

<https://doi.org/10.1038/s41528-024-00315-1>

Ultrasensitive textile strain sensors redefine wearable silent speech interfaces with high machine learning efficiency

Check for updates

Chenyu Tang^{1,6}, Muzi Xu^{1,6}, Wentian Yi^{1,6}, Zibo Zhang², Edoardo Occhipinti³, Chaoqun Dong¹, Dafydd Ravenscroft¹, Sung-Min Jung¹, Sanghyo Lee⁴, Shuo Gao⁵✉, Jong Min Kim¹ & Luigi Giuseppe Occhipinti¹✉

This work introduces a silent speech interface (SSI), proposing a few-layer graphene (FLG) strain sensing mechanism based on thorough cracks and AI-based self-adaptation capabilities that overcome the limitations of state-of-the-art technologies by simultaneously achieving high accuracy, high computational efficiency, and fast decoding speed while maintaining excellent user comfort. We demonstrate its application in a biocompatible textile-integrated ultrasensitive strain sensor embedded into a smart choker, which conforms to the user's throat. Thanks to the structure of ordered through cracks in the graphene-coated textile, the proposed strain gauge achieves a gauge factor of 317 with <5% strain, corresponding to a 420% improvement over existing textile strain sensors fabricated by printing and coating technologies reported to date. Its high sensitivity allows it to capture subtle throat movements, simplifying signal processing and enabling the use of a computationally efficient neural network. The resulting neural network, based on a one-dimensional convolutional model, reduces computational load by 90% while maintaining a remarkable 95.25% accuracy in speech decoding. The synergy in sensor design and neural network optimization offers a promising solution for practical, wearable SSI systems, paving the way for seamless, natural silent communication in diverse settings.

Silent speech interfaces (SSI) have emerged as a cutting-edge solution for scenarios where verbal communication is hindered. These include environments with excessive noise that can significantly interfere with spoken language or cases involving physiological conditions such as stroke, cerebral palsy, Parkinson's disease, or recovery from laryngeal surgeries^{1,2}. By analyzing nonvocal human signals, SSI offers a method for decoding speech in silent conditions. Among the various challenges in SSI research, developing an effective wearable system for real-world applications is a key objective for researchers. To achieve this goal, it is crucial to ensure that the device is comfortable and durable enough for practical use to encourage user

acceptance. Additionally, it is vital that the system operates with high precision and efficiency in distinguishing the speech of different users across a variety of scenarios.

In recent years, researchers have been actively working to develop effective SSI systems suitable for real-world wearable applications. This involves the innovation of devices for capturing human silent speech signals and the design of improved algorithmic models. Human speech-related neural impulses originate in the central nervous system, travel through the peripheral nervous system to the vocal cords, and are then articulated with the help of facial movements, resulting in various speech sounds³. In pursuit

¹Electrical Engineering Division, Department of Engineering, University of Cambridge, Cambridge, UK. ²Department of Electronic and Electrical Engineering, University College London, London, UK. ³UKRI Centre for Doctoral Training in AI for Healthcare, Department of Computing, Imperial College London, London, UK. ⁴School of Materials Science and Engineering, Kumoh National Institute of Technology (KIT), Gumi, South Korea. ⁵School of Instrumentation Science and Optoelectronic Engineering, Beihang University, Beijing, China. ⁶These authors contributed equally: Chenyu Tang, Muzi Xu, Wentian Yi.

✉e-mail: shuo_gao@buaa.edu.cn; lgo23@cam.ac.uk

of decoding this complex process, scientists have developed a range of SSI systems. For instance, techniques such as electroencephalography (EEG)^{4–7} and electrocorticography (ECoG)^{8–10} have been employed to decode speech from brain activity. Additionally, computer vision-based methods have been developed to decode silent speech from lip movements^{11–13}. However, these methods, while innovative, often fall short in practicality for implementation in wearable devices due to their invasive nature and the complexity of their setups.

In the quest to create a more user-friendly SSI, several efforts have been made to analyze mechanical movements in the throat and face, employing sensors such as electromyography (EMG)^{14–17} and strain sensors^{18–22}. These approaches show promise for integration into wearable devices, being noninvasive and adaptable to prolonged use. Compared to EMG sensors, strain sensors are preferred in SSI applications due to their higher signal fidelity and signal-to-noise ratio (SNR). Among them, various textile substrate-based strain sensors, including conductive elastomers, piezoelectric materials, and magnetostrictive materials, have been widely researched in recent years^{23–37}. Although this shift toward physical signal detection has theoretically enhanced wearability, it still faces its own set of challenges, notably the delicate balance between user comfort, signal accuracy, and system efficiency (Supplementary Fig. 1). User comfort requirements often imply the use of fewer sensory channels to limit the impact on the human body, eventually leading to less detailed data capture and reduced accuracy in speech decoding. To mitigate this, an increase in the complexity of the data processing models is needed, such as increasing the system's sampling rate to capture more speech nuances or converting signals into two-dimensional images to enhance data richness, but this solution raises issues of computational load, affecting the overall system efficiency^{38,39}. This interdependence between the three aspects—comfort, accuracy, and efficiency—is a known tradeoff limiting the development of practical, wearable SSI systems. Bridging this gap requires innovative solutions that ensure user comfort without compromising the accuracy and efficiency of the system, a challenge that lies at the heart of current SSI research for effective wearable device applications.

In this work, we address the challenges of wearable SSI with a unique sensor design approach that prioritizes accuracy, user comfort, and computational efficiency. We have developed an ultrasensitive textile-based strain sensor and speech decoding system seamlessly integrated into a wearable choker. This sensor is characterized by the ability to generate high information density signals and complemented by a matched light end-to-end neural network, balancing user comfort with high precision and system efficiency (Fig. 1a). The distinctive sensing mechanism is based upon its unique structure, featuring ordered thorough cracks on graphene-coated textiles, which significantly enhances sensitivity (Fig. 1b). In silent speech scenarios, particularly within small strain ranges, our sensor achieves a gauge factor improvement of 420% over the best results reported in previous works within the same technology area (Fig. 1c). This increase in sensitivity enables the capture of information-rich speech signals, allowing for their efficient processing through our specially designed neural network, with a record accuracy of 95.25% while reducing the network's computational load by 90%. This approach negates the need for high-dimensional, complex model augmentations often associated with traditional SSI algorithms. Our one-dimensional convolutional neural network custom architecture processes this dense information efficiently, reducing the computational load while maintaining high accuracy. The synergy of sensor design and neural network optimization allows the bridging of the gap between user convenience and technical effectiveness and sets a new standard in wearable silent speech communication technologies, forging new avenues with groundbreaking potential for seamless, natural communication across diverse settings. Furthermore, owing to the adopted transfer-learning approach, the proposed system demonstrates a remarkable capability to efficiently generalize the training set from a specific group of users and words to unfamiliar users with diverse genders, geographical, and ethnic backgrounds, as well as to new and potentially ambiguous words encountered in practical applications.

Results

Textile strain sensor based on ordered cracks

To capture abundant information for eliminating the need for laborious multidimensional analyses, high sensitivity within small sensing strain ranges ($\leq 5\%$)^{40,41} is an indispensable characteristic of flexible wearable sensors developed for detecting the throat micromovements associated with speech. It is known that speaking different words is associated with different degrees of stretching or shrinking strains by the throat muscle^{42,43}. Different features of word decoding are hidden in the signals captured by strain sensors intimately connected to the throat muscle, which can be extracted by enhancing the sensitivity of the strain sensors. The more sensitive the sensor is, the more abundant number of features are embedded in the resulting signals. Our proposed ultrasensitive textile strain sensor possesses the ability to detect tiny deformations of throat skin and to distinguish the fundamental signal characteristics even among words with extremely similar pronunciations. Due to its ultrahigh sensitivity resulting from ordered cracks formed on the surface of the textile substrate, high-density information can be obtained as needed for effective and accurate word recognition.

With their unique characteristics, including conformability, breathability, and durability, textiles are considered an ideal substrate for human motion monitoring with extraordinary performance^{44,45}. However, in the current state of the art, the resistance change of traditional textile strain sensors fabricated by printing/coating methods with relatively low gauge factor within a small strain range is insufficient to capture adequate information required for decoding different words, as shown in the inset of Fig. 1c. In this work, we developed a structured graphene sensing layer with ordered cracks, which dramatically improves the sensitivity of the textile strain sensor (Fig. 2a). Such ordered cracks can be formed through a one-step printing. By increasing the number of printing layers of the graphene ink, graphene flakes are not only coated on the surface of a single fiber but form a continuous layer of graphene on the top of the textile substrate. Due to the stiffness mismatch between the top graphene layer and the textile substrate, a series of ordered cracks are created by utilizing the textile matrix as the template after prestretching (Fig. 2b, c). When no strain is applied on the sensor, these ordered crack edges return to contact. As the strain increases gradually, the distance between these ordered cracks becomes larger, and the contact areas decrease rapidly, leading to a sharp change in resistance, which can be reconducted into a percolation network model^{7,46}. Hence, the resulting textile strain sensor shows the unique ability to sense the tiny deformation generated by throat micromovements as the large change in contact areas introduced by ordered cracks magnifies the resistance change with a small strain applied, and the gauge factor can reach 317 within 5% strain. Moreover, ordered cracks obtained with the proposed fabrication method ensure high stability of the resistance response in comparison with the low stability of other graphene-based strain sensors, whereby the nonuniformity of cracks that form randomly in the graphene layer with a certain thickness means that other graphene sensors reported in the literature are less stable and more prone to drift in their performance over prolonged use⁴⁷.

In addition to the ultrahigh sensitivity brought by the ordered cracks, the fabrication method of our textile strain sensors is biocompatible, simple, low cost, and scalable, and the property and performance can be easily controlled by tuning the parameters of the manufacturing process. Owing to its defects that are advantageous for piezoresistivity⁴⁸, graphene nanoplatelets are used in the preparation of functional ink (DI-water based) through high-pressure homogenization (HPH), a straightforward method that weakens the van der Waals forces between graphite layers resulting in few-layer graphene flakes⁴⁹. Figure 2d shows the aspect ratio distribution of graphene flakes we used with a mean value of ~ 45 . By altering the size of the interaction chamber of the homogenizer, the aspect ratio of nanoplatelets, which influences the percolation threshold⁴⁶, can be adjusted (Supplementary Fig. 3). Screen printing is renowned for its customizable pattern, exceptional compatibility with a flexible substrate, affordable cost and scalable fabrication in the field of printing electronics^{50–52}. Diverse patterns

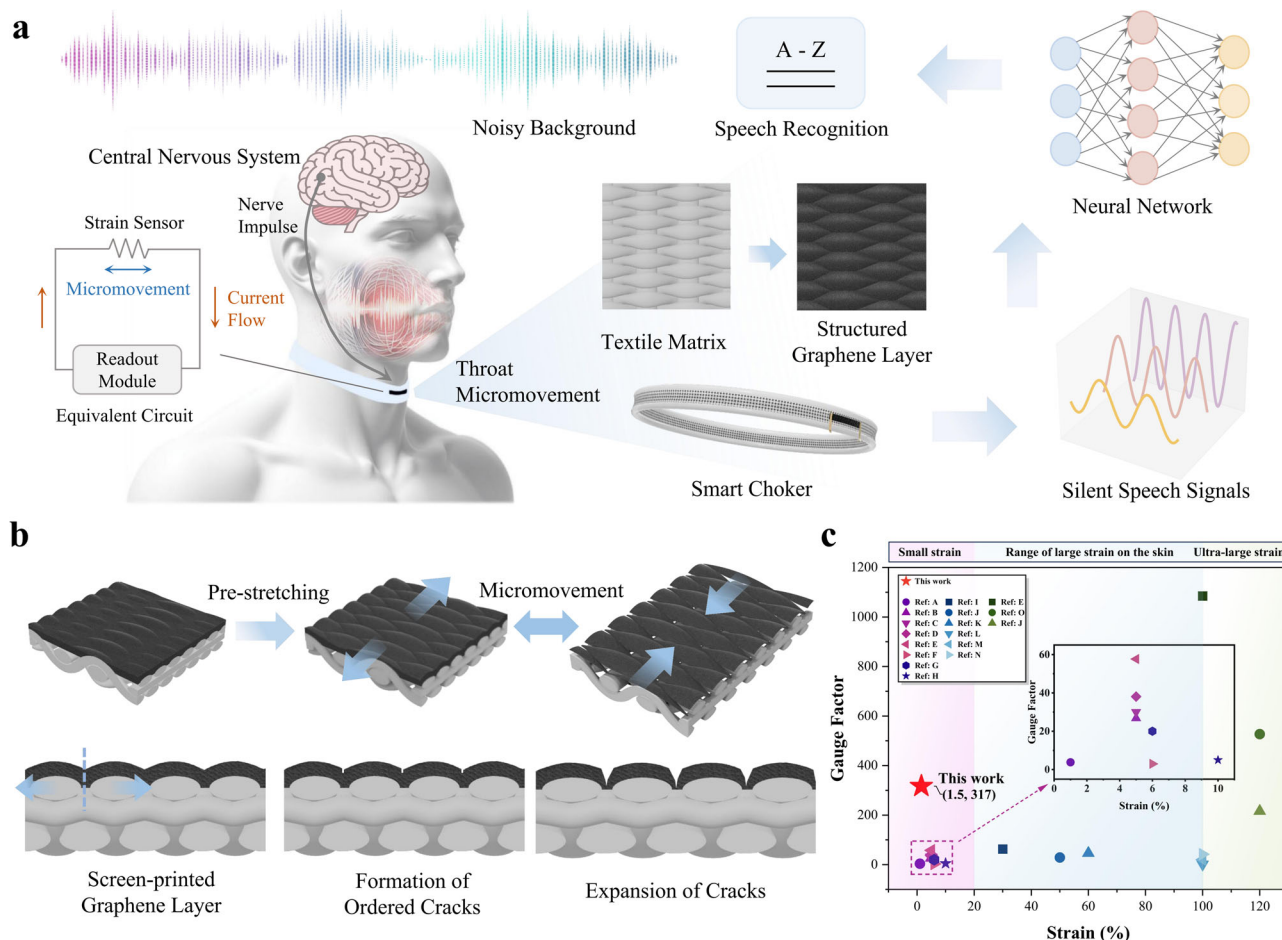


Fig. 1 | Comprehensive overview of the wearable SSI, featuring an ultrasensitive strain sensor and a neural network for efficient speech recognition. **a** The process of speech recognition initiates with nerve impulses from the central nervous system translating into micromovements in the throat. These movements are then captured by an ultrasensitive strain sensor integrated into a smart choker, comprising a textile substrate with an overlying structured graphene layer. The sensor responds by altering its resistance, resulting in a change in the electrical signal, which is then captured and processed by a readout module. The obtained electrical signals are fed into a lightweight end-to-end neural network for processing and speech recognition. The detection of throat micromovement based on orderly-cracked graphene ensures robust performance even in noisy environments, leveraging the high resistance of the sensor to background interference. **b** The sensing mechanism in the textile-based

strain sensor is enhanced with a structured graphene layer. This layer is created through the screen printing of a continuous thick graphene film onto a textile matrix. Following a stretching process, the inherently ordered weaving structure of the textile induces the formation of ordered through cracks in the graphene layer, which are strategically aligned with the weave. The structured graphene layer can dynamically respond to throat micromovements with significant and abrupt changes in electrical resistance. **c** Comparative analysis showcasing the performance metrics of our printed textile-based graphene strain sensor against other reported strain sensors fabricated by printing and coating technologies, focusing on the strain scale and Gauge factor. The exceptional gauge factor of our sensor in the small strain range is critical for capturing rich, information-dense signals. A–N refer to refs. 23–37.

on the printing mesh can be transferred to our textile substrate (made from 95% bamboo fibers and 5% elastane) directly. Varying the number of printing layers can be used to control the thickness of ordered cracks in the graphene-coating layer.

The performance of our textile strain sensor with ordered cracks was evaluated by monitoring the variations in its relative resistance. Within a small sensing range, the textile strain sensor demonstrates a linear relative resistance response with relatively low hysteresis (Fig. 2e). Figure 2f displays the stable stretching-releasing responses under 1%, 1.5%, 2%, 3%, 4%, and 5% strain, and the relative resistance increases linearly with strain ($\leq 5\%$), showing the high reliability of the sensor within a small strain range. Meanwhile, this textile strain sensor exhibits the ability to resist tensile frequency interference (Fig. 2g), which would be useful for identifying the same word spoken with different pitches. The detection limit was tested, as shown in Fig. 2h. Based on ordered cracks, the textile strain sensor realizes an ultralow detection limit (0.05%), which is crucial for tiny strain detection. Durability is crucial for real-world applications of the sensor to determine its lifespan. Our textile strain sensor can withstand over 10,000 stretching-releasing cycles while maintaining stable and reliable electrical functionality

(Fig. 2i). Such excellent durability is mainly attributed to the outstanding adhesion between the graphene ink and the substrate achieved through the careful selection of ink additives and the preprocessing of the textile substrate with plasma treatment. Additionally, the remarkable stability of the ordered cracks formed in the regions of concentrated stress along the textile matrix, which occur under repeated stretching and releasing, contributes significantly to the sensor’s resilience. Overall, the distinctive characteristics of the proposed textile strain sensor with ordered thorough cracks pave the way for its application in real-world silent speech systems.

The lightweight end-to-end neural network for robust speech recognition

In general, various SSI systems based on EMG sensors or strain sensors mainly encounter three types of noise in real-world applications: flicker noise caused by sensor imperfections, sound noise from the external environment, and physiological noise or artefacts arising from users’ body movements, such as breathing, swallowing, or neck movements, when wearing the device. Figure 3a shows a typical signal pattern during speech recognition using our smart choker. Initially, when the user is not wearing

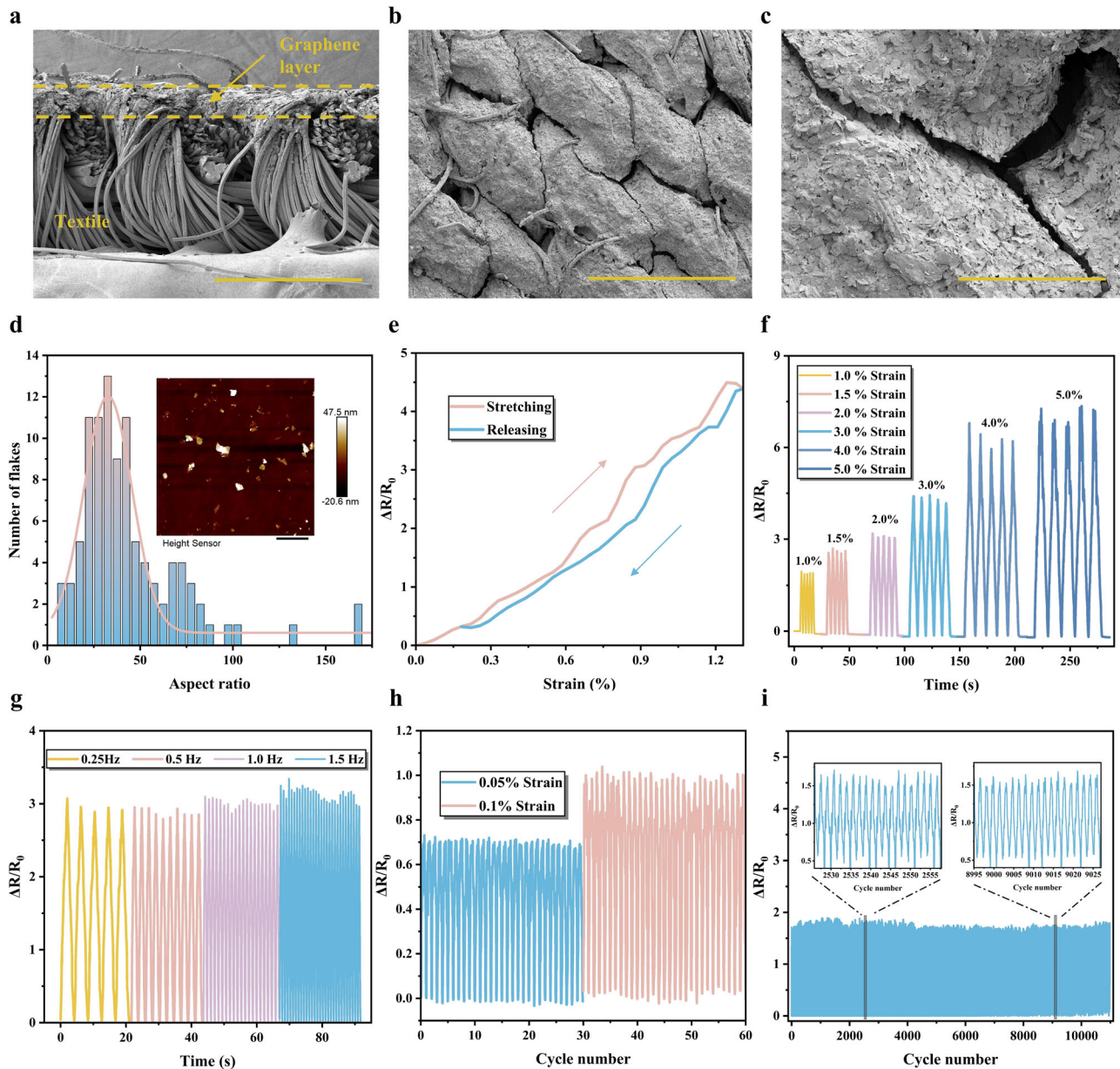


Fig. 2 | Characterization of the device. **a** Cross-sectional SEM image of the textile strain sensor with ordered cracks, showing the top graphene layer and the bottom textile layer, scale bar: 500 μm . **b** Top view SEM image of the ordered cracks formed on the top of the textile substrate, scale bar: 500 μm . **c** SEM image of the surface through-crack structure, scale bar: 100 μm . **d** Aspect ratio distribution of graphene flakes fabricated by the high-pressure homogenizer. The inset shows an AFM image of graphene flakes, scale bar: 4 μm . **e** Hysteresis of the relative resistance change

during a stretching-releasing cycle. **f** Relative resistance responses with 1.0%, 1.5%, 2.0%, 3.0%, 4.0%, and 5.0% cyclic strains. **g** Relative resistance responses with different stretching-releasing rates under 1.5% cyclic strain. **h** Detection limit stability test of the textile strain sensor under 0.05% and 0.1% cyclic strains. **i** Durability test of the textile strain sensor with ordered cracks by multicyclic stretching and releasing with a strain of 1.5% over 10,000 cycles.

the choker, the signal collected by the readout module appears as a superposition of the DC offset, corresponding to the sensor's initial resistance and flicker noise. It is worth noting that at the fifth second, we introduced 100 dB of environmental sound noise. From the response and our subsequent multiple tests on sound noise, it can be concluded that although our smart choker is extremely sensitive to the micromovement of the skin at the throat, it is 100% unresponsive to environmental sound noise. After the choker is worn, the DC offset changes, which is determined by the varying tightness with which the user wears the choker. After wearing, the noise in the signal appears as a superposition of flicker noise and physiological noise. Instead of using filters, we implemented noise injection data augmentation to enhance the system's noise immunity. Although previous methods, such as additive Gaussian noise injections have significantly improved model robustness, we

devised a simple "random noise window" technique to better assist the model in learning real-world noise characteristics (Fig. 3c)⁵³. Initially, users wear the choker silently, engaging in normal activities such as breathing and turning their heads. The signals collected during this time by the readout module represent a noise background without speech. We then randomly select multiple noise windows of the same length as speech samples and overlay the noise from these windows onto the speech samples to create augmented speech samples. This approach, compared to traditional filtering methods, greatly enhances energy efficiency. Such efficiency is vital for wearable systems in real-world applications, as it facilitates extended wearability without compromising performance.

Previous works on SSI often involve the conversion of one-dimensional (1D) time series signals into two-dimensional (2D) images

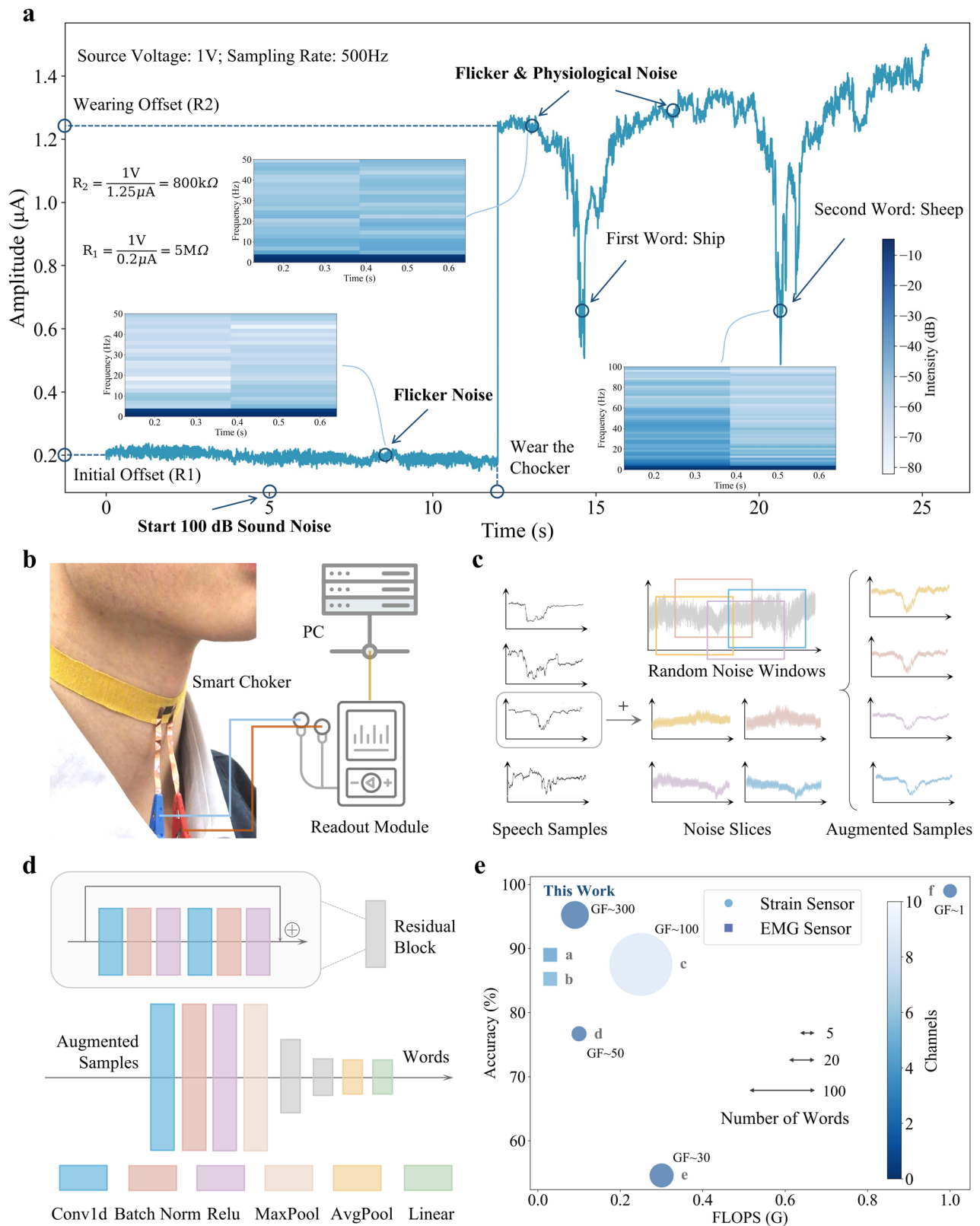


Fig. 3 | System and model architecture. **a** The signal characteristics of an entire silent speech phase are presented. At the 5th second, a sound noise of 100 dB is introduced. Starting at the 12th second, a choker is worn, and signals of two words are collected. Three segments are extracted to visualize the spectrogram, illustrating intensity variations across different frequencies over time. **b** Flowchart of the entire

system, comprising the smart choker, readout module, and the PC for model processing. **c** Flowchart depicting the random noise injection method used for data augmentation. **d** Pipeline of the lightweight end-to-end neural network employed one-dimensional convolutional layers. **e** Comparison of model efficiency (measured in FLOPS), accuracy, and channel usage with relevant works; a–f refer to ^{16,18,19,38,39,56}.

using feature extraction methods, such as Fourier Transform, before feeding them into 2D neural networks for analysis^{38,39}. This approach is primarily driven by two objectives. In cases where the sensing device comprises a multi-channel array, two-dimensional algorithms can extract spatial resolution information between channels. For single-channel devices with lower sensitivity and insufficient signal information density, two-dimensional methods are employed to enhance feature extraction, ensuring accurate speech decoding. However, the use of 2D algorithms significantly increases the computational complexity of the system. This increase makes them less suitable for deployment in edge systems, such as wearable smart devices, which demand high computational and energy efficiency. When input signals lack spatial complexity but possess high information density, employing 1D methods can preserve high system computational efficiency while also maintaining high analysis accuracy. Considering the high information density from our device's exceptional sensitivity, we crafted an end-to-end lightweight one-dimensional neural network for processing and classifying SSI signals. As shown in Fig. 3d, our model unites a series of convolutional layers with fully connected layers, and each component is finely tuned to the subtleties of the SSI data. At the heart of our network are residual blocks, featuring pairs of one-dimensional convolutional layers with a kernel size of 3. This design ensures critical temporal feature capture while optimizing computational efficiency. Each convolutional layer incorporates batch normalization and ReLU activation to bolster stability and learning efficacy. The initial convolution layer, equipped with 64 size-7 filters, followed by batch normalization and ReLU activation, plays a pivotal role in initial feature extraction from input signals. A dropout layer with a 0.2 rate is integrated to mitigate overfitting and maintain robustness across diverse scenarios. Efficient data downsampling is achieved via max-pooling, aligning with our model's focus on handling consistent 3-second, 1500-point signal samples at 500 Hz, which is critical for precise, real-time SSI applications. Concluding the network architecture are the fully connected layers, leading to a classification layer adept at distinguishing specific speech words, reflecting the tailored design of our system for SSI-based communication. A detailed network structure can be found in Supplementary Fig. 9.

In Fig. 3e, our model demonstrates high accuracy in classifying the top 20 frequently used English words with outstanding time and energy efficiency compared to state-of-the-art systems, characterized by low inference floating-point operations per second (FLOPS). This efficiency highlights our network's ability to harness single-channel, high-density data from our sensitive SSI device while minimizing computational demand. Such a streamlined approach promises extended wearability and practicality for daily use, establishing a new benchmark for energy-efficient silent speech recognition, to the best of our knowledge.

Performance in real-world silent speech scenarios

To validate the efficacy of our SSI system in real-world application scenarios, we collected three datasets (based on an English vocabulary) from three participants (see relevant details in Supplementary Table 2) across three of the most common speech communication settings. In Dataset 1, we gathered the ten most frequently used verbs and ten nouns in spoken English, using this collection as a baseline experiment to verify the system's capability to recognize words commonly used in everyday life⁵⁴. For Dataset 2, we compiled a set of ten easily confusable word pairs that differ by only one phonetic element—vowels, consonants, or stress—such as “book” and “look”, “sheep” and “ship”, and the verb and noun pronunciations of “record”. In Dataset 3, we collected five lengthy words at varying reading speeds to test the system's ability to correctly decode the same word across different speech rates. The details of the vocabulary for the three datasets can be found in Supplementary Table 3, and Fig. 4d provides a visualization of the signals for the word “Cambridge” at three different reading speeds.

In each of the three datasets, we collected 100 samples for every example, with 80 designated for the training set and 20 for the testing set. In Dataset 1, our model achieved a classification accuracy of 95.25% for the 20 high-frequency words (see the corresponding confusion matrix in

Fig. 4a); in Dataset 2, we reached a classification accuracy of 93% for the 10 confusable words (see corresponding confusion matrix in Fig. 4b); and in Dataset 3, our model achieved a classification accuracy of 96% for the five long words read at different speeds (see the corresponding confusion matrix in Fig. 4d, and see the reading time length distribution in Supplementary Fig. 12). To highlight the strengths of our network structure, we conducted a model evaluation on Dataset 1 (the baseline dataset), comparing our network with state-of-the-art benchmark backbones (all in 1D mode, results shown in Supplementary Fig. 10). Our network demonstrated advantages in both accuracy and time and energy efficiency, meeting the needs of wearable technology in practical scenarios. Additionally, to investigate whether our lightweight network's simpler architecture could limit performance on larger datasets with more samples per class, we compared the accuracy achieved by models trained with varying numbers of samples (see Supplementary Fig. 11). The results indicated that model accuracy continued to increase with more training samples, without reaching a saturation point, suggesting that the model's performance could be further optimized with the introduction of more data.

To assess whether our model exhibits bias in classification—such as focusing on noise or other irrelevant signal regions—we employed Relevance-Class Activation Mapping (R-CAM) to visualize the signal areas that the model concentrates on during classification (Fig. 4c)⁵⁵. The visualization reveals that the model consistently directs its attention to the key micromovements associated with the words, indicating a targeted and effective recognition process. Moreover, as demonstrated by several word examples in the figure, the DC offsets of the samples vary. This variation arises from our data collection strategy, which embraced the diversity of choker tightness and accounted for slight differences in placement with each wear. This diversity underscores the robustness of our system to the subtle variations in wear positioning and tightness that different users may exhibit in real-world scenarios, ensuring reliability across repeated uses.

To evaluate the system performance on new users and unknown words, we utilized our baseline model trained on Dataset 1 as a pretrained model and transferred it to three new users of different genders, geographical, and ethnic origins (detailed information about the new users can be found in Supplementary Table 2) and ten new words (Fig. 5a). For the new users, we collected the same five words previously gathered from the original three participants. For the new words, we selected the ten confusable words from Dataset 2 as novel entries for the baseline model. We observed that our model could effectively recognize the new user and words with minimal fine-tuning: with only 15 to 20 samples per class, the model achieved an 80% accuracy rate for both new words and users, which is a 43% and 53% improvement, respectively, compared to training directly on new data without a pretrained model. With fine-tuning on just 30 samples per class, the model reached 90% accuracy for both new users and words (Fig. 5b). Figure 5c, d visualize the model's generalization performance on new users and words using t-SNE, showing a significant improvement in the model's classification capabilities after leveraging the learning experience of the baseline pretrained model. Notably, in Fig. 5d, the model's ability to discriminate between confusable words, such as “book” and “look”, is enhanced, indicating the model's feature extraction and generalization capabilities.

Discussion

We introduce an ultrasensitive textile strain sensor technology, integrated into a wearable choker, which has the potential to redefine the field of SSI, enabling real-world applications. The sensing mechanism is based on ordered thorough cracks that form onto graphene-coated textiles, in regions of concentrated stress induced by the textile manufacturing process through weaving, upon an initial prestretching. The thickness of the sensing layer and depth of thorough cracks are optimized and controlled via the set of materials and printing process parameters, which allows the achievement of ultrahigh sensitivity and

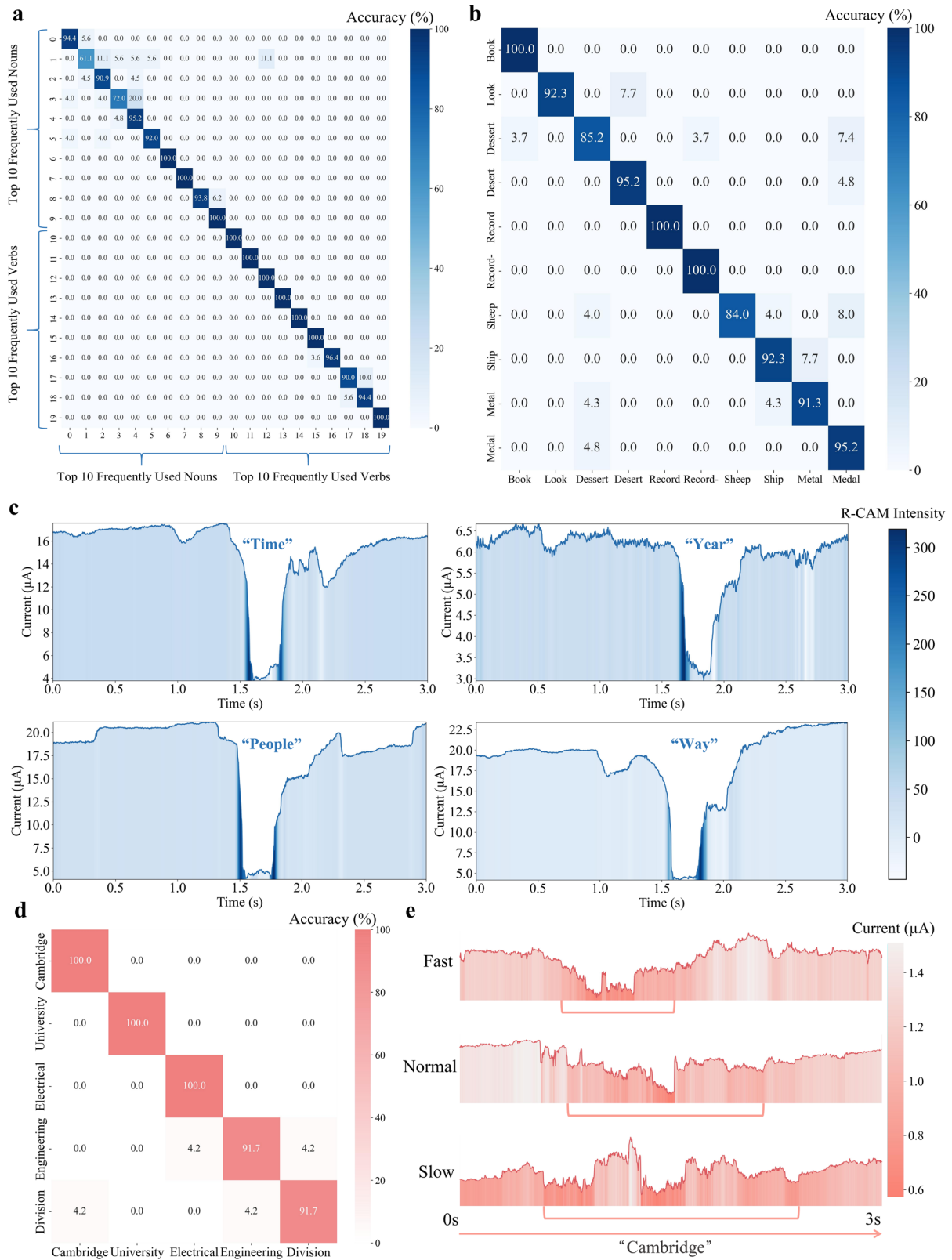


Fig. 4 | Silent speech recognition results. **a** Confusion matrix showing the classification results for the 10 most frequently used verbs and 10 most frequently used nouns, indicating the model’s capability in everyday use. **b** Confusion matrix for the classification of 10 words that are easily confused in terms of vowels, consonants, or stress patterns, demonstrating the model’s ability to discern subtle differences.

c Relevance-Class Activation Mapping (R-CAM) is utilized to highlight the signal areas the model focuses on during word classification. **d** Confusion matrix for the classification of 5 long words read at varying speeds, showcasing the model’s robustness to different reading speeds. **e** Visualization of the long word “Cambridge” read at three different speeds.

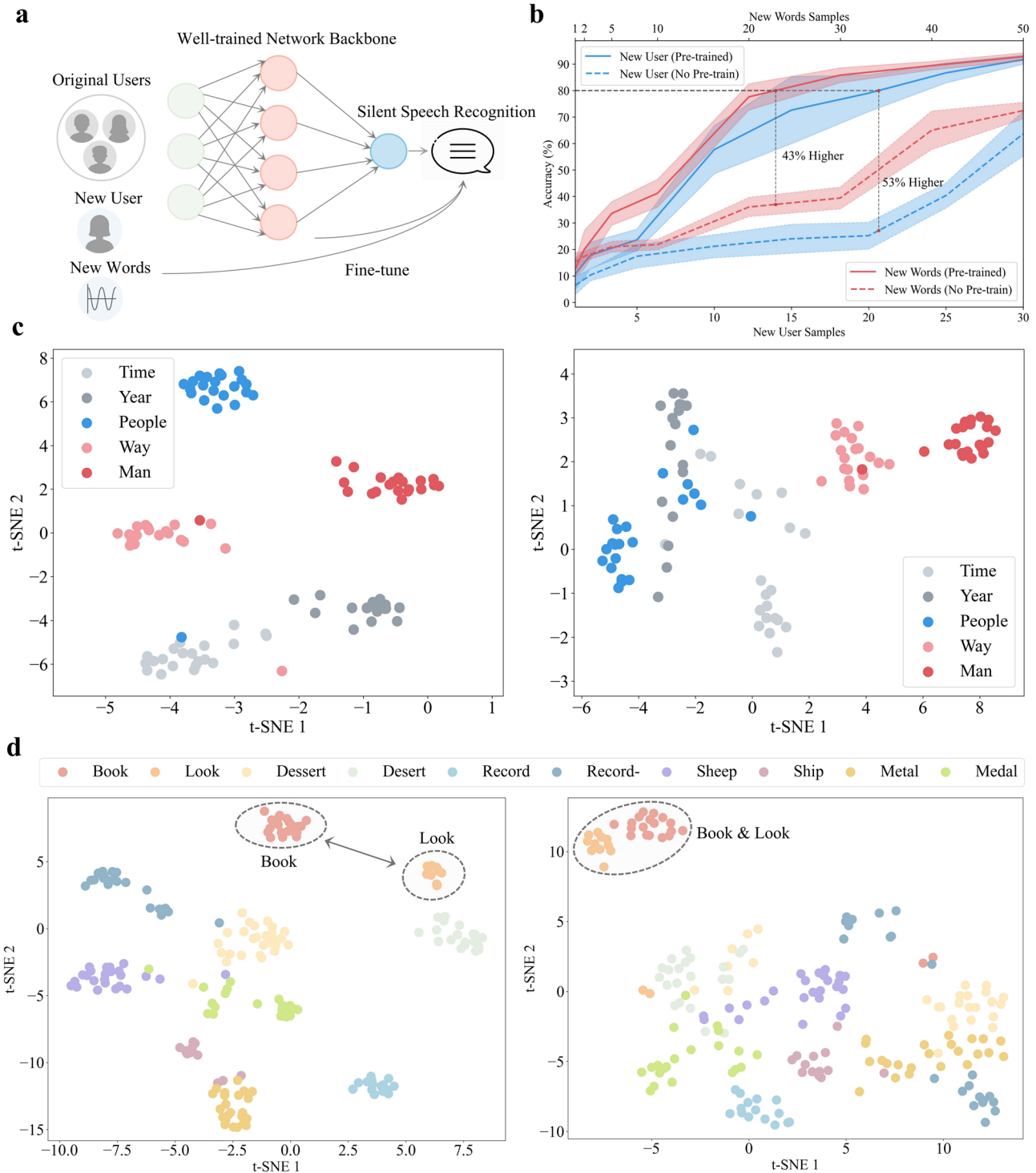


Fig. 5 | Generalization ability. **a** Flowchart depicting the model’s generalization process. **b** Evaluation results of the model’s generalization capabilities: comparison of accuracies when trained from scratch and fine-tuned using a baseline model with samples from new users and new words in varying quantities. **c** T-distributed

stochastic neighbor embedding (t-SNE) visualizations comparing models trained from scratch with new user data (right) to those fine-tuned using a baseline model (left). **d** T-SNE visualizations showing the difference in models trained from scratch with new word data (right) compared to those fine-tuned using a baseline model (left).

durability, simplifying the decoding of speech signals. Coupled with a tailored, energy-efficient neural network architecture, this system demonstrates high accuracy and reduced computational load, meeting the needs of wearable technology in practical scenarios. As a result, the proposed system enables decoding a wide range of words, swiftly adapts to new users and vocabularies, and demonstrates robustness against various noises and physical wear variations.

Methods

Materials

TIMREX KS 25 Graphite (synthetic graphite with a particle size of 25µm) was purchased from IMERYS. The sodium deoxycholate (SDC) (≥97%) and sodium carboxymethyl cellulose (CMC-Na: an average molecular weight of 700,000), as the surfactant and the binder for ink preparation, were both obtained from SIGMA-ALDRICH. Deionized water was provided by

PURELAB Flex Pure Water System. The textile substrate (95% bamboo fibers and 5% elastane) was purchased from Jelly Fabrics Ltd.

Preparation of graphene ink

The functional graphene ink for the sensor fabrication is prepared by HPH, a liquid phase exfoliation (LPE) method to fabricate graphene, and the steps are illustrated as follows. First, as the surfactant, SDC is dissolved in deionized (DI) water (5 g/l) to prevent aggregation of fillers by electrostatic repulsion. Second, TIMREX KS 25 graphite flakes were added to the SDC solution (100 g/l) and mixed by dissolver at 500 rpm for 30 min. Then, the mixtures are exfoliated by a HPH (PSI-40) using a dual-slot deagglomeration chamber (D200D: 200 μ m). It is processed at a pressure of 700 bar and 70 exfoliation cycles. Finally, CMC-Na as a binder was added to the graphene dispersion (10 g/l) to stabilize the flakes and control the viscosity of the printing ink, and the prepared ink was stirred for 3 h at room temperature to fully dissolve CMC-Na.

Fabrication of a textile strain sensor with ordered cracks

The textile graphene-based strain sensor with ordered cracks was fabricated by screen printing to form a functional graphene sensing film on a textile substrate which provides mechanical support and maintains flexibility. The manufacturing process was performed as follows. First, the textile substrate was treated by UV ozone (UV ozone cleaner UVC-1014, NanoBioAnalytics) for 30 min at room temperature to improve the hydrophilicity of the substrate and the adhesion between the graphene ink and textile. Then, the prepared graphene ink at a concentration of 100 g/l was printed onto the textile substrate fixed on the holder of the screen printer with the help of a squeegee forcing the ink through the screen (mesh count 90 T: 230 mesh/inch) with rectangular patterns. To control the formation of ordered thorough cracks, the printing process was repeated 7 times, each time with a 2 ml drop of graphene ink deposited on the screen and printed on the substrate, until the formation of a continuous graphene layer on the top surface of the textile substrate. In between every printing step, the film was let dry at room temperature, assisted with N₂ blow for 1 min. After 7 times printing, the printed sensor was annealed in the oven at 80 degrees for 5 min. Then the substrate was pre-stretched by applying a 5% strain to form ordered thorough cracks in the regions of higher stress. The repeatability results are shown in Supplementary Fig. 14 and Supplementary Fig. 15.

Characterization of the structure and performance of the sensor

The lateral size, thickness and aspect ratio of graphene flakes were assessed by Bruker Icon AFM (Supplementary Fig. 3). One hundred flakes were measured from 3 AFM scans, each with scan area $\sim 20 \mu\text{m} \times 20 \mu\text{m}$. SEM images were obtained using a Magellan 400 to characterize the morphology of the textile strain sensor with ordered cracks. Supplementary Fig. 4 shows the SEM results of the fabrication process. A tensile testing machine (Deben Microtest 200 N Tensile Stage, INSTRON universal testing system) and digital sourcemeter (Keithley 2400 Source Meter Unit) were used to measure the electromechanical properties of the textile strain sensor with ordered cracks. The resistance responses upon repetitive and consecutive strains were recorded to evaluate the sensing performance of the sensor.

Experimental setup of data acquisition

Our strain sensor is printed onto a choker, with copper tape tightly affixed to both ends of the sensor at a 1-centimeter distance, and a potentiostat (EmStat4S, PalmSens) is utilized as the readout module for data acquisition. The readout module inputs a voltage of 1 V and outputs the current passing through the strain sensor. We selected a sampling frequency of 500 Hz for the signal, with each word sample lasting 3 seconds. Supplementary Movie 1 offers a demonstration of the data collection process. Our data collection protocol was designed to reflect real-world usage scenarios where precise positioning and tightness of the choker might vary with each wear. Therefore, during our extensive data collection across various participants, we did not enforce strict calibration of the choker's position nor of its

tightness; participants were instructed to wear the choker comfortably around the neck, roughly positioning it at a medium height. The inherent variability in choker positioning and tightness among different users and experiments means that the collected dataset is representative of a range of different real-life conditions. Despite these variations, our system demonstrated high recognition accuracy, underscoring its robustness to different wearing conditions.

Software environment

The processing of the data and the training of the network were conducted in an environment based on Python 3.8.13, Miniconda 3, and PyTorch 2.0.1, with training acceleration provided by Apple's Metal Performance Shaders (MPS). During the noise injection phase, each original sample was augmented with real-world noise from four different random noise windows, creating four new samples. The optimal hyperparameters for model training can be found in Supplementary Table 4.

Ethics approval and human research participants

The study involving human participants was approved by the Research Ethics Committee of the Department of Engineering at the University of Cambridge and conducted on healthy volunteers following the guidelines approved for this study. All participants were provided with a Participant Information Sheet and asked to complete and sign a Participant Consent Form prior to their participation in the study.

Data availability

The data supporting this study are available at <https://doi.org/10.17863/CAM.104307>.

Code availability

The code used in this study is available at <https://github.com/tcy21414/Ultrasonic-Silent-Speech.git>.

Received: 2 January 2024; Accepted: 26 April 2024;

Published online: 07 May 2024

References

- Denby, B. et al. Silent speech interfaces. *Speech Commun.* **52**, 270–287 (2010).
- Gonzalez-Lopez, J. A., Gomez-Alanis, A., Martin Donas, J. M., Perez-Cordoba, J. L. & Gomez, A. M. Silent speech interfaces for speech restoration: a review. *IEEE Access* **8**, 177995–178021 (2020).
- Kearney, E. & Guenther, F. H. Articulating: the neural mechanisms of speech production. *Lang. Cogn. Neurosci.* **34**, 1214–1229 (2019).
- Anumanchipalli, G. K., Chartier, J. & Chang, E. F. Speech synthesis from neural decoding of spoken sentences. *Nature* **568**, 493–498 (2019).
- Proix, T. et al. Imagined speech can be decoded from low- and cross-frequency intracranial EEG features. *Nat. Commun.* **13**, 48 (2022).
- Nguyen, C. H., Karavas, G. K. & Artemiadis, P. Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features. *J. Neural Eng.* **15**, 016002 (2017).
- Herff, C. et al. Brain-to-text: decoding spoken phrases from phone representations in the brain. *Front. Neurosci.* **9**, 1–11 (2015).
- Angrick, M. et al. Speech synthesis from ECoG using densely connected 3D convolutional neural networks. *J. Neural Eng.* **16**, 036019 (2019).
- Rabbani, Q., Milsap, G. & Crone, N. E. The potential for a speech brain–computer interface using chronic electrocorticography. *Neurother* **16**, 144–165 (2019).
- Martin, S. et al. Word pair classification during imagined speech using direct brain recordings. *Sci. Rep.* **6**, 25803 (2016).
- Akbari, H., Arora, H., Cao, L. & Mesgarani, N. Lip2AudSpec: speech reconstruction from silent lip movements video. *In: Proc. 2018 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP) 2516–2520 (IEEE, 2018).*

12. Chung, J. S., Senior, A. Vinyals, O. & Zisserman A. Lip reading sentences in the wild. *In: Proc. 2017 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Calgary, AB, Canada, 3444–3453 (IEEE, 2017).
13. Pass, A., Zhang, J. & Stewart, D. An investigation into features for multi-view lipreading. *In: Proc. 2010 IEEE Int. Conf. Image Process.* Hong Kong, China, 2417–2420 (IEEE, 2010).
14. Wang, Y. et al. All-weather, natural silent speech recognition via machine-learning-assisted tattoo-like electronics. *npj Flex. Electron.* **5**, 20 (2021).
15. Tian, H. et al. Bioinspired dual-channel speech recognition using graphene-based electromyographic and mechanical sensors. *Cell Rep. Phys. Sci.* **3**, 101075 (2022).
16. Liu, H. et al. An epidermal sEMG tattoo-like patch as a new human-machine interface for patients with loss of voice. *Microsyst. Nanoeng.* **6**, 16 (2020).
17. Wand, M. et al. Tackling speaking mode varieties in EMG-based speech recognition. *IEEE Trans. Biomed. Eng.* **61**, 2515–2526 (2014).
18. Yoo, H. et al. Analysis of directional facial muscle movement. *ACS Appl. Mater. Interfaces* **14**, 54157–54169 (2022).
19. Ravenscroft, D. et al. Machine learning methods for automatic silent speech recognition using a wearable graphene strain gauge sensor. *Sensors* **22**, 299 (2021).
20. Cheng, L. et al. A highly stretchable and sensitive strain sensor for lip-reading extraction and speech recognition. *J. Mater. Chem. C.* **11**, 8413–8422 (2023).
21. Xu, S. et al. Force-induced ion generation in zwitterionic hydrogels for a sensitive silent-speech sensor. *Nat. Commun.* **14**, 219 (2023).
22. Zhou, J., Chen, T., He, Z., Sheng, L. & Lu, X. Stretchable, ultralow detection limit and anti-interference hydrogel strain sensor for intelligent throat speech recognition using Resnet50 neural network. *J. Mater. Chem. C.* **11**, 13476–13487 (2023).
23. Zhang, Y. et al. Cotton fabrics decorated with conductive graphene nanosheet inks for flexible wearable heaters and strain sensors. *ACS Appl. Nano Mater.* **4**, 9709–9720 (2021).
24. Alam, T., Saidane, F., Faisal, A. A., Khan, A. & Hossain, G. Smart-textile strain sensor for human joint monitoring. *Sens. Actuators A: Phys.* **341**, 113587 (2022).
25. Marra, F., Minutillo, S., Tamburrano, A. & Sarto, M. S. Production and characterization of Graphene Nanoplatelet-based ink for smart textile strain sensors via screen printing technique. *Mater. Des.* **198**, 109306 (2021).
26. Xie, X. et al. A spirally layered carbon nanotube-graphene/polyurethane composite yarn for highly sensitive and stretchable strain sensor. *Compos.—A: Appl. Sci. Manuf.* **135**, 105932 (2020).
27. Liu, X. et al. Smart textile based on 3D stretchable silver nanowires/MXene conductive networks for personal healthcare and thermal management. *ACS Appl. Mater. Interfaces* **13**, 56607–56619 (2021).
28. Chen, X. et al. A Single-material-printed, low-cost design for a carbon-based fabric strain sensor. *Mater. Des.* **221**, 110926 (2022).
29. Zhang, Y. et al. Mechanical exfoliation assisted with carbon nanospheres to prepare a few-layer graphene for flexible strain sensor. *Appl. Surf. Sci.* **611**, 155649 (2023).
30. Li, Y. et al. Electronic textile by dyeing method for multiresolution physical kineses monitoring. *Adv. Electron. Mater.* **3**, 1700253 (2017).
31. Yang, S. et al. Facile fabrication of high-performance pen ink-decorated textile strain sensors for human motion detection. *ACS Appl. Mater. Interfaces* **12**, 19874–19881 (2020).
32. Luo, C., Tian, B., Liu, Q., Feng, Y. & Wu, W. One-step-printed, highly sensitive, textile-based, tunable performance strain sensors for human motion detection. *Adv. Mater. Technol.* **5**, 1900925 (2020).
33. Park, H. et al. Dynamically stretchable supercapacitor for powering an integrated biosensor in an all-in-one textile system. *ACS Nano* **13**, 10469–10480 (2019).
34. Chen, G. et al. Superelastic EGaln composite fibers sustaining 500% tensile strain with superior electrical conductivity for wearable electronics. *ACS Appl. Mater. Interfaces* **12**, 6112–6118 (2020).
35. Souri, H. & Bhattacharyya, D. Highly Stretchable Multifunctional Wearable Devices Based On Conductive Cotton And Wool Fabrics. *ACS Appl. Mater. Interfaces* **10**, 20845–20853 (2018).
36. Sadi, M. S. et al. Direct screen printing of single-faced conductive cotton fabrics for strain sensing, electrical heating and color changing. *Cellulose* **26**, 6179–6188 (2019).
37. Peng, J. et al. A highly sensitive, superhydrophobic fabric strain sensor based on polydopamine template-assisted synergetic conductive network. *Appl. Surf. Sci.* **617**, 156535 (2023).
38. Kim, T. et al. Ultrathin crystalline-silicon-based strain gauges with deep learning algorithms for silent speech interfaces. *Nat. Commun.* **13**, 5815 (2022).
39. Yang, Q. et al. Mixed-modality speech recognition and interaction using a wearable artificial throat. *Nat. Mach. Intell.* **5**, 169–180 (2023).
40. Trung, T. Q. & Lee, N.-E. Flexible and stretchable physical sensor integrated platforms for wearable human-activity monitoring and personal healthcare. *Adv. Mater.* **28**, 4338–4372 (2016).
41. Liu, Q., Chen, J., Li, Y. & Shi, G. High-performance strain sensors with fish-scale-like graphene-sensing layers for full-range detection of human motions. *ACS Nano* **10**, 7901–7906 (2016).
42. Wang, Y. et al. Wearable and highly sensitive graphene strain sensors for human motion monitoring. *Adv. Funct. Mater.* **24**, 4666–4670 (2014).
43. Kapur, A., Kapur, S. & Maes, P. AlterEgo: a personalized wearable silent speech interface. *In: Proc. 23rd Int. Conf. on Intelligent User Interfaces (IUI '18), Tokyo, Japan*, 43–53 (ACM, 2018).
44. Libanori, A., Chen, G., Zhao, X., Zhou, Y. & Chen, J. Smart textiles for personalized healthcare. *Nat. Electron.* **5**, 142–156 (2022).
45. Seyedin, S. et al. Textile strain sensors: a review of the fabrication technologies, performance evaluation and applications. *Mater. Horiz.* **6**, 219–249 (2019).
46. Ambrosetti, G., Johner, N., Grimaldi, C., Danani, A. & Ryser, P. Percolative properties of hard oblate ellipsoids of revolution with a soft shell. *Phys. Rev. E* **78**, 061126 (2008).
47. Li, X., Hua, T. & Xu, B. Electromechanical properties of a yarn strain sensor with graphene-sheath/polyurethane-core. *Carbon* **118**, 686–698 (2017).
48. Qiao, Y. et al. Graphene-based wearable sensors. *Nanoscale* **11**, 18923–18945 (2019).
49. Karagiannidis, P. G. et al. Microfluidization of graphite and formulation of graphene-based conductive inks. *ACS Nano* **11**, 2742–2755 (2017).
50. Chen, C. et al. Perovskite solar cells based on screen-printed thin films. *Nature* **612**, 266–271 (2022).
51. Hyun, W. J., Secor, E. B., Hersam, M. C., Frisbie, C. D. & Francis, L. F. High-resolution patterning of graphene by screen printing with a silicon stencil for highly flexible printed electronics. *Adv. Mater.* **27**, 109–115 (2015).
52. Liang, J., Tong, K. & Pei, Q. A water-based silver-nanowire screen-print ink for the fabrication of stretchable conductors and wearable thin-film transistors. *Adv. Mater.* **28**, 5986–5996 (2016).
53. Xie, Q., Dai, Z., Hovy, E., Luong, M.-T. & Le, Q. V. Unsupervised data augmentation for consistency training. *Adv. Neural Inf. Process. Syst.* **33**, 6256–6268 (2020).
54. Leech, G., Rayson, P., Wilson, A. Frequency lists in WFWSE. *In: Word frequencies in written and spoken english: based on the british national corpus.* (Longman, London, companion website: <https://ucrel.lancs.ac.uk/bncfreq/flists.html>. 2001).
55. Lee, J., Kim, S., Park, I., Eo, T. & Hwang, D. Relevance-cam: your model already knows where to look. *In: Proc. 2021 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), Nashville, TN, USA*, 14944–14953 (IEEE, 2021).

56. Jin, Y. et al. Deep-learning-enabled mxene-based artificial throat: toward sound detection and speech recognition. *Adv. Mater. Technol.* **5**, 2000262 (2020).

Acknowledgements

C.T. was supported by Endoenergy Systems (grant no. G119004), M.X. was supported by CSC-Cambridge International Scholarship, W.Y. was supported by Pragmatic Semiconductor (grant no. G117793), E.O. was supported by UKRI Centre for Doctoral Training in AI for Healthcare (grant no. EP/S023283/1), D.R. was supported by EPSRC Center for Doctoral Training in Sensors Technologies and Applications (grant no. EP/L015889/1), S.L. acknowledges funding from National Research Foundation of Korea Grant funded by the Korean Government (NRF-2018R1A6A1A03025761), S.G. acknowledges funding from National Natural Science Foundation of China (grant no. 62171014), L.G.O. acknowledges funding from EPSRC (grants no. EP/K03099X/1, EP/L016087/1, EP/W024284/1, EP/P027628/1), the EU Graphene Flagship Core 3 (grant no. 881603), and Haleon (grant no. G110480). We would like to extend our sincere gratitude to Prof. George Malliaras for his invaluable guidance and mentorship throughout this work as the PhD advisor to C.T. and M.X.

Author contributions

C.T., M.X., W.Y. and L.G.O. conceptualized the work. M.X. and W.Y. fabricated the sensor and performed the characterizations. C.T., M.X. and W.Y. wrote the manuscript and collected the datasets. C.T. built the machine learning models and analyzed the datasets. C.T., M.X., W.Y. and Z.Z. visualized the concept figure. L.G.O. directed the research. S.G. and L.G.O. reviewed the manuscript. All authors analyzed and discussed the results and contributed to the final manuscript. C.T., M.X. and W.Y. contributed equally to the work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41528-024-00315-1>.

Correspondence and requests for materials should be addressed to Shuo Gao or Luigi Giuseppe Occhipinti.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024