


Perspective

Open Access



Electronic skins with multimodal sensing and perception

Jiaqi Tu^{1,2}, Ming Wang^{2,5}, Wenlong Li³, Jiangtao Su², Yanzen Li², Zhisheng Lv², Haicheng Li^{1,2}, Xue Feng^{6,7,*}, Xiaodong Chen^{2,3,4,*} 

¹Institute of Flexible Electronics Technology of THU, Jiaxing 314000, Zhejiang, China.

²Innovative Center for Flexible Devices (iFLEX), Max Planck - NTU Joint Lab for Artificial Senses, School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore.

³Institute of Materials Research and Engineering, Agency for Science, Technology and Research (A*STAR), Singapore 138634, Singapore.

⁴Institute for Digital Molecular Analytics and Science (IDMxS), Nanyang Technological University, Singapore 639798, Singapore.

⁵Frontier Institute of Chip and System, Fudan University, Shanghai 200433, China.

⁶AML, Department of Engineering Mechanics, Tsinghua University, Beijing 100084, China.

⁷Laboratory of Flexible Electronics Technology, Tsinghua University, Beijing 100084, China.

***Correspondence to:** Prof. Xiaodong Chen, Innovative Center for Flexible Devices (iFLEX), Max Planck - NTU Joint Lab for Artificial Senses, School of Materials Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore. E-mail: chenxd@ntu.edu.sg; Prof. Xue Feng, Laboratory of Flexible Electronics Technology, Tsinghua University, Beijing 100084, China. E-mail: fengxue@tsinghua.edu.cn

How to cite this article: Tu J, Wang M, Li W, Su J, Li Y, Lv Z, Li H, Feng X, Chen X. Electronic skins with multimodal sensing and perception. *Soft Sci* 2023;3:25. <https://dx.doi.org/10.20517/ss.2023.15>

Received: 24 Mar 2023 **First Decision:** 25 Apr 2023 **Revised:** 7 Jun 2023 **Accepted:** 13 Jun 2023 **Published:** 11 Jul 2023

Academic Editors: Dae-Hyeong Kim, Zhifeng Ren **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Multiple types of sensory information are detected and integrated to improve perceptual accuracy and sensitivity in biological cognition. However, current studies on electronic skin (e-skin) systems have mainly focused on the optimization of the modality-specific data acquisition and processing. Endowing e-skins with the abilities of multimodal sensing and even perception that can achieve high-level perception behaviors has been insufficiently explored. Moreover, the perception progress of multisensory e-skin systems is faced with challenges at both device and software levels. Here, we provide a perspective on the multisensory fusion of e-skins. The recent progress in e-skins realizing multimodal sensing is reviewed, followed by bottom-up and top-down multimodal perception. With the deepening understanding of neuroscience and the rapid advance of novel algorithms and devices, multimodal perception function becomes possible and will promote the development of highly intelligent e-skin systems.

Keywords: Electronic skins, multimodal sensing, perception fusion



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



INTRODUCTION

Humans and animals are all immersed in a physical environment filled with dynamic and complex sensory cues, such as tactile, visual, auditory, gustatory, and olfactory. These cues are captured and encoded by distinct sensory receptors, each of which is specialized for a specific type of cue and then sent to nervous systems for processing to form senses^[1-3]. In principle, each cue can provide an individual estimate of the same event. However, multiple sensory cues are necessary for high-level perceptual events, such as thinking, planning, and problem-solving, which are integrated and regulated in cortical networks. Multisensory integration will decrease perceptual ambiguity, enabling more accurate detection of events, but it can also improve perceptual sensitivity with the aim of reacting to even slight changes in environments^[4-8].

Skin is the largest sensory organ in the human body, which is responsible for detecting various stimuli. Wearable electronic skin (e-skin) devices are developed to mimic and even go beyond human skin. These devices detect and distinguish different external stimuli tuning them into accessible signals for processing and recognition. The great functionalities and soft mechanical and physical properties altogether provide tremendous application potential for e-skins in fields of healthcare monitoring, human-machine interfaces (HMIs), and sensory skins for robotics^[9-14]. Current e-skin devices mainly emphasize the acquisition and processing of the unimodal sensory cue, involving a myriad of sensors based on nanomaterials/micro-nano structures. These sensors are designed to detect and measure strain, pressure, temperature, and optical and electrophysiological signals^[15-24]. The main concerns are about improving the physical properties of the specific sensor and developing new fabrication methods and signal processing techniques. Although unimodal sensing has been well developed over the past few years, single-functional e-skin systems are insufficient for complex tasks and practical applications, such as robotic hands for detection of spatial distributions of signals and object recognition^[25-27]. Unlike unimodal sensing, multimodal sensing aims to endow e-skins with the same sensing modalities as human skins or even more. Integrating sensors from different modalities, such as physical, electrophysiological, and chemical sensors, forms a multi-parameter sensing network for comprehensive stimuli sensing from the surroundings. Obstacles still exist when trying to simultaneously detect multimodal signals, including the difficulty of differentiating multicomplex signals and the interference between sensing components and the mechanical disturbance. Thus, novel materials and structure designs are urgently needed to overcome these ongoing problems for reliable and accurate measurement.

In addition to multimodal sensing, work on e-skin systems has been undertaken with the aim of realizing multimodal perception. It has been indicated that high-level perceptual behaviors are attributed to the crossmodal synthesis of multimodal sensory information from the aspect of neuroscience^[15,28-31]. The multimodal perception of e-skins takes inspiration from the multisensory integration mechanism of cortical networks, emphasizing the fusion of sensory cues through hardware or algorithms [Figure 1]. Compared to multimodal sensing, studies on multimodal perception are much more limited due to inevitable challenges at both device and software levels. As machine learning is suitable when managing tasks with multi-parameter inputs and without explicit mathematical models, current e-skin systems implement multimodal perception mainly through software-level methods^[27,32,33]. Still, software-level multimodal perception faces difficulties in fusing multimodal signals due to the incompatibility between datasets, including combining the datasets from heterogeneous modalities and dealing with missing data or different levels of noise^[34]. Besides, a sea of raw data collected by sensor networks have to be transmitted to computation units or cloud-based systems, which will bring problems in terms of energy consumption, response time, data storage, and communication bandwidth^[35]. To solve these significant problems, device-level multimodal perception occurs through near-/in-sensor computing, where the data computing is acted close to or even within the sensory units^[36]. However, it requires more advanced computing devices that are suitable for

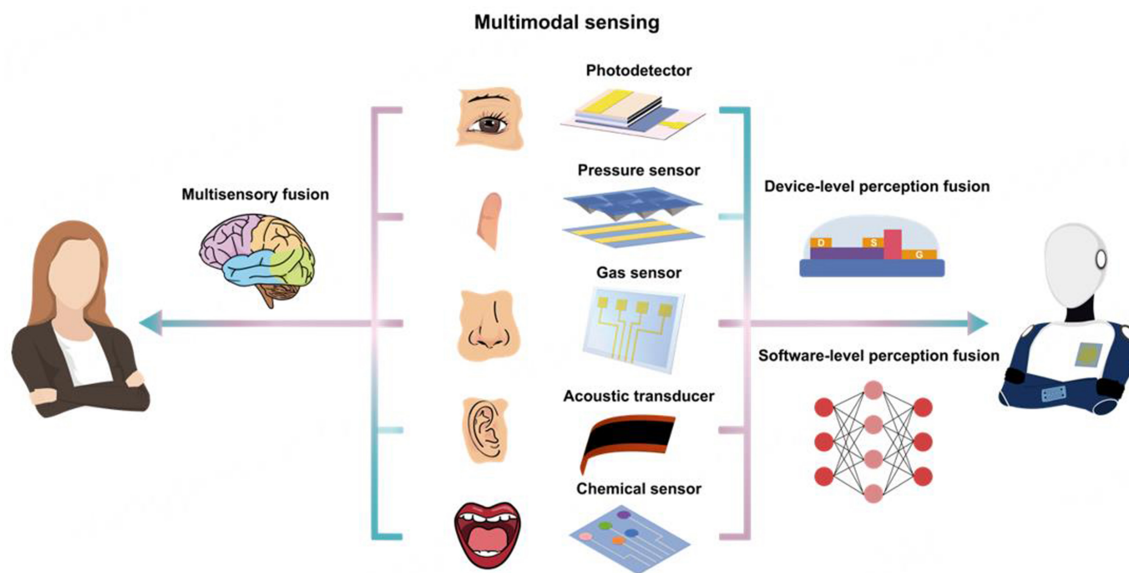


Figure 1. Schematic showing multiple sensing modalities contributing to perception and cognition, indicating the pursuit of e-skin systems toward the next generation.

machine learning algorithms for device-level multimodal perception. Moreover, integrating sensing and computing parts in a planar configuration may reduce the available space for detection of the surrounding physical environment and thus cause disturbance to signals. Novel three-dimensional stacking designs are needed for high communication bandwidth and low latency^[35]. With the deepening understanding of neuroscience and the rapid advance of algorithms and devices, endowing artificial skin with the ability of multimodal perception becomes possible. Therefore, it is necessary to review the progress in this burgeoning field of e-skins at the appropriate time.

This perspective attempts to unfold the recent landscape of e-skins with multimodal sensory fusion and some intriguing future trends for their development. In the first place, we briefly introduce the neurological mechanism of multisensory integration that happens in cerebral cortical networks so as to provide a theoretical basis for the fusion of multimodal sensors in e-skins fields. Burgeoning multifunctional wearable e-skin systems are summarized and categorized into three main subfields: (i) multimodal physical sensor systems; (ii) multimodal physical and electrophysiological sensor systems; and (iii) multimodal physical and chemical sensor systems. Self-decoupling materials and novel mechanisms suppressing the signal interference of multiple sensing modalities are discussed. Then, we discuss some state-of-the-art research on e-skin systems that use bottom-up and top-down approaches to fuse multisensory information. Future trends for e-skin systems with multimodal sensing and perceptual fusion will be explored in the end.

NEUROLOGICAL BASIS OF MULTISENSORY INTEGRATION

Receptors distributed throughout the body could detect and encode multimodal signals in terms of somatosensation (thermoreceptors, touch receptors, and nociceptors), vision (retina), audition (cochlea), olfaction (odorant receptors), and gustatory sensing (taste buds)^[1,3,37,38]. Through afferent pathways, those encoded spike trains from multiple modalities are transmitted into the central nervous system, where the integration of multimodal information takes place^[39,40]. As for multisensory perception fusion in cerebral cortices, bottom-up and top-down multisensory processing are two commonly discussed mechanisms. The bottom-up processing of multisensory stimuli can be further described as three main procedures:

(1) Multimodal inputs first enter sensory cortices networks (including visual, auditory and somatosensory, gustatory, and olfactory cortices) in a modality selective manner. (2) Then, the sensory cortices innervate each other and accept information from other modalities instead of only receiving inputs from their corresponding modalities. These crossmodal projections promote sensory integration processes at an early stage. In this part, temporal coherence of the multisensory stimuli helps to ensure the selection of relevant information and bind different modalities for more robust perception^[15,28,41-43]. (3) Finally, output projections from the sensory cortices reach higher association cortices for further multisensory integration^[44,45].

Besides, the bottom-up information in the sensory cortex can be modulated by the top-down signals representing the internal states of the brain, conveying internal goals or states of the observer^[46,47]. Recent research has shown that task engagement and attention to relevant sensory information can also enhance sensory responses in lower sensory cortices. This response enhancement can be mediated by direct projections from the higher-level cortex areas to the sensory cortices^[48-50].

In summary, multisensory perception fusion begins with modality-specific processing of unisensory inputs with distinctive bottom-up characteristics and is integrated into the higher-order association cortex. On the other hand, the multisensory integration in the brains can be modulated by top-down attention. The multisensory processing of the mammalian brain is dynamically mediated, resulting in a unique and subjective experience of perception. Based on the knowledge of the crossmodal mechanism from the aspect of neuroscience, some research on multimodal sensing and perception is carried out, which is reviewed as follows.

MULTIMODAL E-SKIN SENSING SYSTEMS

Previous to the perceptual fusion of e-skins attracting much research attention, considerable efforts have been put into multimodal sensing integrated wearable systems, holding great potential for applications in the fields of health monitoring, intelligent prosthetics, HMIs, and humanoid robotics. The biological information acquired from wearable skin sensors is generally categorized into three main types: physical, electrophysiological, and chemical signals^[51]. The multimodal sensing e-skin systems can be thus classified into three modes: (1) Integration of multiple physical sensors; (2) Integration of physical and electrophysiological sensors; (3) Integration of physical and chemical sensors [Figure 2]. Most of the multimodal sensing e-skin systems are designed to mimic the functions of human skin, which employ physical sensors to detect a variety of physical signals, including normal force, lateral strain, vibration, temperature, and humidity. Not being limited to that, enduing e-skins with sensing modalities beyond human skin is in great need. To further achieve the next generation of “smart skins”, chemical sensors, electrophysiological sensors, and some physical sensors, such as ultraviolet (UV) light sensors, are integrated into wearable multifunctional e-skin systems^[57,59-61]. Recent works on multimodal sensing of e-skins are reviewed as follows [Table 1].

The most commonly used approach to multimodal sensing systems is integrating different in-plane or out-of-plane sensing units^[65]. In terms of integrating physical sensors, Ho *et al.* presented a transparent multimodal e-skin sensor matrix in which only graphene and its derivatives were used as functional materials^[54]. Through a low-cost and facile lamination process, humidity, thermal, and pressure sensors worked simultaneously and provided output corresponding to the specific external stimulus with negligible responses to other stimuli. A deformable multimodal ionic receptor was recently presented, which is the first e-skin system that managed to differentiate strain and temperature information in a single unit^[53]. The intrinsic features of ion relaxation dynamics (relaxation time and normalized capacitance) were utilized to decouple these two signals and enable monitoring simultaneously without signal interference. Based on

Table 1. Recent progress in multimodal sensing integration for electronic skins

Category	Modes of integration	Sensing materials	Application	Ref.
Physical sensors	Strain+ temperature	Ag NWs/SEBS-EMIM TFSI-Ag NWs/SEBS (strain sensing and temperature sensing)	Tactile motion detection	[53]
	Pressure+ temperature+ humidity	CVD-Graphene (pressure sensing) rGO (temperature sensing) GO (humidity sensing)	Healthcare monitoring, Tactile detection	[54]
	Pressure+ temperature	Vertical array of Te NWs (thermal sensing, pressure sensing)	Object recognition for VR application	[62]
	Strain+ pressure+ proximity+ temperature+ humidity+ UV light+ magnetic field	Pt thin film (temperature sensing) Constantan alloy (strain sensing) PI (humidity sensing) Ag-ZnO-Ag thin films (light sensing) Co/Cu multilayers (magnetic field sensing) Ag-Ecoflex-Ag (Pressure+ proximity sensing)	Intelligent Prosthetics, Human-machine interface, healthcare monitoring	[52]
Physical+ Electrophysiological sensors	Pressure+ MMG+ ECG+EMG	PANi-PVC ionic gel (Blood pressure/ECG/EMG sensing) PVDF-TrFe gel (MMG sensing)	Healthcare monitoring	[63]
	Temperature+ strain+ UV light+ ECG	Ag NPs/SWCNT inks (strain sensing) CNT inks/PEDOT:PSS (temperature sensing) ZnO NW networks (UV light sensing) Ag (ECG sensing)	Healthcare monitoring, Physical activity detection	[55]
	Strain+ hydration+ ECG	Ag NWs/PDMS (hydration sensing) Ag NWs/Dragon skin (strain sensing)	Healthcare monitoring	[64]
	Pressure+ temperature+ humidity +chemical variables	CNT microyarns (pressure sensing, temperature sensing, humidity sensing, chemical variables sensing)	Healthcare monitoring, Humanoid robotic skins	[57]
Physical+ Chemical sensors		Cr/Au metal microwires (temperature sensing)		[58]
	Temperature+ electrolytes+ metabolites	lactate oxidase/chitosan/CNT/Prussian blue/Au electrode (glucose and lactate sensing) Na ionophore X/Na-TFPB/PVC/DOS (Na ⁺ sensing) Valinomycin/NaTPB/PVC/DOS (K ⁺ sensing)	Healthcare monitoring	

CNT: Carbon nanotube; UV: ultraviolet.

that, a multimodal ionic-e-skin was further designed to provide force directions and strain profiles of different tactile motions. Inspired by the structure of human skin (including the epidermis, dermis, and hypodermis), a triple-parameter sensor was produced through an inexpensive and facile method^[66]. Ionic liquid and particular circuit topologies ensure robust stability against mechanical disturbance during real-time sensing tests. Pressure, temperature, and light sensors are integrated into a layer-by-layer structure exhibiting no signal interference, which holds great promise for healthcare monitoring and robotic skins. Hua *et al.* demonstrated a versatile, stretchable, and conformable multilayered matrix network^[52]. Expandable meandering wires have effectively lowered the electrical disturbance induced by mechanical motion. By sensing surrounding strain, pressure, proximity, temperature, humidity, UV light, and magnetic field simultaneously, the capabilities of e-skins are empowered to augment the sensation of humans. In addition, a great deal of e-skin research has focused on merging physical sensors with chemical or electrophysiological sensors, which adds complementary parameters for tracking health conditions and physical activities more effectively. For simultaneously monitoring physical and electrophysiological signals,

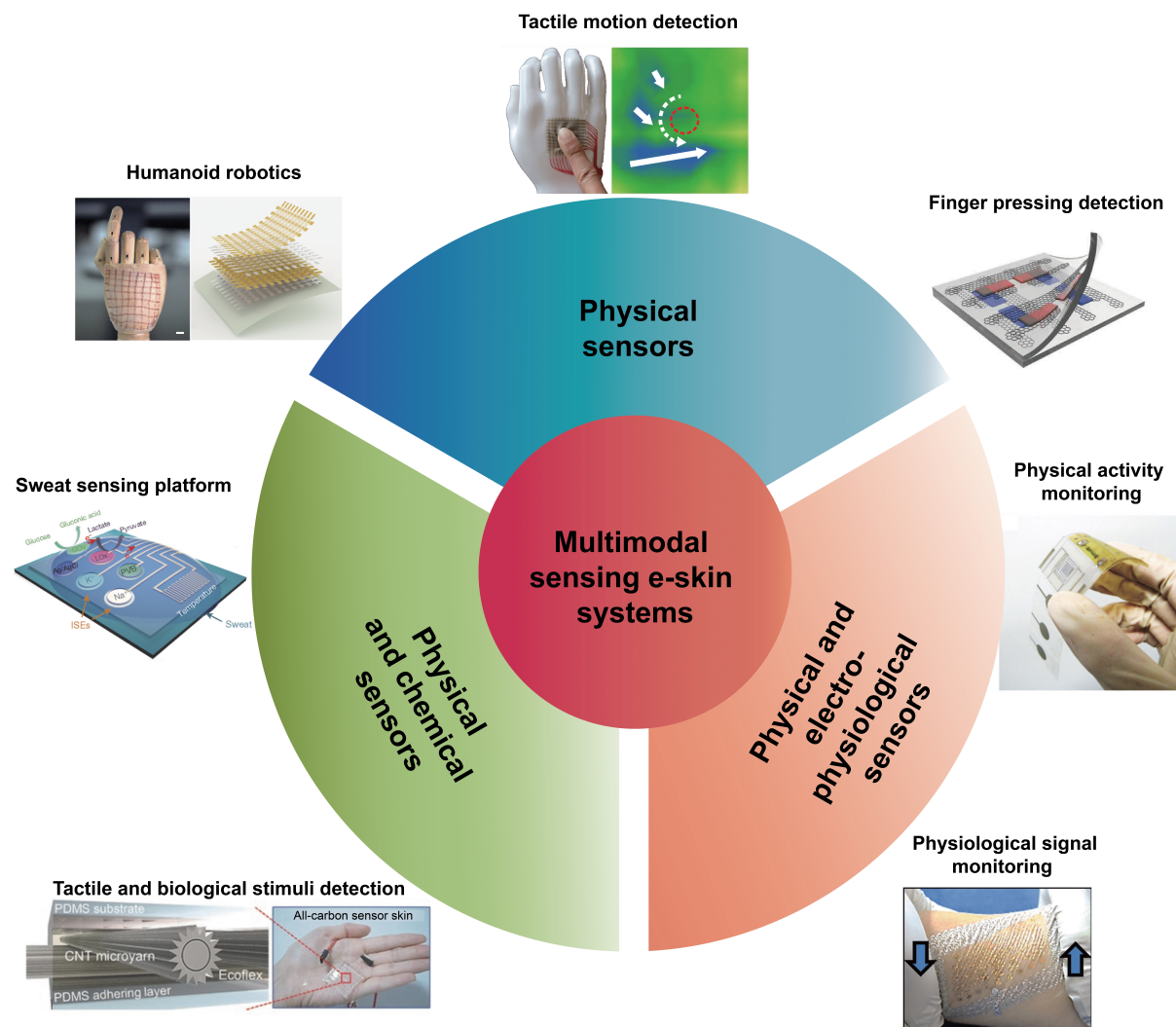


Figure 2. Current state-of-art e-skin systems with multimodal sensing, including (1) Integration of multiple physical sensors; (2) Integration of physical and physiological sensors; and (3) Integration of physical and chemical sensors. Reproduced with permission^[52]. Copyright©2018, Nature Publishing Group. Reproduced with permission^[53]. Copyright©2020, American Association for the Advancement of Science. Reproduced with permission^[54]. Copyright©2016, Wiley-VCH. Reproduced with permission^[55]. Copyright©2016, American Association for the Advancement of Science. Reproduced with permission^[56]. Copyright©2015, Wiley-VCH. Reproduced with permission^[57]. Copyright©2015, Wiley-VCH. Reproduced with permission^[58]. Copyright©2018, Nature Publishing Group. e-skin: Electronic skin.

Yamamoto *et al.* introduced a flexible multifunctional printed sensor system for real-time healthcare monitoring^[55]. This innovative system was equipped with disposable sensing elements, containing a three-axis acceleration sensor, a temperature sensor, a UV light sensor, and an electrocardiogram (ECG) sensor. With this multimodal device attached to the chest of an individual directly, four types of sensory outputs were simultaneously recorded, ensuring comprehensive healthcare tracking ability. Kim *et al.* presented a multifunctional carbon-based e-skin that could monitor touch, temperature, and humidity^[57]. To implement the artificial skin beyond basic human skin capabilities, the e-skin system is also designed to distinguish biological variables with different dipole moments, implying new applications such as electronic nose or tongue. More intriguingly, a fully integrated multimodal sensory platform was developed by Gao *et al.*, realizing in situ and wireless sweat analysis^[58]. This flexible system can simultaneously evaluate sweat metabolites and electrolytes and the skin temperature for calibration of outputs, showing excellent

selectivity when sensing various sweat biomarkers. Through the integration of multiple sensors, e-skins can thus afford the ability to interact with the environment and obtain vital health evaluation indices from human bodies in the meantime.

While massive works have been reported to detect multimodal physical signals simultaneously from skins, one of the challenges is the signal interference between sensing components. The electrical output signals of the flexible electronic device may show motion artifacts due to deformations such as stretching, compressing, and bending^[65]. In the meantime, some multimodal sensing systems include physical signals and thus require decoupling methods to differentiate deformation modes. Self-decoupling materials and novel structural designs are adopted to differentiate multicomplex signals for accurate and reliable measurement.

Self-decoupling materials can intrinsically suppress signal interference through novel sensing mechanisms. Ionic-based materials are suitable for self-decoupling sensing systems with frequency-dependent ion relaxation dynamics^[53,67]. For example, You *et al.* proposed a new artificial receptor that can differentiate thermal and mechanical information without signal interference^[53]. The bulk resistance (R) and capacitance (C) show different behaviors under different frequencies. The charge relaxation frequency (τ^{-1}) does not change with stretching [Figure 3A, ii]. Meanwhile, the normalization of capacitance at the measured temperature can remove the effect of temperature. Thus, the systems can provide complete temperature and force sensing through a self-decoupling ionic conductor. Further, the receptor provides real-time force directions and strain profiles in various tactile motions. In addition, magnetic mechanisms can also be used for force self-decoupling. The force directions can be differentiated by detecting the change of magnetic flux densities. Yan *et al.* introduce a soft tactile sensor that possesses self-decoupling and super-resolution capabilities by utilizing a sinusoidally magnetized flexible film^[68]. In detail, the embedded Hall sensor located at the middle layer can sense deformation, whether it is from the normal or shear direction [Figure 3B, ii]. The normal force and shear force can be decoupled by calculating two different parameters, which are the magnetic field rotation angle and the translational movement of the magnetic field. Subsequently, the sensor converts this deformation into electric signals through the use of a printed circuit board (PCB). Different mechanisms of the same materials are also combined to differentiate multimodal signals. Ferroelectric materials can be candidates for multimodal systems with their triboelectric and pyroelectric effects. Shin *et al.* developed a self-powered multimodal sensor based on^[69]. Based on an interlocked ferroelectric copolymer microstructure, this sensor enables simultaneous detection of mechanical and thermal stimuli without a spacer in a single device, overcoming the drawbacks of conventional sensors. The temperature and pressure are detected through the pyroelectric and triboelectric mechanisms, respectively. The response and relaxation times of the triboelectric and pyroelectric effects are different, as shown in the output signals [Figure 3C, ii]. Herein, this multimodal tactile sensor can intrinsically decouple pressure and temperature information by analyzing the multiple signals based on the response and relaxation times. The above-mentioned self-decoupling mechanisms can be integrated to further develop the design of multimodal sensing systems. For example, Zhang *et al.* proposed a multilayer structure that includes an ionic hydrogel film, a wrinkle-patterned polydimethylsiloxane (PDMS) film, and a carbon nanotube (CNT)/PDMS electrode with self-decoupled pressure, strain, and temperature sensing abilities^[70]. The temperature was decoupled through an ionic hydrogel with an aligned polymer chain structure, which processed an ultrahigh temperature sensitivity in a wide range from 0 °C to 50 °C. In the meantime, it shows surprisingly low strain sensitivity and intrinsic pressure-insensitive properties. The mechanochromic core-shell magnetic nanoparticles with a photonic crystal structure were fast responsive to external strain via interactive color switching. Further, a triboelectric structure comprising a wrinkle-patterned PDMS friction layer with gradient modulus and a CNT-based elastic electrode detected voltage

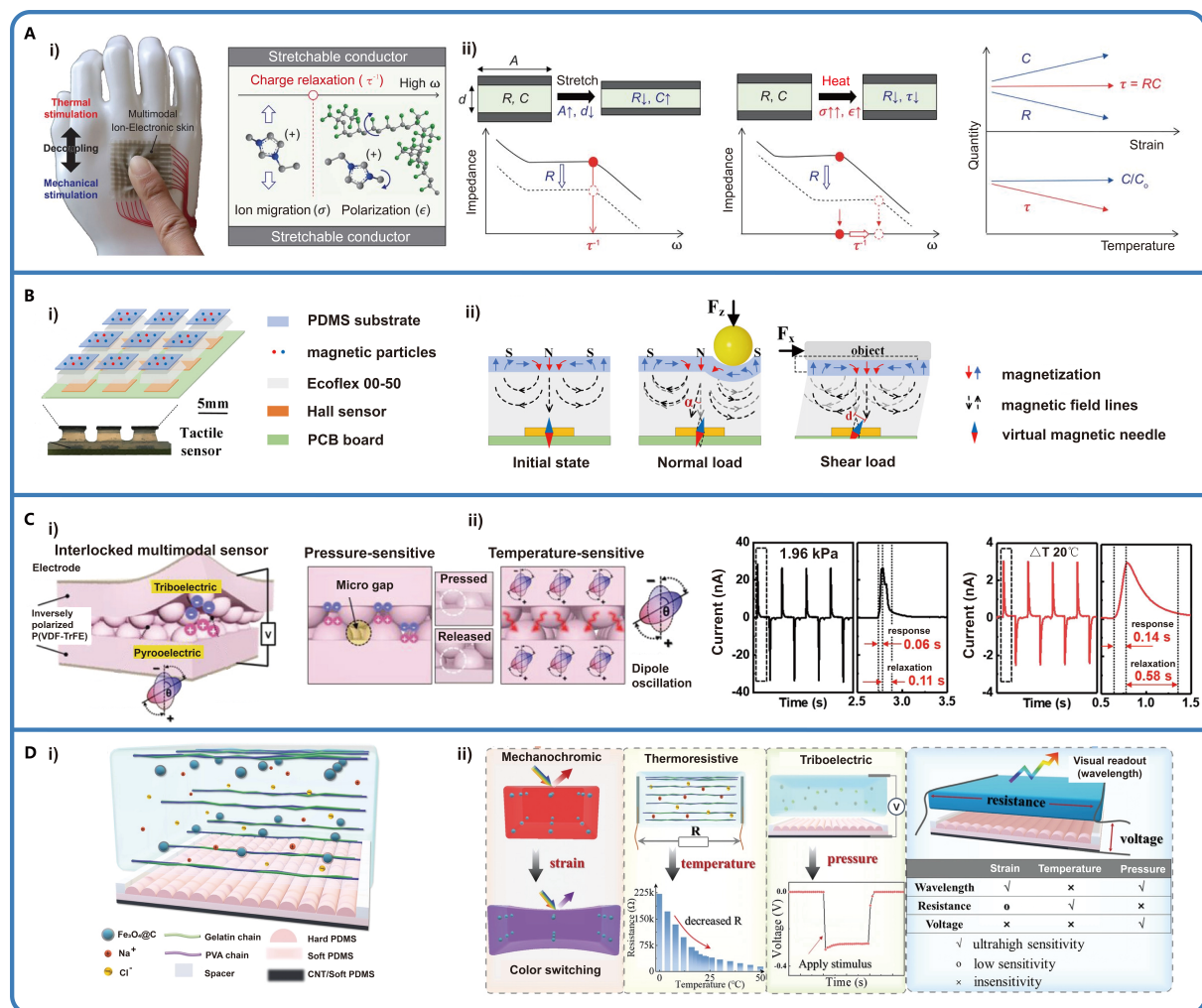


Figure 3. Multimodal sensing systems with self-decoupling mechanisms. (A) Ionic conductor-based multimodal receptors that can intrinsically differentiate strain and temperature. Reproduced with permission^[53]. Copyright©2020, Nature Publishing Group; (B) Artificial skins can decouple the normal or shear force direction with embedded Hall sensors. Reproduced with permission^[68]. Copyright©2021, American Association for the Advancement of Science; (C) A skin-inspired multimodal sensing system and its decoupling mechanism for bimodal signals in a single unit with triboelectric and pyroelectric effects^[69]. (D) A chromotropic ionic skin can differentiate the temperature, pressure, and strain by integrating multiple sensing mechanisms. Reproduced with permission^[70]. Copyright©2022, Wiley-VCH. e-skin: Electronic skin; PCB: printed circuit board; PDMS: polydimethylsiloxane.

output for strain-unperturbed and temperature-insensitive pressure sensing [Figure 3D].

RECENT LANDSCAPE OF E-SKINS WITH MULTIMODAL PERCEPTION FUSION

Bottom-up multimodal perception fusion

Recent progress in processing multimodal e-skin information mainly uses the bottom-up modulation approach, which can be further categorized into two modes: fusion at the device level and fusion at the software level [Figure 4]. In the former, multisensory perception fusion is realized by utilizing innovative hardware, where crossmodal signals are integrated close to or even within sensing devices before being transmitted to the exterior software, mimicking the procedure of multisensory fusion in primary sensory cortices. The latter multisensory fusion strategy is accomplished by using mathematical algorithms corresponding to the view of the multisensory processing in different cortical network levels.

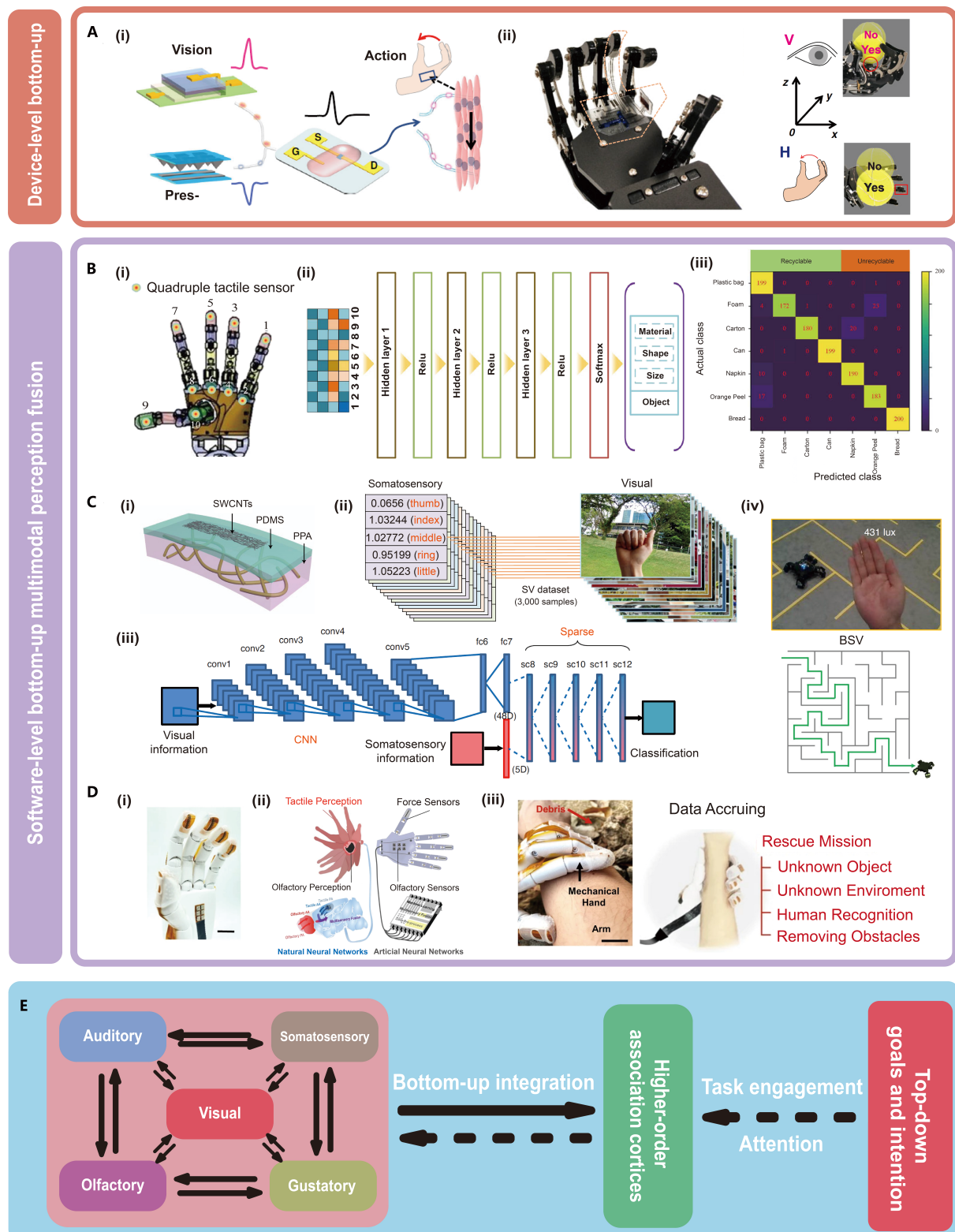


Figure 4. Recent progress in bottom-up multimodal perception fusion of e-skin systems and schematic diagram of multisensory fusion. (A) Multimodal perception fusion at the device level. Reproduced with permission^[71]. Copyright©2020, Nature Publishing Group; (B-D) Multimodal perception fusion at the software level. Reproduced with permission^[27]. Copyright©2020, American Association for the Advancement of Science. Reproduced with permission^[33]. Copyright©2020, Nature Publishing Group. Reproduced with permission^[72].

Copyright©2022, Nature Publishing Group; (E) Schematic diagram of bottom-up and top-down multisensory fusion.

Emerging neuromorphic computing devices hold great potential for bottom-up multisensory fusion at the device level. A bimodal artificial sensory neuron was developed to achieve the sensory fusion processes^[71] [Figure 4A]. Pressure sensors and photodetectors are integrated to transform tactile and visual stimuli into electrical signals. Then the combined signals are transmitted via an ion cable to the synaptic transistor, where they are integrated to produce an excitatory postsynaptic current. As a result, the somatosensory and visual information are fused at the device level, achieving multimodal perception integration after further data processing. In a multi-transparency pattern recognition task, robust recognition confirms potential application in neurorobotics and artificial intelligence, even with smaller datasets. However, the issue remains that the visual-haptic fusion matrix was just implemented as feature extraction layers of artificial neural networks (ANNs). In other words, the device part alone cannot realize multimodal perception tasks without additional algorithms.

For software-level perception fusion, various machine learning algorithms, such as k-nearest-neighbor classifiers^[73], supporting vector machines (SVMs)^[73,74], and convolutional neural networks (CNNs)^[75,76], are common strategies for data fusion. Among innovative e-skin systems, these advanced algorithms are implemented to achieve multimodal perception. Li *et al.* integrated flexible quadruple tactile sensors onto a robot hand to realize precise object identification [Figure 4B]. This novel skin-inspired quadruple tactile sensor was constructed in a multilayer architecture, which enables the perceiving of the grasping pressure, environment temperature, and temperature and thermal conductivity of objects with no interference. To realize accurate object recognition, the multimodal sensory information collected through this smart hand was fused as a 4×10 signal map at the dataset level. After being trained using multilayer perception networks (also known as ANNs), the smart robotic hand achieves a classification accuracy of 94% in a garbage sorting task^[27]. Feature-based cognition fusion is also a common strategy, which involves extracting features from multisensory signals and concatenating them into a single feature vector. The feature vector is then fed into pattern recognition algorithms, such as neural networks, clustering algorithms, and template methods^[77]. Wang *et al.* proposed a bio-inspired architecture for data fusion that can recognize human gestures by fusing visual data with somatosensory data from skin-like stretchable strain sensors [Figure 4C]. For early visual processing, the learning architecture uses a sectional CNN and then implements a sparse neural network for sensor data fusion and feature-level recognition, resembling the somatosensory-visual (SV) fusion hierarchy in the higher association cortices of brains. Using stacked soft materials, the sensor section was designed to be highly stretchable, conformable, and adhesive, enabling the sensor to adhere tightly to the knuckle for precise monitoring of finger movement. This bioinspired algorithm can achieve a recognition accuracy of 100% in its own dataset and even maintain high recognition results when texting non-ideal conditions images^[33]. Liu *et al.* reported a tactile-olfactory sensing system [Figure 4D]. The bimodal sensing array was integrated with mechanical hands. Olfactory and tactile data fusion was then achieved through a machine-learning strategy for robust object recognition in rough situations. This artificial bimodal system could classify 11 objects with an accuracy of 96.9% in a simulated fire scenario^[72]. Although more studies should be carried out on perception fusion models and near/in-sensor fusion devices, both types of bottom-up multimodal perception fusion still motivate the next generation of e-skins.

Top-down attention-based multimodal perception fusion

Sensory responses in lower sensory cortices are modulated by attention and task engagement for the efficient perception of relevant sensory stimuli [Figure 4E]. In the scenario of multimodal stimuli competing for processing resources, the saliency for individual stimuli in the potentially preferred modality may

remain at a low level and thus affect accurate perception and cognition^[46]. To solve this, an attention-based mechanism engages in and conditionally selects a salient modality between different signals. Although it is still blank in the field of e-skins about the top-down multisensory fusion mechanism, some research on the attention-based fusion mechanism provides future e-skin systems with algorithm models for reference. There have been many attention-based fusion models being constructed in other fields, such as video descriptions^[78,79], event detection^[80,81], and speech recognition^[79,82]. For example, Zhou *et al.* presented a robust attention-based dual-modal speech recognition system. In virtue of the multi-modality attention-based method, the system can strike a balance between visual and audio information by fusing representations of them based on their importance. In addition, the attention of different modalities can be mediated over time by modeling temporal variability for each modality using a long short-term memory^[82]. Considering further exploration in neuroscience and developing advanced algorithm models, a top-down attention-based fusion technique can push forward the progress of smart skins.

CONCLUSIONS AND FUTURE PERSPECTIVE

Collectively, we overviewed the recent works in the intriguing field of e-skins with multimodal sensing and perception fusion. Although considerable progress in multimodal sensing integration has been made over the last few years, challenges remain and need to be addressed. As a fast-growing research interest, multimodal perception fusion deserves much deeper investigation. To realize the next generation of e-skins, more attention should be paid to the following aspects:

(i) **Decoupled sensing modalities without signal interference.** It is worthy of in-depth research to endow e-skins with sensing abilities, which are the same as or even beyond the basic functions of human skin. In order to achieve higher-level perception, integrating other sensing parts, such as chemical, sound, and light sensors, with the existing e-skin systems enables more accurate detection of events. Nevertheless, the same sensor can respond to different stimuli where interference comes along. This will affect the accuracy of the signal outputs for each sensing mode. Signal processing can sometimes minimize the effect of interference but also come along with processing complexity. Therefore, multimodal sensing systems with self-decoupled mechanisms are desired for the superiority of simplified data processing and higher accuracy of signals with less interference. Self-decoupling materials can remove signal interference intrinsically through novel sensing mechanisms. Ionic-based materials are suitable for self-decoupling sensing systems with frequency-dependent ion relaxation dynamics. An ionic-based conductor differentiates thermal and mechanical information without signal interference. In addition, ferroelectric materials can be candidates for multimodal systems with their triboelectric and pyroelectric effects. The different response and relaxation times of the triboelectric and the pyroelectric effect can decouple the pressure and thermal signals. With the superiority of direction differentiation, magnetic mechanisms can also be used for force self-decoupling. Strain and pressure can be distinguished by detecting the change of magnetic flux densities. With the advantage of these novel self-decoupling materials, the next-generation multimodal sensing systems will fulfill practical demands for healthcare, HMIs, and robotics. Eliminating interferences caused by external stimuli is a significant challenge for next-generation e-skins and demands more effort in finding novel materials and integrating multiple sensing mechanisms.

(ii) **High-density, high-fidelity, and large-area integration.** A highly integrated e-skin system with multimodal sensing abilities will provide device-level foundations for further research on multimodal perception fusion and surely contributes to a wider range of applications in smart healthcare, soft robotics, and HMIs. However, highly integrated e-skin systems with various sensors, electrical interconnectors, and signal processing units are faced with great challenges. Growing density of and decreasing spaces between interconnect lines and the lower signal intensity caused by the miniaturization of sensors induce signal

interference (crosstalk). Large-area fabrication and integration on irregular three-dimensional surfaces also bring huge difficulties in sensor resolution, layer-to-layer registration, and large-area uniformity. In addition, a high level of integration inevitably results in a short distance between sensors and processing units. The signal-to-noise ratio is thus affected by the smaller spaces for sensing external stimuli. Novel electrode materials, device architecture, and large-area fabrication techniques are required to solve these problems.

(iii) **Wireless communication.** In e-skin systems, wireless communication deserves more research attention. This technique can get rid of additional wiring in order to alleviate the spatial limits and disturbance. So far, most wireless e-skin systems have been based on conventional wireless techniques, such as Bluetooth and near-field communication. The need for flexibility and stretchability gives rise to electromagnetic coupling, where signals are transmitted between internal and external coils. However, signal interference caused by other working electronics, the permittivity of the surrounding environment, and motions restricts the wide application of electromagnetic coupling in e-skin systems. These issues should be fully addressed, and novel wireless communication techniques are in need to construct wireless e-skin systems with the growing demand for Internet-of-Things.

(iv) **Optimum of bottom-up multimodal algorithms.** More effort remains to be put into e-skins based on bottom-up multimodal perceptual fusion. Near-/in-sensor computing requires more advanced neuromorphic computing devices. These devices are suitable for neural network algorithms to realize device-level multimodal perception. Thus, problems, such as power efficiency and fault tolerance, can be suppressed when the size of the data is highly increased. Apart from algorithms, including SVMs, clustering methods, CNNs, and ANNs, the software-level bottom-up multimodal perception requires more advanced algorithms models and architectures to overcome the challenges of fusing the datasets from heterogeneous modalities and dealing with missing data or different levels of noise.

(v) **Development of top-down multimodal algorithms.** Top-down selective attention is necessary to function for more efficient multisensory integration processes in this situation. However, the area of top-down attention-based multimodal perception fusion of e-skins is still blank but highly worthy of being discovered. As growing attention-based fusion research on other areas, such as speech recognition and video captioning, keeps arousing, there will be a better chance for e-skin systems to accomplish brain-like perception and cognition.

DECLARATIONS

Acknowledgments

The authors gratefully acknowledge the support from the National Natural Science Foundation of China (Grant Nos. U20A6001 and 11921002).

Authors' contributions

Conceptualization: Tu J, Wang M

Methodology: Tu J, Wang M

Writing - Original Draft: Tu J, Wang M

Writing - Review & Editing: Li W, Su J, Li Y, Lv Z, Li H, Feng X, Chen X

Supervision: Feng X, Chen X

Availability of data and materials

Not applicable.

Financial support and sponsorship

Tu J acknowledges the research scholarship awarded by the Institute of Flexible Electronics Technology of Tsinghua, Zhejiang (IFET-THU), Nanyang Technological University (NTU), and Qiantang Science and Technology Innovation Center, China (QSTIC).

Conflicts of interest

The authors declare no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2023.

REFERENCES

1. Lumpkin EA, Caterina MJ. Mechanisms of sensory transduction in the skin. *Nature* 2007;445:858-65. [DOI](#) [PubMed](#)
2. Ohyama T, Schneider-Mizell CM, Fetter RD, et al. A multilevel multimodal circuit enhances action selection in drosophila. *Nature* 2015;520:633-9. [DOI](#) [PubMed](#)
3. Tan H, Zhou Y, Tao Q, Rosen J, van Dijken S. Bioinspired multisensory neural network with crossmodal integration and recognition. *Nat Commun* 2021;12:1120. [DOI](#) [PubMed](#) [PMC](#)
4. Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 2002;415:429-33. [DOI](#) [PubMed](#)
5. Macaluso E, Driver J. Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci* 2005;28:264-71. [DOI](#) [PubMed](#)
6. Green AM, Angelaki DE. Multisensory integration: resolving sensory ambiguities to build novel representations. *Curr Opin Neurobiol* 2010;20:353-60. [DOI](#) [PubMed](#) [PMC](#)
7. Ohshiro T, Angelaki DE, DeAngelis GC. A normalization model of multisensory integration. *Nat Neurosci* 2011;14:775-82. [DOI](#) [PubMed](#) [PMC](#)
8. Hagmann CE, Russo N. Multisensory integration of redundant trisensory stimulation. *Atten Percept Psychophys* 2016;78:2558-68. [DOI](#) [PubMed](#) [PMC](#)
9. Zhu B, Wang H, Liu Y, et al. Skin-inspired haptic memory arrays with an electrically reconfigurable architecture. *Adv Mater* 2016;28:1559-66. [DOI](#) [PubMed](#)
10. Chen S, Jiang K, Lou Z, Chen D, Shen G. Recent developments in graphene-based tactile sensors and e-skins. *Adv Mater Technol* 2018;3:1700248. [DOI](#)
11. Jeon S, Lim S, Trung TQ, Jung M, Lee N. Flexible multimodal sensors for electronic skin: principle, materials, device, array architecture, and data acquisition method. *Proc IEEE* 2019;107:2065-83. [DOI](#)
12. Xu K, Lu Y, Takei K. Multifunctional skin-inspired flexible sensor systems for wearable electronics. *Adv Mater Technol* 2019;4:1800628. [DOI](#)
13. Li H, Ma Y, Liang Z, et al. Wearable skin-like optoelectronic systems with suppression of motion artifacts for cuff-less continuous blood pressure monitor. *Natl Sci Rev* 2020;7:849-62. [DOI](#) [PubMed](#) [PMC](#)
14. Wu X, Ahmed M, Khan Y, et al. A potentiometric mechanotransduction mechanism for novel electronic skins. *Sci Adv* 2020;6:eaba1062. [DOI](#) [PubMed](#) [PMC](#)
15. Choi I, Lee JY, Lee SH. Bottom-up and top-down modulation of multisensory integration. *Curr Opin Neurobiol* 2018;52:115-22. [DOI](#) [PubMed](#)
16. Li H, Xu Y, Li X, et al. Epidermal inorganic optoelectronics for blood oxygen measurement. *Adv Healthc Mater* 2017;6:1601013. [DOI](#) [PubMed](#)
17. Boutry CM, Negre M, Jorda M, et al. A hierarchically patterned, bioinspired e-skin able to detect the direction of applied pressure for robotics. *Sci Robot* 2018;3:eaau6914. [DOI](#) [PubMed](#)
18. Choi S, Han SI, Jung D, et al. Highly conductive, stretchable and biocompatible Ag-Au core-sheath nanowire composite for wearable and implantable bioelectronics. *Nat Nanotechnol* 2018;13:1048-56. [DOI](#) [PubMed](#)
19. Wang M, Wang W, Leow WR, et al. Enhancing the matrix addressing of flexible sensory arrays by a highly nonlinear threshold switch. *Adv Mater* 2018;30:e1802516. [DOI](#) [PubMed](#)

20. Yang JC, Mun J, Kwon SY, Park S, Bao Z, Park S. Electronic skin: recent progress and future prospects for skin-attachable devices for health monitoring, robotics, and prosthetics. *Adv Mater* 2019;31:e1904765. [DOI](#) [PubMed](#)
21. Lee S, Franklin S, Hassani FA, et al. Nanomesh pressure sensor for monitoring finger manipulation without sensory interference. *Science* 2020;370:966-70. [DOI](#) [PubMed](#)
22. Wang Y, Lee S, Yokota T, et al. A durable nanomesh on-skin strain gauge for natural skin motion monitoring with minimum mechanical constraints. *Sci Adv* 2020;6:eabb7043. [DOI](#) [PubMed](#) [PMC](#)
23. Yang X, Li L, Wang S, et al. Ultrathin, stretchable, and breathable epidermal electronics based on a facile bubble blowing method. *Adv Electron Mater* 2020;6:2000306. [DOI](#)
24. Wang M, Luo Y, Wang T, et al. Artificial skin perception. *Adv Mater* 2021;33:e2003014. [DOI](#)
25. Tien NT, Jeon S, Kim DI, et al. A flexible bimodal sensor array for simultaneous sensing of pressure and temperature. *Adv Mater* 2014;26:796-804. [DOI](#)
26. Won SM, Wang H, Kim BH, et al. Multimodal sensing with a three-dimensional piezoresistive structure. *ACS Nano* 2019;13:10972-9. [DOI](#) [PubMed](#)
27. Li G, Liu S, Wang L, Zhu R. Skin-inspired quadruple tactile sensors integrated on a robot hand enable object recognition. *Sci Robot* 2020;5:eabc8134. [DOI](#) [PubMed](#)
28. Senkowski D, Schneider TR, Foxe JJ, Engel AK. Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 2008;31:401-9. [DOI](#) [PubMed](#)
29. Stein BE, Stanford TR. Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci* 2008;9:255-66. [DOI](#) [PubMed](#)
30. Fetsch CR, DeAngelis GC, Angelaki DE. Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nat Rev Neurosci* 2013;14:429-42. [DOI](#) [PubMed](#) [PMC](#)
31. Wang J, Wang C, Cai P, et al. Artificial sense technology: emulating and extending biological senses. *ACS Nano* 2021;15:18671-8. [DOI](#) [PubMed](#)
32. Li G, Zhu R. A multisensory tactile system for robotic hands to recognize objects. *Adv Mater Technol* 2019;4:1900602. [DOI](#)
33. Wang M, Yan Z, Wang T, et al. Gesture recognition using a bioinspired learning architecture that integrates visual data with somatosensory data from stretchable sensors. *Nat Electron* 2020;3:563-70. [DOI](#)
34. Baltrušaitis T, Ahuja C, Morency L. Challenges and applications in multimodal machine learning. *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition* 2018;2:17-48. [DOI](#)
35. Zhou F, Chai Y. Near-sensor and in-sensor computing. *Nat Electron* 2020;3:664-71. [DOI](#)
36. Wang M, Wang T, Luo Y, et al. Fusing stretchable sensing technology with machine learning for human-machine interfaces. *Adv Funct Mater* 2021;31:2008807. [DOI](#)
37. Carleton A, Accolla R, Simon SA. Coding in the mammalian gustatory system. *Trends Neurosci* 2010;33:326-34. [DOI](#) [PubMed](#) [PMC](#)
38. Svechtarova MI, Buzzacchera I, Toebes BJ, Lauko J, Anton N, Wilson CJ. Sensor devices inspired by the five senses: a review. *Electroanalysis* 2016;28:1201-41. [DOI](#)
39. Keat J, Reinagel P, Reid RC, Meister M. Predicting every spike: a model for the responses of visual neurons. *Neuron* 2001;30:803-17. [DOI](#) [PubMed](#)
40. Bean BP. The action potential in mammalian central neurons. *Nat Rev Neurosci* 2007;8:451-65. [DOI](#) [PubMed](#)
41. Fries P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci* 2005;9:474-80. [DOI](#) [PubMed](#)
42. Womelsdorf T, Schoffelen JM, Oostenveld R, et al. Modulation of neuronal interactions through neuronal synchronization. *Science* 2007;316:1609-12. [DOI](#) [PubMed](#)
43. Atilgan H, Town SM, Wood KC, et al. Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron* 2018;97:640-655.e4. [DOI](#) [PubMed](#) [PMC](#)
44. Beauchamp MS. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr Opin Neurobiol* 2005;15:145-53. [DOI](#) [PubMed](#)
45. Kayser C, Petkov CI, Augath M, Logothetis NK. Integration of touch and sound in auditory cortex. *Neuron* 2005;48:373-84. [DOI](#) [PubMed](#)
46. Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG. The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci* 2010;14:400-10. [DOI](#) [PubMed](#) [PMC](#)
47. Arnal LH, Giraud AL. Cortical oscillations and sensory predictions. *Trends Cogn Sci* 2012;16:390-8. [DOI](#) [PubMed](#)
48. Atiani S, David SV, Elgueda D, et al. Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron* 2014;82:486-99. [DOI](#) [PubMed](#) [PMC](#)
49. Makino H, Komiyama T. Learning enhances the relative impact of top-down processing in the visual cortex. *Nat Neurosci* 2015;18:1116-22. [DOI](#) [PubMed](#) [PMC](#)
50. Manita S, Suzuki T, Homma C, et al. A top-down cortical circuit for accurate sensory perception. *Neuron* 2015;86:1304-16. [DOI](#) [PubMed](#)
51. Someya T, Amagai M. Toward a new generation of smart skins. *Nat Biotechnol* 2019;37:382-8. [DOI](#) [PubMed](#)

52. Hua Q, Sun J, Liu H, et al. Skin-inspired highly stretchable and conformable matrix networks for multifunctional sensing. *Nat Commun* 2018;9:244. DOI PubMed PMC
53. You I, Mackanic DG, Matsuhisa N, et al. Artificial multimodal receptors based on ion relaxation dynamics. *Science* 2020;370:961-5. DOI PubMed
54. Ho DH, Sun Q, Kim SY, Han JT, Kim DH, Cho JH. Stretchable and multimodal all graphene electronic skin. *Adv Mater* 2016;28:2601-8. DOI PubMed
55. Yamamoto Y, Harada S, Yamamoto D, et al. Printed multifunctional flexible device with an integrated motion sensor for health care monitoring. *Sci Adv* 2016;2:e1601473. DOI PubMed PMC
56. Yang S, Chen YC, Nicolini L, et al. "Cut-and-paste" manufacture of multiparametric epidermal sensor systems. *Adv Mater* 2015;27:6423-30. DOI PubMed
57. Kim SY, Park S, Park HW, Park DH, Jeong Y, Kim DH. Highly sensitive and multimodal all-carbon skin sensors capable of simultaneously detecting tactile and biological stimuli. *Adv Mater* 2015;27:4178-85. DOI PubMed
58. Gao W, Emaminejad S, Nyein HYY, et al. Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. *Nature* 2016;529:509-14. DOI PubMed PMC
59. Wang C, Xia K, Zhang M, Jian M, Zhang Y. An all-silk-derived dual-mode e-skin for simultaneous temperature-pressure detection. *ACS Appl Mater Interfaces* 2017;9:39484-92. DOI PubMed
60. Zhao S, Zhu R. Electronic skin with multifunction sensors based on thermosensation. *Adv Mater* 2017;29:1606151. DOI PubMed
61. Yu Y, Nassar J, Xu C, et al. Biofuel-powered soft electronic skin with multiplexed and wireless sensing for human-machine interfaces. *Sci Robot* 2020;5:eaaz7946. DOI PubMed PMC
62. Li L, Zhao S, Ran W, et al. Dual sensing signal decoupling based on tellurium anisotropy for VR interaction and neuro-reflex system application. *Nat Commun* 2022;13:5975. DOI PubMed PMC
63. Chun KY, Seo S, Han CS. A wearable all-gel multimodal cutaneous sensor enabling simultaneous single-site monitoring of cardiac-related biophysical signals. *Adv Mater* 2022;34:e2110082. DOI PubMed
64. Yao S, Myers A, Malhotra A, et al. A wearable hydration sensor with conformal nanowire electrodes. *Adv Healthc Mater* 2017;6:1601159. DOI PubMed
65. Yang R, Zhang W, Tiwari N, Yan H, Li T, Cheng H. Multimodal sensors with decoupled sensing mechanisms. *Adv Sci* 2022;9:e2202470. DOI PubMed PMC
66. Gui Q, He Y, Gao N, Tao X, Wang Y. A skin-inspired integrated sensor for synchronous monitoring of multiparameter signals. *Adv Funct Mater* 2017;27:1702050. DOI
67. Scaffaro R, Maio A, Citarrella MC. Ionic tactile sensors as promising biomaterials for artificial skin: review of latest advances and future perspectives. *Eur Polym J* 2021;151:110421. DOI
68. Yan Y, Hu Z, Yang Z, et al. Soft magnetic skin for super-resolution tactile sensing with force self-decoupling. *Sci Robot* 2021;6:eabc8801. DOI PubMed
69. Shin YE, Park YJ, Ghosh SK, Lee Y, Park J, Ko H. Ultrasensitive multimodal tactile sensors with skin-inspired microstructures through localized ferroelectric polarization. *Adv Sci* 2022;9:e2105423. DOI PubMed PMC
70. Zhang H, Chen H, Lee J, et al. Bioinspired chromotropic ionic skin with in-plane strain/temperature/pressure multimodal sensing and ultrahigh stimuli discriminability. *Adv Funct Mater* 2022;32:2208362. DOI
71. Wan C, Cai P, Guo X, et al. An artificial sensory neuron with visual-haptic fusion. *Nat Commun* 2020;11:4602. DOI PubMed PMC
72. Liu M, Zhang Y, Wang J, et al. A star-nose-like tactile-olfactory bionic sensing array for robust object recognition in non-visual environments. *Nat Commun* 2022;13:79. DOI PubMed PMC
73. Ehatisham-ul-haq M, Javed A, Azam MA, et al. Robust human activity recognition using multimodal feature-level fusion. *IEEE Access* 2019;7:60736-51. DOI
74. Ahmad Z, Khan N. Human action recognition using deep multilevel multimodal (M^2) fusion of depth and inertial sensors. *IEEE Sensors J* 2020;20:1445-55. DOI
75. Dawar N, Ostadabbas S, Kehtarnavaz N. Data augmentation in deep learning-based fusion of depth and inertial sensing for action recognition. *IEEE Sens Lett* 2019;3:1-4. DOI
76. Dawar N, Kehtarnavaz N. Action detection and recognition in continuous action streams by deep learning-based sensing fusion. *IEEE Sensors J* 2018;18:9660-8. DOI
77. Hall D, Llinas J. An introduction to multisensor data fusion. *Proc IEEE* 1997;85:6-23. DOI
78. Xu J, Yao T, Zhang Y, Mei T. Learning multimodal attention LSTM networks for video captioning. Proceedings of the 25th ACM international conference on Multimedia; 2017. p. 537-45. DOI
79. Shon S, Oh T, Glass J. Noise-tolerant audio-visual online person verification using an attention-based neural network fusion. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019. p. 3995-9. DOI
80. Shyu M, Xie Z, Chen M, Chen S. Video semantic event/concept detection using a subspace-based multimedia data mining framework. *IEEE Trans Multimedia* 2008;10:252-9. DOI
81. Yang Z, Li Q, Lu Z, Ma Y, Gong Z, Pan H. Semi-supervised multimodal clustering algorithm integrating label signals for social event detection. 2015 IEEE International Conference on Multimedia Big Data; 2015 Apr 20-22; Beijing, China. IEEE; 2015. p. 32-9. DOI
82. Zhou P, Yang W, Chen W, Wang Y, Jia J. Modality attention for end-to-end audio-visual speech recognition. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019. p. 6565-9. DOI