roots of the above equations. Since fiber optic microphones (FOMs) are placed at varied locations, the minimum distance (L_m) between them determines the frame rate of sound localization. More details are shown in Supplementary Note S2.

4.2 Fabrication and optimization of diverse fiber optical microphones

An FOM is fabricated as follows: first, we prepare a capillary glass tube, with inner diameter $d_0 = 0.127 \pm 0.001$ mm (suitable for fixing and calibrating a single-mode fiber (SMF)). Then we put an SMF with a flat-cut end-face into the capillary glass tube, and fix its position in depth using glue. We control the distance between the fiber end and the capillary glass tube end is L_1. Afterwards, we use a large-diameter glass sleeve to attach the MEMS sensing film. The distance between the sleeve end and the end of the capillary glass tube is L_2. The inner diameter of the glass sleeve is $d_2 = 2.8 \pm 0.01$ mm. Finally, we put the MEMS film on the sleeve and optimize the total cavity length ($L = L_1 + L_2$), the optimized L is ≈ 100 μm. Specifically, the silicon nitride MEMS diaphragm working as the acoustic oscillator is fabricated via plasma-enhanced chemical vapor deposition (PECVD) and lithography, the MEMS film has a size of 1.9×1.9 mm^2, a thickness of 400 nm. More details are provided in Supplementary Note S3.

4.3 Generation and stabilization of the on-chip Kerr soliton dual microcombs

First, two auxiliary laser diodes are separately tuned into two resonances (both around 1533.6 nm) of two microresonators on chip (Si_3N_4, 25 GHz repetition), and both positioned in the blue-detuned region. This setting ensures thermal stability when exciting soliton microcombs later. Then, an external cavity diode laser acts as the optical pump, directly stimulating solitons microcombs in the two distinct silicon nitride microrings on-chip through frequency scanning. The microrings display slightly different repetition rates, with a frequency difference of 4.1 MHz. Because of the high Q factor (≈ 4.6×10^6) of our microring, there is no need for external fiber amplification. After integrated amplification, output power of the 1550 nm pump laser is 400 mW, and the coupling loss from the laser to the microring chip is 1.4 dB. Therefore, we can ensure optical power launched into each microring > 160 mW. The dual microcomb generation process is automatically controlled using our FPGA module. Upon achieving a single soliton state, we fully stabilize both the pump frequency and comb intervals using an optoelectronic feedback technique. Specifically:1) We use an ultra-stable vacuum Fabry–Perot (F-P) cavity (customized, $Q > 10^9$) to stabilize the pump frequency and the 20th comb line of comb#1. 2) We select 22nd comb line of comb #1 and comb #2, detecting their beat note with another on-chip photodetector. Using a clock reference in the FPGA (Infineon S2F44T, 500 MHz), this signal is fed to auxiliary laser #2 to stabilize the frequency difference between comb #1 and comb #2, thereby locking the repetition of comb #2. In the full stabilization process, optical operations such as filtering and coupling are realized in our silicon optoelectronic chip, while electronic operations such as low-pass filtering, frequency mixing, and proportional-integral-derivative control are all handled within our FPGA compactly. More detailed performance metrics are provided in Supplementary Note S3.

4.4 Automatically optoelectronic control

The on-chip microcomb device enters soliton state, controlled by our automatic scanning program. In the

tuning process, the on-line detected output power triggers the pumping frequency via electrical feedback. In sensing operation, signals from each FOM are separately collected by a photodetector on-chip, and demodulated in our FPGA core. In electronics, we can use fast Fourier transform to analyze the spectra of different targets. Leveraging dynamic filtering, we can achieve simultaneous detection and localization of different acoustic targets.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43593-025-00099-5.

Additional file 1.

Acknowledgements
The authors thank Dr. Jinlong Xiang from Shanghai Jiao Tong University and Dr. Sipei Liu from the Institute of Zoology, Chinese Academy of Sciences for their helpful discussions.

Author contributions
T.T., X.Y.H., B.C., X.H.G. and H.Z. contributed equally to this work. B.C.Y. led the general study while X.H.G. and B.C.Y. led the research on integrated photonics. X.H.G. and Y.K.S. contributed the passive chip integration. X.H.G. contributed the chip tests. T.T., X.Y.H., B.C., C.W.W. and H.Z. built the on-chip dual-comb devices. Y.W. and X.Y.H. designed and fabricated the biomimetic fiber microphones. X.Y.H., T.T., Y.J.H. and Y.W. contributed the acoustic detection and localization. Z.P.W. and Y.P.L. programmed the FPGA and contributed the signal processing. X.Y.H., T.T., and Y.P.L. performed the out-field experiment. S.Q.G. contributed the biomimetic design, characterization and theoretical verification. T.T., X.Y.H. and H.Z. performed the theoretical analysis. Y.J.R. supervised this work. All authors processed and analyzed the results. B.C.Y., T.T., B.C., X.Y.H., and X.H.G. prepared the manuscript.

Funding
National Natural Science Foundation of China,U24A20311,BAICHENG YAO,62305050,Teng Tan,National Key Research and Development Program of China,2023YFB2806200,BAICHENG YAO,2023YFB2805600,Teng Tan,National Postdoctoral Innovation Talent Support Program of China,BX20220056,Teng Tan

Data availability
All data are available in the main text or the supplementary materials. The source data that support the plots within this paper and other findings of this study are available from the corresponding author on reasonable request.

Declarations

Competing interests
Authors declare that they have no competing interests.

Received: 31 July 2025 Revised: 2 August 2025 Accepted: 6 August 2025
Published: 8 September 2025

References

1. A.C. Mason, M.L. Oshinsky, R.R. Hoy, Hyperacute directional hearing in a microscale auditory system. Nature **410**, 686–690 (2001)
2. Y. Liu et al., Epidermal mechano-acoustic sensing electronics for cardiovascular diagnostics and human-machine interfaces. Sci. Adv. (2016). https://doi.org/10.1126/sciadv.1601185
3. K. Lee et al., Mechano-acoustic sensing of physiological processes and body motions via a soft wireless device placed at the suprasternal notch. Nat. Biomed. Eng. **4**, 148–158 (2019)
4. Y. Gao et al., Hydrogel microphones for stealthy underwater listening. Nat. Commun. **7**, 12316 (2016)
5. A.G. Krause, M. Winger, T.D. Blasius, Q. Lin, O. Painter, A high-resolution microchip optomechanical accelerometer. Nat. Photonics **6**, 768–772 (2012)
6. S. Basiri-Esfahani, A. Armin, S. Forstner, W.P. Bowen, Precision ultrasound sensing on a chip. Nat. Commun. **10**, 132 (2019)
7. S.-J. Tang et al., Single-particle photoacoustic vibrational spectroscopy using optical microresonators. Nat. Photonics **17**, 951–956 (2023)
8. H. Wu et al., Beat frequency quartz-enhanced photoacoustic spectroscopy for fast and calibration-free continuous trace-gas monitoring. Nat. Commun. **8**, 15331 (2017)
9. N.J. Lindsey, T.C. Dawe, J.B. Ajo-Franklin, Illuminating seafloor faults and ocean dynamics with dark fiber distributed acoustic sensing. Science **366**, 1103–1107 (2019)
10. B. Fischer, Optical microphone hears ultrasound. Nat. Photonics **10**, 356–358 (2016)
11. A.N. Popper, R.R. Fay, *Springer handbook of auditory research : sound source localization* (Springer Handpoo, New York, 2005)
12. F. Montealegre-z, T. Jonsson, K.A. Robson-brown, M. Postles, D. Robert, Convergent evolution between insect and mammalian audition. Science **80**(968), 1–5 (2013)
13. W. Yan et al., Single fibre enables acoustic fabrics via nanometre-scale vibrations. Nature **603**, 616–623 (2022)
14. K. Ma et al., A wave-confining metasphere beamforming acoustic sensor for superior human-machine voice interaction. Sci. Adv. **8**, 1–11 (2022)
15. X. Sun et al., Sound localization and separation in 3D space using a single microphone with a metamaterial enclosure. Adv. Sci. (2020). https://doi.org/10.1002/advs.201902271
16. J. Pan et al., Parallel interrogation of the chalcogenide-based micro-ring sensor array for photoacoustic tomography. Nat. Commun. **14**, 3250 (2023)
17. Y. Rao, Z. Wang, H. Wu, Z. Ran, B. Han, Recent advances in phase-sensitive optical time domain reflectometry (Φ-OTDR). Photonic Sensors **11**, 1–30 (2021)
18. J.-T. Li et al., Coherently parallel fiber-optic distributed acoustic sensing using dual Kerr soliton microcombs. Sci. Adv. (2024). https://doi.org/10.1126/sciadv.adf8666
19. B. Chang et al., Dispersive Fourier transform based dual-comb ranging. Nat. Commun. **15**, 1–10 (2024)
20. X. Lu, Y. Wu, Y. Gong, Y. Rao, A miniature fiber-optic microphone based on an annular corrugated MEMS diaphragm. J. Light. Technol. **36**, 5224–5229 (2018)
21. M. Yao et al., Ultracompact optical fiber acoustic sensors based on a fiber-top spirally-suspended optomechanical microresonator. Opt. Lett. **45**, 3516 (2020)
22. G. Wu et al., Development of highly sensitive fiber-optic acoustic sensor and its preliminary application for sound source localization. J. Appl. Phys. (2021). https://doi.org/10.1063/5.0044997
23. S. Lorenzo, O. Solgaard, Acoustic localization with an optical fiber silicon microphone system. IEEE Sens. J. **22**, 9408–9416 (2022)
24. T.J. Kippenberg, A.L. Gaeta, M. Lipson, M.L. Gorodetsky, Dissipative Kerr solitons in optical microresonators. Science (2018). https://doi.org/10.1126/science.aan8083
25. L. Chang, S. Liu, J.E. Bowers, Integrated optical frequency comb technologies. Nat. Photonics **16**, 95–108 (2022)
26. Y. Sun et al., Applications of optical microcombs. Adv. Opt. Photonics **15**, 86 (2023)
27. P. Marin-Palomo et al., Microresonator-based solitons for massively parallel coherent optical communications. Nature **546**, 274–279 (2017)
28. Y. Geng et al., Coherent optical communications using coherence-cloned Kerr soliton microcombs. Nat. Commun. **13**, 1070 (2022)
29. J. Feldmann et al., Parallel convolutional processing using an integrated photonic tensor core. Nature **589**, 52–58 (2021)

30. Y. Li et al., Nonlinear co-generation of graphene plasmons for optoelectronic logic operations. Nat. Commun. **13**, 1–7 (2022)
31. M.G. Suh, K.J. Vahala, Soliton microcomb range measurement. Science **359**, 884–887 (2018)
32. J. Riemensberger et al., Massively parallel coherent laser ranging using a soliton microcomb. Nature **581**, 164–170 (2020)
33. B. Yao et al., Gate-tunable frequency combs in graphene–nitride microresonators. Nature **558**, 410–414 (2018)
34. H. Zhang et al., Soliton microcombs multiplexing using intracavity-stimulated Brillouin lasers. Phys. Rev. Lett. **130**, 153802 (2023)
35. C. Qin et al., Electrically controllable laser frequency combs in graphene-fibre microresonators. Light Sci. Appl. **9**, 185 (2020)
36. C. Qin et al., Co-generation of orthogonal soliton pair in a monolithic fiber resonator with mechanical tunability. Laser Photon. Rev. **17**, 2200662 (2023)
37. T. Tan et al., Multispecies and individual gas molecule detection using Stokes solitons in a graphene over-modal microresonator. Nat. Commun. **12**, 8–15 (2021)
38. D.T. Spencer et al., An optical-frequency synthesizer using integrated photonics. Nature **557**, 81–85 (2018)
39. I. Coddington, N. Newbury, W. Swann, Dual-comb spectroscopy. Optica **3**, 414 (2016)
40. M.-G. Suh, Q.-F. Yang, K.Y. Yang, X. Yi, K.J. Vahala, Microresonator soliton dual-comb spectroscopy. Science **354**, 600–603 (2016)
41. B.C. Yao et al., Interdisciplinary advances in microcombs : bridging physics and information technology. eLight (2024). https://doi.org/10.1186/s43593-024-00071-9
42. Y. Geng et al., Phase noise of Kerr soliton dual microcombs. Opt. Lett. **47**, 4838 (2022)
43. J.A. Bucaro, N. Lagakos, B.H. Houston, J. Jarzynski, M. Zalalutdinov, Miniature, high performance, low-cost fiber optic microphone. J. Acoust. Soc. Am. **118**, 1406–1413 (2005)
44. C. Wang et al., High energy and low noise soliton fiber laser comb based on nonlinear merging of Kelly sidebands. Opt. Express **30**, 23556 (2022)
45. A. Francl, J.H. McDermott, Deep neural network models of sound localization reveal how perception is adapted to real-world environments. Nat. Hum. Behav. **6**, 111–133 (2022)

Supplementary information of

Biomimetic acoustic perception via chip-scale dual-soliton microcombs

Supplementary Notes

Note S1. Extended data figures.

The extended data figures provide information related to figures in maintext. Specifically, **Figs. S1-1** to **S1-3** showcase the extended characterizations and measurements about on-chip and on-fiber devices, the comb stabilization setup, and the acoustic localization scheme.

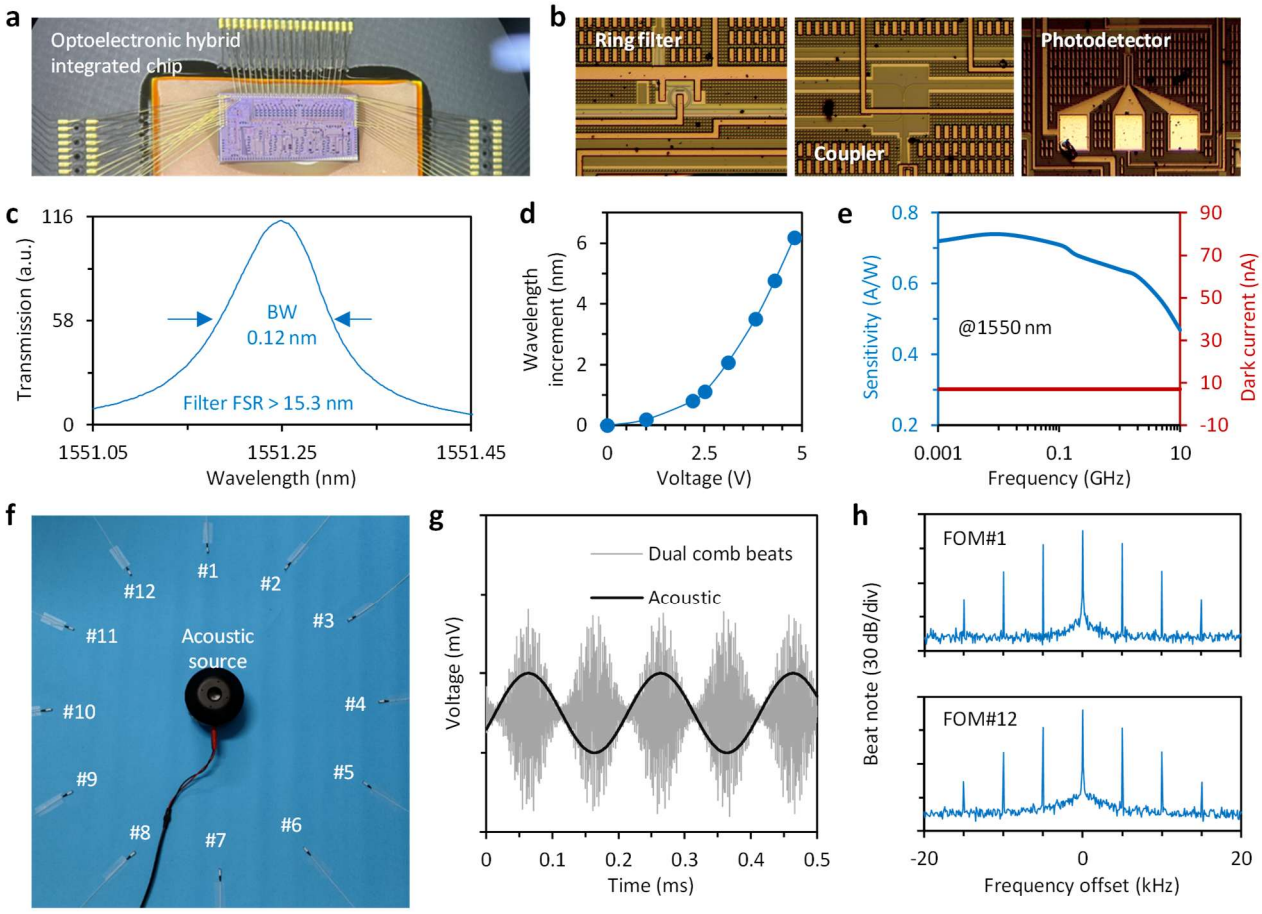

Fig. S1-1. Characterizations and performances of the devices on chip. **a,** Picture of the integrated chip. **b,** Close ups of major components, including silicon ring filter, coupler, and Ge-photodetector. **c,** Transmission of a ring filter. FSR of every ring filter is > 15.3 nm. Bandwidth of the ring resonance is 0.12 nm. **d,** The ring filters can be thermally tuned through the application of heaters. By adjusting the heating voltage from 0 to 5 V, the central wavelength of each ring filter can be shifted over a range of 6 nm. This allows for precise filtering of each comb line. **e,** Sensitivity (blue curve) and dark current

(bias -1 V, red curve) of a photodetector on-chip. **f,** A picture shows that comb driven 12 FOMs detect the sample acoustic signal (f_A = 5 kHz) for example. **g,** Measured temporal traces. Black: acoustic signal, grey: acoustically modulated dual comb signal. **h,** Spectra shows the acoustic harmonics in FOM #1 and FOM #12.

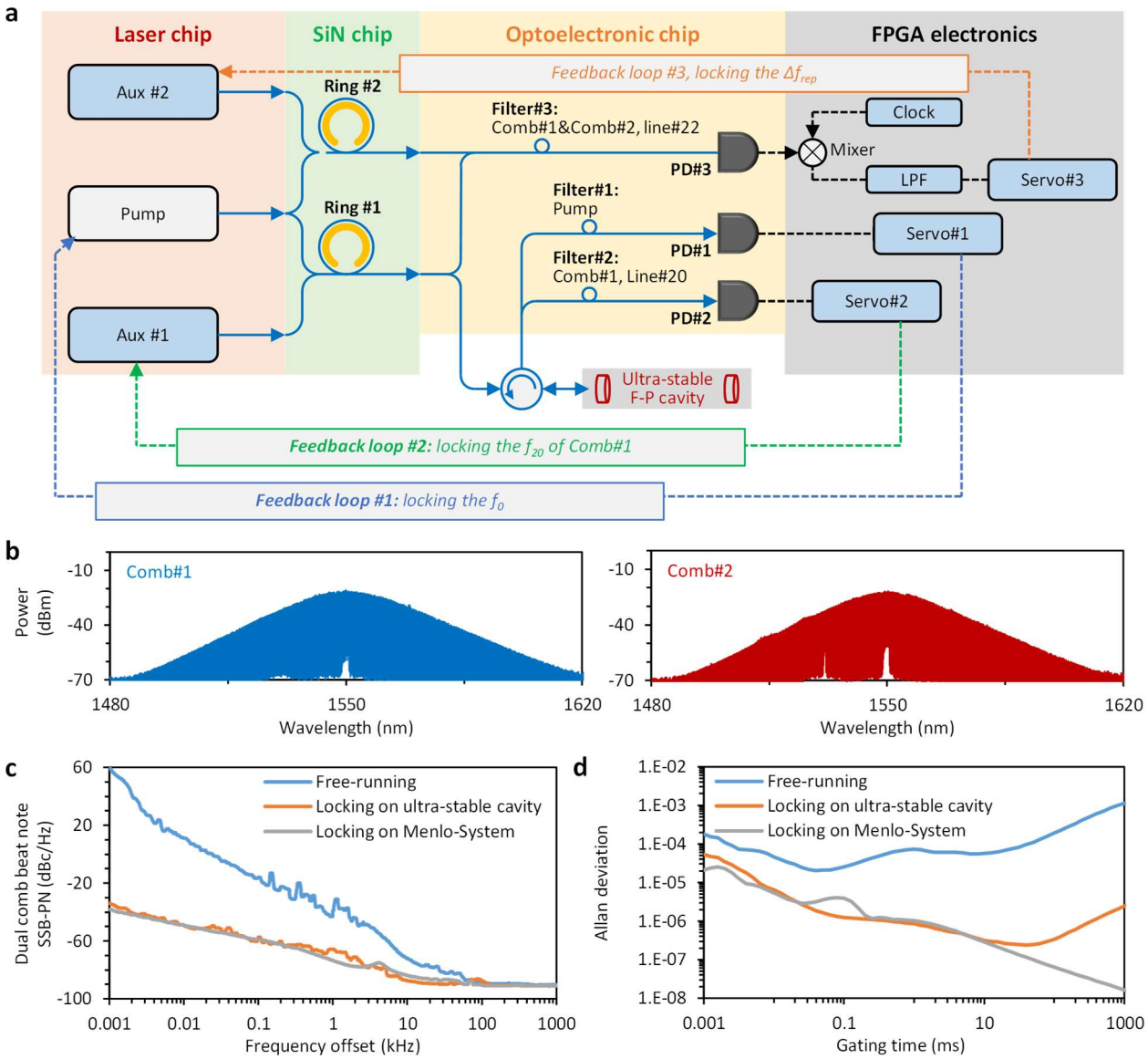

Fig. S1-2. Setup of the dual microcomb module and beating performance. **a,** Experimental setup for dual comb generation and stabilization. We use 3 feedback loops locking the shared carrier frequency, and the repetition frequencies of two microcombs. The whole system is compact. **b,** Measured optical spectra of the two soliton microcombs, in the band 1480 nm to 1620 nm. **c,** Measured single-sideband phase noises (SSB-PNs) of the dual-comb beating signal, before and after locking. **d,** Allan deviations. In **c-d**, we test the 108th beating line. Here the orange and grey curves are based on the same locking

scheme via optical frequency division. It verifies that our scheme showcases comparable performance to the strategy using Menlo-system.

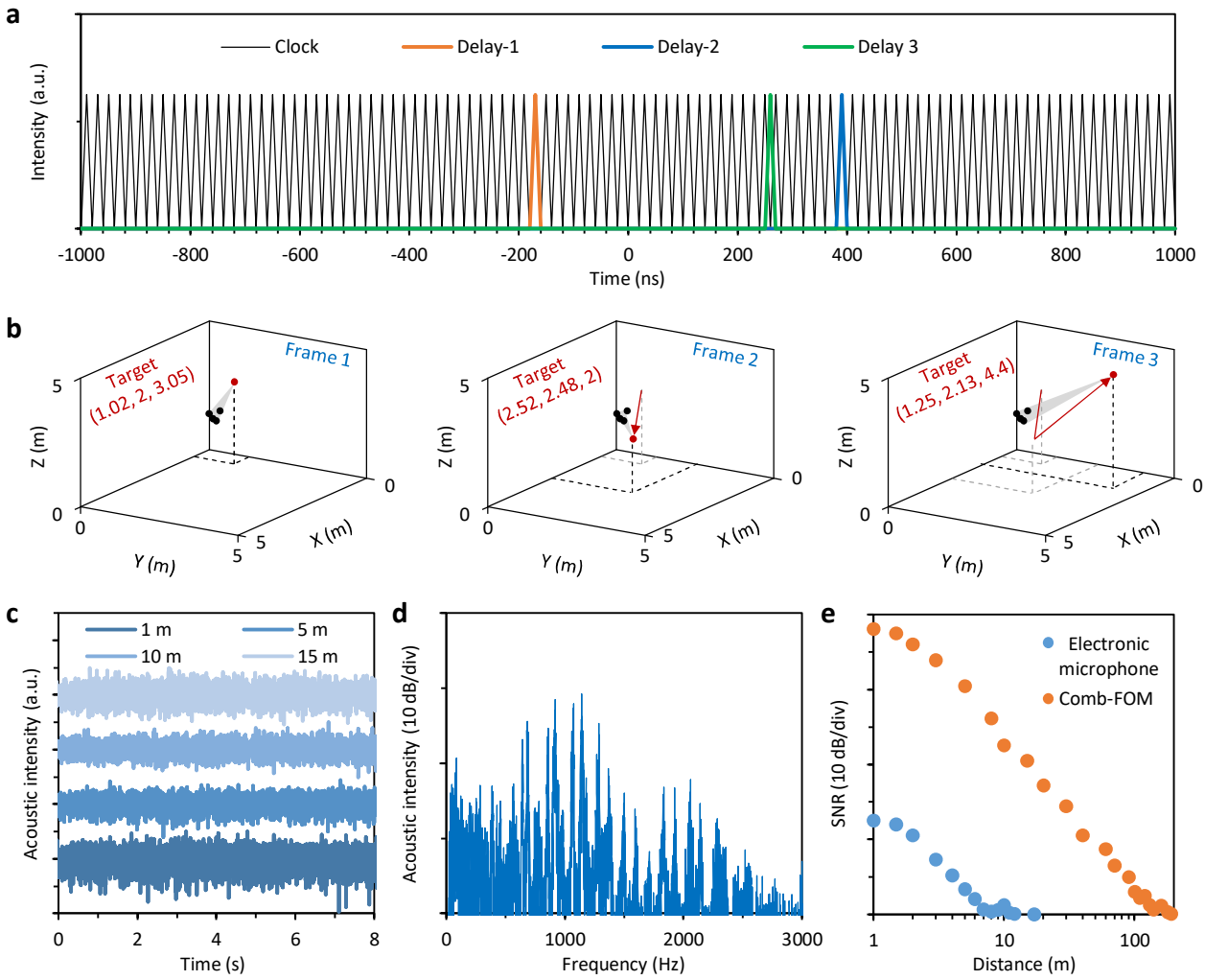

Fig. S1-3. Real-time demodulation of target's spatial position and acoustic characteristics of the UAV. a, In digital processing, we use a clock (100 MHz) to localize the arrival coherences between the target and sensors. Here we show an example: orange, blue and green curve respectively shows the cross-delay of <target-FOM1, target-FOM2>, <target-FOM1, target-FOM3>, and <target-FOM2, target-FOM3>. **b,** 3D interface displays the dynamic spatial position of an acoustic target. Here black dots show the FOMs, red dots show the target. Arrow: target movement. **c,** Temporal traces, measured by using an electronic microphone, at 1, 5, 10, 15 m away. **d,** Acoustic spectrum of the UAV, measured by using an electronic microphone, at 1 m away. **e,** SNR versus distance. Blue dots show the results got from an electronic microphone, orange dots show measured data from a comb-driven FOM.

Note S2. Supplementary theoretical analysis and simulations.

S2.1. Soliton generation in an on-chip micro-ring.

Silicon-nitride (Si_3N_4) microring cavities with a width-height cross-section of 2480 nm × 840 nm and a cavity length 6.4 mm are utilized for high-density Kerr soliton dual-microcomb generation. According to the design parameters of the microring, the transverse electrical mode field distribution and transmission spectrum of the microring cavity are simulated by COMSOL Multiphysics, as shown in **Fig. S2-1a** and **Fig. S2-1b**. The longitudinal mode interval obtained by our simulation is 0.2 nm (≈ 25 GHz). The Si_3N_4 microring parameters obtained by simulation and experiment are summarized in **Table S1**.

Table S1. Parameters used in analytical calculations for soliton microcombs

Parameters	Value
n_{eff}	1.86
L	6.4 mm
D_1	2π×25 GHz
γ	0.779 $W^{-1}\cdot m^{-1}$
β_2	-8.46×10^{-26} s^2 m^{-1}
Q	4.6×10^6
α	0.0053
θ	0.0053
$(\alpha_T Q)/(Q_{abs} C_P)$	5.3×10^{-6} J^{-1}
K/C_P	7×10^{-4} s^{-1}

Theoretical models of soliton Kerr comb generation have been extensively discussed [1]. A significant challenge in soliton comb formation is managing the thermal balance during nonlinear excitation, which impacts the stability and robustness of a soliton comb device in practical applications. In this study, we utilize an auxiliary laser heating scheme to achieve soliton generation in two microcavities [2]. This approach enables deterministic single soliton generation and facilitates repetition rate locking [3], enhancing the reliability and applicability of our dual-comb source as a practical tool outside the laboratory. In the context of the auxiliary laser heating scheme, the thermal effects within the microcavity are managed by injecting the auxiliary laser into the microring. Consequently, the influence of the auxiliary light must be considered alongside the standard Lugiato–

Lefever equation (LLE). As the auxiliary light propagates in the opposite direction to the pump light, it induces only a linear phase shift on the forward pump light and Kerr comb through the cross-phase-modulation (XPM) effect, with the magnitude of this shift depending on the intracavity auxiliary light power. Additionally, to accurately simulate the regulation of the thermal effects by the auxiliary light on the microcavity, a thermal nonlinearity term must be introduced. Therefore, the LLE model incorporating both the cavity thermal effect and the auxiliary laser can be represented as follows:

$$T_R \frac{\partial E_c(t,\tau)}{\partial t} = \left[-\alpha - i\left(\delta_p - \delta_{th}\right) + iL\sum_{k\geq 2} \frac{\beta_k}{k!}\left(i\frac{\partial}{\partial \tau}\right)^k + i\gamma L\left(\left|E_c\right|^2 + 2P_a\right)\right] E_c(t,\tau) + \sqrt{\theta P_{pump}} \tag{S1}$$

$$T_R \frac{\partial E_a(t,\tau)}{\partial t} = \left[-\alpha - i\left(\delta_a - \delta_{th}\right) + iL\sum_{k\geq 2} \frac{\beta_k}{k!}\left(i\frac{\partial}{\partial \tau}\right)^k + i\gamma L\left(\left|E_a\right|^2 + 2P_c\right)\right] E_a(t,\tau) + \sqrt{\theta P_{aux}} \tag{S2}$$

$$\frac{\partial \delta_{th}}{\partial t} = \frac{1}{C_p}\left[\alpha_T \frac{Q}{Q_{abs}}\left(P_c + P_a\right) - K\delta_{th}\right] \tag{S3}$$

Here, T_R is the roundtrip time of the Si_3N_4 microring; t and τ are denote the slow time at the scale of the cavity photon lifetime and the fast time defined in a reference frame moving at the light group velocity in the cavity, respectively, which are used to describe the evolution of the intracavity comb field $E_c(t, \tau)$ and auxiliary laser field $E_a(t, \tau)$; α is the cavity decay per roundtrip; δ_p and δ_a are the detuning of the pump and auxiliary laser, while δ_{th} is the thermal drifting of the resonance; L is cavity length; γ is the nonlinear coefficient; θ is the coupling rate from the bus waveguide to the microring; β_k is the k-th order dispersion; P_a and P_c are the intracavity average power of pump-comb field and auxiliary laser-comb field, respectively; P_{pump} and P_{aux} are the pump laser power and auxiliary laser power, respectively; α_T is the temperature coefficient, C_p is the thermal capacity, K is the thermal conductivity, Q and Q_{abs} are the loaded and intrinsic quality factors.

Based on the parameters shown in **Table S1**, and utilizing the thermal-assisted LLE model, we have numerically simulated the generation process of a soliton microcomb within a microring. **Fig. S2-1c** illustrates the evolution of intracavity pump laser power alongside auxiliary laser power. As the intracavity pump light power increases, the auxiliary light power decreases, maintaining a balanced intracavity total power. This balance mitigates the thermal effects caused by intracavity power fluctuations and aids in the deterministic generation of a single soliton. **Fig. S2-1d** and **S2-1e** depict the comb evolution in both the time and frequency domains as the pump frequency is red-detuned. These figures demonstrate the typical time-frequency characteristics of a soliton frequency comb [4]. Throughout this simulation, a distinct soliton state emerges. The appearance of a single soliton follows the progression through states such as primary comb, chaotic state, and multi-soliton state.

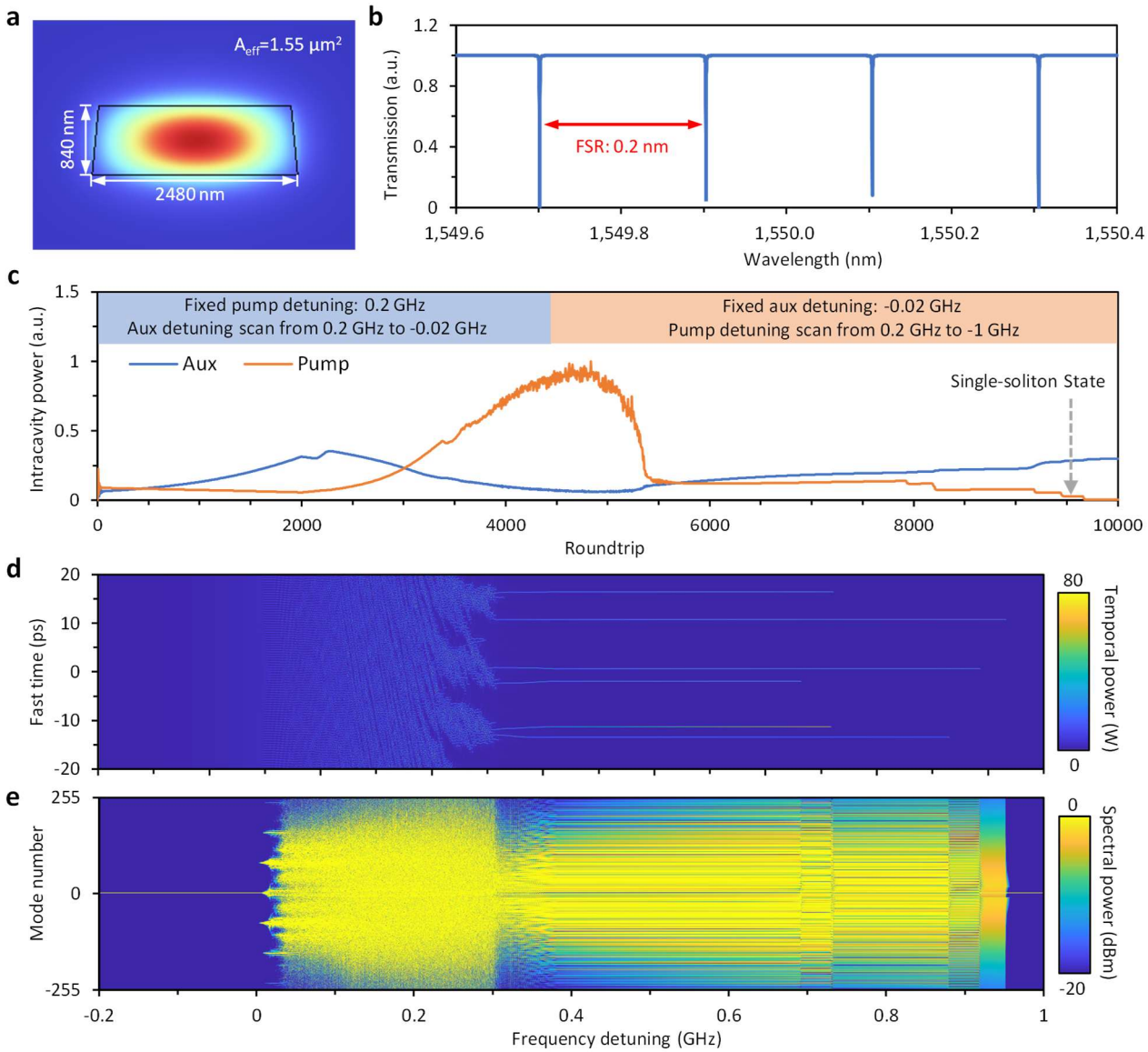

Fig. S2-1. Numerical simulations of an integrated Kerr soliton frequency comb formation using an auxiliary laser. a, Transverse mode field distribution of microring resonator. **b,** Longitudinal resonances of the microring, FSR ≈ 0.2 nm. **c-e,** The evolution of intracavity power, optical field in time domain and in frequency domain during the pump wavelength red-detuning.

S2.2. Optical frequency division based dual comb locking.

Here, we delve into the principle of two-point locking for stabilizing Kerr soliton frequency combs more in details. The frequency noise in a Kerr soliton frequency comb, once generated, is governed by two primary factors: 1) Pumping frequency drift, which is linked to the uncertainty of the carrier-envelope-offset frequency (f_{ceo}). 2) Repetition rate variations, which are induced by thermal instability within the microcavity.

Typically, the kth comb line exhibits a frequency given by $f_k = f_p + kf_{rep}$[1], where f_p represents the

pumping frequency, k is the order of the comb line, and f_{rep} is the repetition rate. The frequency noise of a line located at f_k can generally be expressed as:

$$n_k = n_p + kn_r \tag{S4}$$

Here, n_p denotes the noise originating from the pump, and n_r represents the noise in the repetition rate. Typically, as the comb number increases meanwhile f_k becomes further away from the pump, the frequency instability of an individual comb line increases linearly. Now, let's consider noise suppression process, as illustrated in **Fig. S2-2a**. By utilizing a reference, one can initially suppress n_p to n_p'. In this scenario, we obtain:

$$n_k' = n_p' + kn_r \tag{S5}$$

Then one can use another reference to lock the kth line, suppressing its noise from n_k' to n_k''. Now total noise of this line becomes:

$$n_k'' = n_p' + kn_r' \tag{S6}$$

That means, this operation not only modifies n_p to n_p' but also alters n_r to n_r'. Specifically, after implementing this two-point locking, n_r' can be expressed as $(n_k''\text{-}n_p')/k$. Compared to the case where only the pump is locked, the repetition rate noise suppression can be calculated as $(n_k'\text{-}n_k'')/k$. This indicates that a larger k results in a lower repetition rate noise. The technique of stabilizing a comb line distant from the pumping frequency is referred to as "locking based on optical frequency division" [2]. **Fig. S2-2b** schematically illustrates this process. For example, in a free-running comb (red line), n_p is 1 Hz/Hz$^{1/2}$ and n_r is 2 Hz/Hz$^{1/2}$. When k = 100, n_k is 201 Hz/Hz$^{1/2}$. After locking the 100th comb line and suppressing n_k to 2 Hz/Hz$^{1/2}$ using a reference, n_r reduces to 0.01 Hz/Hz$^{1/2}$, as shown by the blue line. In another scenario, if we lock the 1000th line (with k = 1000) and use the same reference to suppress n_k to 2 Hz/Hz$^{1/2}$, we can achieve a very small n_r of 10^{-3} Hz/Hz$^{1/2}$ (yellow line).

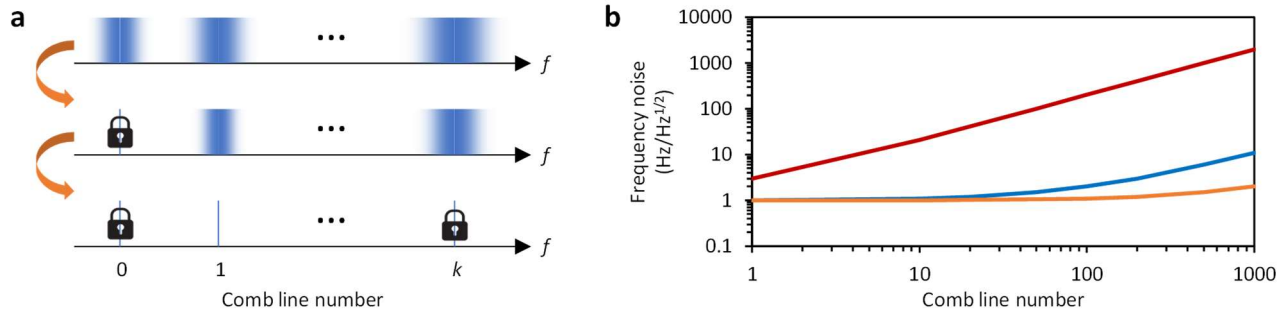

Fig. S2-2. Optical frequency division based comb stabilization. a, Schematic diagram that one can lock a comb at two-points, the pump and a kth line. **b,** Schematic calculation, locking a comb at a larger k leads to a better stabilization.

S2.3. Fabry-Perot cavities based on MEMS films and their acoustic sensing properties.

We use optical fiber Fabry-Perot (F-P) microcavities as opto-acoustic sensors. For enhancing acoustic response, these microcavities are incorporated by metal film fabricated via micro electromechanical system (MEMS). **Fig S2-3a** shows the model of our F-P acoustic sensor. When external sound pressure acts on the propagation film, the deformation of the MEMS film will cause changes in the length of the optical fiber F-P cavity. This leads to a spectral shift of the optical resonance inside [3]. In our experiment, we use one frequency comb line to detect the resonance shift of every F-P cavity. The intensity alteration of the reflected light refers to the acoustic magnitude. Specifically, the cavity length L of an F-P cavity can be detected in the following equation:

$$L = \frac{\lambda_1 \lambda_2}{2n|\lambda_1 - \lambda_2|} \tag{S7}$$

Here, $\lambda_1$and λ_2 are the two central wavelengths of adjacent resonance peaks in the Fabry-Perot cavity interference spectrum. They can be directly measured from the interference spectrum; n = 1 is the refractive index of the medium in the F-P cavity, as the F-P cavity is filled with air.

In sensing application, the two reflective end faces of the fiber F-P cavity (MEMS film and single-mode fiber end face) are weakly reflective surfaces (< 5%), we can approximately equate it in double-beam interference model [4]. Accordingly, the total reflected intensity of the fiber F-P cavity at a wavelength λ can be expressed as:

$$I(\lambda) = I_{\text{Fiber}}(\lambda) + I_{\text{MEMS}}(\lambda) + 2\sqrt{I_{\text{Fiber}}(\lambda) I_{\text{MEMS}}(\lambda)} \cos(\frac{4\pi L}{\lambda}) \tag{S8}$$

Here, $I_{\text{Fiber}}(\lambda)$ is the reflected intensity of the single-mode fiber (SMF) facet, $I_{\text{MEMS}}(\lambda)$ is the reflected intensity from the MEMS diaphragm. In dual comb demodulation, $\cos(4\pi L/\lambda)$ mainly demonstrates acoustic modulation.

When an external sound pressure P is applied to the MEMS diaphragm, the deformation of the diaphragm will change the F-P cavity length (ΔL), as **Fig. S2-3b** illustrates. Within the dynamic range of the MEMS diaphragm, such a change of cavity length ΔL caused by the sound pressure is linear:

$$P = \alpha \Delta L \tag{S9}$$

In this equation, α is a constant. Therefore, it can modulate the reflection intensity $\Delta I(\lambda)$ by shaping in ΔL. When changing the ΔL, $I(\lambda)$ become:

$$I'(\lambda) = I_{\text{Fiber}}(\lambda) + I_{\text{MEMS}}(\lambda) + 2\sqrt{I_{\text{Fiber}}(\lambda) I_{\text{MEMS}}(\lambda)} \cos[\frac{4\pi}{\lambda}(L - \Delta L)] \tag{S10}$$

The intensity modulation writes:

$$\Delta I(\lambda) = I(\lambda) - I'(\lambda) = 2\sqrt{I_{\text{Fiber}}(\lambda) I_{\text{MEMS}}(\lambda)} \cos(\frac{4\pi L}{\lambda}) - 2\sqrt{I_{\text{Fiber}}(\lambda) I_{\text{MEMS}}(\lambda)} \cos[\frac{4\pi}{\lambda}(L - \Delta L)] \tag{S11}$$

For getting the highest sensitivity, we choose a λ satisfying $\cos(4\pi L/\lambda) = 0$, so that:

$$\Delta I(\lambda) = -2\sqrt{I_{\text{Fiber}}(\lambda) I_{\text{MEMS}}(\lambda)} \sin\left(\frac{4\pi \Delta L}{\lambda}\right) = -2\sqrt{I_{\text{Fiber}}(\lambda) I_{\text{MEMS}}(\lambda)} \sin\left(\frac{4\pi P}{\alpha \lambda}\right) \quad \text{(S12)}$$

Such an effect is schematically shown in **Fig. S2-3c**. This optical modulation can be detected by a photodetector (PD) and demonstrated in electronics. Therefore, the sensitivity of the MEMS diaphragm-based fiber optic F-P acoustic wave sensor is mainly affected by the following two factors: 1) the displacement of the elastic diaphragm at the center of the diaphragm under unit sound pressure; 2) the reflected optical intensity.

When assuming $I_{\text{Fiber}}(\lambda) = I_{\text{MEMS}}(\lambda)$, D can be further simplified as $|\sin(4\pi \Delta L/\lambda)|$. When ΔL is far smaller than λ, the response is approximately linear, $\approx 4\pi \Delta L/\lambda$. For a circular elastic diaphragm with a fixed periphery, its acoustic response sensitivity (S) can be expressed as the maximum diaphragm displacement caused by unit sound pressure change ΔL, written in the following equation [5]:

$$S = \frac{\Delta L}{P} = \frac{3}{16} \frac{(1-\mu^2) r^4}{Y h^3} \quad \text{(S13)}$$

In these equations, r is the effective radius of the film, that is, the inner radius of the glass sleeve of the fiber optic F-P acoustic sensor, h is the thickness of the film, μ is the Poisson's ratio of the film, and Y is the Young's modulus of the film, ρ = 3.17 g/cm^3 is the material density. When the external sound pressure of the film is constant, a larger radius of the film or a smaller thickness of the film can enable a higher sensitivity. Besides, a smaller Young's modulus can also promote sensitivity. In practice, the above parameters which determine the sensitivity of the diaphragm cannot be freely designed. For instance, in fabrication, due to the thermal preparation, a diaphragm will suffer initial stress. As a result, h cannot be infinitely small. In this work, we use a thin film with a designed micro-structure instead of a pure flat film.

Based on biomimetic structures of insects, we designed 3 types of microstructured MEMS films. The structures are shown in **Fig. 3** in the maintext. For all the 3 types, r = 0.9 mm, h = 400 nm. Related to flat geometry, microstructures can reduce rigid resistance and residual stress by using the annular corrugated geometry. This can increase the μ and decrease the Y. Specifically, μ of a flat silicon nitride film (made via PECVD) is 0.2, Y of a flat silicon nitride film is 280 GPa [6].

Figure S2-3d illustrates the modeling of three biomimetic film types using CST Studio, a commercial software. Type 1 features an annular corrugated geometry with 5 concentric rings, where the minimum and maximum radii are 575 μm and 800 μm, respectively. Type 2 consists of 32 radial stripes, with an inner radius of 475 μm and an outer radius of 825 μm. Type 3 includes a large hexagon surrounded by 6 sets of smaller hexagons, with side lengths of 225 μm for the large hexagon and 75 μm for the small hexagons. For all types, the groove depth and width are consistently set at 25 μm

each. The film oscillation is stimulated by an acoustic pressure of 1 Pa (static), and its deformation in the vertical direction is depicted in **Fig. S2-3e**. Specifically, under quasi-static excitation, the acoustic sensitivity of the three types is 0.78 μm/Pa, 0.81 μm/Pa, and 0.84 μm/Pa, respectively. In **Fig. S2-3f**, the frequency responses of these diaphragms are analyzed. When varying the acoustic frequency from 0 to 20 kHz, Type 1 achieves its maximum sensitivity (0.9 μm/Pa) at frequencies above 10 kHz, Type 2 reaches a peak sensitivity (0.89 μm/Pa) around 1 kHz, and Type 3 displays its highest sensitivity (0.87 μm/Pa) at frequencies below 100 Hz.

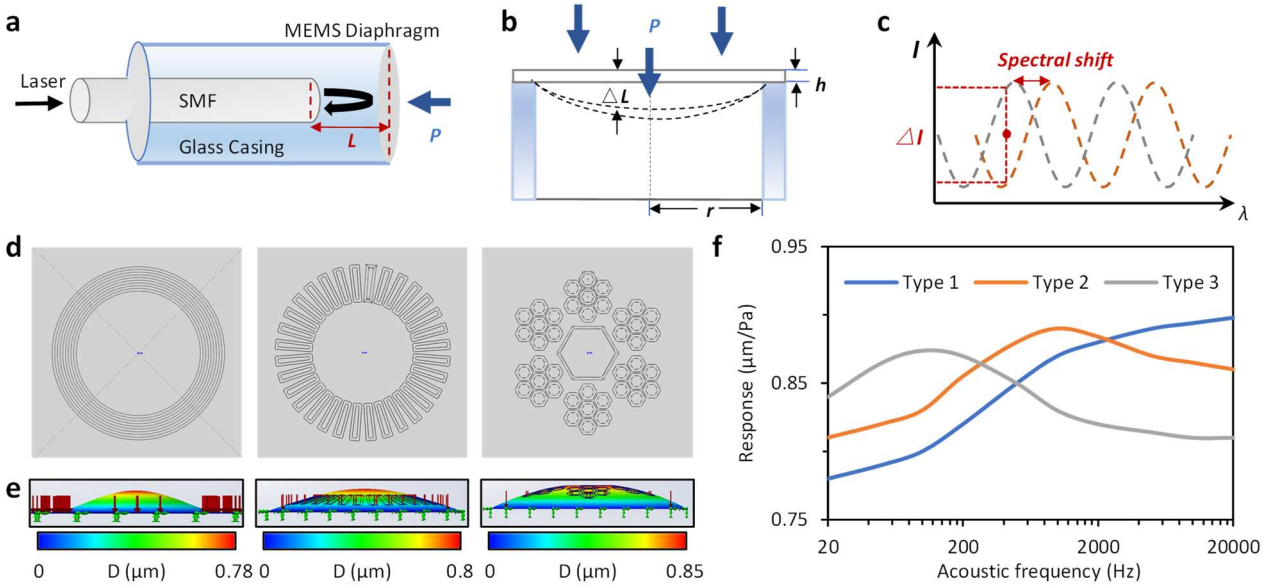

Fig. S2-3. Acoustic sensing using a MEMS diagram-based F-P cavity. a, Schematic diagram of an F-P cavity, which is sensitive to external pressure, SMF: single mode fiber. **b,** Principle of the response, external pressure changes the shape of the MEMS film, and tunes the cavity length of the F-P resonator. **c,** Alteration of the cavity length changes the optical resonance, which leads to reflected intensity modulation at a specific wavelength. **d,** Structures of diverse MEMS films. **e,** Acoustic pressure induced diaphragm displacements. **f,** Simulated sensitivity for different acoustic frequencies.

S2.4. Qualitative analysis of the influences from the laser instability.

In schematics, **Fig. S2-4a** shows the model in which we use a fixed laser line (e.g. a comb line) to detect the resonance shift. Since the reflectivity of the fiber/air facet and the air/MEMS film is pretty small (<5%), such an F-P resonator has a very low finesse (≈ 1). The resonant spectrum of our FOM could be approximately written in:

$$R(f) = r\left\{\exp[j\frac{4\pi fL}{c}]+1\right\} \tag{S14}$$

Here f is the optical frequency, r is the resonant magnitude, c is the light velocity, $L \approx 10^{-4}$ m is the

cavity length. Free-spectral-range (FSR) of every FOM is on 1.475 THz level around 1550 nm. This equation shows the case of dual-beam interference. When Q factor of the F-P resonator increases, the resonant dips would be narrower, as $Q = f/\Delta f$. Here Δf is the 3-dB resonant linewidth. On the other hand, power spectrum of the laser can be simplified as:

$$L(f) = l\,\mathrm{sech}^2\left\{\frac{f - f_0}{M}\right\} \tag{S15}$$

Full width at half maximum (FWHM) of $L(f)$ is $1.76M$ in fitting. The reflected laser power is $\int L(f)R(f)df$. Here l is the laser's peak power density, f_0 is the central frequency of the laser, and M determines the linewidth of the laser. For a fixed $L(f)$, when the reflection curve changes from $R(f)$ to $R'(f)$, the reflected laser power alters, as the red area shows in schematics. **Fig. S2-4b** shows the zoomed in curve in linear approximation, here $a \neq b$ suggests the resonance shift. In this case, the acoustic pressure induced reflected power alteration is:

$$\Delta P = l(b-a)\int \mathrm{sech}^2\left\{\frac{f - f_0}{M}\right\}df \tag{S16}$$

Here b-a is determined by the cavity length alteration. Now we consider that there exists power fluctuation Δl and frequency drift Δf in the laser between two measurements, i.e. the $L(f)$ writes:

$$L(f) = (l + \Delta l)\,\mathrm{sech}^2\{\frac{f + \Delta f - f_0}{M}\} \tag{S17}$$

Therefore, the acoustic pressure induced reflected power alteration in this case is:

$$\Delta P' = \Delta l m\int \mathrm{sech}^2\{\frac{f + \Delta f - f_0}{M}\}fdf + l(b - a + \Delta l)\int \mathrm{sech}^2\{\frac{f + \Delta f - f_0}{M}\}df \tag{S18}$$

We show this alteration schematically in the figure, $\Delta P'$ is the difference between the area of Zone B and Zone A. The noise-induced reflected power uncertainty is $N_P = \Delta P' - \Delta P$, and the final SNR is $\Delta P/N_P$. Typically, the FSR of the FOM (> 1 THz) is much larger than the linewidth of the laser (<< 1 GHz), we can linearly approximate the resonance curve and obtain a relation:

$$\mathrm{SNR} \propto \frac{l(a-b)}{\Delta l b + (l + \Delta l)\Delta f m} \tag{S19}$$

We summarize the relation as $\mathrm{SNR} \approx \zeta_1/[\zeta_2 \mathrm{RIN} + m\Delta f]$. Here ζ_1 and ζ_2 are constants determined by the FOM. Specifically, $\zeta_1 \propto a$-b, $\zeta_2 \propto m$. In practice, RIN and Δf can be directly measured by using noise analyzers. In **Fig. S2-4c**, by fixing $\zeta_1 = 10^{-6}$, $\zeta_2 = 10^{-1}$, we map the SNR varying with Δf and RIN. In sum, using a laser line with higher RIN and lower frequency drift is a key point to improve the accuracy of acoustic detection based on FOMs. In our experiment, when using a stabilized comb line, the frequency drifting is at 10^2 level, while its total RIN is at -90 dB level, meeting the expectation that SNR > 90 dB.

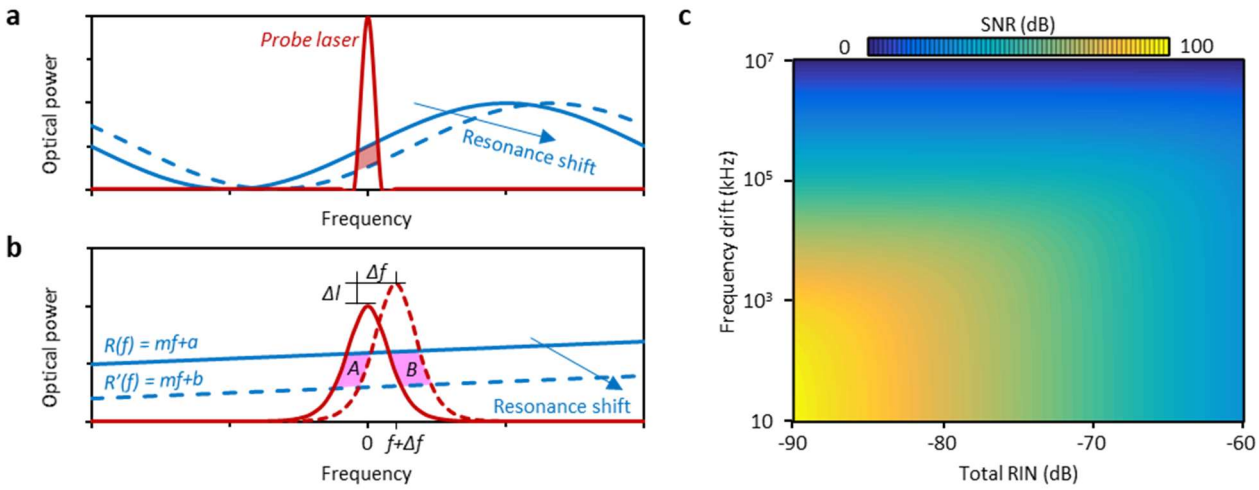

Fig. S2-4. Stabilization of the laser line plays a role in improving the SNR for acoustic detection. **a,** Schematic principle. **b,** Uncertainties from the laser line, including power fluctuation Δl and frequency drift Δf. **c,** Calculated SNR when fixing $\zeta_1 = 10^{-6}$, $\zeta_2 = 10^{-1}$.

S2.5. Acoustic positioning algorithm.

The 3-dimensional acoustic positioning relies on solving the propagation paths of sound waves. Leveraging multiplexed sensors, the times when the sound waves reach different sensors are distinct. Related to beam synthesis [7], this method is simple, cheap, easy to transport, and convenient to set up on-site. Moreover, it can offer high resolution when the arrival time analysis is accurate [8]. In our design, fiber optic microphones (FOMs) are placed at varied locations, the minimum distance (L_m) between them determines the sampling rate of sound localization. Typically, we use a L_m = 40 cm in the experiment, suggesting that the maximum sampling rate for acoustic waves is 850 Hz.

When using this sensor array system to measure the three-dimensional coordinates of an acoustic target in the open air, we introduce our positioning method based on the time difference of arrival (TDOA). Each FOM has an independent coordinate M_i (x_i, y_i, z_i), here i = 1, 2, 3 or 4, speed of sound is v_A = 340 m/s, $t_{Mi,Mj}$ signifies arrival time difference between M_i and M_j ($i \neq j$). Considering we have N acoustic sensors, we can solve $N(N\text{-}1)/2$ acoustic paths. For example, when N = 4 and the sensors' coordinates are $M_1(x_1, y_1, z_1)$, $M_2(x_2, y_2, z_2)$, $M_3(x_3, y_3, z_3)$, $M_4(x_4, y_4, z_4)$, and the coordinate of the target is *Target* (x, y, z), the sound arrival time difference between every two array elements is $t_{M2,M1}$, $t_{M3,M1}$, $t_{M4,M1}$, $t_{M3,M2}$, $t_{M4,M2}$, $t_{M4,M3}$, the following positioning equation can be obtained:

$$\begin{cases} F_1(x,y,z) = \sqrt{(x-x_2)^2+(y-y_2)^2+(z-z_2)^2} - \sqrt{(x-x_1)^2+(y-y_1)^2+(z-z_1)^2} - t_{\mathrm{M2,M1}} v_{\mathrm{A}} = 0 \\ F_2(x,y,z) = \sqrt{(x-x_3)^2+(y-y_3)^2+(z-z_3)^2} - \sqrt{(x-x_1)^2+(y-y_1)^2+(z-z_1)^2} - t_{\mathrm{M3,M1}} v_{\mathrm{A}} = 0 \\ F_3(x,y,z) = \sqrt{(x-x_4)^2+(y-y_4)^2+(z-z_4)^2} - \sqrt{(x-x_1)^2+(y-y_1)^2+(z-z_1)^2} - t_{\mathrm{M4,M1}} v_{\mathrm{A}} = 0 \\ F_4(x,y,z) = \sqrt{(x-x_3)^2+(y-y_3)^2+(z-z_3)^2} - \sqrt{(x-x_2)^2+(y-y_2)^2+(z-z_2)^2} - t_{\mathrm{M3,M2}} v_{\mathrm{A}} = 0 \\ F_5(x,y,z) = \sqrt{(x-x_4)^2+(y-y_4)^2+(z-z_4)^2} - \sqrt{(x-x_2)^2+(y-y_2)^2+(z-z_2)^2} - t_{\mathrm{M4,M2}} v_{\mathrm{A}} = 0 \\ F_6(x,y,z) = \sqrt{(x-x_4)^2+(y-y_4)^2+(z-z_4)^2} - \sqrt{(x-x_3)^2+(y-y_3)^2+(z-z_3)^2} - t_{\mathrm{M4,M3}} v_{\mathrm{A}} = 0 \end{cases} \quad \text{(S20)}$$

Based on numerical optimization methods, iterative algorithms are usually used to approximate the roots of the above equations. First, we provide an initial guess point. These nonlinear equations return a function, whose value is a vector, representing the residual of the system. This function would be called at each iteration step and calculate the residual of the system of equations based on the current guess point. Then, the optimization algorithm iteratively updates the current guess point. In each iteration step, it calculates the gradient (or approximate gradient) of the objective function and the Hessian matrix. With decreasing the residuals, the algorithm terminates when the solutions meet the convergence criterion.

Estimate the time delay between the arrival of the sound wave signal of the sound source and the four array elements, and then combine the positional relationship of the array elements based on obtaining the time delay to obtain the relative coordinates of the sound source relative to the origin of the array. Although the principle is simple and the cost is low, compared with manual methods, the efficiency is much higher, but there are still certain errors.

As mentioned, such a localization method is automatic, but relies on estimating the errors. The distance difference between the target and any two acoustic sensors is written $\Delta d_{i,j}=v_{\mathrm{A}}\times t_{\mathrm{M}i,\mathrm{M}j}$. When the sound path changes due to unknown external interference (e.g. noise, echo, wind noise, and other influences), a new tiny delay variable τ is introduced, and the distance calculated by the TDOA method becomes $\Delta d_{i,j}{}'=v_{\mathrm{A}}\times (t_{\mathrm{M}i,\mathrm{M}j}+\tau)$. Here, we simulate a scenario, in which we add random noises. Here four sensors with coordinates $N_1(-\frac{1}{2}L_{\mathrm{m}}, -\frac{1}{2\sqrt{3}}L_{\mathrm{m}}, 0)$, $N_2(\frac{1}{2}L_{\mathrm{m}}, -\frac{1}{2\sqrt{3}}L_{\mathrm{m}}, 0)$, $N_3(0, \frac{1}{\sqrt{3}}L_{\mathrm{m}}, 0)$, $N_4(0, 0, \frac{\sqrt{2}}{\sqrt{3}}L_{\mathrm{m}})$ are used, L_{m} = 0.4 m, and the target's coordinate *Target* (*x, y, z*) is arbitrarily set. Assuming four cases, SNR = 3 dB, 5 dB, 10 dB and 15 dB, **Fig. S2-5** shows the simulated results. Through 1000 calculations, we confirm that the error of localization becomes smaller when the SNR is higher. **Fig. S2-5a** shows the calculated Δd, where $\Delta d = [(x-x_{\mathrm{m}})^2+(y-y_{\mathrm{m}})^2+(z-z_{\mathrm{m}})^2]^{1/2}$. Here ($x$, y, z) and (x_{m}, y_{m}, z_{m}) are the set coordinate and the calculated coordinate of the target, respectively. According to statistics, **Fig. S2-5b** shows the distribution intervals of the measurement error, while **Fig. S2-5c** shows when we want an error < 2 cm, how many times we need to measure.

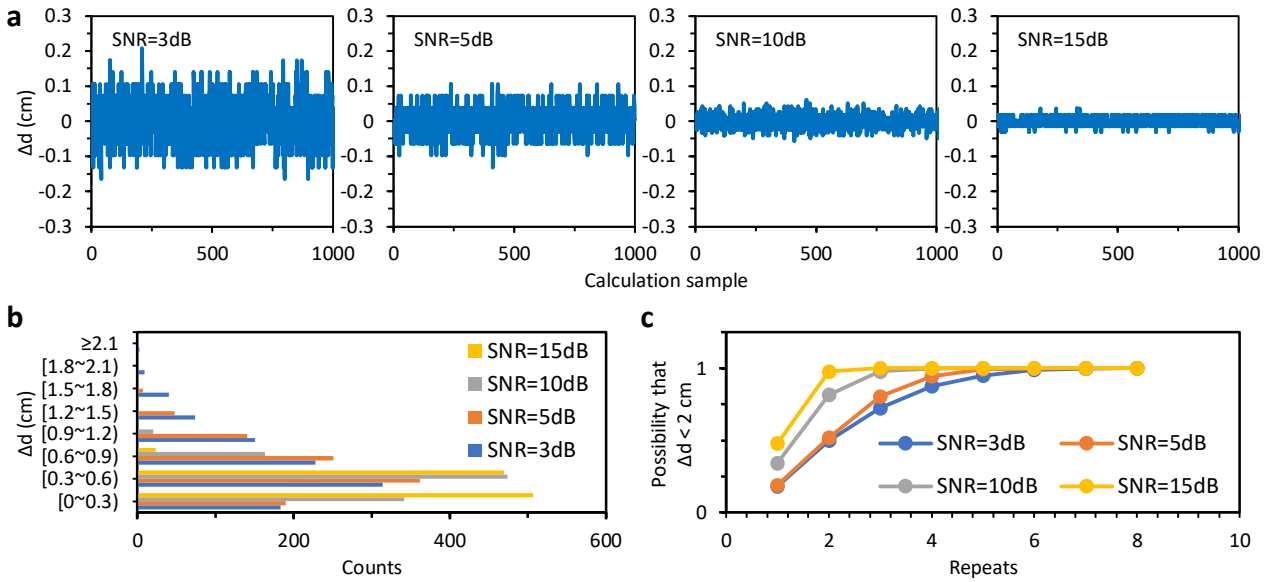

Fig. S2-5. Simulated accuracy when using the TDOA. a, 1000 measured Δd numbers when the SNR is 3 dB, 5 dB, 10 dB, and 15 dB. **b,** Statistical histogram of TDOA error. **c,** Probability to reach an error < 2cm after repeated measurements.

Note S3. Supplementary experimental details.

S3.1. Preparation of the on-chip silicon-nitride microrings.

The generation of optical frequency combs using silicon nitride (Si_3N_4) microrings demands precise fabrication techniques to create high-Q resonators with minimal losses. The Damascus process is particularly effective, offering a method that produces ultra-smooth waveguides with high geometrical accuracy, essential for frequency comb applications. Fabrication starts with depositing a silicon nitride layer onto a silicon dioxide substrate through low-pressure chemical vapor deposition (LPCVD), which ensures a low-stress film suitable for high-Q applications. The precise thickness of the SiN layer is critical, as it determines the waveguide's dispersion, directly influencing the formation of soliton microcombs. To achieve the necessary anomalous dispersion conditions for bright solitons, we used a silicon nitride thickness of 840 nm.

After deposition, a patterned hard mask is applied using advanced photolithography. The microring resonator structures are etched into the silicon nitride layer using reactive ion etching (RIE), ensuring steep sidewalls with minimal surface roughness. This is crucial to reduce scattering losses, as any imperfections in the waveguide can degrade the quality factor (Q) and impact the stability of soliton states in frequency comb generation. Following the etching process, chemical-mechanical planarization (CMP) is performed, a crucial step in the Damascus process to achieve a flat and smooth top surface. This step minimizes propagation losses and ensures the long-term stability of the optical

frequency combs. Planarization also provides better control over the resonator's geometry, vital for dispersion engineering and the formation of broadband combs.

The final structure, illustrated in **Fig. S3-1a**, boasts a high *Q*-factor and low propagation loss, making it ideal for pumping with continuous-wave lasers to generate Kerr soliton frequency combs. **Fig. S3-1b** and **Fig. S3-1c** show our microring chip post-fabrication, here we also show more details of the ring. Repetition rate of each ring cavity is 25 GHz, the cross-sectional area of the ring waveguide is 2480 nm × 840 nm, ensuring single-mode transmission, with a gap of 680 nm between the ring and the bus. This ensures quasi critical coupling.

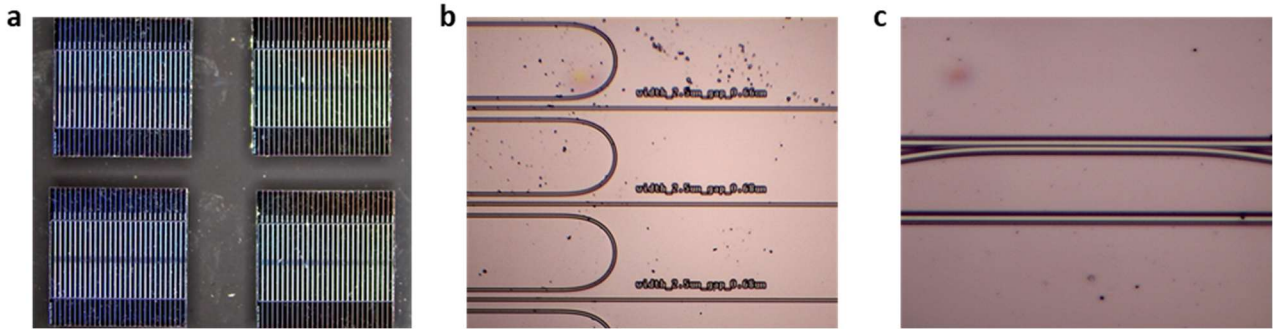

Fig S3-1. Fabrication and characterization of on-chip Si_3N_4 microring. a, Photo of on-chip microrings. **b-c,** Optical microscopic pictures, showing the waveguide details.

S3.2. Preparation of micro-structured MEMS films.

The manufacturing process flow of microstructured MEMS films is shown in **Fig. S3-2**. The fabrication steps can be summarized as follows (**Fig. S3-2a**): First, we cover a layer of evenly coated photoresist on the silicon wafer substrate (i). Then a photolithography mask with designed pattern is prepared (ii). After photoresist lithography, the pattern on the photoresist mask is transferred to the photoresist coating on the silicon wafer surface (iii). After the pattern is successfully transferred, the silicon wafer with the shape of the photoresist coating is etched using the reactive ion etching (RIE) method [9], see step (iv). Afterwards, the photoresist coating on the surface of the silicon wafer is immersed in pure acetone and washed away via ultrasonic vibration cleaning. Then, a microstructured silicon nitride film was deposited on the surface of the silicon substrate by plasma-enhanced chemical vapor deposition (PECVD), as step (v) shows. Finally, the silicon substrate is etched away to obtain the silicon nitride film with a thickness of several hundred nanometers (vi). In **Fig. S3-2b**, we show the samples of the fabricated MEMS silicon nitride films, verifying that this component can be prepared on a large scale, with high consistency. The size of every sample is 1.9×1.9 mm^2. Thickness of the Si_3N_4 MEMS film is 400 nm.

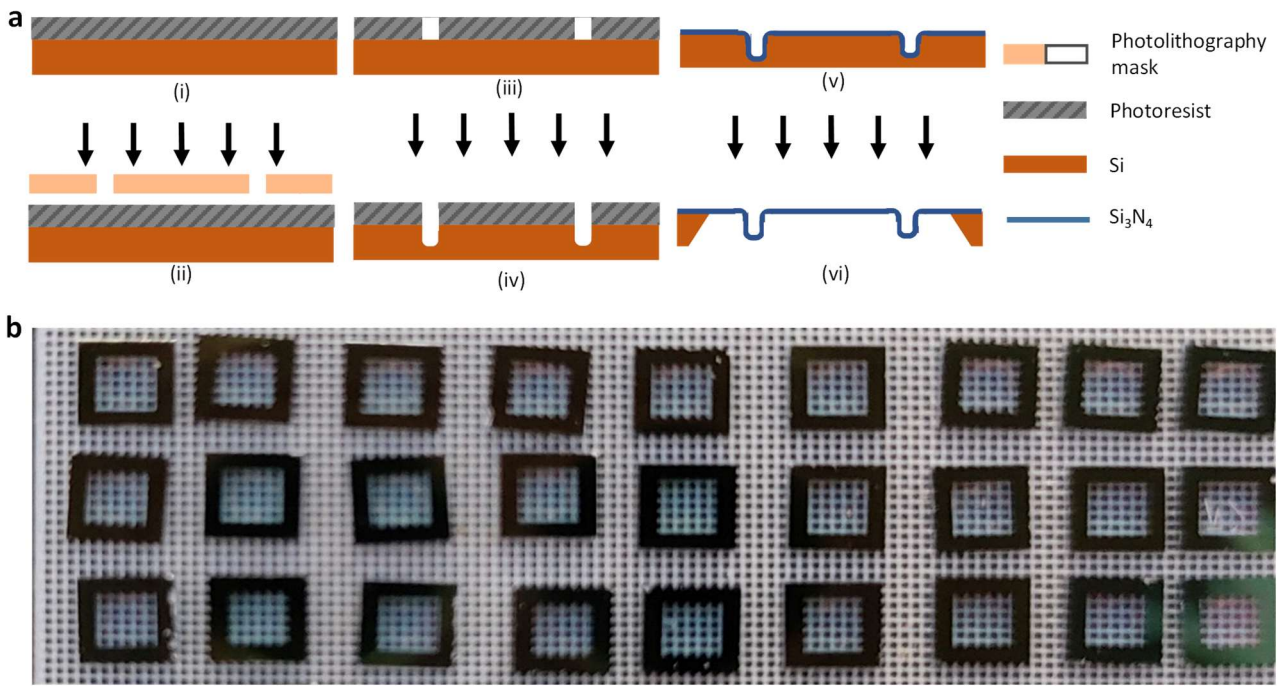

Fig.S3-2. Preparation and characterization of the micro-structured MEMS films. a, Manufacturing process flow of the MEMS diaphragms. (i) Photoresist coating; (ii) Photolithography mask; (iii) Photolithography treatment; (iv) Reactive ion etching; (v) Depositing the silicon nitride film on the surface of the silicon wafer via PECVD; (vi) Removing the silicon substrate. **b,** Picture of MEMS film samples.

S3.3. Fabrication and characterization of fiber optical microphones.

The specific manufacturing process of fiber optic acoustic wave sensors is shown in **Fig. S3-3a**. First, we prepare a capillary glass tube (i), whose inner diameter is $D_0 = 0.127\pm0.001$ mm (suitable for fixing and calibrating a single mode fiber, or SMF). Its outer diameter is $D_1 = 1.8\pm0.01$ mm. Then we put an SMF with a flat-cut end-face into the capillary glass tube, and fix its position in depth using glue (ii). The distance between the fiber end and the capillary glass tube end is L_1. Afterwards, we use a large-diameter glass sleeve to attach the MEMS sensing film. This forms another reflective surface of the F-P cavity (iii). The distance between the sleeve end and the end of the capillary glass tube is L_2. The inner diameter of the glass sleeve is $D_2 = 2.8\pm0.01$mm, and its outer diameter is 4.0 ± 0.01 mm. Then, we put the MEMS film on the sleeve (iv), and optimize the total cavity length ($L = L_1 + L_2$). All the structures are finally fixed by using UV glue. **Fig. S3-3b** shows the picture of the FOM device, during fabrication. Its spatial parameters are precisely controlled by using a displacement table. The final product is packaged with an outer diameter of 4 mm. Optimizing the number $L = L_1 + L_2$ is significant to achieve a higher extinction ratio (*ER*). Specifically, reflections of light from the end face of a single-mode fiber can be equivalent to the Fresnel reflection under normal incidence. Typically,

the reflection coefficient of the single-mode fiber end face is R_{Fiber} = 3.614%. And, the surface reflection coefficient of the MEMS film is approximately R_{MEMS} = 18.237%.

To achieve higher reflectivity, a method is gold (Au) thin films on the two reflective surfaces of the cavity. **Fig. S3-3c** shows the picture of our MEMS film before and after Au coating, maximum thickness of the Au film is 20 nm. In **Fig. S3-3d**, we show that in 1520 nm to 1580 nm band, after reflection enhancement, reflection loss of the fiber end is 0.043, while reflection loss of the MEMS facet is 0.032. According to the light intensity loss theory of an optical fiber F-P cavity, the transmission loss ε of the Faber-Perot cavity can be expressed as the following:

$$\varepsilon=4[1+(\frac{2\lambda L}{\pi n_0 r_0^2})^2]/[2+(\frac{2\lambda L}{\pi n_0 r_0^2})^2]^2 \tag{S21}$$

Here λ is the wavelength of the incident light, L is the length of the F-P cavity, n_0 is the refractive index of the F-P cavity medium, and r_0 is the beam mode field radius. Here, for single-mode fiber, n_0=1, r_0 = 3.2 μm, λ = 1550 nm, $L \approx$ 100 μm, ε = 0.04. Therefore, total loss per roundtrip of the F-P cavity is l = 0.115. Referring the cavity $Q = 2\pi f T_r/l$, we obtain Q factor of our F-P cavity can reach 7185. Here f = 193.5 THz is the optical frequency, T_r = 0.68 ps is the roundtrip time.

In our acoustic mapping system, to achieve the optimal modulation effect of the F-P cavity on the comb teeth, it is essential to precisely control the resonance position. This can be conveniently realized by accurately tuning the cavity length. During fabrication, the reflected spectrum of our F-P cavity is precisely adjustable on-line during the packaging process [10]. As shown in **Fig. S3-3e**, we can accurately tune the resonance wavelength with a resolution exceeding 10 GHz. By considering the maximum Q factor of our F-P microcavity, this tuning ensures that the F-P resonance aligns precisely at the desired wavelength when synchronizing with the comb output. In **Fig. S3-3f**, we compare the Q factor of the F-P microresonator before and after Au coating. Typically, Resonance linewidth before and after Au coating is 0.78 THz and 26.9 GHz, this suggests that the high reflectivity coating scheme enables Q improvement from 248 to 7185.

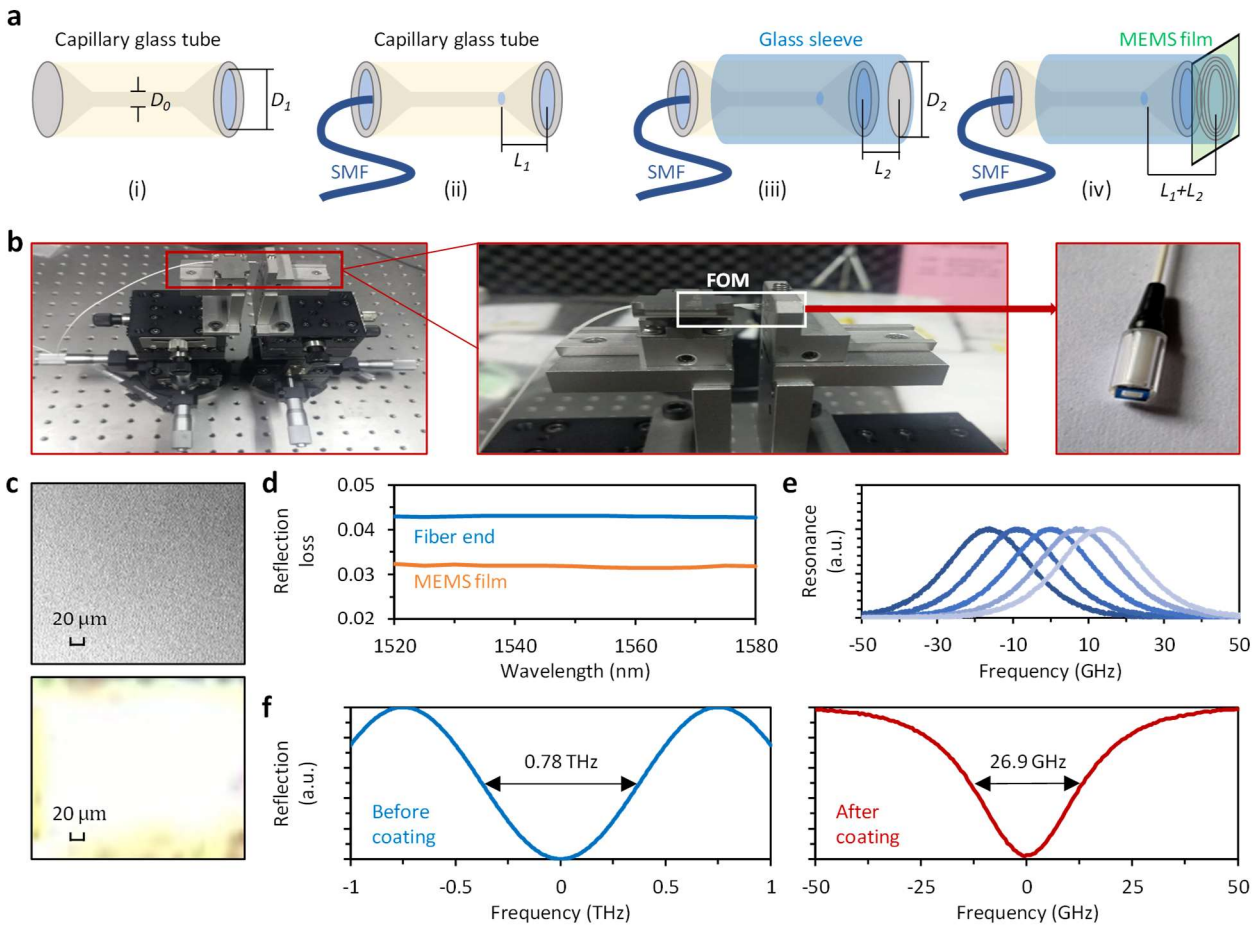

Fig.S3-3. Fabrication and characterization of FOM microcavity. a, Manufacturing process flow of the FOM device. **b,** Experimental picture shows how we calibrate the fiber in an F-P microresonator. **c,** Pictures demonstrate the MEMS film before and after Au coating. **d,** Reflection loss of the fiber end and the MEMS facet. **e,** Controllable resonance of our F-P cavity. **f,** Measured linewidth of a F-P microcavity before and after Au coating.

S3.4. Characterization of the reference for acoustic sensing.

We use a standard low-noise free-field TEDS Microphone Bruel & Kajar Type 4955 (B&K 4955) as the reference for calibrating the performances of our FOMs. **Fig. S3-4a** displays the picture of the B & K 4955, whose acoustic probe volume is much larger than our FOM. By using a standard acoustic source with bandwidth 20 Hz to 20 kHz, we show the broadband response spectrum of the B & K 4955 in **Fig. S3-4b**. In the band 100 Hz to 10 kHz, the B&K 4955 shows a response higher than 118 dB. Its typical noise base in sound pressure level (*SPL*) is 5.5 dB, while its dynamic range reaches 110 dB (in *SPL*). Here, $SPL = 20\times\log(P_A/20\ \mu Pa)$, P_A is the actual sound pressure. Accordingly, in principle, a B & K 4955 can detect acoustic pressure from 0 ~ 6.3 Pa, and has a noise limited detectable acoustic pressure 37.6 μPa. In **Fig. S3-4c**, we test the noise base of the B & K 4955, using two fixed acoustic

frequencies 1 kHz and 10 kHz, with acoustic pressure 37 mPa. Signal to noise ratio (SNR) of the B & K 4955 approaches 89.2 dB. Referring the resolution bandwidth of our audio analyzer is 2 Hz, the minimum detectable pressure (MDP) of the B&K 4955 is 0.91 μPa/Hz$^{1/2}$. Here MDP = $[P_A^2/(BW*SNR)]^{1/2}$. In **Fig. S3-4d**, we measure the sensitivity of the B & K 4955, it shows a linear sensitivity of 1.046 V/Pa.

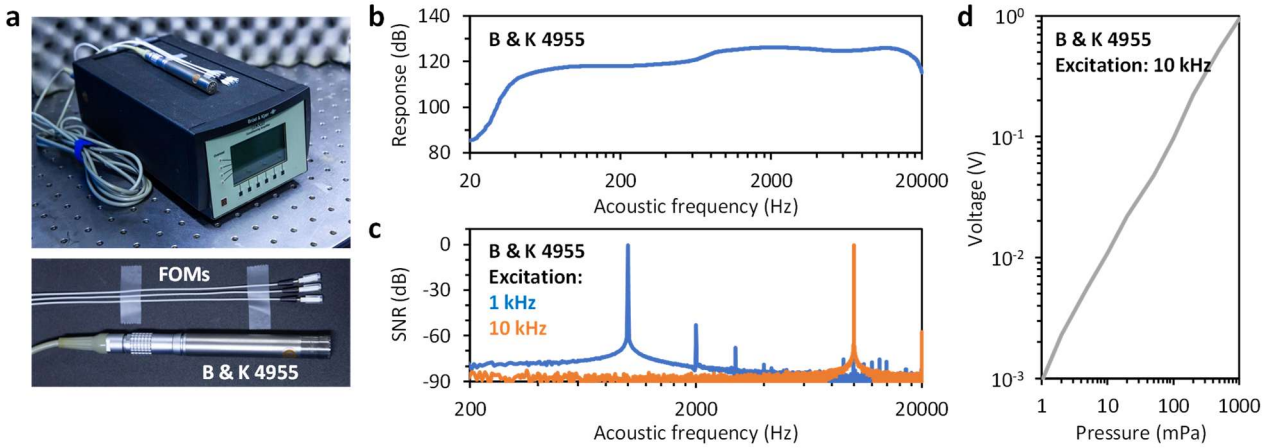

Fig.S3-4. Test of the B & K 4955. a, Picture of the acoustic reference (B & K 4955). **b,** Response of the B & K 4955. **c,** Signal to noise ratio of the B & K 4955. **d,** Sensitivity of the B & K 4955.

S3.5. Characterization of the fiber optic microphones.

Figure S3-5a shows the experimental setup and environment measuring the response of an FOM. We use the microcomb device to drive FOMs, the circulator (CIR) is used to collect the reflected light, and the PD is used to filter out the optical frequency (≈ 193 THz) while detecting the acoustic wave (with frequency < 20 kHz). A frequency tunable speaker provides an acoustic signal in the band 20 Hz ~ 20 kHz. The speaker, FOM and the reference are fixed in a silent chamber to cancel environment noises. Here, we also show the picture in the silent chamber.

Figure S3-5b shows the measured responses of several FOM samples. Top panel displays that measured noise floors 12 FOMs. In the acoustic band 20 Hz ~ 20 kHz, the maximum noise floor ≈ -150 dBc/Hz @ 600 Hz, the minimum noise floor ≈ -170 dBc/Hz @ 4 kHz. Bottom panel shows that the maximum response is ≈ 140 dB @ 12 kHz, the minimum response is ≈ 75 dB when the acoustic frequency < 80 Hz. Response of a typical FOM is flat in the range of 200 Hz to 2 kHz. **Fig. S3-5c** demonstrates measured acoustic directionality of our FOM based on F-P cavity. Here the 0 degree marks the position directly facing the FOM. Using an acoustic wave with frequency 2 kHz, the blue dots show the dual-comb beating intensity, while the red dots show the acoustic modulation intensity.

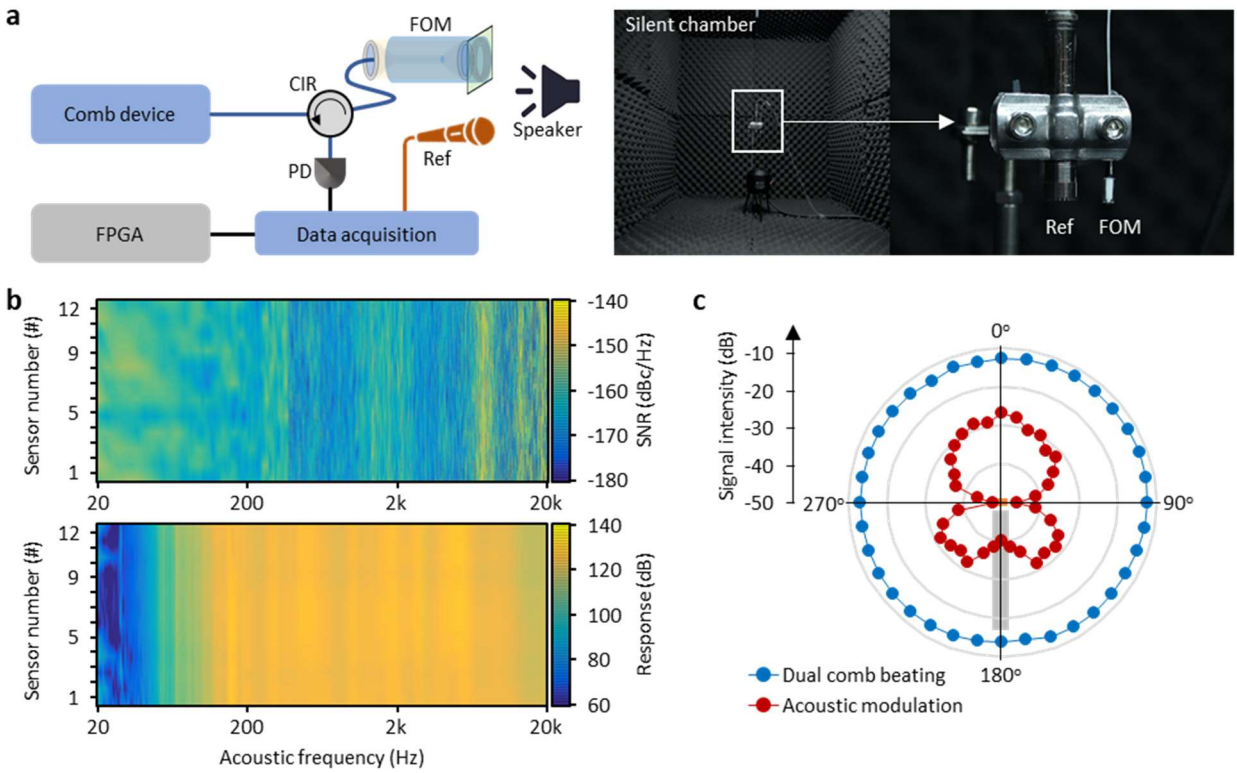

Fig.S3-5. Response and SNR of our FOMs. a, Setup to measure the response and SNR of an acoustic sensor. Here we also show picture of the measurement environment. **b,** Measured SNR (top) and response (bottom) of our F-P cavities based FOMs. **c,** Measured acoustic directionality of a FOM.

S3.6. Devices for stabilizing our on-chip dual microcombs.

In the extended data figure and methods of our maintext, we show the experimental setup for the generation and stabilization of dual Kerr soliton microcomb. Two laser diodes are used as the optical pumps. These two lasers drive two separate silicon nitride microrings on chip. The two microrings have slightly different repetition rates. Their difference in repetition frequencies (Δf_{rep}) is 4.1 MHz. Thanks to the high Q factor (4.6×10^6), soliton threshold of each microring is below 80 mW [11]. We achieve full stabilization of the dual comb outputs compactly via optoelectronic feedback to optical and electrical references. An ultra-stable microcavity is used for stabilizing the pump laser and the line#20 of comb#1, while a RF reference stabilizes the Δf_{rep}.

Figure S3-6a shows the picture that our ultra-stable F-P microcavity is in test. **In Fig. S3-6b**, we illustrate the ring-down curve of the ultra-stable cavity, suggesting a loaded Q factor 6.05×10^9. Here the spectral scan speed is 125 GHz/s. In **Fig. S3-6c**, we demonstrate the long-term stability of the ultra-stable F-P microcavity, in which the temperature is controlled by a TEC with resolution 0.01 K. In a 5-minute period, resonance drift is less than 100 Hz. The typical spectrum of our electrical reference integrated in FPGA is shown in **Fig. S3-6d**. The instantaneous linewidth of this reference is less than

1 Hz, while its integrated linewidth is less than 10 Hz.

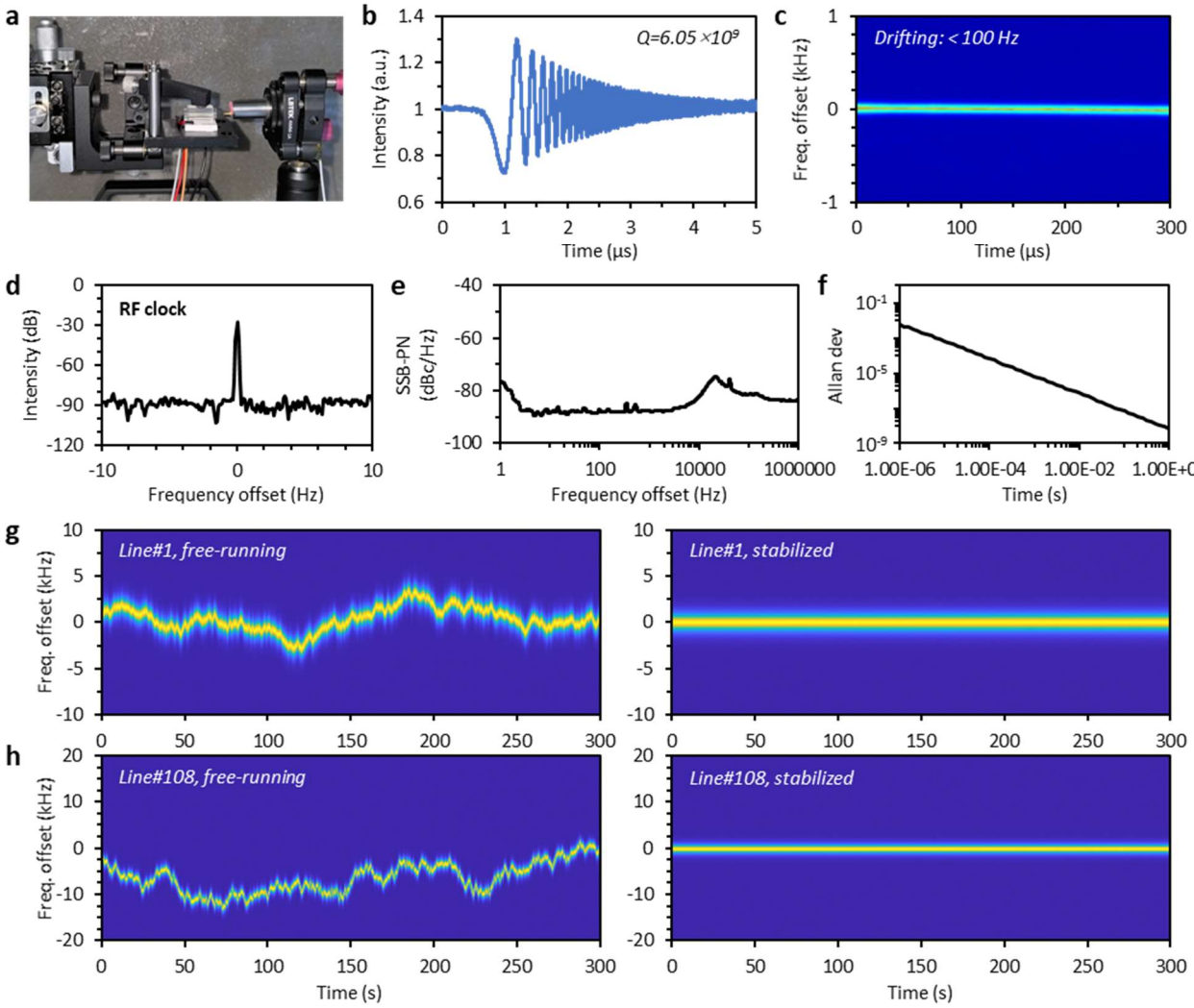

Fig.S3-6. Stabilization of the microcomb devices. a, Picture of the ultra-stable F-P cavity in test. **b,** Measured ring-down curve of the ultra-stable cavity. **c,** Measured long-term stability of the cavity. **d,** Measured spectrum of the clock reference in our FPGA. **e-f,** Measured SSB-PN and Allan deviation of the reference. **g,** Measured long term frequency drifting of the comb line#1, before and after stabilization. **h,** Measured long term frequency drifting of the comb line#108, before and after stabilization.

In **Fig. S3-6e and S3-6f**, we show single-sideband phase noise (SSB-PN) and Allan deviation of the reference. During the stabilization process, frequency down-converted beat notes are collected by an integrated data acquisition card and analyzed in the FPGA. Within the FPGA, low-pass filters are employed to eliminate high-frequency noise, and the beat signals are processed through proportional–integral–derivative (PID) controllers. Using an optical time and frequency standard (Menlo system FC1500-ULNnova-ORS), we can characterize the spectral uncertainties of the pump laser, the first

comb line of comb#1, and the first comb line of comb#2. **Figs. S3-6g** to **S3-6h** illustrate the stability of the comb#1 (1st line), and the stability of the comb#1 (108th line). The measured results suggest that all optical frequencies of the two microcombs have been well locked. In the 4 colored maps, the left panels display the free-running case, while the right panels show spectral drift after stabilization. The results indicate that before locking, the linewidths of the pump laser and comb lines in the optical band are pretty large. After full stabilization, their linewidths are reduced to single Hz level. Meanwhile, the locking operation significantly enhances long-term stability.

Note S4. Extended discussions.

S4.1. Discussion of the trade-off between sensitivity and dynamic range of a FOM.

When using FOMs as acoustic probes, there is a trade-off between sensitivity and dynamic range. Increasing the Q factor of a FOM enhances its sensitivity, but reduces its dynamic range due to the narrower resonance linewidth. **Fig. S4-1a** schematically illustrates this trade-off. The acoustic wave-induced free spectral range shift of an F-P cavity is given by $\Delta FSR = c/2L_2 - c/2L_1$, where L_1 and L_2 are the cavity lengths before and after exposure to acoustic pressure, and c is the speed of light. For our 400 nm thick MEMS film, the typical response to acoustic pressure is 0.8×10^{-14} m/μPa, resulting in an acoustic pressure-induced ΔFSR of 1.2 kHz/μPa. Assuming the measured resonance corresponds to the Nth optical resonance, where $N = f_o/FSR$, and $f_o \approx 193.5$ THz is the optical frequency, the cavity length in our experiment is approximately 100 μm, typically yielding $N = 129$ at a testing wavelength around 1550 nm. Consequently, the frequency sensitivity S_F is approximately 155 kHz/μPa. When normalizing the resonance intensity to 1, the intensity sensitivity of the FOM is $S_I = QS_F/f_o$ under linear approximation. The dynamic range of a FOM is given by f_o/QS_F. In **Table S2**, we present theoretical calculations of the S_I and dynamic range with varying parameters.

Table S2. Calculated MDP and dynamic range based on different parameters.

	Q	***S_I***	***Dynamic range***
Case 1: F-P cavity with highly reflective films	7185	5.76×10^{-6} /μPa	0~174 mPa
Case 2: F-P cavity with weakly reflective films	248	1.99×10^{-7} /μPa	0~5.03 Pa

Figure S4-1b shows measured data for two F-P microcavities: one with a high Q factor (7185) and another with a low Q factor (248, two-beam interference, f_o = 193.5 THz, resonance width 0.78 THz). We note that during fabrication, the Q factor of our F-P cavities can be adjusted by modifying the coating parameters of the high-reflection films. For instance, using a FOM with Q = 7185 results in a sensitivity of 15.1 V/Pa, but with a smaller dynamic range up to approximately 150 mPa. In

contrast, a FOM with $Q = 248$ offers a sensitivity of 0.52 V/Pa but a significantly larger dynamic range. In practice, fine designing a FOM offers selectivity for different sensitivities and dynamic ranges.

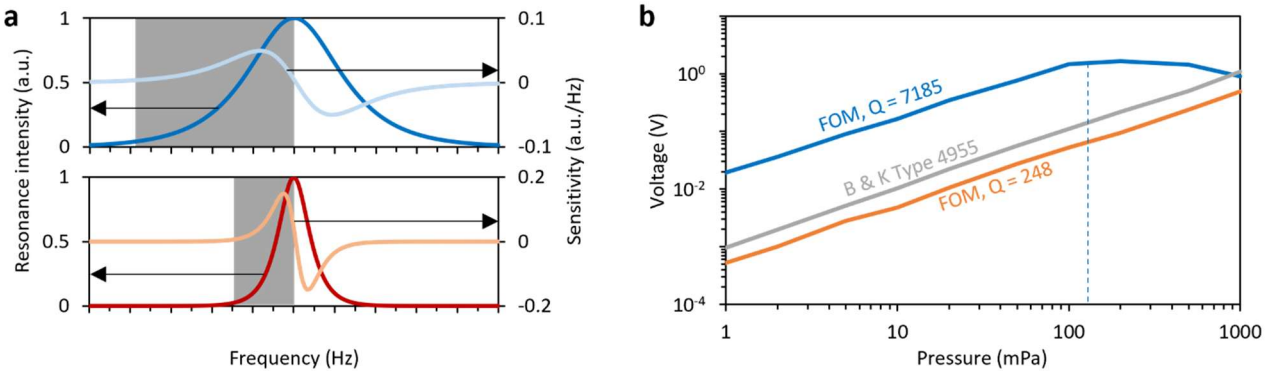

Fig.S4-1. Sensitivity and dynamic range of a FOM. a, Schematic diagram showing that a higher Q factor enables a higher sensitivity, but smaller dynamic range for acoustic sensing. Here the grey regions mark the half widths of the resonances, which scales with the dynamic range. **b,** Measured responses of different microphones: blue: FOM with high Q factor, orange: FOM with low Q factor, grey: B & K Type 4955 as a reference. Here the blue dashed line marks the dynamic range of the FOM with high Q factor.

S4.2. Discussion of the noises using FOMs.

In acoustic sensing process, sensitivity is one of the most important parameters, which determines the ability to detect weak acoustic waves. It is defined as the minimum detectable sound pressure (MDP). When using light to read out the signal, in principle, sensitivity is typically characterized by the noise equivalent pressure density (NEPD) [12]. Typically, thermal noise plays the major role in acoustic sensing, therefore one can write the NEPD of an acoustic sensor when ignoring the shot noise:

$$NEPD = \frac{1}{r\zeta A}\sqrt{\frac{2\gamma k_B T}{m\left[\left(\omega_M{}^2 - \omega^2\right)^2 + \omega^2\gamma^2\right]}} \tag{S22}$$

Here $r \approx 1$ is represents the ratio of the pressure difference between the upper and lower surfaces of the film to the peak pressure at the antinode of the incident acoustic wave, ζ is the spatial overlap between the incident sound and the mechanical displacement profile, $A = 3.6\times10^{-6}$ m^2 is the sensor area. Besides, the parameters m and γ represent the effective mass and damping rate of the FOM, ω_M and ω are the mechanical resonant frequency of the sensing film and the acoustic frequency, while k_B is the Boltzmann's constant, T is the temperature. In mechanics, $\omega_M = \mu^2(Y/\rho h^2)^{1/2}/2R$, here h is the thickness of the film, $\mu = 0.25$ is the Poisson's ratio of the film, and $Y = 280$ GPa is the Young's modulus, $\rho = 3.17$ g/cm^3 is the material density, R is the radius.

Based on this physical model, we see that a larger film area, a lower temperature, a higher material density and a detecting acoustic frequency far away from the intrinsic frequency can be helpful for improving the SNR, or decreasing the NEPD. In **Fig. S4-2a**, we calculate the ω_M related to geometric parameters. Since our silicon nitride film has very large Y while very small h, the ω_M is on tens of Grad level, far beyond acoustic frequencies. **Fig. S4-2b** calculates the NEPD, in which we mainly vary the parameters γ and T, as $\omega_M^2 - \omega^2 \approx \omega_M^2$. We find that a smaller T or smaller γ can induce a lower NEPD. In these calculations, we mark our experimental parametric space using the white dot. In measurement, we test that noise base of our comb driving FOM sensor is on -110 dBc/Hz level, approaching the thermal noise limit.

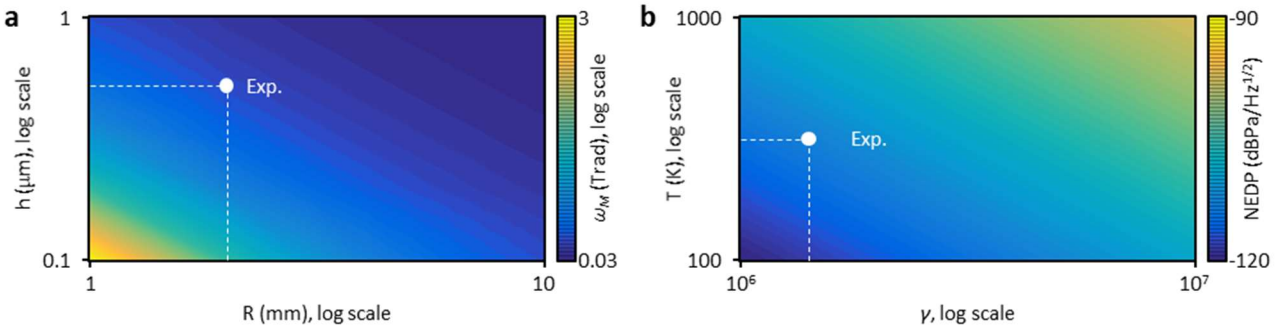

Fig. S4-2. Parameters determine the sensor noise. a, Calculated map shows that the mechanical frequency varies with film radius and thickness. **b,** Calculated map demonstrates that the NEPD is determined by damping rate and temperature.

S4.3. Extended data in acoustic localization experiment.

Before the application out-of-lab, we demonstrate that our on-chip dual-comb based FOM array is able to achieve high-precision passive acoustic target localization in-door. The concept is illustrated in the top panel of **Fig. S4-3a.** First, each FOM has an independent coordinate Mi (x_i, y_i, z_i), here i = 1, 2, 3 or 4. When an object situated at the coordinate (x, y, z) emits sound waves, the unique distances (e.g. D_1, D_2, D_3 and D_4) between every FOM and the target will result in variable pairwise arrival time delays for each FOM pair: $v_A \times t_{M2,M1} = D_2 - D_1$; $v_A \times t_{M3,M1} = D_3 - D_1$; $v_A \times t_{M4,M1} = D_4 - D_1$; $v_A \times t_{M3,M2} = D_3 - D_2$; $v_A \times t_{M4,M2} = D_4 - D_2$; $v_A \times t_{M4,M3} = D_4 - D_3$. Here v_A = 340 m/s signifies the acoustic velocity, $t_{Mi,Mj}$ signifies arrival time difference between M$_i$ and M$_j$ ($i \neq j$). Meanwhile, D_1, D_2, D_3 and D_4 can be represented in the coordinate system, such as $D_1 = |(x_1\text{-}x)^2 + (y_1\text{-}y)^2 + (z_1\text{-}z)^2|^{1/2}$. The spatial position of the target can thus be determined. Typically, a minimum of four acoustic detectors are required to locate a target in three dimensions. Employing additional acoustic detectors at different locations will add redundancy but enhance the accuracy of the positioning. In this measurement, we use 8 FOMs (M$_1$

~ M_8), enabled by dual comb technology, to localize an indoor sound source. **Table S3** displays the locations of the FOMs (M_1 ~ M_8) and the target (T). The bottom panel of **Fig. S4-3a** demonstrates this design.

Table S3. Locations of the FOMs and the acoustic target.

Location	x (m)	y (m)	z (m)
M1	0	0	0
M2	3	0	0
M3	3	3	0
M4	0	3	0
M5	0	0	3
M6	3	0	3
M7	3	3	3
M8	0	3	3
T	2.2	1.4	0.8

As a result, the distances between the target and detector are as follows: D_1 = 2.728 m, D_2 = 1.8 m, D_3 = 1.96 m, D_4 = 2.835 m, D_5 = 3.412 m, D_6 = 2.728 m, D_7 = 2.835 m, D_8 = 3.499 m. While playing classical piano music from the sound source, **Fig. S4-3b** presents the acoustic traces detected in different FOMs. There are noticeable temporal misalignments in waveform. Subsequently, we determine the delay difference of these detected acoustic traces across all pairwise FOMs by employing the equation $R(\tau) = \int_L M_i(t)M_j(t+\tau)dt$. Here L signifies temporal length of the sampled trace and $i \neq j$. The maximum value of $R(\tau)$ identifies the delay difference denoted by τ. **Fig. S4-3c** illustrates the cross-correlations between M_i and M_j ($i \neq j$), which are calculated based on the measured outcomes from our FOMs.

As demonstrated in **Fig. S4-3d**, we can retrieve the distances from D_1 to D_8. The specific measurements obtained are as follows: D_1 = 2.723 m, D_2 = 1.8 m, D_3 = 1.96 m, D_4 = 2.831 m, D_5 = 3.41 m, D_6 = 2.731 m, D_7 = 2.829 m, D_8 = 3.493 m. By solving the matrix equations $Di = |(x_i\text{-}x)^2 + (y_i\text{-}y)^2 + (z_i\text{-}z)^2|^{1/2}$, we can determine the spatial location of the target: T (2.19 m, 1.42 m, 0.8 m). This result aligns well with the actual number. **Fig. S4-3e** compares the measured coordinates and the actual coordinates, showing that the average error in the x, y, or z direction is less than 1.5%. **Fig. S4-3f** shows the outcomes of 100 repeated measurements, where the Root Mean Square (RMS) errors for x, y and z reach 0.62 cm, 0.56 cm and 0.6 cm respectively.

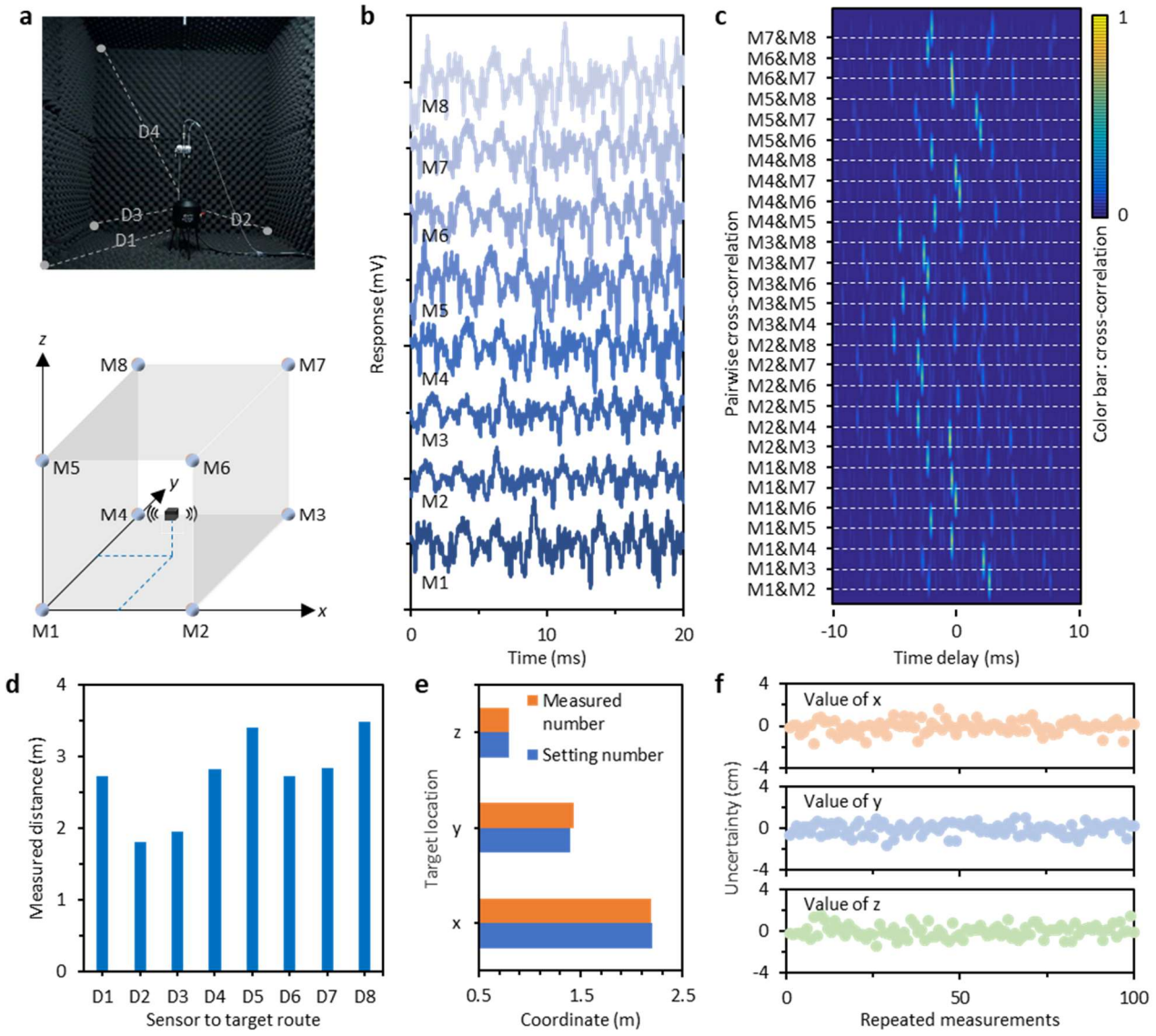

Fig. S4-3. Localization of an acoustic target in 3D space. a, Top: Conceptual mechanism in a picture that ≥ 4 acoustic sensors can localize a target via solving the acoustic path matrix. Bottom: experimental design that uses 8 comb-driven FOMs to localize an acoustic target. M1 ~ M8 are coordinates of the FOMs, *T* is coordinate of the target. **b,** Measured acoustic traces from M1 to M8. **c,** Measured cross-correlations. These traces reveal the target-sensor distances. **d,** Measured target-sensor distances *D*1 *to D*8. **e,** Retrieved target location, with measured coordinates x = 2.19 m, y = 1.42 m, z = 0.8 m (orange columns). In comparison, the blue columns show the true numbers. f, Repeatedly measured uncertainties of x, y, and z. Maximum RMS error is 0.62 cm.

S4.4. Influences of the Doppler effect.

When utilizing multiple sensor probes to localize a mobile target, it is important to consider the Doppler effect. This phenomenon describes how relative motion between a wave source and a receiver

affects the frequency of the wave. In situations where there is relative motion between the wave source and the receiver, the relationship between the received wave frequency (f') and the original wave frequency (f) can be expressed as follows:

$$f' = f \times \frac{v_A \pm v_r}{v_A \mp v_s} \tag{S23}$$

Here f' is the frequency received by the receiver, f is the original frequency emitted by the wave source, v_A is the acoustic speed, and v_r is the speed of the receiver relative to the medium. And, v_s is the velocity of the wave source relative to the medium. Now we discuss several factors that Doppler effect may contribute in the system. **Fig. S4-4a** shows the case schematically. We consider a simple case: there are two sensors, FOM#1 (x_1, y_1), FOM#2 (x_2, y_2), and a moving target M (x, y). Acoustic frequency of the target is f. Relative speeds are written in:

$$v_1 = v\cos\left[arctg(\frac{y-y_1}{x-x_1})\right]; v_2 = v\cos\left[arctg(\frac{y-y_2}{x-x_2})\right] \tag{S24}$$

Therefore, the detected acoustic frequencies are $f_1 = fv_A/(v_A-v_1)$; $f_2 = fv_A/(v_A-v_2)$. Traces detected by the two FOMs are $\{M_1, M_2\} = DFT\{f_1, f_2\}$. The spectral alteration may slightly influence the mapping accuracy.

On the other hand, **Fig. S4-4b** illustrates another potential influence of the Doppler effect. When detecting multiple targets with diverse characteristic frequencies, spectral drift may cause aliasing, which can affect the recognition of different targets. In practice, our system primarily operates in air, where the speed of sound (v = 340 m/s, at temperature 288 K) is constant. The biomimetic hexapod robot moves quasi-statically, with a maximum speed of less than 1 m/s. The fastest target is the UAV (DJI Mavic 2), with a maximum velocity of 20 m/s. Assuming the frequency of our UAV is approximately 600 Hz, we simulate scenarios in which the UAV's speed increases, using coordinates FOM#1 (-0.2 m, 0 m), FOM#2 (0.2 m, 0 m), and UAV (22.3 m, 14.7 m). The calculated cross-correlation traces are shown in **Fig. 4-4c**. Compared to detecting a static UAV, when v_{UAV} = 20 m/s, the temporal error in cross-correlation calculation is 0.03 ms, indicating a spatial mapping error of 2.3 mm. Since our system has a spatial resolution of ±5 cm, such an impact can be neglected. In **Fig. 4-4d**, we present the measured results. For v_{UAV} = 0 m/s, 10 m/s and 20 m/s, the mapping accuracy remains nearly unchanged. To further enhance positioning accuracy, improved algorithms can be employed in signal processing to approach the Cramér-Rao Lower Bound (CRLB) [13].

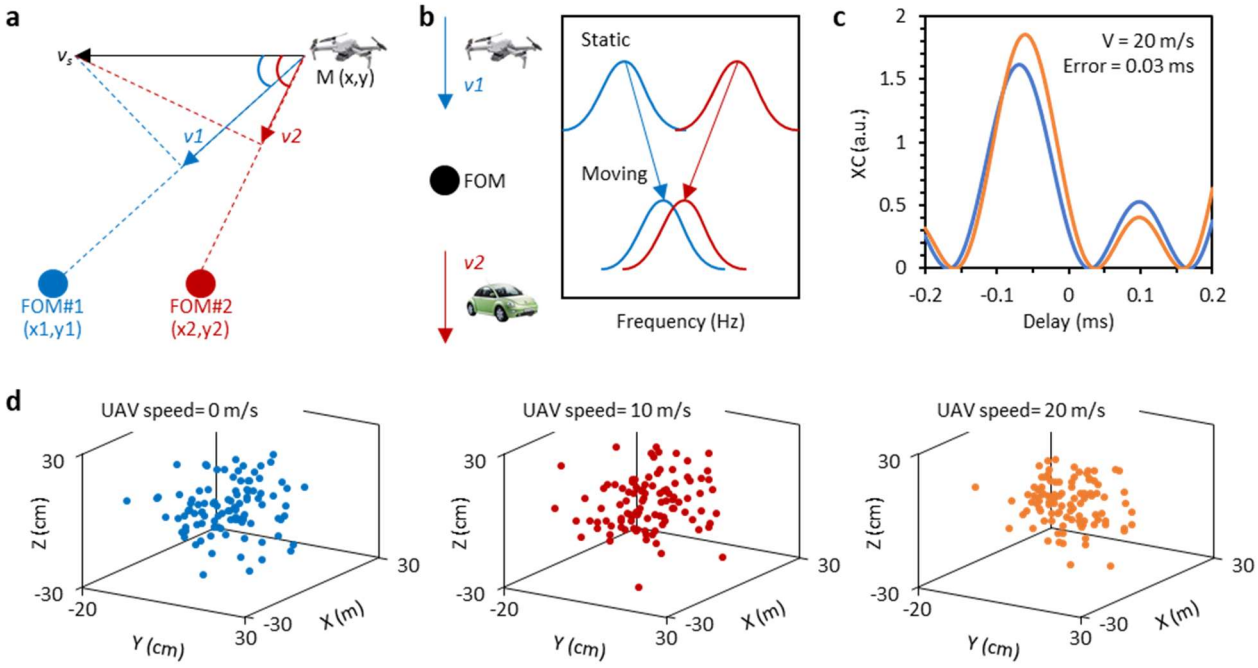

Fig. S4-4. Doppler effect. a, Conceptual illustration that a target moves towards two FOMs. **b,** Schematic diagram that the Doppler's effect changes the detected frequencies. **c,** Calculated cross-correlation for a static target and a moving target. **d,** Measured mapping results for mobile targets.

S4.5. Sound recognition based on CAM++ convolutional neural network algorithm.

Using electronic filtering alone would be insufficient for identifying acoustic targets with similar frequencies. **Fig. S4-5a** illustrates the architecture and operation of our CAM++ based sound recognition system. The system comprises the following key components. 1) Front-End Feature Extraction Module: This module converts the high-fidelity sound signal output by the FOM acoustic sensor into a feature vector. 2) Two-Dimensional Fast Convolution Module (FCM) with Residual Connections: This module extracts multi-scale voiceprint features in the time-frequency domain. 3) Improved Context-Aware Module (CAM): This module enhances the Deep Time-Delay Neural Network (D-TDNN) layer with the ability to dynamically allocate feature weights. It combines a multi-granularity pooling strategy to effectively aggregate contextual information. **Fig. S4-5b** displays the monitored parametric convergence curve during the machine learning process. After 60 iterations of training, our acoustic mapping and recognition system achieved an equal error rate (EER) of 0.15 and a minimum detection cost function (min DCF) of 0.70. Additionally, **Fig. S4-5c** shows the time-frequency characteristics of the three voice samples used in the main text. Here the maps demonstrate records of sound i, ii, iii and their mixture. It is clear that all the sound waves demonstrate main

frequencies distributed in the band 200 Hz ~ 500 Hz.

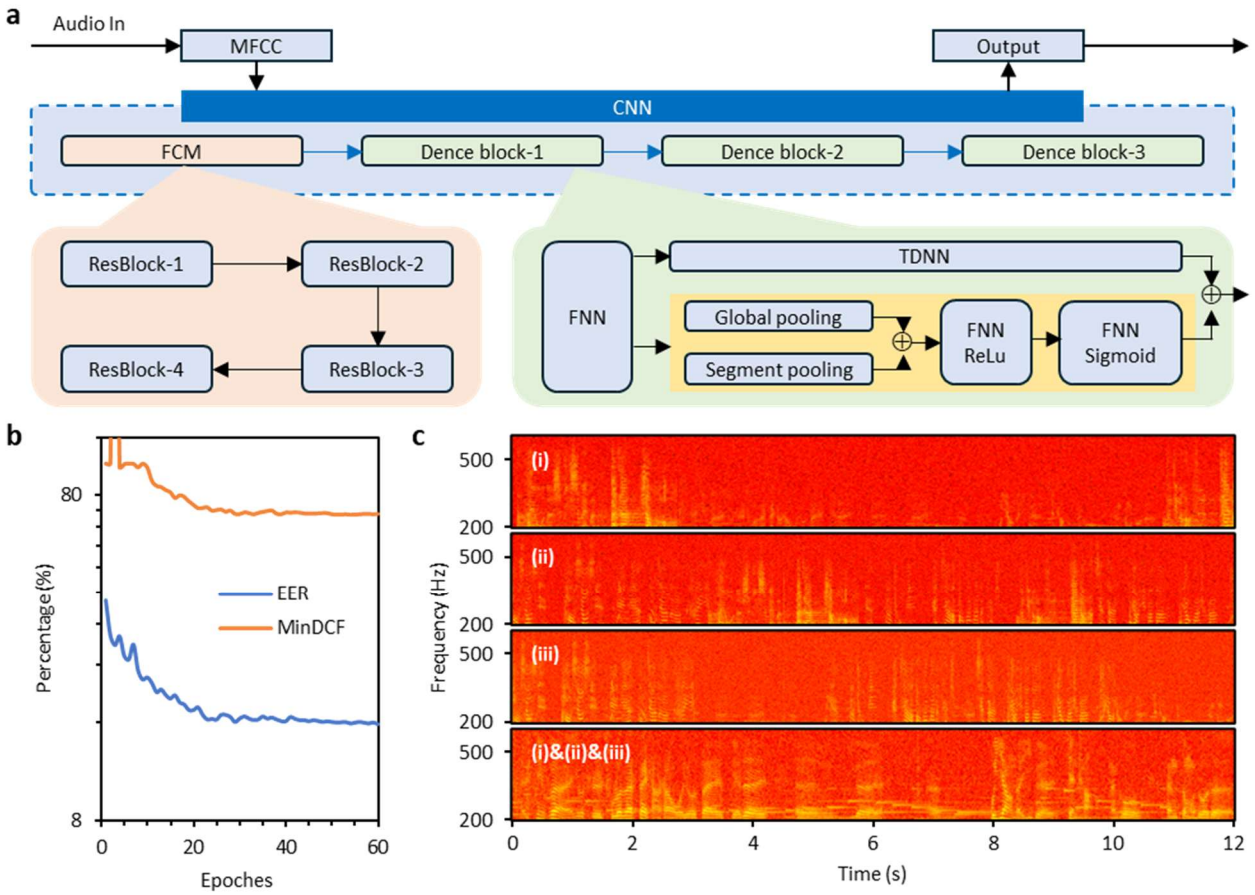

Fig. S4-5. Sound recognition based on convolution neural network. a, Architecture of the software model. **b,** EER and MinDCF parameters decreases during the network training. **c,** Measured time-frequency characterizations for the human speaking samples. From top to bottom: speaker i, ii, iii, and the mixed signal including i, ii, iii.

S4.6. Performance comparison of comb driven FOMs and single frequency laser driven FOMs.

In **Fig. S4-6**, we compare the performance of comb-driven FOMs with single frequency laser (SFL) driven FOMs. **Fig. S4-6a** illustrates the setups. Utilizing our fully stabilized dual microcomb, it is possible to drive over 100 FOMs in parallel, facilitated by an integrated arrayed waveguide grating (AWG) for de-multiplexing the comb channels. In contrast, the individual SFL scheme requires 100 SFLs to drive 100 FOMs, along with 100 photodetectors for data acquisition. We focus on comparing the sensing performance between the line#100 of our comb (whose stability is worse than line#1 ~ line#99) and an SFL (NKT-E15). **Fig. S4-6b** displays the long-term integrated linewidths of the SFL and the comb line. The 3-dB linewidth of the SFL reaches 12 kHz, whereas the comb line maintains a

linewidth of less than 100 Hz. Here the linewidths are calibrated by using heterodyne measurement based on Menlo System. **Fig. S4-6c** depicts their single sideband phase noises (SSB-PN), revealing that the comb line exhibits significantly lower phase noise, especially in the acoustic band. Specifically, at a 20 Hz offset, the SSB-PN of the SFL is 13 dBc/Hz, while that of our comb line approaches -19 dBc/Hz. Using these two light sources in acoustic sensing, **Fig. S4-6d** presents the minimum detectable pressures (MDPs) for FOMs in type 1, type 2, and type 3. The MDP of an SFL-driven FOM is primarily constrained by laser instability, whereas the MDP of a comb-driven FOM is mainly determined by the sensor device itself (< 200 nPa/Hz$^{1/2}$). It is noteworthy that once the free-running SFL is fully stabilized (based on the same feedback loop), it can achieve comparable performance; however, the cost and complexity of stabilizing hundreds of SFLs is prohibitive.

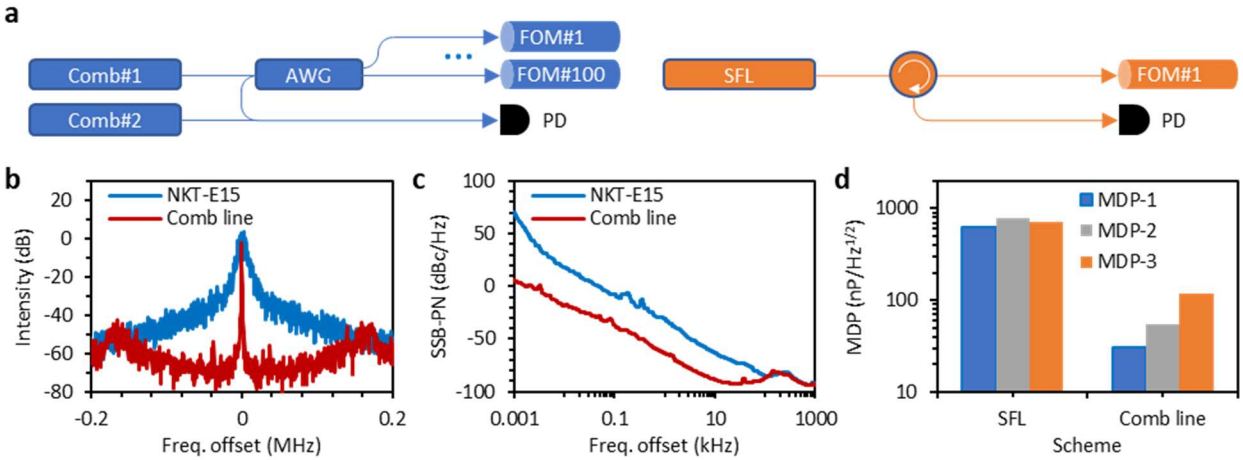

Fig. S4-6. Comparison of comb scheme and SFL scheme. a, Setups. **b,** Measured integrated spectra. **c,** Measured SSB-PN of the two optical laser lines. **d,** MDPs of FOMs in type 1, 2 and 3, which are driven by the SFL and our comb line, respectively.

S4.7. Technical advantages of on-chip Kerr soliton microcombs.

Table S4 summarizes that on-chip Kerr soliton microcomb can offer unique technical advantages for opto-acoustic mapping [[14]]. (1) Multiple photon-acoustic sensors should be driven by a single light source to create an effective sound detection system, necessitating the use of various independent optical wavelengths and optical filtering. Unlike low repetition rate mode-locked laser frequency combs, which operate at tens of MHz-level repetition rates, on-chip soliton combs offer much higher repetition rates (with comb spacing at the 100 GHz level or higher), enhancing their suitability for filtering operations. (2) To accommodate the future need for driving more optical sensors using larger-

scale wavelength division multiplexing, there is a requirement for increased spectral bandwidth and a greater number of comb teeth. An on-chip soliton comb allows for easier control of dispersion and nonlinear parameters. In comparison to electro-optic combs, soliton combs can generate a larger number of comb teeth (over 100) and cover a broader wavelength range (greater than 100 nm). (3) Kerr soliton optical combs, when compared to mode-locked fiber laser frequency combs (MLFL combs) and electro-optic combs (EO combs), offer higher integration levels and do not need expensive high-frequency signal sources. This makes them more advantageous for industrial production.

Table S4. Properties of different comb schemes

	Typical repetition	Typical bandwidth	Notes
MLFL comb	10 ~ 100 MHz	< 10 nm	Not integrated, hard to filter one comb line
EO comb	< 40 GHz	< 10 nm	Needs high-speed signal driver
Soliton Kerr comb	20 ~ 200 GHz	> 100 nm	/

S4.8. Comparison with other acoustic mapping schemes.

In **Table S5**, we share more acoustic mapping schemes and demonstrate the technical advances in this work.

Table S5. Acoustic sensing and mapping schemes

Method	Unique advance	Sensitivity	Mapping capability	Reference
Electrical	Voice recognition	MDP 0.9 Pa	No	[15]
	44 dB gain	/	No	[16]
	Using 2D material	MDP 0.1 Pa	No	[17]
	Sensor array	SNR 72 dB	Yes, acoustic orientation, 60 degrees	[18]
	Multiple sound source detection	/	Yes, acoustic orientation	[19]
Optical	3 FOM array with a simple setup	MDP 2 μPa/Hz$^{1/2}$	Yes, spatial localization, accuracy ± 5 cm	[20]
	DFB-LD driving 4 EFPIs, 300 m detection range	MDP 126.2 μPa/Hz$^{1/2}$	Yes, acoustic orientation, 5 degrees	[21]
	Fiber DAS	1 rad/μPa	Yes, acoustic orientation, 1.47 degree	[22]
This work	Biomimetic design, dual comb driving, and flexible deployment	MDP 36.9 nPa/Hz$^{1/2}$	Yes, spatial localization, accuracy ± 2 cm	/

S4.9. Parameters and cost of our acoustic mapping scheme.

In **Table S6**, we offer detailed information about the key devices in our system and compare their complexity and cost to conventional methods. For simultaneously driving 108 FOMs with comparable performance, this work demonstrates unparalleled advantages in terms of reliability, cost, and volume when compared to complex systems that rely on the collaboration of numerous fiber optic components.

Table S6. Devices and their information for acoustic mapping

	Our scheme	**Conventional scheme using divided devices**
Optical source	Dual microcomb chip ($ 2k)	108 single frequency lasers ($ 1k each)
Reference	Ultra-stable F-P cavity ($ 3k)	Ultra-stable laser ($ 20k)
Feedback loop	3, in FPGA (Infineon S2F44T) ($ 0.1k each)	108 (at least $ 10k in total)
Filters	Silicon chip ($ 5k)	108 fiber FBGs ($ 0.2k each)
MUX/DEMUX	PIC AWG (Shijia Photonics, customized) ($ 2k)	108 couplers + AWGs ($ 20k)
Photodetector	One (on chip) (free)	108 (e.g. Thorlabs PDA50B2) ($ 0.6k each)
Total volume	6×10^{-3} m^3 level	> 1 m^3 level
Total cost	≈$ 15 k	>$ 250 k
Robustness	High	Low

In **Table S7**, we compare different schemes for the full stabilization of a dual-microcomb light source. Specifically, our strategy using an ultra-stable F-P microcavity as the optical reference showcases the lowest cost, highest compactness and best performance.

Table S7. Dual microcomb stabilization schemes

	Our scheme (Optical frequency division)	**Stabilization based on RF references**	**Stabilization based on Menlo-System**
Optical reference	Ultra-stable FP microcavity (customized)	Ultra-stable laser	Menlo-system
Electrical reference	Standard clock in FPGA (Infineon S2F44T)	3 RF generators	3 RF generators
Modulator	/	5 GHz modulator	/
Electrical mixers	1, in FPGA	3	3

Min I-L (First line)	0.17 Hz	0.41 Hz	0.54 Hz
Min I-L (108th line)	16.3 Hz	733 Hz	141.7 Hz
Min 1 s stability	7×10^{-13}	1.4×10^{-12}	10^{-14}
Total volume	6×10^{-3} m^3 level	10^{-2} m^3 level	> 1 m^3 level
Total cost	≈ $ 15 k	≈ $ 55 k	≈ $ 700 k

*I-L: Instantaneous linewidth.

Supplementary references

1. Kippenberg, T. J., Holzwarth, R. & Diddams, S. A. Microresonator-based optical frequency combs. *Science (80-.).* **332**, 555–559 (2011).

2. Sun, S. *et al.* Integrated optical frequency division for microwave and mmWave generation. *Nature* **627**, 540–545 (2024).

3. Lu, X., Wu, Y., Gong, Y. & Rao, Y. A miniature fiber-optic microphone based on an annular corrugated MEMS diaphragm. *J. Light. Technol.* **36**, 5224–5229 (2018).

4. Rao, Y.-J. Recent progress in fiber-optic extrinsic Fabry–Perot interferometric sensors. *Opt. Fiber Technol.* **12**, 227–237 (2006).

5. Wu, Y. *et al.* A Highly Sensitive Fiber-Optic Microphone Based on Graphene Oxide Membrane. *J. Light. Technol.* **35**, 4344–4349 (2017).

6. Khan, A., Philip, J. & Hess, P. Young's modulus of silicon nitride used in scanning force microscope cantilevers. *J. Appl. Phys.* **95**, 1667–1672 (2004).

7. Zhang, P. *et al.* Generation of acoustic self-bending and bottle beams by phase engineering. *Nat. Commun.* **5**, 4316 (2014).

8. Cao, S., Chen, X., Zhang, X. & Chen, X. Effective Audio Signal Arrival Time Detection Algorithm for Realization of Robust Acoustic Indoor Positioning. *IEEE Trans. Instrum. Meas.* **69**, 7341–7352 (2020).

9. Jansen, H., Gardeniers, H., Boer, M. de, Elwenspoek, M. & Fluitman, J. A survey on the reactive ion etching of silicon in microtechnology. *J. Micromechanics Microengineering* **6**, 14 (1996).

10. Qin, C. *et al.* Co-Generation of Orthogonal Soliton Pair in a Monolithic Fiber

Resonator with Mechanical Tunability. *Laser Photon. Rev.* **17**, 2200662 (2023).

11. Yi, X., Yang, Q.-F., Yang, K. Y., Suh, M.-G. & Vahala, K. Soliton frequency comb at microwave rates in a high-Q silica microresonator. *Optica* **2**, 1078–1085 (2015).

12. Cao, X., Yang, H., Wu, Z. L. & Li, B. B. Ultrasound sensing with optical microcavities. *Light Sci. Appl.* **13**, (2024).

13. Ma, F., Guo, F. & Yang, L. Direct Position Determination of Moving Sources Based on Delay and Doppler. *IEEE Sens. J.* **20**, 7859–7869 (2020).

14. Yao, B. C. *et al.* Interdisciplinary advances in microcombs : bridging physics and information technology. *eLight* (2024) doi:10.1186/s43593-024-00071-9.

15. Zhao, X. *et al.* A self-filtering liquid acoustic sensor for voice recognition. *Nat. Electron.* **7**, (2024).

16. Lenk, C. *et al.* Neuromorphic acoustic sensing using an adaptive microelectromechanical cochlea with integrated feedback. *Nat. Electron.* **6**, 370–380 (2023).

17. Ma, K. *et al.* A wave-confining metasphere beamforming acoustic sensor for superior human-machine voice interaction. *Sci. Adv.* **8**, 1–11 (2022).

18. Sun, X. *et al.* Sound Localization and Separation in 3D Space Using a Single Microphone with a Metamaterial Enclosure. *Adv. Sci.* **7**, 1–7 (2020).

19. Jung, I. J. & Ih, J. G. Combined microphone array for precise localization of sound source using the acoustic intensimetry. *Mech. Syst. Signal Process.* **160**, 107820 (2021).

20. Lorenzo, S. & Solgaard, O. Acoustic Localization With an Optical Fiber Silicon Microphone System. *IEEE Sens. J.* **22**, 9408–9416 (2022).

21. Wu, G. *et al.* Development of highly sensitive fiber-optic acoustic sensor and its preliminary application for sound source localization. *J. Appl. Phys.* **129**, (2021).

22. Fang, J., Li, Y., Ji, P. N. & Wang, T. Drone Detection and Localization Using Enhanced Fiber-Optic Acoustic Sensor and Distributed Acoustic Sensing Technology. *J. Light. Technol.* **41**, 822–831 (2023).