Sensing technologies for silent speech interfaces

Chenyu Tang [1], Liang Qi[2], Shuo Gao [2,3] ✉, Zibo Zhang [1], Wentian Yi [1], Muzi Xu [1], Edoardo Occhipinti[4], Yu Pan[5] & Luigi G. Occhipinti [1] ✉

Silent speech interfaces decode speech intent without audible sound, enabling communication in settings where voice is inaccessible, or for individuals with speech impairments. Here we examine how sensing technologies shape the capabilities of silent speech interfaces. We compare off-, on- and in-body sensing modalities, identifying how proximity, coupling stability and invasiveness govern signal fidelity, robustness and user comfort. We highlight key trends, including the rise of flexible bioelectronics, multimodal sensor fusion for artefact resilience, and the growing role of edge artificial intelligence in real-time, low-power decoding. We show that on-body systems currently offer the best balance between accuracy and deployability, whereas in-body approaches provide unmatched neural access for individuals with complete loss of articulation. Looking ahead, advances in multimodal sensing, embedded intelligence and closed-loop architectures are poised to expand silent communication across rehabilitation, daily interaction and human–machine interfaces.

Silent speech interfaces (SSIs) aim to unlock a new dimension of human communication by decoding speech-related intent without relying on vocalized sound[1,2]. Rather than viewing speech solely as an audible output, these systems conceptualize it as a complex neuromotor process that originates in the brain, propagates through articulatory musculature and can be sensed and reconstructed through a variety of physiological pathways[3]. This transformative class of technologies redefines how individuals interact with machines and with each other, offering fundamentally new modes of expression in contexts where acoustic speech is inaccessible, impractical or undesired. From enabling silent interaction in noise-sensitive or privacy-critical environments to restoring communication for individuals affected by stroke, neurodegenerative disease or laryngectomy, SSIs address both everyday and medically underserved needs across society[2,4].

At the core of any SSI lies the sensing interface, which governs which signal can be accessed, how reliably it can be recorded and under which constraints it can be deployed. Sensing strategies fall into three major categories—off body, on body and in body—defined not solely by physical placement, but also by the nature and stability of coupling between the sensor and the physiological source of speech-related information (Fig. 1a). Off-body interfaces, such as optical and acoustic sensors, prioritize user comfort and deployability through non-contact or loosely coupled approaches, but are limited by indirect or distal coupling to the articulatory system[5,6]. On-body sensors, including electromyography (EMG), strain and inertial modules, offer a closer and more stable connection to neuromuscular activity through tight skin contact, but introduce wearability considerations[7–9]. In-body neural interfaces, such as electrocorticography (ECoG) or intracortical microelectrodes, provide direct access to the brain's speech-generating regions and enable decoding even in the complete absence of peripheral muscle control. However, this advantage comes at the cost of invasiveness and a need for custom-designed solutions or intervention procedures due to patient diversity in their anatomical structure, pathological conditions, biological responses and functional goals, alongside ethical constraints[10,11].

These sensing modalities present diverse form factors tailored to different user contexts (Fig. 1b). Off-body sensing has been integrated

[1]Department of Engineering, University of Cambridge, Cambridge, UK. [2]School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, China. [3]Hangzhou International Innovation Institute, Beihang University, Hangzhou, China. [4]Department of Mechanical Engineering, University College London, London, UK. [5]Department of Rehabilitation Medicine, Beijing Tsinghua Changgung Hospital, Tsinghua University, Beijing, China. ✉e-mail: shuo_gao@buaa.edu.cn; lgo23@cam.ac.uk
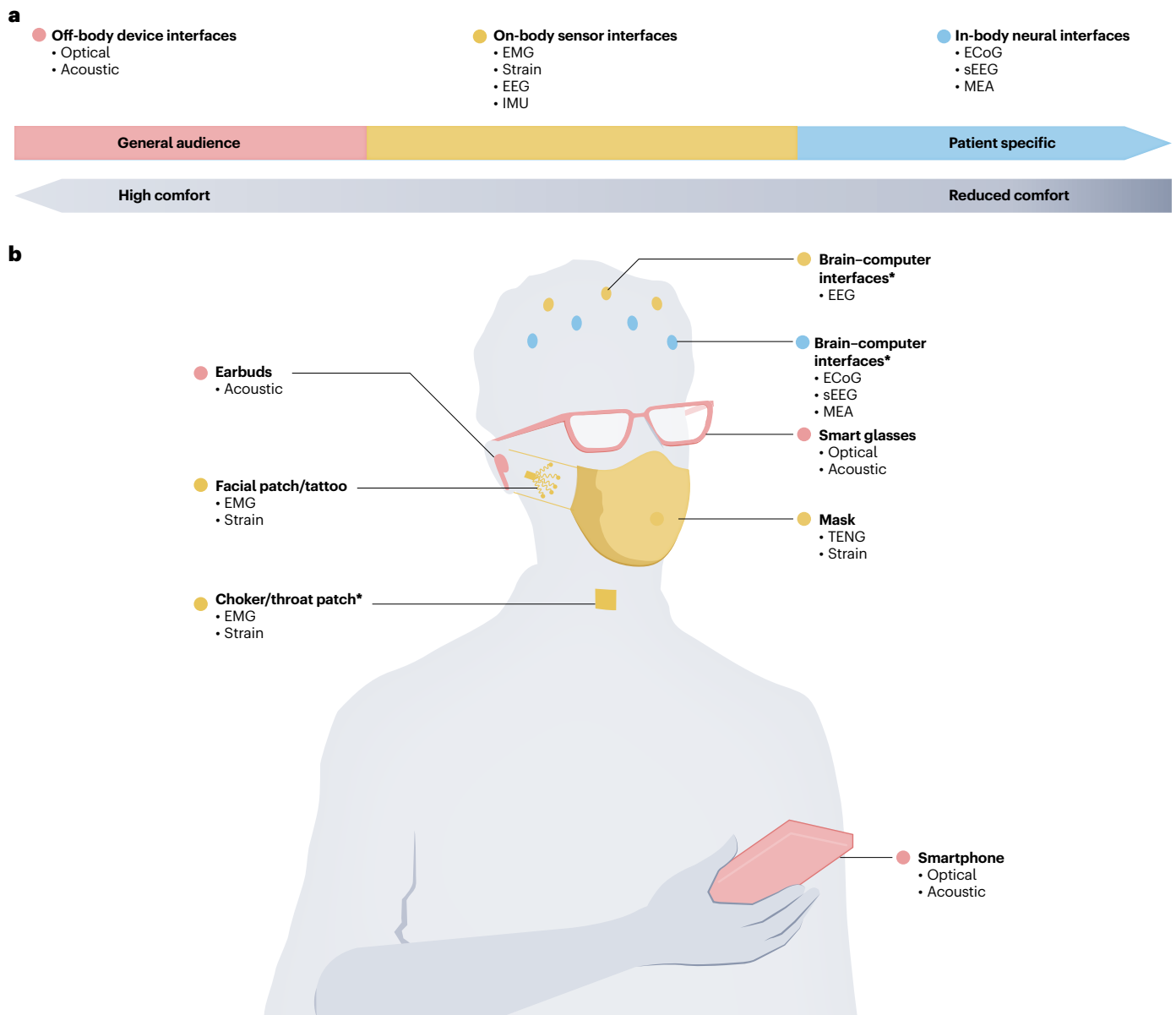
Fig. 1 | Sensor-based classification of SSIs and their deployment form factors.
a, Sensor modalities in SSIs can be categorized, by the proximity of signal acquisition to the human body, as either off-body device interfaces (for example, optical cameras or ultrasound probes), on-body sensor interfaces (for example, surface EMG, strain sensors, EEG and IMUs) or in-body neural interfaces (for example, ECoG, sEEG or MEAs). These categories represent a continuum between user comfort and signal specificity, ranging from general-purpose wearable devices to highly personalized clinical systems. **b**, Representative deployment examples for each category, aligned with the same taxonomy: off body (smartphones, earbuds and smart glasses), on body (facial patch/tattoo, choker/throat patch, mask and EEG hat) and in body (ECoG, sEEG and MEA implants). The asterisks denote technologies with clinical validation.

into earbuds, mobile phones and smart glasses, offering non-contact solutions with high comfort but lower signal specificity. On-body approaches, leveraging biomechanical and bioelectrical sensors, are implemented in smart chokers, facial patches, masks and head-mounted wearables, balancing wearability with stable neuromuscular access. In-body strategies are represented by invasive brain–computer interfaces, such as ECoG grids and intracortical electrodes, some of which have already been applied in clinical speech decoding trials.

Overall, these categories delineate trade-offs among comfort, invasiveness and system complexity. Signal fidelity introduces a further dimension: while not dictated by proximity alone, it depends on modality, device properties (for example, impedance and bandwidth), placement stability and downstream processing. Yet, a broad trend remains—closer interfaces to articulatory or neural sources generally yield higher fidelity, manifested as greater information throughput

and error resilience under realistic use. These gradients also map onto application domains: off-body sensors support general-purpose interaction (such as silent command input or privacy-preserving communication); on-body sensors enable early-stage restoration of speech in patients with residual motor function; and in-body systems remain the only current method for decoding continuous, near-real-time speech in completely locked-in individuals.

The timeline of the development of SSI technologies mirrors this stratification (Fig. 2). Early studies explored both camera-based lipreading and surface EMG[12,13], with intracortical recording of speech-related brain activity also beginning to show feasibility in the late 1990s[14]. Since 2020, the field has witnessed the proliferation of flexible sensing materials, integration with consumer electronics, and initial clinical deployments of wearable and implanted systems. Across modalities, advances in spatial and temporal resolution, signal
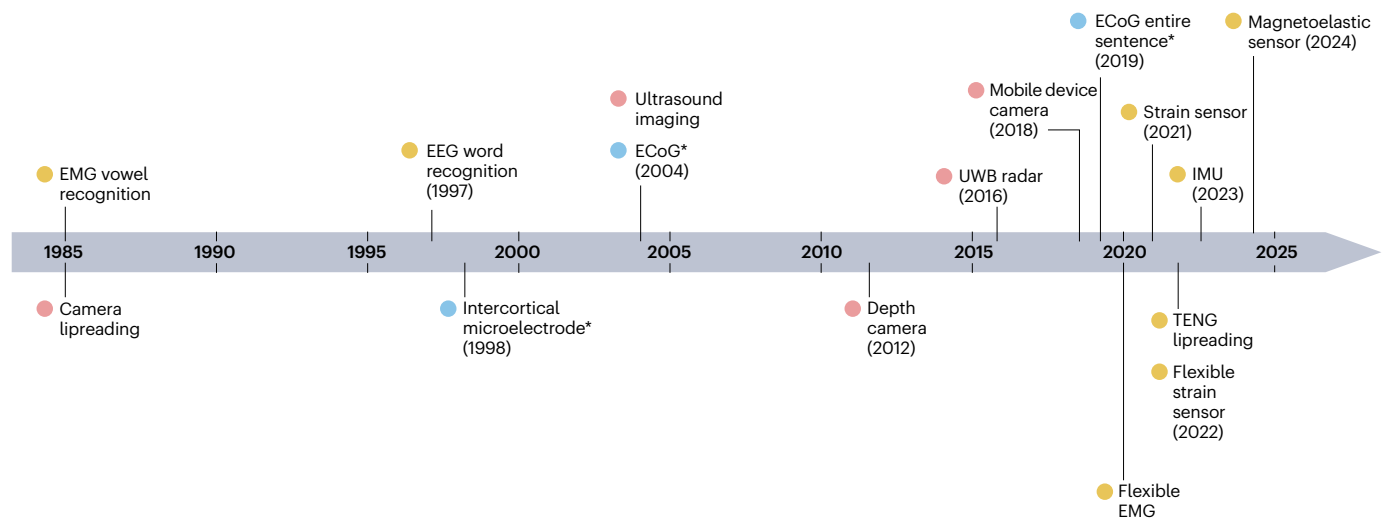
Fig. 2 | Timeline of sensor-driven innovations in SSIs. Presented is a timeline of SSI innovations, from the earliest to the latest: EMG-based vowel recognition (1985)[12], camera-based lipreading (1985)[13], EEG word decoding (1997)[14], intracortical microelectrodes (1998)[81], ECoG speech decoding (2004)[52], ultrasound imaging (2004)[82], depth cameras (2012)[83], ultra-wideband (UWB) radar (2016)[84], camera SSIs for mobile devices (2018)[85], full-sentence decoding via ECoG (2019)[86], flexible EMG sensors (2020)[7], wearable strain sensors (2021)[87], TENG-based lipreading (2022)[33], flexible strain sensors (2022)[8], IMU-based decoding (2023)[30] and magnetoelastic silent speech sensors (2024)[88]. Each milestone is categorized as off body (red), on body (yellow) or in body (blue) based on the specific sensing implementation used in the corresponding silent speech study, rather than the full spectrum of technological variants. The asterisks denote technologies with clinical validation.

robustness and integration with learning algorithms are reshaping what is feasible (Table 1).

As silent speech technologies approach a translational inflection point, a sensing-led perspective becomes essential to reframe system capability, usability and societal reach. The field is rapidly expanding with the introduction of novel sensor modalities, advances in flexible bioelectronics, and early-stage clinical studies. However, it still lacks a comprehensive framework that integrates these developments through the lens of sensing. By positioning sensing as both a constraint and a catalyst, this Review redefines the foundations of silent speech systems. Trade-offs in comfort, fidelity and invasiveness are not just technical considerations but strategic levers that shape adoption and impact.

We classify existing approaches into off-, on- and in-body sensing strategies, each presenting distinct trade-offs in signal fidelity, comfort and clinical relevance. We compare sensor modalities across spatial and temporal resolution, invasiveness and integration potential and trace their development from early vision-based systems to recent advances in flexible bioelectronics and neural interfaces. Finally, we outline emerging directions in sensor–artificial intelligence (AI) integration, real-time decoding and closed-loop systems that are poised to transform communication, rehabilitation and human–machine interaction at scale.

SSIs using off-body sensors
Off-body sensors enable non-intrusive SSIs by capturing articulatory and physiological signals without direct skin contact. These systems typically rely on optical and acoustic modalities integrated into external devices such as cameras, smartphones, glasses or headsets. Their appeal lies in high user comfort, ease of deployment and compatibility with commodity hardware, making them attractive for scalable and low-burden interaction. However, signal quality in off-body sensing is often susceptible to environmental conditions, such as lighting variation or acoustic interference, and may suffer from indirect coupling to the user's intent.

Recent advances have broadened the landscape of off-body silent speech systems, transitioning from laboratory-grade setups to increasingly wearable and context-aware platforms. For instance, optical systems have evolved from static RGB cameras to mobile and multi-angle vision modules, whereas acoustic approaches have moved beyond medical ultrasound towards integrated earbud and headphone solutions that leverage subtle biomechanical cues. These innovations mark an important step towards accessible, privacy-preserving and device-integrated SSIs, but challenges remain in achieving consistent performance across diverse real-world settings and user profiles.

Optical sensing
Optical sensing enables silent speech input by visually capturing articulatory movements, including lip and facial dynamics, through off-body modalities. The most straightforward approach involves RGB cameras, which offer high-resolution visual data for articulator tracking[5,15,16]. However, this method is inherently sensitive to occlusion, head pose variation and lighting conditions and typically requires a fixed setup, limiting portability and real-world applicability.

To improve deployment flexibility, optical sensing has been integrated into mobile platforms. Front-facing smartphone cameras enable real-time interaction without auxiliary hardware (Fig. 3a)[17–19], whereas depth-sensing modules enhance robustness against environmental variation by capturing three-dimensional motion profiles, achieving 91.3% within-user accuracy and 74.9% cross-user accuracy on a 30-command vocabulary[20]. Nevertheless, both solutions remain constrained by frontal positioning and the instability of handheld use.

To overcome these limitations, alternative camera placements have been explored. Side-[21] and chin-mounted[22] optical systems provide more stable tracking and allow for more natural user movements during interaction, with reported performance exceeding 90% accuracy on vocabularies of approximately 50 words. Although challenges remain in ensuring consistent performance across diverse usage scenarios, optical sensing remains a user-friendly and hardware-light strategy that is particularly suited for applications prioritizing convenience and accessibility.

Acoustic sensing
Acoustic sensing has gained marked traction in SSIs due to its off-body implementation, strong resilience to occlusion and poor lighting, and inherent advantages for preserving user privacy. Although

Table 1 | Comparison of sensing modalities used in SSIs

Family	Modality	Sensing principle	Spatial resolution	Temporal resolution	Signal fidelity	Strengths	Limitations
Off body (non-contact or loose coupling; high comfort, lower specificity)	Optical	Visible or infrared	Sub-millimetre (lips) or millimetre scale (face)	30–60 Hz	Medium	Non-contact; commodity hardware	Lighting variability; occlusion; camera angle sensitivity
	Ultrasound	Air-coupled acoustic time-of-flight	Millimetre scale	30–100 Hz (frame based)	Medium	Access to deep vocal-tract geometry; robust to lighting	Bulky probe placement; needs calibration; airflow noise
On body (stable skin contact; balanced wearability and fidelity)	IMU	MEMS	Motion scale	100–200 Hz	Low to medium	Captures micro-motions; low cost and small size	Motion artefacts; head movement coupling; signal drift
	Strain	Piezoresistive or capacitive deformation	Millimetre scale	1–1,000 Hz (material dependent)	High	Skin conformal; comfortable	Thermal or mechanical drift; nonlinear response
	EMG	Surface myopotential	Millimetre scale	1–2 kHz (physiological band: 20–450 Hz)	Medium to high	Mature technology; direct muscle intent	Electrode impedance drift; motion/ sweat artefacts; placement sensitivity
	TENG	Triboelectric charge transfer	Millimetre scale	Event-triggered pulses (roughly microsecond scale)	Medium to high	No external power; high SNR pulses	Session-to-session variability; long-term stability
	EEG	Scalp neural potentials	Centimetre scale	1–2 kHz (physiological band: 0.5–100 Hz)	Low to medium	Non-invasive brain intent	Low SNR; setup time; hair interference
In body (implanted; highest fidelity with surgical invasiveness)	ECoG	Subdural cortical LFP	Millimetre scale	1–5 kHz (LFP band: 1–500 Hz)	High	High bandwidth; low noise	Requires craniotomy; patient specific
	sEEG	Depth-electrode sEEG	Millimetre scale	1–5 kHz (LFP band: 1–500 Hz)	High	3D access to speech network	Haemorrhage or infection risk
	MEA	Intracortical spikes	Cellular scale	30 kHz (spike band: 300–5,000 Hz)	Very high	Highest information rate; single-neuron decoding	Foreign-body response; long-term signal loss; most invasive

The reported values represent typical ranges across SSI studies and may vary depending on sensor configuration, body location, signal processing pipeline and application scenario. Modality classification (off body, on body or in body) reflects typical implementations within the SSI literature. 3D, three dimensional; LFP, local field potential; MEMS, micro-electro-mechanical systems; SNR, signal-to-noise ratio.

conventional speech decoding relies on external microphones to capture audible voice signals, such methods fall outside the scope of SSIs, which aim to decode speech intent in the absence of vocalized sound. One early-studied silent approach utilizes ultrasound imaging, wherein probes placed beneath the jaw capture fine-grained tongue kinematics with high spatial fidelity (Fig. 3b)[23]. This approach achieves a 3.6-s decoding speed with ~33% word error rate (WER) on vocabularies of several dozen commands, but its dependence on specialized imaging hardware restricts its practicality for daily use.

To circumvent such hardware constraints, several strategies have turned to commodity devices. Smartphone-based methods, for instance, emit inaudible acoustic signals via built-in speakers and analyse their reflections using onboard microphones[24–26]. The reflected echoes encode articulatory movements of the lips and tongue, allowing the systems to reach >90% word-level accuracy and <10% sentence-level WER across vocabularies spanning from simple commands to short conversational sentences. These systems offer a hardware-light solution, yet often face performance degradation in noisy or dynamic environments.

To enhance robustness and integrate seamlessly into everyday settings, researchers have embedded acoustic sensors into wearable devices. Glasses-mounted systems detect perioral skin deformation[27],

earbuds capture air-pressure variations within the ear canal[6,28] and headphones monitor temporomandibular joint motion[29]. Each configuration taps into distinct biomechanical cues, collectively enabling silent command recognition with minimal user effort and improved tolerance to ambient noise. Reported implementations have repeatedly achieved >90% accuracy on vocabularies of more than 100 words, delivering performance comparable to smartphone-based methods while offering substantially greater portability. Together, these diverse implementations underscore the versatility of acoustic sensing and highlight its growing relevance as a scalable, non-intrusive pathway for silent speech decoding.

Developmental trends and outlook

Off-body sensing has evolved from static, laboratory-bound cameras and ultrasound probes to wearable and device-integrated platforms such as smartphones, glasses and earbuds. Reported accuracies above 90% demonstrate feasibility, yet performance remains fragile under real-world lighting, acoustic noise and inter-user variability. Future progress hinges on environmental robustness, cross-user adaptation and edge AI integration, and socially acceptable and privacy-preserving form factors are likely to define scalability. Together, these directions frame off-body systems as the most accessible entry point for widespread SSI adoption.

SSIs using on-body sensors

Although off-body sensors offer high user comfort and ease of integration, their indirect coupling to physiological signals and sensitivity to environmental noise often limit decoding accuracy in unconstrained scenarios. To address these limitations, on-body sensing has emerged as a compelling alternative, providing closer physical proximity to the articulatory system and enabling more direct access to neuromuscular or biomechanical activity. By attaching directly to the body surface through either flexible materials or compact rigid modules, on-body sensors can achieve motion-resolved and often higher-fidelity capture of speech-related signals, even under motion, occlusion or low-light conditions.

This improvement in signal quality comes with trade-offs. Compared with off-body methods, on-body systems require physical contact, which introduces considerations around comfort, attachment stability and long-term usability. Nevertheless, these challenges are often outweighed by the benefits in scenarios that demand precision and robustness. On-body sensors have therefore gained traction not only in silent communication for daily use but also in clinical contexts, where they assist individuals with impaired speech. These systems have been explored in the decoding of residual muscle activity in patients with dysarthria, laryngectomy or neurodegenerative diseases, offering an accessible and responsive interface where conventional acoustic speech is unavailable. The enhanced signal access and application versatility position on-body sensing as a critical component in the development of both consumer and healthcare-grade SSIs.

Inertial measurement unit sensing

Inertial measurement units (IMUs), long used in gait and gesture recognition, have only recently gained traction for SSIs due to difficulty resolving fine-scale articulatory kinematics amid head motion. Early systems used facial accelerometer arrays to detect speech-induced vibrations, achieving high accuracy (94.65 ± 2.54% in classifying 40 English words) but suffering from head-motion artefacts that constrained real-world use[30]. Differential sensing paradigms were developed to tackle this challenge, where signals from articulator-mounted IMUs are fused with reference sensors on stable regions such as the forehead or ears to isolate speech-specific motion (Fig. 3c), achieving an average accuracy of 92% across seven users for actual continuous lip-speech recognition on 93 English sentences[9]. The method offers strong environmental robustness, with immunity to lighting, occlusion and noise, and supports ultra-low-power, consumer-ready form factors. Despite sacrificing spatial specificity compared with bioelectric methods, their motion resilience and wearability position IMUs as scalable solutions for mobile SSI deployment.

Triboelectric nanogenerator sensing

Triboelectric nanogenerators (TENGs) convert mechanical deformation into electrical signals via contact electrification and electrostatic induction. Their self-powered, low-cost and flexible design makes them attractive for wearable sensing[31,32]. In SSIs, TENGs enable energy-autonomous detection of articulatory motion. Recent studies have shown that soft, skin-mounted TENG arrays can capture lip dynamics and decode phrases using machine learning (Fig. 3d), yielding an accuracy of 94.5% on 20 words[33]. Compared with optical methods, which depend on ambient illumination, TENGs provide greater privacy and robustness in variable environments. They can also be seamlessly embedded into daily wear, such as face masks or patches, making them unobtrusive for long-term use. However, their outputs depend on tribo-charge state and contact regime and are sensitive to humidity, sweat and material ageing. Combined with the ultra-high source impedance and load-dependent readout, this leads to non-stationarity and session-to-session gain drift that often requires charge management, high-impedance front ends and periodic calibration.

EMG sensing

Surface EMG provides non-invasive direct access to muscle activation underlying articulation, enabling robust decoding of speech intent in silent contexts. Compared with inertial or force-based methods, EMG captures signals at the neuromuscular source, offering higher specificity but requiring stable skin contact and being susceptible to motion artefacts[34,35]. Early systems used rigid electrodes and benchtop acquisition setups, limiting usability. Recent advances in conformal bioelectronics have addressed this. Tattoo-like EMG sensors affixed to facial muscles enabled high-accuracy word-level decoding under dynamic, real-world conditions, recognizing up to 110 daily-use words with an average accuracy of 92.6% (Fig. 3e)[36]. Additionally, textile-based EMG electrodes integrated into headphone earmuffs captured neuromuscular activity from periauricular and jaw-adjacent muscles and leveraged a multi-channel adaptive decoding network to dynamically weight signal quality. This setup demonstrated high usability and robustness in mobile contexts, achieving 96% accuracy on ten commonly used voice-free control words[37]. Together, these studies exemplify a convergence of high-fidelity sensing with ergonomic form factors.

Strain sensing

Strain sensors capture subtle deformations of facial and laryngeal tissue during articulation, providing a direct mechanical interface between user intent and silent speech decoding. Foundational work in wearable strain sensors established the feasibility of stretchable, skin-conformal materials for physiological monitoring across dynamic body surfaces[38,39]. Translating this concept to SSI, researchers have explored a range of material strategies to balance comfort, signal stability and deployment readiness.

Early demonstrations using ultrathin silicon gauges achieved high sensitivity and fast relaxation times, enabling robust word-level decoding under skin strain (87.53% accuracy among 100 words)[8]. Building on bioinspired mechanisms, ionic hydrogel sensors mimicked mechanoreceptor transduction to detect throat vibrations without requiring electrical contact, achieving an average accuracy of 95% in the 26-instruction test[40]. Hybrid systems integrating facial deformation and subcutaneous vibration cues revealed that combined motion signatures carry rich, decodable linguistic information even in noisy settings, demonstrating an average accuracy of 99.05% in classifying basic speech elements (phonemes, tones and words)[41].

Textile-based strain sensors further advanced the field by embedding high-resolution sensing into wearable form factors optimized for daily use (Fig. 3f), achieving 95.25% accuracy on 20 frequently used

Fig. 3 | Representative sensor modalities used in SSIs. a–i, Schematics illustrating key sensing technologies enabling silent speech decoding, categorized by sensing principle: optical sensing (**a**; a smartphone-based front camera detects articulatory motion, such as lip and jaw movements); ultrasonic sensing (**b**; an ultrasound imaging probe beneath the jaw visualizes tongue motion in real time); IMU sensing (**c**; IMUs distributed at the head, lip and chin track multi-point facial kinematics during articulation); triboelectric sensing (**d**; self-powered wearable sensors detect facial motion through contact electrification); EMG (**e**; tattoo-like epidermal electrodes acquire facial myopotentials associated with silent articulation); strain sensing (**f**; textile strain sensors embedded in a smart choker capture throat deformation patterns); EEG (**g**; in-ear conformal bioelectronics measure brain activity associated with speech imagery); ECoG (**h**; implanted cortical arrays decode neural activity from speech-generating regions in patients with brainstem injury); and MEA (**i**; intracortical electrodes implanted in the motor cortex record neural activity associated with fine motor intention, enabling high-resolution decoding of attempted handwriting and speech imagery). PI, polyimide; PVC, polyvinyl chloride. Panels adapted from: **d**, ref. 33, CC BY 4.0; **e**, ref. 36, CC BY 4.0; **f**, ref. 42, CC BY 4.0; **g**, ref. 47, CC BY 4.0.
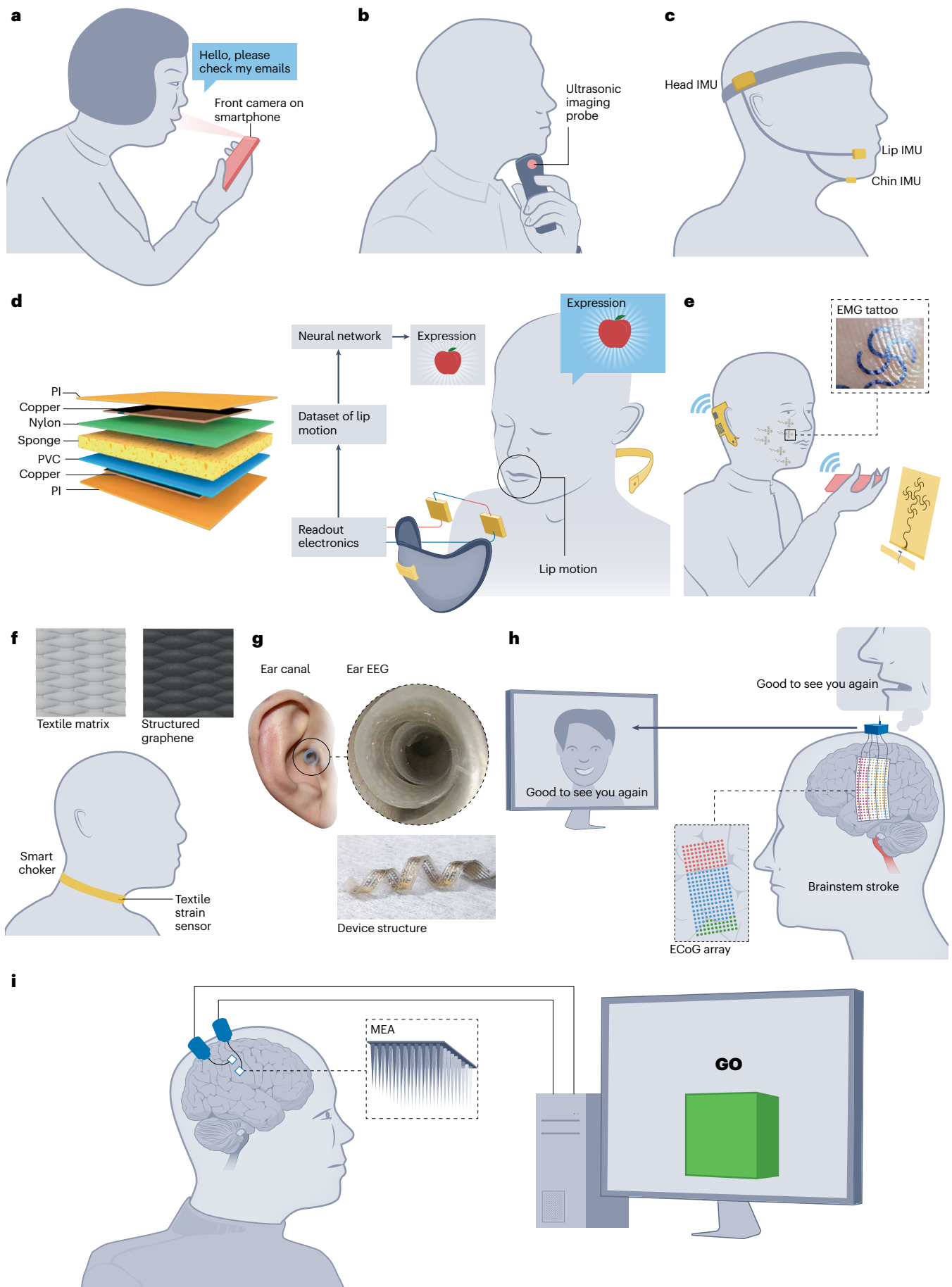
English words[42]. These systems have evolved from controlled experiments to real-world trials in patients who have suffered strokes, where strain signals were leveraged not only for word decoding but also for capturing emotional context, achieving a 4.2% WER and improving daily communication satisfaction by 55% in five patients recovering from strokes[43]. Across these developments, strain sensing highlights how material and structural innovations can directly shape the usability and expressiveness of silent speech technologies.

Electroencephalography sensing

Electroencephalography (EEG) enables non-invasive access to cortical activity via scalp or ear-adjacent electrodes, offering a wearable, surface-based approach to capture neural correlates of speech intent upstream of articulation. However, EEG signals are inherently noisy and spatially diffuse, posing major challenges for speech decoding. Foundational studies demonstrated that event-related potentials[44] and modulated sensorimotor rhythms[45] can support basic intent detection, laying the groundwork for EEG-based communication.

Recent efforts have redefined the role of EEG in silent speech decoding by innovating at the sensor and system levels. Ear-centred[46] and in-ear EEG devices (Fig. 3g)[47] have matched traditional scalp setups in decoding accuracy while enhancing comfort and form factor. Meanwhile, to mitigate physiological and motion artefact limitations[48], recent studies have explored integrated multimodal platforms combining EEG with additional modalities[49,50]. By providing complementary spatial and physiological signals (for example, inertial, haemodynamic or muscular activity), these systems enable artefact identification and weighting in decoding pipelines, thereby improving robustness. Moving beyond command classification, large-scale studies now leverage contrastive learning to align EEG signals with deep speech representations[51], enabling zero-shot decoding of perceived sentences without retraining on specific vocabularies, with an average accuracy of ~41% (up to 80% in some individuals) over more than 1,000 candidate segments.

Although EEG is constrained by low signal fidelity and high inter-subject variability, its distinct advantage lies in the ability to access pre-articulatory neural activity, offering a uniquely scalable non-invasive modality for early-stage intent decoding. Although current accuracies remain modest compared with peripheral sensing methods, ongoing advances in sensor design and learning architectures highlight the potential of EEG to enable predictive and generalized SSI systems.

Developmental trends and outlook

On-body sensors have evolved from rigid electrodes to soft, skin-conformal bioelectronics and energy-harvesting textiles, aiming to balance signal fidelity with everyday wearability. However, this proximity introduces vulnerability to motion artefacts, skin impedance drift and long-term comfort challenges. Recent progress increasingly hinges on functionally complementary multimodal fusion. For example, EMG signals capture neuromuscular intent but are motion sensitive, whereas strain sensors track tissue deformation and sensor displacement. Fusing the two allows decoding models to contextualize signal quality, improving robustness under movement.

This shift reframes sensor fusion not as redundancy, but as a deliberate strategy to resolve the fidelity–stability trade-off inherent in wearable decoding. Combined with advances in stretchable substrates and low-power design, on-body platforms are positioned as the most deployable SSI solution, balancing accuracy with resilience in real-world conditions.

SSIs using in-body sensors

In-body sensing offers the most direct access to the neural substrates of speech, capturing activity from cortical regions responsible for planning and articulation. By implanting electrodes either onto the brain surface or within its depths, these systems bypass peripheral musculature and enable speech decoding in individuals who have lost all voluntary motor output. They are uniquely positioned to support communication in cases of locked-in syndrome or advanced neurodegeneration, where no other interface is viable.

Although in-body approaches provide exceptional signal fidelity and decoding precision, they require invasive procedures, patient-specific adaptation and long-term clinical support. As such, their application remains limited to research settings and highly selected clinical scenarios. Nevertheless, recent progress in electrode miniaturization, signal processing and neurosurgical techniques has expanded the scope of implanted speech interfaces. These systems not only advance assistive communication but also serve as platforms for probing the neural basis of language and developing future brain-centred interaction technologies.

ECoG sensing

ECoG acquires high-fidelity local field potentials via subdural electrode grids placed over cortical speech regions. Compared with non-invasive EEG, ECoG offers superior spatial resolution and bandwidth with reduced signal attenuation, enabling direct access to the neural substrates of articulation[52,53]. Its semi-invasive nature positions it between scalp-based and intracortical approaches, making it a clinically viable modality for patients with severe motor speech impairments.

Over the past decade, ECoG-based silent speech systems have progressed from decoding isolated phonemes to reconstructing full sentences in real time[54,55]. Further evolution reflects a broader shift from offline, trial-based studies to continuous, streaming paradigms that prioritize naturalistic communication (Fig. 3h), achieving large-vocabulary decoding at up to 78 words per minute with ~25% WER[56] and enabling online fluent synthesis in 80-ms increments[57]. Recent efforts emphasize low-latency, high-intelligibility speech synthesis directly from neural activity, marking a transition towards closed-loop brain-to-speech systems. In parallel, the development of high-density micro-electrocorticography arrays has enabled finer spatial resolution and improved signal fidelity, reinforcing a device-level trend towards more precise, information-rich neural interfaces[58]. As decoding architectures mature and deployment barriers narrow, ECoG stands as the most clinically advanced in-body interface for restoring communication in individuals with profound speech loss.

Stereo-EEG sensing

Stereo-EEG (sEEG) records intracranial neural activity via depth electrodes implanted in cortical and subcortical regions, offering three-dimensional spatial coverage. Compared with ECoG, which requires craniotomy to place grid or strip electrodes directly on the cortical surface, sEEG electrodes are introduced through stereotactically guided burr holes. This minimally invasive surgical approach generally carries lower perioperative morbidity, despite the deeper implantation sites[59,60]. This accessibility to deeper structures makes sEEG particularly valuable for capturing both cortical and subcortical elements of speech planning and tone modulation.

Recent advances have enhanced sEEG as a sensing modality by improving both stimulation and recording strategies. Intermediate-frequency protocols increase mapping sensitivity while reducing afterdischarges, enabling more stable and spatially specific probing of language circuits[61]. Complementing these protocol-level refinements, the release of structured sEEG datasets capturing vocalized, mimed and imagined speech provides a richer basis for modelling the sensor-to-signal relationship in diverse linguistic contexts[62]. Together, these developments position sEEG as a minimally invasive and increasingly optimized sensing platform for speech decoding.

Microelectrode array sensing

Intracortical microelectrode arrays (MEAs) offer direct access to the spiking activity of neurons, providing unparalleled temporal
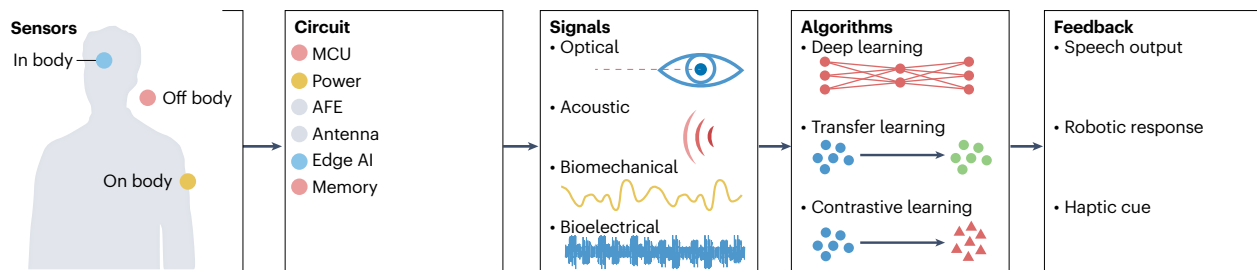
Fig. 4 | Conceptual integration of sensing, processing and feedback in future SSIs. SSIs follow a multi-stage pipeline beginning with the acquisition of articulatory or neural signals via off-, on- or in-body sensors. These signals are first conditioned by hardware circuits comprising analogue front ends (AFEs), microcontrollers (MCUs), wireless modules (such as Bluetooth or WiFi) and power and memory units. Although most systems transmit pre-processed data to external devices for inference, a subset integrates edge AI processors to enable local, low-latency decoding. The captured signals span optical, acoustic, biomechanical and bioelectrical domains, each reflecting distinct aspects of speech-related activity. Decoding is achieved through algorithmic frameworks, including deep learning, transfer learning and contrastive learning, enabling context-aware interpretation across users and settings. The outputs are translated into feedback modalities, such as synthesized speech, robotic actions or haptic cues, forming a closed-loop interface for assistive and interactive applications.

resolution and spatial specificity for speech decoding[63,64]. By penetrating cortical tissue, these sensors can capture fine-scale dynamics of speech motor planning that are not accessible through surface-level recordings.

Initial demonstrations showed that MEAs could support phoneme-level decoding in open-loop paradigms, laying the groundwork for intracortical speech interfaces[65]. More recently, their integration into closed-loop systems has enabled real-time synthesis of intelligible words and phrases in individuals with severe speech loss (Fig. 3i)[66]. These advances mark a shift from offline analysis to continuous decoding pipelines aimed at restoring functional communication.

Despite their exceptional signal fidelity, MEAs face practical limitations including surgical invasiveness, long-term biocompatibility and signal degradation over time. Nevertheless, they remain the benchmark for understanding the upper bounds of neural resolution in brain-to-speech interfaces.

Developmental trends and outlook
In-body interfaces represent the frontier of SSI potential, offering direct access to cortical speech networks and enabling decoding capabilities that are fundamentally beyond the reach of peripheral sensing. Although current systems have yet to match the real-time accuracy of advanced on-body solutions, particularly in healthy users, their unique strength lies in restoring communication when peripheral musculature is no longer viable. This transformative promise, however, is inseparable from profound biological and ethical tensions. Biologically, craniotomy, long-term biocompatibility challenges and foreign-body responses threaten signal stability, and the absence of fully implantable wireless systems increases infection risk and hinders long-term deployment. Ethically, these interfaces engage directly with the neural substrates of language, raising unprecedented concerns around mental privacy, data ownership and informed consent for the continuous decoding of inner speech. Far from ancillary, these risks will fundamentally shape the pace, scope and social acceptability of in-body SSI technologies.

Looking ahead, progress in minimally invasive electrodes, fully sealed wireless closed-loop platforms and adaptive learning architectures may help to reconcile the demands of high decoding fidelity with the requirements of long-term biological safety. At the same time, robust governance frameworks addressing privacy, data rights and user agency will be essential to uphold ethical standards and ensure responsible deployment. Taken together, these technological and ethical developments may allow in-body systems to serve not only as clinical tools for individuals with severe impairments, but also as scientific instruments for exploring and interacting with the neural mechanisms underlying human language.

Conclusions
SSIs are reshaping the landscape of human communication by enabling speech decoding without audible output. This Review examines the field through the lens of sensing technologies, revealing how sensing configurations fundamentally shape system fidelity, comfort and usability. By categorizing SSIs into off-, on- and in-body modalities, we have highlighted the trade-offs that govern their deployment—from ambient, non-contact interaction to implantable neuroprosthetics.

These sensing choices are not merely technical parameters but strategic levers that determine who can benefit, where systems can operate and how effectively silent speech can be translated into actionable outputs. With the convergence of flexible electronics, neuromuscular decoding and intelligent feedback, SSIs are emerging not only as assistive tools for individuals with speech impairments but also as scalable platforms for future human–machine interaction.

Yet, sensing alone does not determine system capability. As SSIs transition towards real-world deployment, the co-evolution of sensing, embedded hardware and learning algorithms is becoming increasingly central. In this context, sensing acts not only as an input modality, but as the foundation of tightly coupled signal processing pipelines that enable responsive, low-latency interaction.

Signals acquired from off-, on- or in-body sensors must be routed through analogue front ends, microcontrollers and wireless modules. Modern SSI systems increasingly incorporate edge AI processors that support real-time, low-power inference directly on wearable or mobile platforms, minimizing latency and safeguarding privacy[67]. This hardware–algorithm co-design paradigm underpins recent advances, including throat-mounted acoustic and biomechanical sensors that drive robotic control via silent commands[68].

Figure 4 provides a conceptual overview of this integrated pipeline—from acquisition to decoding to feedback. The captured signals span optical, acoustic, biomechanical and bioelectrical domains, each offering distinct trade-offs between fidelity and comfort. For instance, wearable ultrasonic sensors and strain gauges have demonstrated strong signal-to-noise ratios and resilience to ambient interference[8,24], in some cases outperforming surface EMG in specificity[7]. These signals are standardized through pre-processing pipelines and increasingly fused across modalities to improve robustness across users and settings.

The decoding stage is primarily driven by lightweight deep learning models—convolutional and transformer-based architectures that support near-real-time performance on embedded systems[42,69]. Transfer learning and few-shot personalization enable rapid adaptation to new vocabularies or users[11,70], whereas contrastive and cross-modal learning increasingly bridge silent and vocal speech domains using

large-scale audio datasets[17,51]. These algorithmic advances substantially reduce data requirements and enhance generalization.

The final outputs of SSIs range from text and synthesized speech to direct robotic or digital commands[43,68]. Real-time feedback, such as haptic cues or voice synthesis, is crucial for enabling closed-loop, interactive communication. Although challenges remain in generalization, energy efficiency and vocabulary breadth, the synergistic development of sensing hardware, embedded inference and learning architectures is rapidly moving SSIs towards widespread real-world application.

Looking ahead, the next wave of hardware advances will centre on sensor evolution, miniaturization and integration. Flexible and stretchable materials will continue to transform sensors into conformal, skin-like interfaces capable of robust acquisition under motion and daily wear[71–73]. The emergence of hybrid systems, merging bioelectrical, biomechanical and acoustic domains, will offer richer signal streams and redundancy to counteract artefacts and variability[74]. Real-time decoding will be increasingly enabled by edge AI processors embedded in wearable circuits, unlocking low-latency inference and feedback without cloud reliance[75]. These developments are complemented by ultra-low-power design and energy-autonomous sensors such as TENGs, paving the way for continuous, passive sensing ecosystems. As sensing hardware becomes more discreet, adaptive and multimodal, it will not only reshape how silent speech is captured, but enlarge where and by whom it can be used.

Beyond hardware, the software layer of SSIs is entering a new era driven by embodied intelligence and large-scale models. Algorithmically, we anticipate a shift from task-specific neural networks towards multimodal foundation models that integrate audio, EMG, strain and even neural data through shared embeddings and attention mechanisms[76]. This will enable cross-modal learning, few-shot personalization and generalization across vocabularies and users. Embedded inference will increasingly adopt neuromorphic or event-driven architectures, optimizing latency and energy consumption for real-world use[77,78]. Furthermore, SSIs are positioned to serve as a key interface for human body digital twins, where physiological signals are continuously mapped onto real-time avatars for speech, emotion and motor intent, enabling closed-loop interaction across healthcare, social robotics and communication prosthetics[79]. As algorithms evolve to reflect not just learned data but the embodied dynamics of human expression, the boundary between input and interface will blur.

In the broader societal context, the future of SSIs extends far beyond clinical assistive technology. As silent interfaces become embedded in daily life, supporting unobtrusive interaction in public spaces, privacy-preserving commands in shared environments and silent collaboration in noisy or sensitive contexts, they will extend the boundary of human–human and human–machine communication. More importantly, for individuals with profound speech or motor disabilities, SSIs might offer not just a tool but a restoration of agency and identity, helping to alleviate psychological distress and foster more effective rehabilitation[80]. By grounding these technologies in inclusive design and responsible innovation, we can ensure that silent speech systems contribute not only to technical progress but also to human dignity, empowerment and equitable access to communication.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

References
1. Denby, B. et al. Silent speech interfaces. *Speech Commun.* **52**, 270–287 (2010).
   This comprehensive review formalizes SSIs as a research field.
2. Gonzalez-Lopez, J. A. et al. Silent speech interfaces for speech restoration: a review. *IEEE Access* **8**, 177995–178021 (2020).
3. Guenther, F. H. et al. in *Speech Motor Control in Normal and Disordered Speech* 29–49 (Oxford Univ. Press, 2004).
4. Silva, A. B. et al. The speech neuroprosthesis. *Nat. Rev. Neurosci.* **25**, 473–492 (2024).
5. Afouras, T. et al. Deep audio-visual speech recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 8717–8727 (2022).
6. Jin, Y. et al. EarCommand: "hearing" your silent speech commands in ear. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **6**, 1–28 (2022).
7. 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).
   This study introduces flexible surface EMG sensors to silent-speech-related human–machine interfaces.
8. Kim, T. et al. Ultrathin crystalline-silicon-based strain gauges with deep learning algorithms for silent speech interfaces. *Nat. Commun.* **13**, 5815 (2022).
9. Liu, S. et al. A data-efficient and easy-to-use lip language interface based on wearable motion capture and speech movement reconstruction. *Sci. Adv.* **10**, eado9576 (2024).
10. Card, N. S. et al. An accurate and rapidly calibrating speech neuroprosthesis. *N. Engl. J. Med.* **391**, 609–618 (2024).
11. Willett, F. R. et al. A high-performance speech neuroprosthesis. *Nature* **620**, 1031–1036 (2023).
12. Sugie, N. & Tsunoda, K. A speech prosthesis employing a speech synthesizer: vowel discrimination from perioral muscle activities. *IEEE Trans. Biomed. Eng.* **32**, 485–490 (1985).
13. Petajan, E. D. Automatic lipreading to enhance speech recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 40–47 (IEEE, 1985).
   This conference paper describes an off-body silent speech decoding system using camera-based automatic lipreading.
14. Suppes, P., Lu, Z.-L. & Han, B. Brain wave recognition of words. *Proc. Natl Acad. Sci. USA* **94**, 14965–14969 (1997).
15. Assael, Y. M., Shillingford, B., Whiteson, S. & de Freitas, N. LipNet: end-to-end sentence-level lipreading. Preprint at https://doi.org/10.48550/arXiv.1611.01599 (2016).
16. Wand, M., Koutník, J. & Schmidhuber, J. Lipreading with long short-term memory. In *Proc. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing* 6115–6119 (IEEE, 2016).
17. Su, Z., Fang, S. & Rekimoto, J. LipLearner: customizable silent speech interactions on mobile devices. In *Proc. 2023 CHI Conference on Human Factors in Computing Systems* 1–21 (ACM, 2023).
18. Pandey, L. & Arif, A. S. LipType: a silent speech recognizer augmented with an independent repair model. In *Proc. 2021 CHI Conference on Human Factors in Computing Systems* 1–19 (ACM, 2021).
19. Pandey, L. & Arif, A. S. MELDER: the design and evaluation of a real-time silent speech recognizer for mobile devices. In *Proc. CHI Conference on Human Factors in Computing Systems* 1–23 (ACM, 2024).
20. Wang, X., Su, Z., Rekimoto, J. & Zhang, Y. Watch your mouth: silent speech recognition with depth sensing. In *Proc. CHI Conference on Human Factors in Computing Systems* 1–15 (ACM, 2024).
21. Chen, T. et al. C-Face: continuously reconstructing facial expressions by deep learning contours of the face with ear-mounted miniature cameras. In *Proc. 33rd Annual ACM Symposium on User Interface Software and Technology* 112–125 (ACM, 2020).
22. Zhang, R. et al. SpeeChin: a smart necklace for silent speech recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **5**, 1–23 (2021).
23. Kimura, N., Kono, M. & Rekimoto, J. SottoVoce: an ultrasound imaging-based silent speech interaction using deep neural networks. In *Proc. 2019 CHI Conference on Human Factors in Computing Systems* 1–11 (ACM, 2019).

24. Tan, J., Nguyen, C.-T. & Wang, X. SilentTalk: lip reading through ultrasonic sensing on mobile phones. In *Proc. IEEE INFOCOM 2017—IEEE Conference on Computer Communications* 1–9 (IEEE, 2017).

25. Gao, Y. et al. EchoWhisper: exploring an acoustic-based silent speech interface for smartphone. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **4**, 1–27 (2020).

26. Zhang, Q. et al. SoundLip: enabling word and sentence-level lip interaction for smart devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **5**, 1–28 (2021).

27. Zhang, R. et al. EchoSpeech: continuous silent speech recognition on minimally-obtrusive eyewear powered by acoustic sensing. In *Proc. 2023 CHI Conference on Human Factors in Computing Systems* 1–18 (ACM, 2023).

28. Dong, X. et al. ReHEarSSE: recognizing hidden-in-the-ear silently spelled expressions. In *Proc. CHI Conference on Human Factors in Computing Systems* 1–16 (ACM, 2024).

29. Zhang, R. et al. HPSpeech: silent speech interface for commodity headphones. In *Proc. 2023 International Symposium on Wearable Computers* 60–65 (ACM, 2023).

30. Kwon, J. et al. Novel three-axis accelerometer-based silent speech interface using deep neural network. *Eng. Appl. Artif. Intell.* **120**, 105909 (2023).

31. Zhou, Q. et al. Triboelectric nanogenerator-based sensor systems for chemical or biological detection. *Adv. Mater.* **33**, 2008276 (2021).

32. Wang, Z. L. On Maxwell's displacement current for energy and sensors: the origin of nanogenerators. *Mater. Today* **20**, 74–82 (2017).

33. Lu, Y. et al. Decoding lip language using triboelectric sensors with deep learning. *Nat. Commun.* **13**, 1401 (2022).
    This paper reports on a triboelectric-sensor-based silent speech decoding system.

34. De Luca, C. J. The use of surface electromyography in biomechanics. *J. Appl. Biomech.* **13**, 135–163 (1997).

35. Phinyomark, A. et al. Feature extraction and selection for myoelectric control based on wearable EMG sensors. *Sensors* **18**, 1615 (2013).

36. Wang, Y. et al. All-weather, natural silent speech recognition via machine-learning-assisted tattoo-like electronics. *npj Flex. Electron.* **5**, 20 (2021).

37. Tang, C. et al. Wireless silent speech interface using multi-channel textile EMG sensors integrated into headphones. *IEEE Trans. Instrum. Meas.* **74**, 1–10 (2025).

38. Lipomi, D. J. et al. Skin-like pressure and strain sensors based on transparent elastic films of carbon nanotubes. *Nat. Nanotechnol.* **6**, 788–792 (2011).

39. Amjadi, M. et al. Stretchable, skin-mountable, and wearable strain sensors and their potential applications: a review. *Adv. Funct. Mater.* **26**, 1678–1698 (2016).

40. Xu, S. et al. Force-induced ion generation in zwitterionic hydrogels for a sensitive silent-speech sensor. *Nat. Commun.* **14**, 219 (2023).

41. Yang, Q. et al. Mixed-modality speech recognition and interaction using a wearable artificial throat. *Nat. Mach. Intell.* **5**, 169–180 (2023).

42. Tang, C. et al. Ultrasensitive textile strain sensors redefine wearable silent speech interfaces with high machine learning efficiency. *npj Flex. Electron.* **8**, 27 (2024).

43. Tang, C. et al. Wearable intelligent throat enables natural speech in stroke patients with dysarthria. Preprint at https://doi.org/10.48550/arXiv.2411.18266 (2024).
    This on-body sensing study demonstrates generalized silent speech decoding directly in patients with speech impairment.

44. Farwell, L. A. & Donchin, E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* **70**, 510–523 (1988).

45. Wolpaw, J. R. & McFarland, D. J. Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans. *Proc. Natl Acad. Sci. USA* **101**, 17849–17854 (2004).

46. Kaongoen, N., Choi, J. & Jo, S. Speech-imagery-based brain–computer interface system using ear-EEG. *J. Neural Eng.* **18**, 016023 (2021).

47. Wang, Z. et al. Conformal in-ear bioelectronics for visual and auditory brain–computer interfaces. *Nat. Commun.* **14**, 4213 (2023).

48. Occhipinti, E., Davies, H. J., Hammour, G. & Mandic, D. P. Hearables: artefact removal in ear-EEG for continuous 24/7 monitoring. In *Proc. 2022 International Joint Conference on Neural Networks* 1–6 (IEEE, 2022).

49. Mandic, D. P. et al. In your ear: a multimodal hearables device for the assessment of the state of body and mind. *IEEE Pulse* **14**, 17–23 (2023).

50. Cooney, C. et al. A bimodal deep learning architecture for EEG-fNIRS decoding of overt and imagined speech. *IEEE Trans. Biomed. Eng.* **69**, 1983–1994 (2021).

51. Défossez, A. et al. Decoding speech perception from non-invasive brain recordings. *Nat. Mach. Intell.* **5**, 1097–1107 (2023).

52. Leuthardt, E. C. et al. A brain–computer interface using electrocorticographic signals in humans. *J. Neural Eng.* **1**, 63–71 (2004).

53. Schalk, G. & Leuthardt, E. C. Brain–computer interfaces using electrocorticographic signals. *IEEE Rev. Biomed. Eng.* **4**, 140–154 (2011).

54. Angrick, M. et al. Speech synthesis from ECoG using densely connected 3D convolutional neural networks. *J. Neural Eng.* **16**, 036019 (2019).

55. Moses, D. A. et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *N. Engl. J. Med.* **385**, 217–227 (2021).

56. Metzger, S. L. et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature* **620**, 1037–1046 (2023).

57. Littlejohn, K. T. et al. A streaming brain-to-voice neuroprosthesis to restore naturalistic communication. *Nat. Neurosci.* **28**, 1–11 (2025).

58. Duraivel, S. et al. High-resolution neural recordings improve the accuracy of speech decoding. *Nat. Commun.* **14**, 6938 (2023).

59. Munari, C. et al. Stereo-electroencephalography methodology: advantages and limits. *Acta Neurol. Scand.* **89**, 56–67 (1994).

60. Mullin, J. P. et al. Is SEEG safe? A systematic review and meta-analysis of stereo-electroencephalography-related complications. *Epilepsia* **57**, 386–401 (2016).

61. Abarrategui, B. et al. New stimulation procedures for language mapping in stereo-EEG. *Epilepsia* **65**, 1720–1729 (2024).

62. He, T. et al. VocalMind: a stereotactic EEG dataset for vocalized, mimed, and imagined speech in tonal language. *Sci. Data* **12**, 657 (2025).

63. Schwartz, A. B. et al. Brain-controlled interfaces: movement restoration with neural prosthetics. *Neuron* **52**, 205–220 (2006).

64. Flint, R. D. et al. Accurate decoding of reaching movements from field potentials in the human motor cortex. *J. Neural Eng.* **9**, 046006 (2012).

65. Mugler, E. M. et al. Direct classification of all American English phonemes using signals from functional speech motor cortex. *J. Neural Eng.* **11**, 035015 (2014).

66. Willett, F. R. et al. High-performance brain-to-text communication via handwriting decoding. *Nature* **593**, 249–254 (2021).

67. Matsumura, G. et al. Real-time personal healthcare data analysis using edge computing for multimodal wearable sensors. *Device* **3**, 100597 (2025).

68. Liu, T. et al. Machine learning-assisted wearable sensing systems for speech recognition and interaction. *Nat. Commun.* **16**, 2363 (2025).

69. Cai, D. et al. SILENCE: protecting privacy in offloaded speech understanding on resource-constrained devices. *Adv. Neural Inform. Proc. Syst.* **37**, 105928–105948 (2024).

70. Deng, Z. et al. Silent speech recognition based on surface electromyography using a few electrode sites under the guidance from high-density electrode arrays. *IEEE Trans. Instrum. Meas.* **72**, 1–11 (2023).

71. Pang, C. et al. A flexible and highly sensitive strain-gauge sensor using reversible interlocking of nanofibres. *Nat. Mater.* **11**, 795–801 (2012).

72. Tang, C. et al. A deep learning-enabled smart garment for accurate and versatile monitoring of sleep conditions in daily life. *Proc. Natl Acad. Sci. USA* **122**, e2420498122 (2025).

73. Xu, M. et al. Simultaneous isotropic omnidirectional hypersensitive strain sensing and deep learning-assisted direction recognition in a biomimetic stretchable device. *Adv. Mater.* **37**, 2420322 (2025).

74. Gong, S. et al. Hierarchically resistive skins as specific and multimetric on-throat wearable biosensors. *Nat. Neurosci.* **18**, 889–897 (2023).

75. Xue, C. et al. A CMOS-integrated compute-in-memory macro based on resistive random-access memory for AI edge devices. *Nat. Electron.* **4**, 81–90 (2021).

76. Fei, N. et al. Towards artificial general intelligence via a multimodal foundation model. *Nat. Commun.* **13**, 3094 (2022).

77. Roy, K. et al. Towards spike-based machine intelligence with neuromorphic computing. *Nature* **575**, 607–617 (2019).

78. Wang, S. et al. Memristor-based adaptive neuromorphic perception in unstructured environments. *Nat. Commun.* **15**, 4671 (2024).

79. Tang, C. et al. A roadmap for the development of human body digital twins. *Nat. Rev. Elect. Eng.* **1**, 199–207 (2024).

80. Zinn, S. et al. The effect of poststroke cognitive impairment on rehabilitation process and functional outcome. *Arch. Phys. Med. Rehab.* **85**, 1084–1090 (2004).

81. Kennedy, P. R. & Bakay, R. A. E. Restoration of neural output from a paralyzed patient by a direct brain connection. *Neuroreport* **9**, 1707–1711 (1998).
    This paper reports on a patient-level in-body neural interface demonstrating direct brain-connected communication in a paralysed individual.

82. Denby, B. & Stone, M. Speech synthesis from real-time ultrasound images of the tongue. In *Proc. 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing* 685–688 (IEEE, 2004).
    This conference paper demonstrates silent speech synthesis using real-time ultrasound imaging of the tongue.

83. Galatas, G., Potamianos, G. & Makedon, F. Audio-visual speech recognition incorporating facial depth information captured by the Kinect. In *Proc. 20th European Signal Processing Conference* 2714–2717 (IEEE, 2012).

84. Shin, Y. H. & Seo, J. Towards contactless silent speech recognition based on detection of active and visible articulators using IR-UWB radar. *Sensors* **16**, 1812 (2016).

85. Sun, K. et al. Lip-Interact: improving mobile device interaction with silent speech commands. In *Proc. 31st Annual ACM Symposium on User Interface Software and Technology* 581–593 (ACM, 2018).

86. Anumanchipalli, G. K., Chartier, J. & Chang, E. F. Speech synthesis from neural decoding of spoken sentences. *Nature* **568**, 493–498 (2019).
    This paper reports a neural decoder that synthesizes full spoken sentences from ECoG recordings.

87. Ravenscroft, D. et al. Machine learning methods for automatic silent speech recognition using a wearable graphene strain gauge sensor. *Sensors* **22**, 299 (2022).

88. Che, Z. et al. Speaking without vocal folds using a machine-learning-assisted wearable sensing-actuation system. *Nat. Commun.* **15**, 1873 (2024).

Acknowledgements

Author contributions

C.T., S.G. and L.G.O. discussed and conceived of the idea for the paper. C.T., L.Q. and S.G. drafted the manuscript. C.T. and Z.Z. visualized the figures. S.G. and L.G.O. supervised the work. C.T., S.G., W.Y., M.X., E.O., Y.P. and L.G.O. contributed to discussing, writing and revising the article.

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/s44460-025-00010-2.

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

Peer review information *Nature Sensors* thanks Jun Chen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Corresponding author(s):   Shuo Gao, Luigi G. Occhipinti

Last updated by author(s):   Nov 24, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	This Review Article does not involve new data collection, experiments, or acquisition of primary datasets.
Data analysis	No new data analyses, statistical pipelines, or computational workflows were performed. All results discussed in the manuscript are derived from previously published studies.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

This Review Article does not generate or analyse any new datasets. All data discussed in the manuscript are derived from previously published studies, which are fully cited in the References. No new experimental, clinical, or computational data were produced for this work.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](). See also policy information about [sex, gender (identity/presentation), and sexual orientation]() and [race, ethnicity and racism]().

Reporting on sex and gender	This Review Article does not involve new human participant research or the analysis of sex- or gender-disaggregated data. All human studies discussed are derived from previously published work, and sex or gender information is reported as available within those original publications.
Reporting on race, ethnicity, or other socially relevant groupings	This Review Article does not involve new human participant research and does not collect or analyse data related to race, ethnicity, or other socially relevant groupings. Any such information mentioned in the manuscript is drawn from previously published studies and reported as available in those original sources.
Population characteristics	No new participant data were collected or analysed in this Review. All population characteristics described in the manuscript refer to previously published studies, which are fully cited.
Recruitment	This Review does not involve recruitment of human participants. All referenced findings are based on studies conducted and ethically approved by the original investigators.
Ethics oversight	As a literature-based Review Article with no new human or animal research, this work did not require institutional ethics approval. All studies discussed were conducted under the ethical oversight reported in the respective original publications.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This manuscript is a Review Article and does not involve the generation, analysis, or collection of new experimental or clinical data; therefore, no sample sizes were determined.
Data exclusions	No new data were collected or analysed, and no data were excluded.
Replication	No experiments, computational analyses, or replication studies were performed as part of this Review.
Randomization	No allocation or assignment of participants, samples, or conditions was carried out, as the article synthesizes previously published literature only.
Blinding	No experimental blinding procedures were required, as no new data, samples, or human/animal subjects were involved.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Clinical data
☒ ☐	Dual use research of concern
☒ ☐	Plants

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.