sensors

MDPI

Review

Intelligent, Flexible Artificial Throats with Sound Emitting, Detecting, and Recognizing Abilities

Junxin Fu [1,2], Zhikang Deng [1,2], Chang Liu [1,2], Chuting Liu [1,2], Jinan Luo [1,2], Jingzhi Wu [1,2], Shiqi Peng [1,2], Lei Song [1,2], Xinyi Li [1,2], Minli Peng [1,2], Houfang Liu [3], Jianhua Zhou [1,2,*] and Yancong Qiao [1,2,*]

[1] School of Biomedical Engineering, Shenzhen Campus of Sun Yat-sen University, No. 66, Gongchang Road, Guangming District, Shenzhen 518107, China
[2] Key Laboratory of Sensing Technology and Biomedical Instruments of Guangdong Province, School of Biomedical Engineering, Sun Yat-sen University, Guangzhou 510275, China
[3] School of Integrated Circuits and Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China
* Correspondence: zhoujh33@mail.sysu.edu.cn (J.Z.); qiaoyc3@mail.sysu.edu.cn (Y.Q.)

Abstract: In recent years, there has been a notable rise in the number of patients afflicted with laryngeal diseases, including cancer, trauma, and other ailments leading to voice loss. Currently, the market is witnessing a pressing demand for medical and healthcare products designed to assist individuals with voice defects, prompting the invention of the artificial throat (AT). This user-friendly device eliminates the need for complex procedures like phonation reconstruction surgery. Therefore, in this review, we will initially give a careful introduction to the intelligent AT, which can act not only as a sound sensor but also as a thin-film sound emitter. Then, the sensing principle to detect sound will be discussed carefully, including capacitive, piezoelectric, electromagnetic, and piezoresistive components employed in the realm of sound sensing. Following this, the development of thermoacoustic theory and different materials made of sound emitters will also be analyzed. After that, various algorithms utilized by the intelligent AT for speech pattern recognition will be reviewed, including some classical algorithms and neural network algorithms. Finally, the outlook, challenge, and conclusion of the intelligent AT will be stated. The intelligent AT presents clear advantages for patients with voice impairments, demonstrating significant social values.

Keywords: artificial throat; sound sensor; thermoacoustic effect; machine learning

Citation: Fu, J.; Deng, Z.; Liu, C.; Liu, C.; Luo, J.; Wu, J.; Peng, S.; Song, L.; Li, X.; Peng, M.; et al. Intelligent, Flexible Artificial Throats with Sound Emitting, Detecting, and Recognizing Abilities. *Sensors* **2024**, *24*, 1493. https://doi.org/10.3390/s24051493

Academic Editor: Giorgio Biagetti

Received: 23 January 2024
Revised: 22 February 2024
Accepted: 22 February 2024
Published: 25 February 2024

1. Introduction

Verbal communication is the basic communication method of human beings. However, some patients in the world have deficiencies in language ability. In China, oral and oropharyngeal cancer accounts for approximately 307,000 new cases each year, constituting over half of the 572,000 new cases identified worldwide [1–3]. In addition to laryngeal cancer, diseases such as esophageal cancer and other unexpected accidents will also seriously affect patients' linguistic ability. The interpersonal communication and quality of life of the mute people will also be seriously spoiled, leading to a negative impact on their mental and physical health [4]. Therefore, how to effectively reconstruct the vocal function of speech-impaired people, aiming at minimizing the detrimental effects of speech impairment, has become the focus of the whole society.

Nowadays, one of the most widely used pronunciation reconstruction methods is still the conventional electrolarynx [5]. The conventional electrolarynx (Figure 1a) [6], composed of motorized transducers with large rigidity, volume, and complex structure, can assist mute people to emit sound [7]. The working principle of the conventional electrolarynx is initially creating vibrations of the oral or pharyngeal at a constant fundamental frequency. Then, these vibrations will be transmitted to the throat or mouth. Following this, after interacting with the vocal cord tissue, the patient can produce audible speech [5,8]. The

usage is shown in Figure 1b. Currently, most of the currently available electrolarynx has been adapted, designed, and modified on this working principle. For example, Isshiki et al. proposed an electrolarynx with better performance in voice [9]. Wu et al. put forward a method of solving the problem to eliminate the abnormal acoustic properties [10].

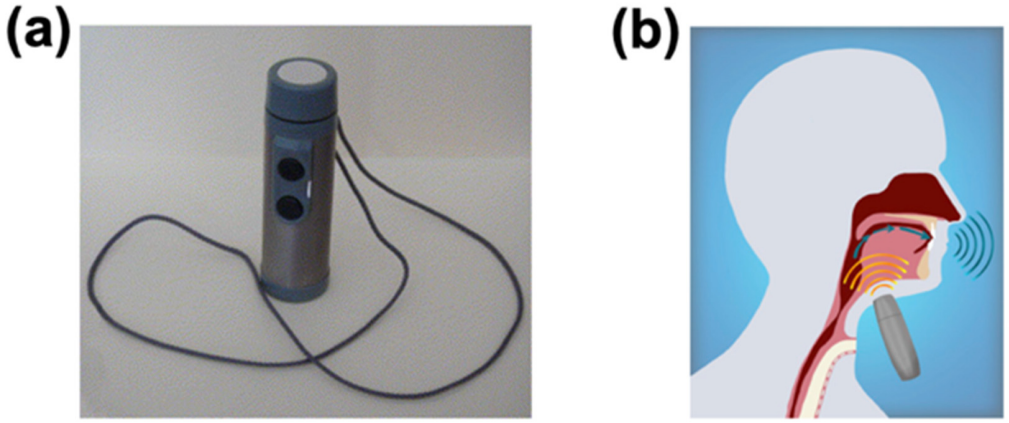

Figure 1. The conventional electrolarynx. (**a**) The overview of the conventional electrolarynx. Reproduced with permission [6]. (**b**) The usage of the conventional electrolarynx [8].

However, despite continuous changes and improvements, the conventional electrolarynx still has many shortcomings that have not been effectively addressed. First, the conventional electrolarynx is a hand-held device that takes up a person's hand and restricts the normal movement of the hand. Second, managing the conventional electrolarynx is challenging. Initially, the patients often need to spend much time finding the most suitable site to attach the electrolarynx to the neck muscles. In addition to the proper site, the interface between the conventional electrolarynx and the skin significantly influences its vocalization. Consequently, achieving the appropriate tightness to adhere the electrolarynx to the neck muscle is also a time-consuming process for patients. Third, the conventional electrolarynx is only capable of producing a mechanized and monotonous sound, which lacks variation compared to natural sounds. This poor quality of sound may seriously affect the patient's voice expression and communication experience, limiting their ability to communicate fluently. Therefore, the conventional electrolarynx has significant deficiencies in sound quality and learning difficulty, which affects speech recovery in patients with laryngeal cancer or laryngectomy.

In recent years, with the continuous development of materials science, solid-state physics, and electronic engineering, new methods have been provided to solve the difficult problems plaguing the development of the conventional electrolarynx [11]. In conventional electrolarynx, it is limited to being used as a sound emitter. However, the latest developing artificial throat (AT) can not only be used as a sound emitter but also as an intelligent sound sensor, integrating sound perception technology with the assistance of algorithms. As for the sound sensor, capacitive, piezoelectric, electromagnetic, and piezoresistive materials have been widely leveraged in the field. Some sound sensors are self-powered and can be combined with other physiological signs. In terms of sound emitters, conventional sound emitters typically rely on the electromagnetic effect with rigid and solid materials. In contrast, thermoacoustic materials are often flexible and thin, which could be fabricated as thin-film sound emitters. This characteristic makes them exceptionally well-suited for wearable applications. In contrast to the conventional electrolarynx, the latest AT developed based on new materials is lightweight, thin in thickness, simple in structure, soft in use, and comfortable to wear. These attributes collectively contribute to an enhanced patient experience.

As mentioned before, with the assistance of a machine learning algorithm, the AT has progressively gained intelligence in speech detection and recognition, addressing the deficiency of the electrolarynx in this regard. In 2020, Jin et al. developed a model

with a large amount of data to recognize the long vowels and short vowels of human pronunciations, and the recognition accuracy can reach 83.6% in the long vowels, and 88.9% in the short vowels. Afterward, many other models have also been used in the speech recognition of AT, such as SR-CNN [12], AlexNet [13,14], Inception V3 [13], SCNN [15], etc. Obviously, machine learning algorithms have greatly assisted AT in breaking through the deficiencies of the electrolarynx in receiving voices. Moreover, AT combined with a machine learning algorithm brings these speech-impaired patients more effective assistance than the conventional electrolarynx.

In this review, we aim to give an overview of the intelligent flexible AT, which consists of the sound emitter, sensor, and recognition algorithm (Figure 2). The concept and composition of the AT are initially elucidated, and compared with the conventional electrolarynx. The second part focuses on the sound sensor, which covers not only sound but also other physiological signals, including electromyographic (EMG) signals that reflect voice information. In the third section, a thin-film sound emitter based on the thermoacoustic effect will be discussed, which can help the user to emit sound. In the fourth section, a detailed review of the intelligent AT, supported by various algorithms for sound wave recognition, will be stated, such as some digital signal processing techniques, classical machine learning algorithms, deep learning algorithms, and so on. Finally, the outlooks and limitations will be given. This review is helpful for the researchers who intend to study the AT devices.

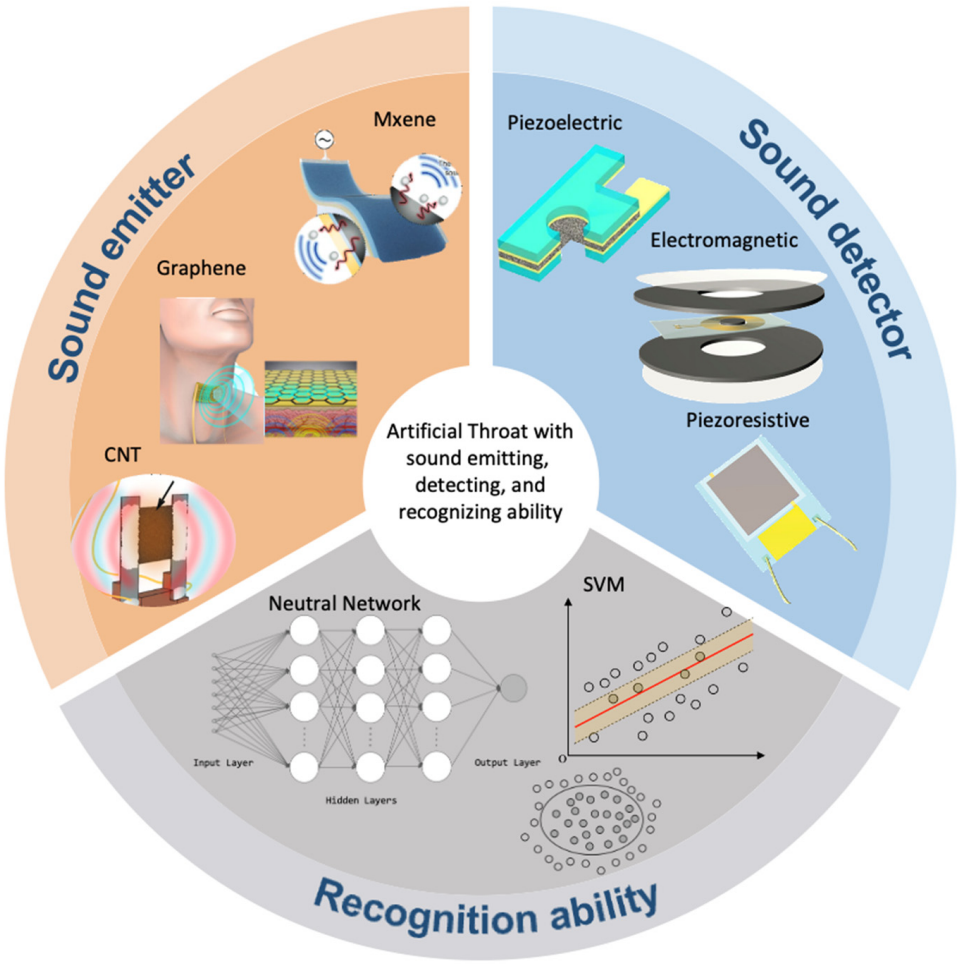

Figure 2. Intelligent flexible AT serves as the sound emitter, detection, and recognition devices [13,16–20].

2. Sound Sensor

When a sound emitter vibrates, such as the vocal cords of a person, the strings of a musical instrument, or other objects, it induces the surrounding medium to generate alternating zones of compression and rarefaction. This phenomenon gives rise to the

production of sound. Sound can be regarded as a combination of simple harmonic waves, which can propagate through a solid, liquid, or gas [21–24]. When a sound wave strikes the human's ear membrane, the different frequencies of sound result in various levels of vibration; only the frequencies ranging from 20 Hz to 20 kHz can stimulate the human nervous response to produce the hearing sensation [25,26]. Within this process, the human ear plays a key role, acting as a sound sensor with a sophisticated structure. Owing to the sophisticated design, human ears have a high sensitivity to voice. In the realm of sound sensors, the material and structure also need to be specially designed to detect sound, especially weak sound.

Recent decades have witnessed the rapid development of sound sensors. Capacitive, piezoelectric, electromagnetic, and piezoresistive sensors have been widely studied. Moreover, the design of sensing materials and the preparation of sensors have significantly improved. Furthermore, when integrated with other human physiological signals, sound detection has achieved high levels of accuracy.

2.1. Capacitive Sound Sensor

The capacitance for the common plain plates can be expressed as

$$C = \varepsilon \frac{S}{d}, \tag{1}$$

where ε is the dielectric coefficient of the medium between the plates, depending on the physical property of the material between the plates. S stands for the area of one plate, and d is the distance between the plates. Derived from this formula, the capacitive sensor can be divided into three categories: the first category is based on a variable dielectric coefficient, when the dielectric changes, the capacitance will change. Many humidity sensors are developed on this basis. The second category considers the changes in area [27–29]. The third category is building on the shift in distance. Considering the characteristics of the sound waves, many capacitive sensors are based on the distance changes between movable plates and other fixed plates [30].

The principle of the third category of capacitive sound sensors can be stated as follows: A bias voltage is initially applied to load the plates. Subsequently, the voltage between the two plates remains constant unless an incoming sound wave induces vibrations in the movable plates. This vibration leads to a change in capacitance and, consequently, a variation in voltage [31]. In this manner, the capacitive sound sensor senses the acoustic signal, transforming it into an electrical signal with a flat frequency response [32].

Lee et al. realized a sound sensor with a sophisticated capacitive structure [33]. When the device is attached to the neck skin, the vibration of the neck muscles will cause changes in the capacitance between the movable plates and the fixed plates (Figure 3a). When connected to the capacitance sensing circuit that effectively converts capacitance changes to voltage variations, the vibration of the neck muscles will be transformed into an electrical signal (Figure 3b). Compared with the conventional microphone, the device could effectively resist noise interference, owing to it recognizing the voice just by the skin vibration (Figure 3c).

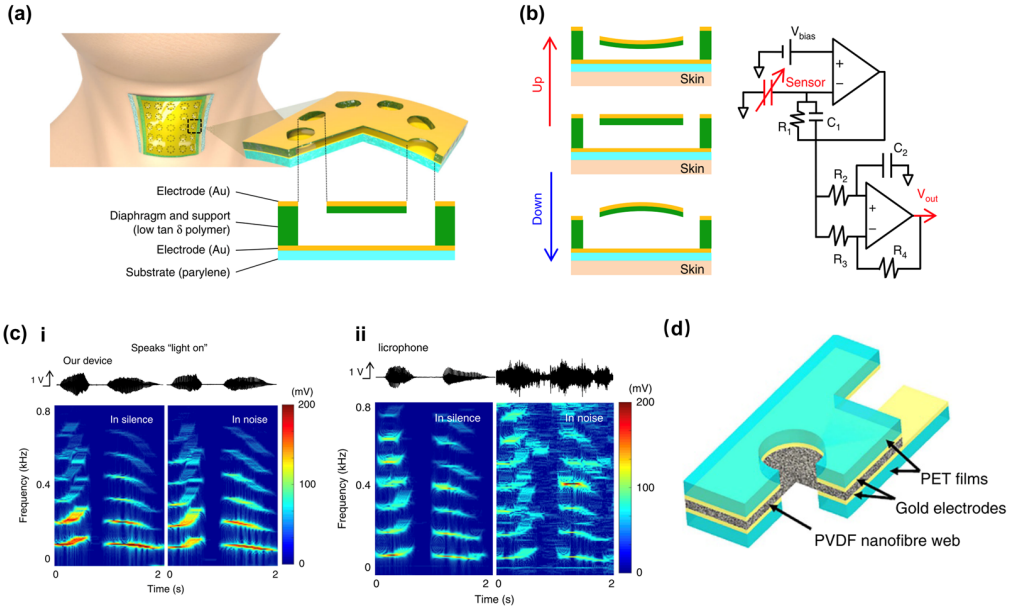

Figure 3. Sound sensor based on capacitive and piezoelectric effect. (**a**) Illustration of the capacitive sound sensor attached to the neck and the diaphragm structure. (**b**) The circuit diagram within the sensor. (**c**) Comparison of waveform and frequency spectrum in silent and noisy environments when a person speaks 'light on' with the capacitive sound sensor and licrophone. (**i**) utilizes the capacitive sound sensor, while (**ii**) utilizes the licrophone [33]. (**d**) Sound sensor structure based on the piezoelectric materials [17].

2.2. Piezoelectric Sound Sensor

The ability to generate an electrical charge by applying mechanical stress is called the piezoelectric effect [34], which can be found in some materials, such as polyvinylidene difluoride (PVDF) [17,35–37], lead zirconium titanate (PZT) [36,38–40], zinc oxide (ZnO) [41–43]. Compared with capacitive sensors, piezoelectric sensors do not require an additional bias voltage to be supplied and do not require additional circuit design. Furthermore, a quantity of sensors can also be self-powered.

Lang et al. developed a PVDF-based sound sensor using electrospinning technology (Figure 3d) [17]. Figure 4a schematically illustrates a proposed sound-sensing mechanism. When the sound wave hits the sensor, the sound absorption induces vibration of the nanofiber network, the Au layer, and the polyethylene terephthalate (PET) sheet. Part of the nanofiber mesh is covered by a PET sheet and Au layer, but part of the nanofiber mesh is directly exposed to the sound absorption. The directly exposed part vibrates more intensively than those covered, causing asymmetric vibrations on the propagation along the fiber and a heightened sensitivity in sound perception. In addition, the piezoelectric sound sensor has good sound perception at low frequencies. Figure 4b-ii demonstrates its ability to effectively distinguish between two different low frequencies, approximately 190 Hz and 260 Hz. However, its measured sound pressure after 400 Hz will rapidly drop to 0, so its performance at high frequencies is notably inferior to that of the capacitive sensor [32,37]. Recent studies have shown another key physical property of this material: the thickness of the piezoelectric material plays a key role in sound detection. Lim et al. fabricated a piezoelectric sound sensor with single-walled carbon nanotubes (SWCNTs) and a PVDF network with varying thicknesses. They discovered that the output voltage and impedance would show a nearly linear relationship within a proper thickness. If the thickness is too small (below 200 μm), the impedance will drop rapidly, possibly linked to a short circuit inside the electrostatic spinning filament [37].

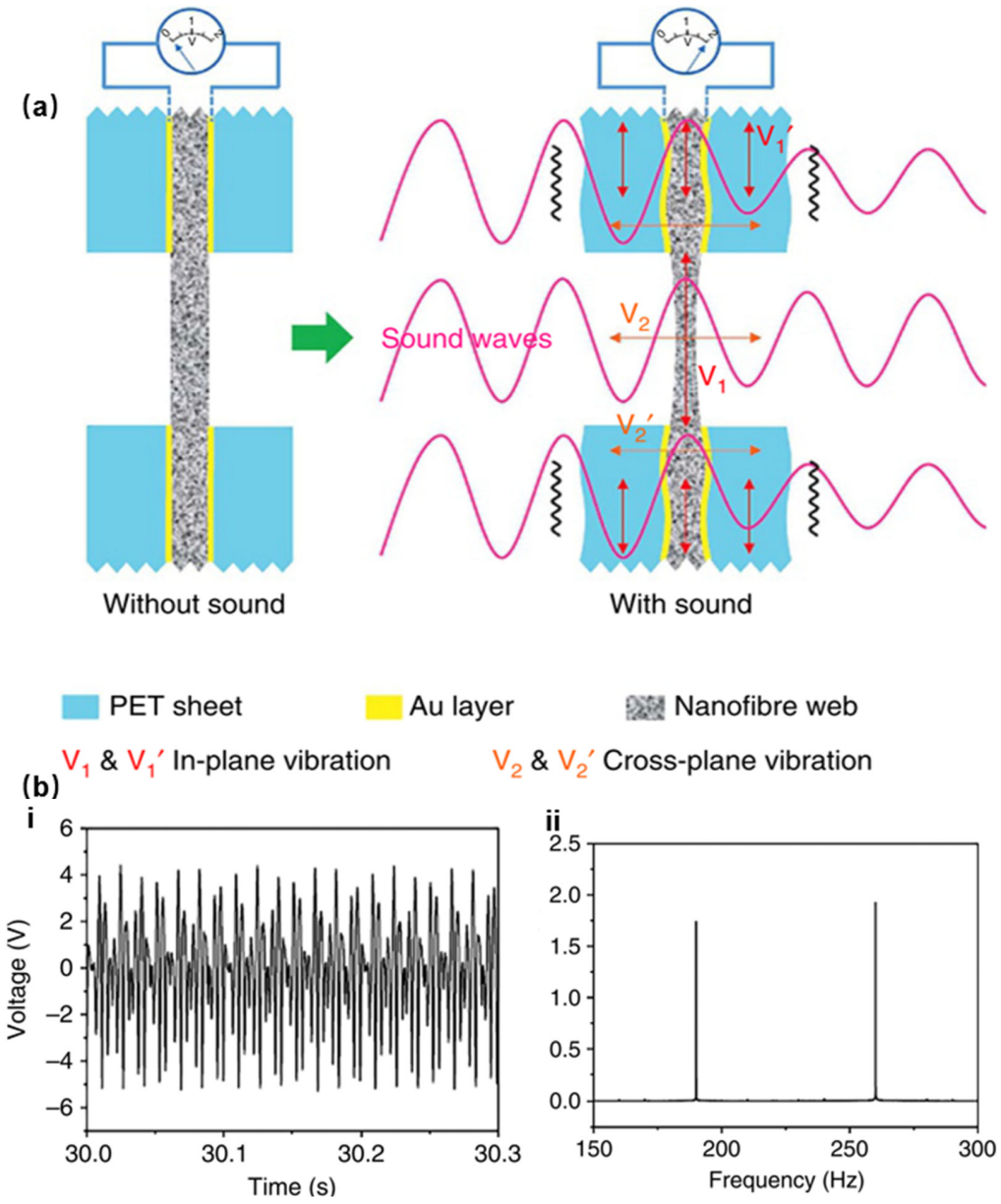

Figure 4. Sound sensor based on piezoelectric effect. (**a**) When sound waves hit the piezoelectric nanofibers, vibration of the piezoelectric materials takes place. (**b-i**) is the voltage spectrum under double-frequency sound waves, while (**b-ii**) is the frequency under double-frequency sound waves [17].

Owing to the fact that in piezoelectric sensors, the mechanical input can be converted directly into an electrical output with no external power source required, this type of sensor is considered to be self-powered, which is promising in fields like sound energy harvesting [44,45]. The working process of the self-powered device can be explained in Figure 5a [32,46]. Figure 5a-i schematically illustrates the structure of the spring-substrate nanogenerator, which is composed of a metal spring. The metal spring is composed of an Ag electrode and a quantity of ZnO nanowires which are passivated with polymethyl methacrylate (PMMA). When a tiny plate weighing 15.2 N is placed on the self-powered sensor, the piezoelectric sensor will produce an output voltage of 0.23 V. Furthermore, Figure 5a-iii shows an almost linear relationship between the output voltage and current, exhibiting a sensitivity of 2.8 nA $\times$ kg^{-1} and 45 mV $\times$ kg^{-1}. This remarkable sensitivity indicates the sensor's performance which has been employed in piezoelectric sound detection.

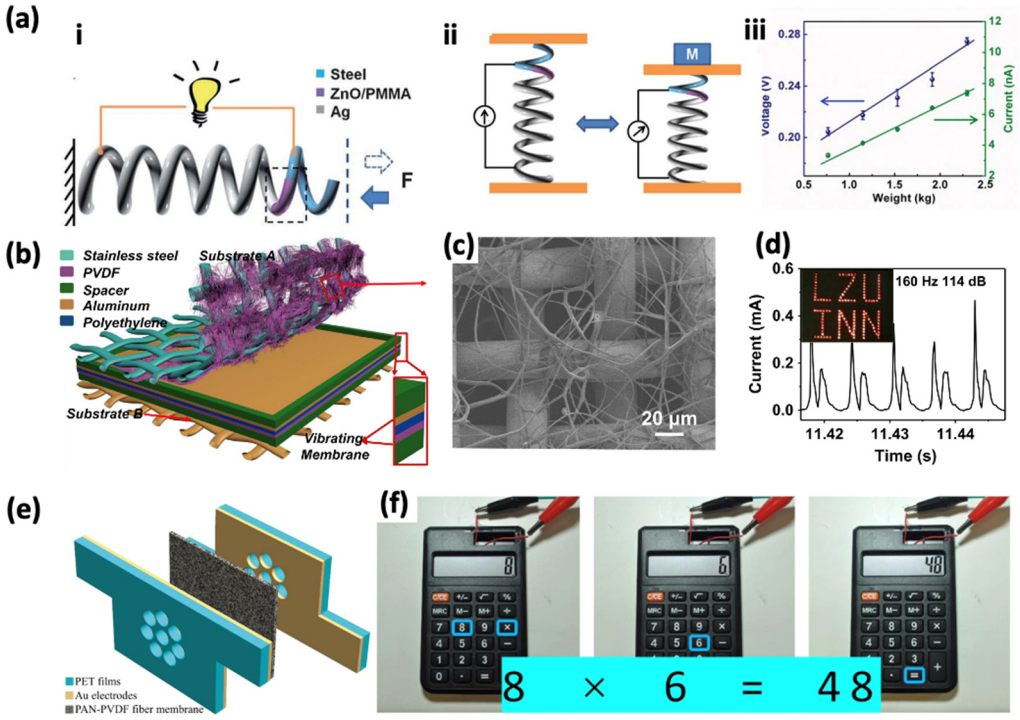

Figure 5. Self-powered piezoelectric sound sensors. (**a-i**) Schematic structure of nanogenerator based on ZnO. (**a-ii**,**a-iii**) The voltage and current vary when weight is put on the nanogenerator sensor [32]. (**b**) Schematic of a fabricated sound TENG. (**c**) SEM image of the PVDF nanofibers. (**d**) 138 LEDs were driven by the sound TENG with the sound of 144 dB and 160 Hz [47]. (**e**) Structure of the PAN-PVDF noise harvester structure. (**f**) The sound sensor powers the calculator to perform the calculation process [48].

Cui et al. achieved a sound-driven triboelectric nanogenerator (TENG) based on the piezoelectric material PVDF (Figure 5b,c). Their device demonstrates the capability to instantaneously illuminate 138 LEDs, as shown in the inset of Figure 5d, in response to a 114 dB/160 Hz sound [47]. The energy generated by this nanogenerator can not only light the basic electronic components, like LEDs, but also power other commercial products. This indicates that these sound sensors have an exceptionally wide range of applications. Shao et al. fabricated a single-layer piezoelectric nanofiber sound sensor utilizing PET films, Au electrodes, and PAN-PVDF fiber membrane (Figure 5e). This sensor can power the calculator to perform the calculation process and charge the capacitor (Figure 5f) [48].

2.3. Electromagnetic Sound Sensor

The classic structure of electromagnetic sensors is often composed of a coil, a diaphragm, and a permanent magnet. When the diaphragm is vibrated by sound waves, the diaphragm will drive the coil to move in the magnetic field, thus generating output current. However, this classic structure is mostly rigid and only partially flexible due to limitations such as coils and magnets [49–51].

As a result of advancements in manufacturing methods for achieving flexible magnetic membranes, Huang et al. successfully produced flexible neodymium magnet (NdFeB) membranes in the origami approaches in 2019 [52]. In 2020, they manufactured an additional fully flexible electromagnetic sensor incorporating NdFeB (Figure 6a), enabling repeated bending and twisting for attachment to the body [18]. Moreover, this fully flexible structure can serve as a sound sensor by utilizing the electromagnetic induction between the copper coil and magnetic membrane to detect the vibration of the vocal cords (Figure 6b,c).

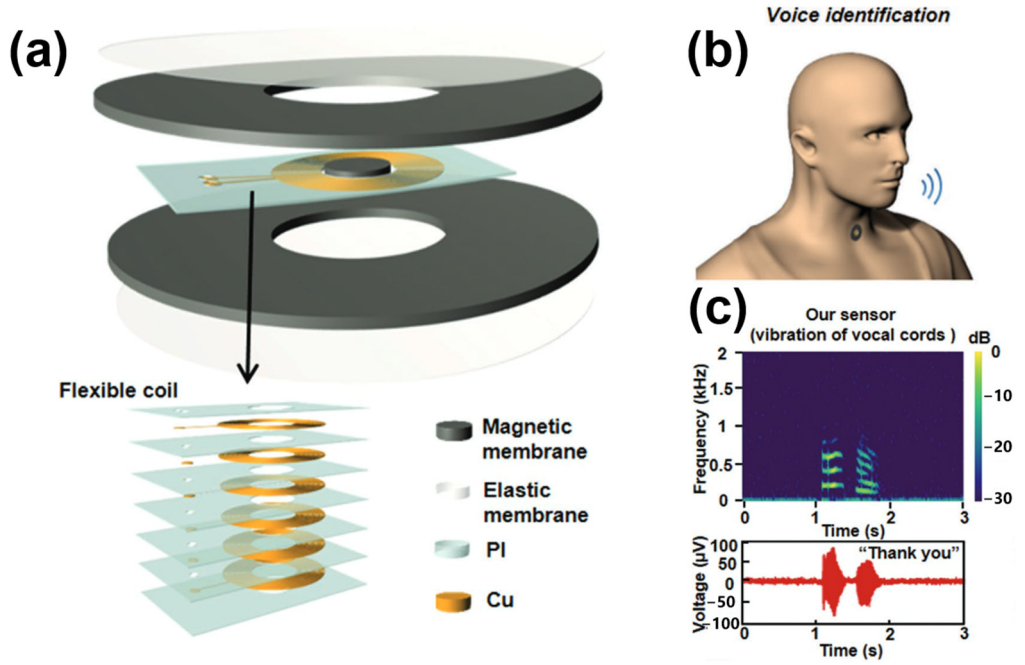

Figure 6. Electromagnetic effect-based sound sensor. (**a**) The structure of the electromagnetic sensor. (**b**) The sensor is attached to the neck for voice identification. (**c**) The time-frequency diagram measured by a sensor attached to the neck and the frequency spectrum converted by a fast Fourier transform [18].

2.4. Piezoresistive Sound Sensor

The piezoresistive effect refers to the change in resistance of a material of semiconductor or metal when mechanical strain is applied [53–56]. The piezoresistive effect in metals and semiconductors was discovered in the 18th century and 19th respectively. For some electrical conductors with the same physical property when measured in different directions, the relative resistance change can be derived as

$$\frac{\Delta R}{R} = \frac{\Delta l}{l}(1 + 2v) + \frac{\Delta \rho}{\rho}, \tag{2}$$

where l describes the length of the electrical conductor, v is Possion's ration of the electrical conductor, and ρ is the resistivity. As described in this formula, the change of the length (Δl) and the variance of the ρ $(\Delta \rho)$ determine the change of resistance (ΔR) [54,57].

As depicted in Figure 7a, the application of an external force induces compressive deformation in the sensor within increasing material contact, which creates additional conductive paths and varying resistance. In recent years, studies have indicated that strain sensors fabricated with nanomaterials, when exposed to sound waves, generally exhibit the piezoresistive effect and are widely recognized as sound sensors. These nanomaterials include graphene [55,58,59], carbon nanotube (CNT) [60–62], MXene [19,63–66], and so on. Tao et al. fabricated a sound sensor with graphene, derived from the graphene resistance changes in response to applied forces [58]. Ma et al. developed a piezoresistive sensor capitalizing on highly ordered hierarchical architectures of hybrid 3D MXene/reduced graphene oxide (MXene/rGO) (Figure 7b). This design combines the large specific surface area of graphene oxide with the excellent conductivity of MXene, enabling the sensor to recognize a wider range of pressures and detect the vibration when attached to the throat (Figure 7c) [19]. Gong et al. proposed an ultrathin gold nanowires (AuNWs) impregnated tissue paper sandwiched between a blank PDMS sheet and a patterned PDMS sheet (Figure 7d), which could achieve the sensitivity of 1.14 kPa^{-1}. As illustrated in Figure 7e,f, when a voltage is fixed, the application of external pressure results in a decrease in resistance and an increase in current [67].

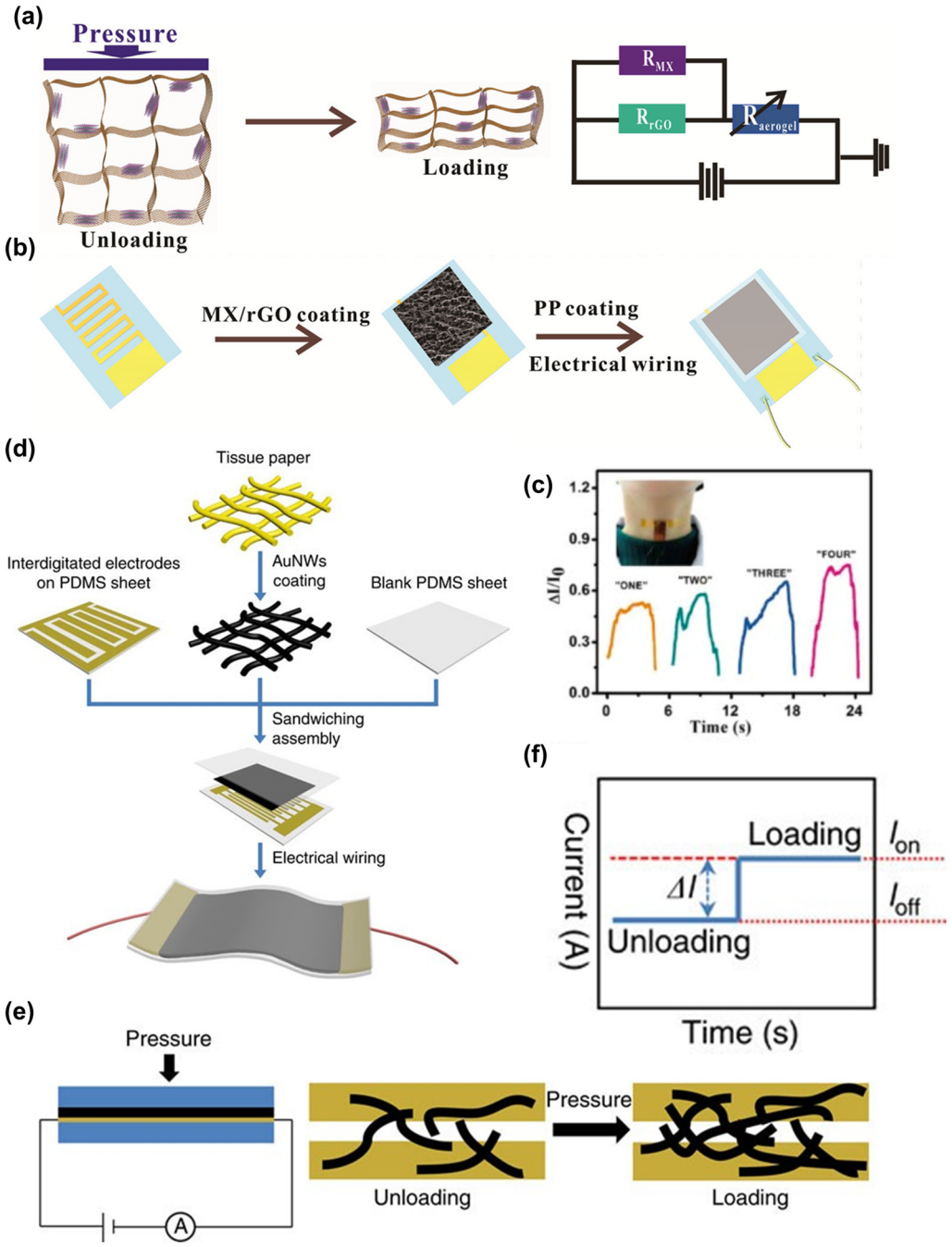

Figure 7. Piezoresistive effect-based sound sensor. (**a**) The schematic illustration of the piezoresistive material sensing mechanism. (**b**) The fabrication process of the MX/rGO sensor. (**c**) The continuous monitoring of the tiny strain and human voice using MX/rGO sensors [19]. (**d**) Schematic illustration of the fabrication of the piezoresistive sensor based on AuNWs. (**e**,**f**) The illustration of the sensing mechanism and current changes when applying pressure [67].

2.5. Silent Speech Interfaces in Sound Recognition

Alternative methods for sound recognition exist for scenarios where sound signals are unavailable. One such method is the silent speech interface, a system capable of generating a digital representation of speech by acquiring sensor data during the human speech production process [68]. For instance, some physiological signals during speaking are relevant to vocalization information, which can serve as a sensor in the silent speech interface [69–72]. Tanja Schultz et al. obtained a high recognition rate of oral word signals by using EMG signals [73]. Liu et al. fabricated a tattoo-like patch to acquire the EMG from

three muscle channels to recognize the instructions [74]. Compared with the acquisition of only a single resistance or voltage signal, the simultaneous acquisition of the EMG signal and other physiological signals will bring great help to the subsequent signal processing and speech recognition [75,76]. Tian et al. proposed a dual-channel speech recognition system based on the EMG and mechanical sensors. In comparison to utilizing the mechanical signal of the neck muscles, the signals including the movement of the neck muscles and EMG exhibit superior performance in sound detection and recognition [77]. In addition to EMG, electroencephalographic (EEG) also plays a role in speech recognition. Pradeep Kumar et al. developed a speech recognition framework with the help of EEG signals with high accuracy in the recognition task of 30 text and not-text classes [78]. Anne Porbadnigk et al. also investigated the use of EMG by utilizing 16 EEG channels with a 128-cap montage for speech recognition [79]. Apart from the physiological signals, silent speech interfaces also include some real-time characterization of the vocal tract methods such as using ultrasound and optical imaging of the tongue and lips for speech recognition.

As stated before, the sensing principles for the latest generation of flexible and wearable sound sensors including capacitive, piezoresistive, piezoelectric, etc. The present sensors have been gradually miniaturized and flexible with soft, highly curved properties, playing a crucial role in the voice recognition capabilities of AT. However, the current AT still have problems such as sensitivity and accuracy, which need to be further improved. Additionally, optimal performance and durability during use also need to be focused [57].

3. Sound Emitter

The sound emitter is an important component of the AT. The successful restoration of patients' voices with the help of AT relies on the performance of the sound emitter [80]. A sound emitter is a transducer that converts electrical signals into sound signals. Taking the most common moving-coil sound emitter as an example, audio-electrical signals are transmitted through an electromagnetic effect. This effect induces vibrations in its diaphragm, resonating with the surrounding air and generating sound. However, the electromagnetic effect requires a permanent magnet, a coil, and a diaphragm to create vibrations in the air and then produce sound. As a result, moving-coil speakers are typically larger in size [6]. In addition to utilizing the traditional vibrating sound emitter, the AT can also produce sound by means of the thermoacoustic (TA) effect [81]. This sound emitter is only a thin film with a small size. Due to this principle, the AT can be worn directly on the patient's larynx as an electronic skin [82].

The sound emitters based on the TA effect is a device that generates sound using heat. The physical process of the TA effect can be described as follows: when an alternating current signal passes through a thin metal film, the film generates Joule heat, which is rapidly transferred to the surrounding air medium. Due to the periodic rise and fall of the temperature of the surface of the metal, the air molecules in the thin layer of the surface of the metal are constantly expanding and contracting, thus generating sound waves. By controlling the rate of heating and cooling, the frequency of the sound produced can be modulated, allowing for the generation of different sound intensities and tones [11,83].

3.1. Development of the TA Sound Emitter

The TA effect was discovered more than 200 years ago. In the 18th century, Byron Higgins experimented with a hydrogen flame placed in the proper position in a vertical tube with openings at both ends, and sound was produced in the tube. This is historically known as a singing flame and was the first discovery of the thermoacoustic effect.

In 1917, Arnold and Crandall proposed a TA sound emitter made of suspended 700-nm platinum film (Figure 8a) [84]. Then, they analyzed its sound-emitting mechanisms theoretically. When an AC current with the sound frequency passes through the surface of a platinum film with a low heat capacity, the heat is transferred to the ambient air, causing

the air to expand periodically, thus producing sound. In their theory, the formula of the SP can be derived as

$$P_{rms} = \frac{\sqrt{\alpha}\rho_0}{2\sqrt{\pi}T_0} \times \frac{1}{r} \times P_{input} \frac{\sqrt{f}}{C_s}, \quad (3)$$

where C_s is the heat capacity per unit area (HCPUA) of the thermoacoustic thin film, and f is the frequency of the excitation frequency. P_{input} and r are the input power and the distance between the thin film with the microphone, respectively. α, ρ_0, and T_0 are the thermal diffusivity, density, and temperature of the ambient gas. This equation indicates that the sound pressure produced by the TA sound emitters increases with smaller HCPUA, higher frequency, and input power.

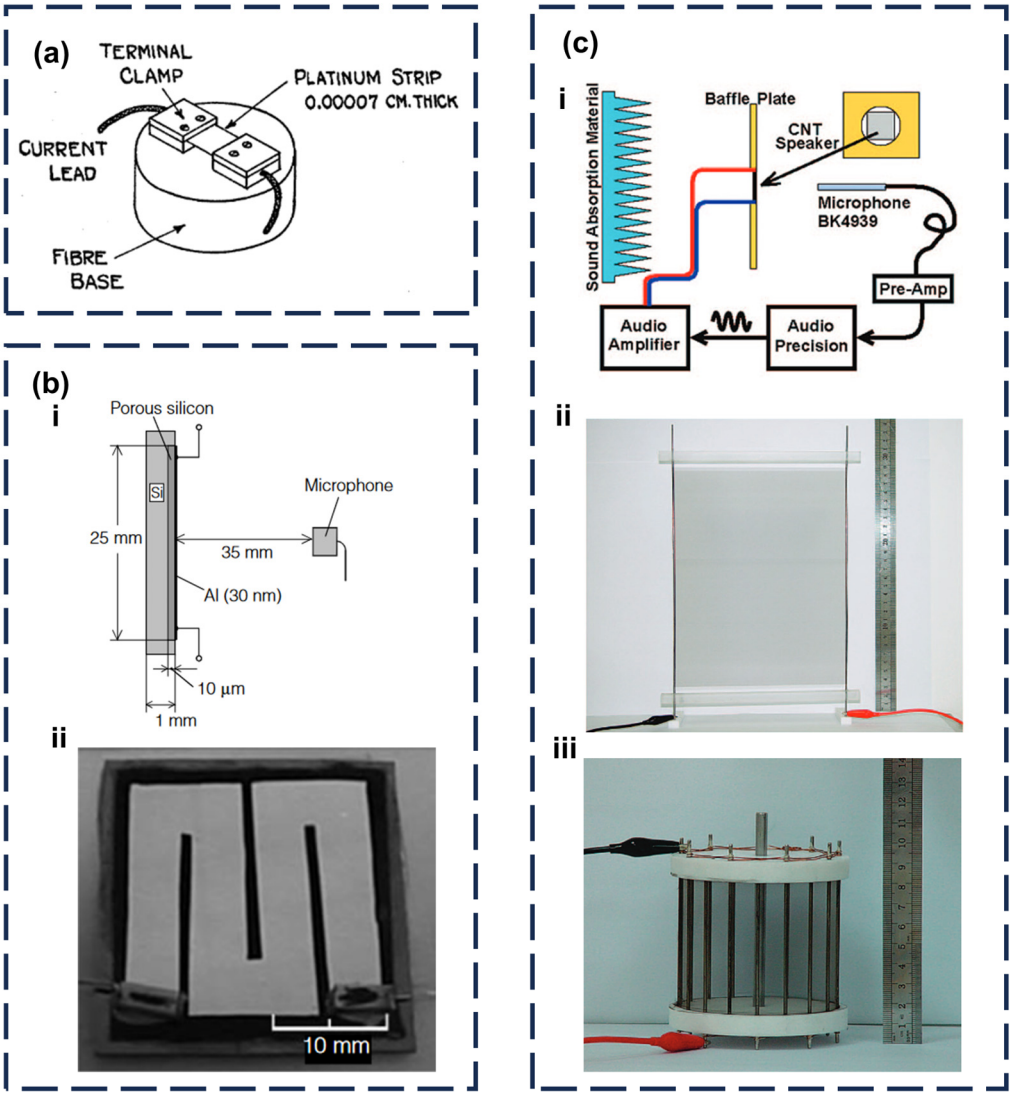

Figure 8. Development of the TA sound emitter. (**a**) Simple TA sound emitter made of the platinum strip [84]. (**b-i**) Cross-sectional view of the fabricated device and set-up for sound measurement. (**b-ii**) Photograph of a top view of the device [81]. (**c-i**) Schematic illustration of the experimental setup for CNT thin film sound emitters. (**c-ii**) A4 paper size CNT thin film sound emitter. (**c-iii**) the cylindrical cage shape CNT thin film sound emitter [83].

Their theoretical model led to the derivation of the basic sound generation equation. However, over the subsequent 100 years, the TA sound emitter has been overlooked, due to the specific properties of materials and the limitation of signals being only at 4 kHz, coupled

with low sound pressure. Nonetheless, it is crucial to acknowledge that this theoretical groundwork provides a foundation for the subsequent development of the thermal.

It was not until 1999 that H. Shinoda et al. extended Amold et al.'s surface-sounding theory. They introduced a porous-silicon-based sound emitter in Nature (Figure 8b) [81]. This approach involved applying a 30-nm-thick aluminum film on top of a 10-µm-thick, porous silicon layer, resulting in a wide-band sound emitter capable of achieving a notable sound pressure of 0.1 Pa (1–100 kHz). In this work, they enhanced and refined the previous module. In their theory, the SP can be described as

$$P(x,\omega) = \sqrt{\frac{\gamma\alpha_a}{C_a}}\frac{P_A}{vT_A} \times \frac{\exp(-jkx)}{\sqrt{\alpha C}} \times q(\omega), \tag{4}$$

where P_A is atmospheric pressure, T_A is room temperature, v is the sound velocity, $\gamma = \frac{C_p}{C_v} = 1.4$, C_p is the heat capacity at constant pressure, C_v is the heat capacity at constant volume, k is the wavenumber of sound in free space. α_a is the thermal conductivity in air, and C_a is the HCPUA.

In the 21st century, the development of nanotechnology has led to breakthroughs in T_A sound emitter devices. In 2008, Xiao et al. achieved a groundbreaking in thermoacoustic theory [83,85]. They fabricated a sound emitter utilizing CNT (Figure 8c). This device boasts a wide frequency response range and a high SPL, due to the low HCPUA of CNT. However, their experimental results did not align with Arnold and Crandall's theory, then they identified that Arnold's theory neglected the rate of heat loss per unit area of the thin film and the instantaneous heat exchange per unit area. Based on this observation, they proposed their own model as follows:

$$P_{rms} = \frac{\sqrt{\alpha}\rho_0}{2\sqrt{\pi}T_0} \times \frac{1}{r} \times P_{input} \times \frac{\sqrt{f}}{C_s} \times \frac{\frac{f}{f_2}}{\sqrt{\left(1+\sqrt{\frac{f}{f_1}}\right)^2 + \left(\frac{f}{f_2}+\sqrt{\frac{f}{f_1}}\right)^2}}, \tag{5}$$

In their new module, two constants f_1 and f_2 were added, $f_1 = \frac{\alpha\beta^2}{\pi k^2}$ and $f_2 = \frac{\beta_0}{\pi C_s}$. Additionally, the previous Arnold's theory is only suitable for higher HCPUA and is not applicable to smaller HCPUA. Xiao et al. introduced a modified model that overcame these limitations. Based on their theory's findings, they fabricated a CNT thin film TA sound emitter, which possesses the merits of nanometer thickness and are transparent, flexible, and stretchable [83].

In 2010, Hu et al. modeled a TA sound emitter in the low and high-frequency bands on the basis of H. Shinoda [86], confirming that there exists a very wide range of constant amplitude-frequency response mostly in the ultrasonic region for TA emission from any solid. In the same year, V. Vesterinen et al. verified the theoretical model by using nanoscale aluminum as a sound-emitting layer (Figure 9a) [87]. Then, they concluded that the primary factor influencing sound pressure in the low-frequency band is the properties of the substrate, whereas in the high-frequency band, the material's heat capacity is the predominant major. However, there are still some defects in Hu's model. In 2011, Tian et al. prepared graphene as a thermoacoustic device by means of chemical vapor deposition (CVD) (Figure 9b) [88]. Then, they elucidated the relationship between the surface temperature of the sound-emitting layer and the applied energy. Their experimental results are not in line with the previous module. They found that in Hu's module, they omitted the 30 nm aluminum which functioned as the heat source. They assumed that the conductor was thin enough for this aspect and could be neglected. Owing to these findings, they modified their module as follows:

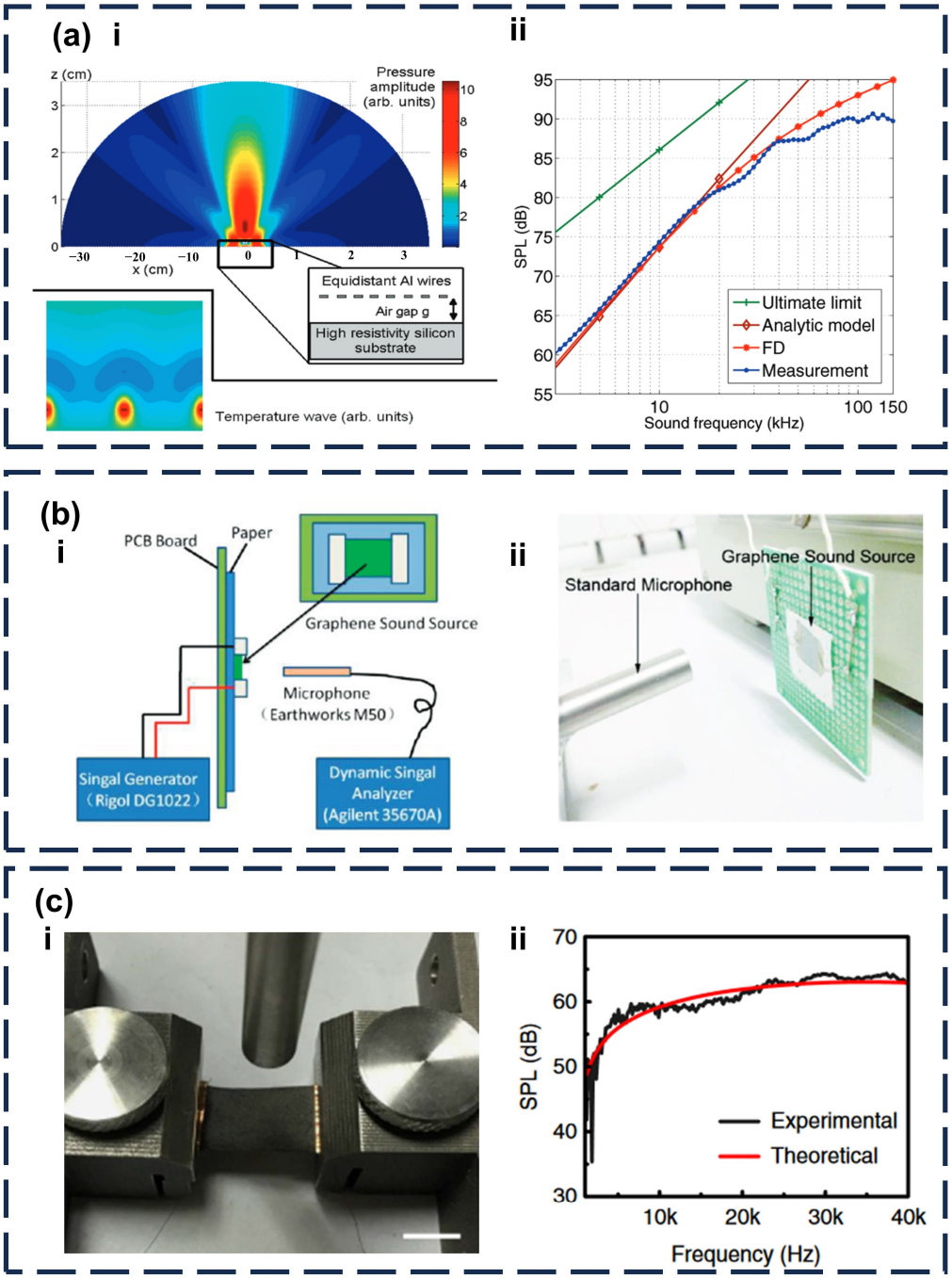

Figure 9. Development of the TA sound emitter. (**a-i**) An illustration of sound radiation from array of metal wires in modeling and experiments of TA sound emitters. (**a-ii**) comparisons between measurement and analytic model [87]. (**b-i**) Schematic diagram of test platform for graphene sound emitter. (**b-ii**) Onsite photo of the experimental setup. [88] (**c-i**) onsite photo of the experimental setup for graphene-based intelligent AT. (**c-ii**) the SPL versus the frequency showing that the model agrees well with experimental results [58].

For $f < \frac{a_s}{4\pi L_S^2}$ at low frequencies in far-field, the SP could be derived as

$$P_{rms} = \frac{R_0}{\sqrt{2r_0}} \times \frac{\gamma - 1}{v_g} \times \frac{e_g}{M(e_s + a_c) + e_g} \times q_0, \quad (6)$$

for $f > \frac{a_s}{4\pi L_S^2}$ at low frequencies in far-field, the SP could be derived as

$$P_{rms} = \frac{R_0}{\sqrt{2r_0}} \times \frac{\gamma - 1}{v_g} \times \frac{e_g}{(e_s + a_c) + e_g} \times q_0, \quad (7)$$

where f is the frequency of voice; α_s and L_s is the thermal diffusivity and thickness of the substrate, respectively; r_0 is the distance between the TA sound emitters and the microphone for the test; γ is the heat capacity ratio in gas; v_g is the velocity of voice in gas; e_i is the thermal effusivity of material which is determined by material; q_0 is the input power density; M is a frequency-related factor.

Xie et al. also proposed a new model [89], based on the energy conversation, which is easy to analyze and calcite. They module can be displayed as follows:

$$p_{rms} = \frac{m_{air} \cdot f \cdot (\dot{Q_{air}})}{2\sqrt{2} C_p T_0 r}, \quad (8)$$

where m_{air} is the molecular weight of air, f is the frequency of the acoustic, Q_{air} is the thermal energy diffused into the air, C_p is the heat capacity at constant pressure, T_0 is the room temperature and r is the measuring distance from the source. Tao et al. verified the correctness of the experiment (Figure 9c) [58].

3.2. TA Sound Emitter Made of Different Materials

A high-performance TA sound emitter needs to efficiently conduct heat into the air and convert it into sound. This imposes elevated requirements on the sound-generating material, which needs to have a very low specific HCPUA. In order to make high-performance TA sound emitter devices, three conditions should be satisfied. First, the conductor should be thin enough with a low HCPUA. Second, the conductor should be suspended to prevent thermal leakage from the substrate. Third, the conductor area should be large enough to build a sufficient sound field [90]. Various materials can be used in the construction of TA sound emitters, each with its own set of characteristics and applications. There are some common materials including graphene [90–92], MXene [11,93], CNT [94–96], metallic nanowires [97,98], and so on.

3.2.1. Graphene

Graphene is an emerging two-dimensional material with high electromobility, high flexibility, and low heat capacity. It is very suitable to be applied in TA sound emitters. Graphene-based TA devices combine the advantages of graphene and TA sound emitter, exhibiting unique and excellent performance. The sonic frequency required will change as the frequency of the excitation voltage is altered.

CVD is a common technique for graphene preparation. In 2012, J. Suk et al. prepared a graphene film with excellent light transmission by CVD and fabricated it into a TA sound emitter (Figure 10a) [91]. Then, they demonstrated the effect of different substrates and areas of substrates on the sound pressure through experimental studies. For the first time, they improved the influence of the sound pressure from the membrane material to the flexible substrate, such as PET. Meanwhile, they experimented with the TA sound emitter with different curvatures, which opened a new application of the TA sound emitter in flexible devices. Using the same fabrication technique, CVD, Tian et al. prepared monolayer graphene, which has a defect-free structure and excellent light transmission and can be controlled in terms of the number of layers [90]. The monolayer graphene was then fabricated into graphene headphones (Figure 10b). Then they tested its delay, flatness, and power linearity. Due to its ultra-high frequency response, TA sound emitter headphones have been utilized in animal studies as signal transmitters, facilitating the future exploration of animal communication.

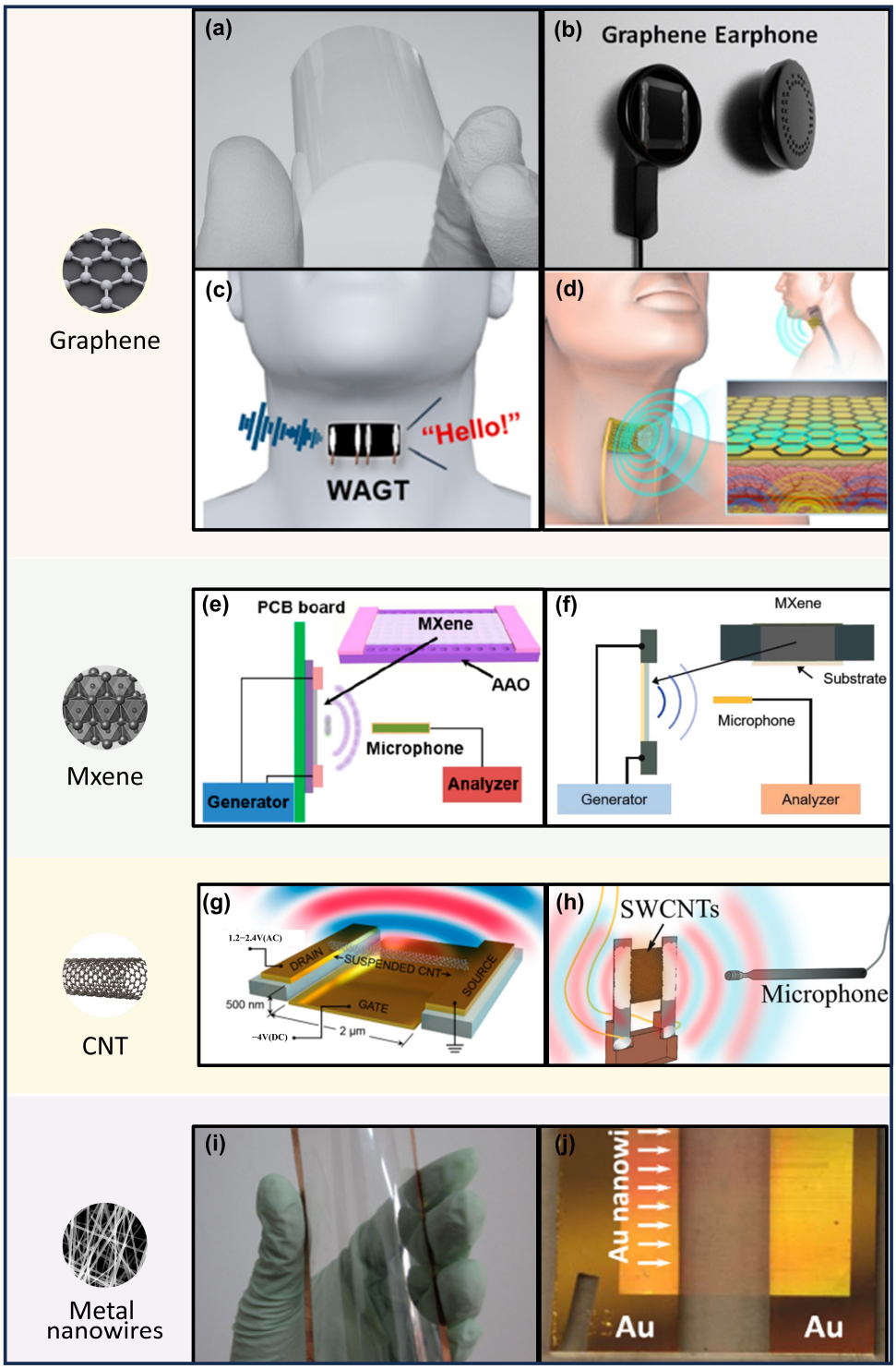

Figure 10. TA sound emitter made of different materials. (**a**) Monolayer graphene on PET as transparent and flexible sound emitters [91]. (**b**) Graphene earphone in a commercial earphone casing [90]. (**c**) Schematic of graphene sound emitter when attached to throat [80]. (**d**) Schematic diagram of the interaction paradigm of the intelligent artificial graphene throat [13]. (**e**) Schematic structure of MXene-based TA sound emitter [93]. (**f**) Schematic of the MXene-based TA sound measurement setup [20]. (**g**) Schematic diagram of suspended CNT-based TA sound emitter geometry [99]. (**h**) Schematic structure of SWCNTs-based TA sound emitter [16]. (**i**) Photograph of flexible and transparent silver nanowire-based sound emitter [100]. (**j**) Optical image of gold nanowire-based TA sound emitter [101].

The frequency of graphene TA sound emitters is linked to the voltage and current applied. M.S. Heath et al. proposed a graphene-based ultrasonic TA sound emitter by combining various frequencies of alternating current applied to a thermoacoustic device to generate sound waves of different frequencies. The TA device was then made into a field effect tube, and the bias voltage was controlled to switch the TA sound emitter on and off and adjust the volume of the TA sound emitter [92].

The graphene sound emitter exhibits outstanding electromobility and flexibility, enabling its attachment to a person's skin. In this capacity, it serves as a sound emitter in AT. In 2019, Wei et al. proposed a wearable skinlike ultrasensitive artificial graphene, which can serve as a sound emitter and can be directly attached to the larynx of the aphasic person (Figure 10c) [80]. In 2023, Yang et al. also fabricated an AT within a graphene sound emitter (Figure 10d) [13].

3.2.2. MXene

MXene is 2D transition metal carbides or carbonitrides with the composition $M_{n+1}X_nT_x$, where M is a transition metal; X is carbon or nitrogen; T represents surface functional groups such as -OH, =O, and -F; and n is an integer from one to four [85,102–105]. In particular, the abundant surface functional groups on MXene enable strong adhesion to various substrates. This capability allows the fabrication of mechanically stable flexible TA sound emitters, ensuring resistance to delamination from substrates during mechanical deformations [20].

In comparison to graphene, the MXene-based sound emitter device has a higher SP than that of graphene with the same thickness. Gou et al. fabricated MXene-based TA sound emitters using anodic aluminum oxide (AAO) and polyimide (PI) substrates (Figure 10e) [93]. These Ti_3C_2 MXene exhibits a higher SPL of 68.2 dB (f = 15 kHz) and displays a very stable sound output spectrum when the frequency varies from 100 Hz to 20 kHz.

The property of TA sound emitters based on MXene is stable. In a study conducted in 2023, Kim et al. successfully fabricated an ultrathin MXene-based TA sound emitter exhibiting consistent sound performance for 14 days (Figure 10f) [20]. Moreover, these sound emitters exhibit deformability in various configurations such as bent, twisted, cylindrical, and stretched-kirigami. They can be manipulated into diverse 2D and 3D shapes under different mechanical deformations.

3.2.3. CNT

CNTs are cylindrical structures composed of carbon atoms with extraordinary electrical and mechanical properties. CNTs exist in various forms, including SWCNTs and multi-walled carbon nanotubes (MWCNTs), depending on the number of layers of carbon atoms [106]. CNTs have low HCPUA and high surface area per unit volume, which helps to generate high-level TA sound. In addition, the aerogel structure of CNT films facilitates the permeation of gas molecules, boosting its efficiency remarkably in sound emitting [16].

In 2015, Mason et al. observed the thermoacoustic transduction process at the single-molecule level, as illustrated in Figure 10g [99]. Leveraging this minimal length scale, they tested the assumptions made in previous models used to describe 2D thermoacoustic films. Additionally, they sought to establish correlations between the thermoacoustic efficiencies of these nanotube devices and their electrical impedance, aiming to gain insights into underlying loss mechanisms.

Similar to the previously discussed graphene TA sound emitters, when the HCPUA of CNT films is so low, the CNT TA sound emitters can also achieve very high SP. Romanov et al. fabricated TA sound emitters made of thin and freestanding films of randomly oriented SWCNTs (Figure 10h) [16] with a small HCPUA, the maximum frequency of the emitting sound can reach as high as 100 kHz.

3.2.4. Metallic Nanowires

Metallic nanowires are extremely thin wires with diameters on the nanoscale, with many unique behaviors that have not been seen in bulk materials. [100,107–110]. Ag nanowires (AgNWs) are one kind of metallic nanowires, and there is high conductivity and transmittance in random networks. Utilizing this property, Tian et al. fabricated flexible, ultrathin, and transparent sound-emitting devices with a low driving voltage, as illustrated in Figure 10i [100]. However, the presence of nanowire–nanowire junctions within these devices poses challenges in precisely defining their lateral dimensions. In contrast to AgNWs, AuNWs exhibit distinct properties. They can be precisely defined lateral dimensions. Consequently, AuNWs allow for experimental performance comparison with theoretical predictions. By employing AuNWs, Dutta et al. prepared TA sound emitters consisting of arrays (Figure 10j) [101]. Their results fit with the classical theory proposed by Vesterinen et al. [87].

Due to the high intrinsic electrical conductivity of copper, copper nanowires (CuNWs) also represent a promising future in the TA sound emitter. Bobinger et al. fabricated TA sound emitters utilizing CuNWs [110], featuring an exceptional HCPUA of 1.9×10^{-2} J/(m^{2}K), rendering them well-suitable for applications of TA sound emitters.

In summary, TA sound emitters have been invented and discussed since the early 20th century. Since then, these emitters have undergone significant evolution and refinement, paralleling the continuous advancements in material preparation techniques. In the process, materials have evolved from nanoscale aluminum layers to carbon nanotubes, and finally to graphene, which is now the dominant material. Furthermore, applications have also transitioned from the simplicity of basic TA sound emitters to their integration and expanded use, including sophisticated roles such as serving as sound emitters in intelligent AT.

4. Post-Processing and Recognition Algorithm

Based on sound detect devices mentioned in Section 2, vibration signals and other physiological signals can be collected directly. Semantic analysis of these signals is the ultimate purpose of AT, as these signals contain rich and crucial infseormation for communication [111–114]. The simplest way for recognition and distinction is directly observing the electrical output wave or capturing the wave with a microcontroller in the time domain [80,115,116]. Nevertheless, some throat vibration signals with similar pronunciations can be challenging to distinguish in the time domain. To accurately analyze semantic information, machine learning is an appropriate solution [72,117–119].

Depending on whether the input data is labeled, the machine learning algorithm can mainly be devised into supervised learning, where the input data is labeled, and unsupervised learning, where the input data is not labeled. The training set for the semantic recognition needs to be labeled, so most of the machine learning algorithm utilized is supervised learning algorithms, such as neutral network [120–122], support vector machine (SVM) [123], Bayes classification [124], and so on.

4.1. SVM

SVM is a supervised machine learning algorithm used for classification and regression tasks. SVM classifies data by constructing hyperplanes in a high-dimensional space. It represents samples as points, maximizing the gap between distinct categories. SVM adapts to complex patterns using kernel functions, making it suitable for diverse applications like image recognition and text classification. Moreover, due to its ability to identify decisive support vectors and eliminate numerous redundant training samples, SVM is a helpful tool for avoiding the "dimension disaster". Fang et al. fabricated a PVDF flexible piezoelectric sensor to collect the throat vibration signals, utilizing the SVM to recognize and process the signals [114]. During the machine learning process, the number of training sets and test sets is very important which determines the accuracy and training cost. In this work, they discovered that when the number of training set samples and test set samples was 50 and 100, a very small sample, the training sets and test sets can represent the best performance.

As for the hyper parameter, they adopted a heuristic method Grid Search-Support Vector Machine (GSSVM), which finds the appropriate hyper parameters through a grid search with a specified range and step size. As depicted in Figure 11a,b, the 3D viewer illustrates that when the penalty factor 'c' was set to 22.6274, the recognition accuracy reached its optimum level. The result of accuracy for speaker recognition and semantic recognition can reach as high as 95.97% and 97.5%, respectively.

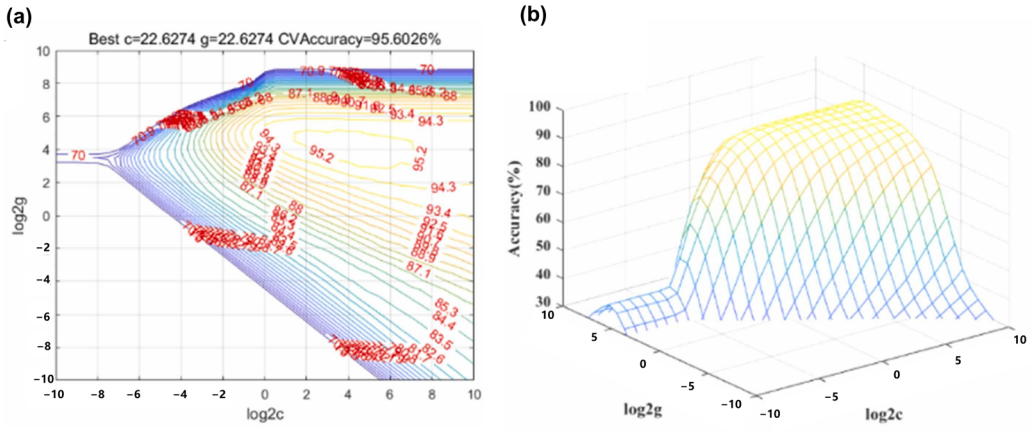

Figure 11. Post-processing and recognition of the detected signals. (**a**,**b**) 3D view and the contour view of SVM parameter selection [114].

4.2. Neural Network

Neural network is a computational model, inspired by the structure and function of the human brain, particularly the work principle of neurons. A convolutional neural network (CNN) is a specialized type of neural network that learns feature engineering by itself. The CNN with deep structures is adept at uncovering concealed intrinsic connections within the data and extracting abstract features effectively. The structure of the CNN consists of an input layer, convolutional layers, pooling layers, fully connected layers, and output layers. As a fundamental building block, the convolutional layers apply filters or kernels to extract the features in the input data. Moreover, the pooling layers follow the convolutional layers and are used to downsample the spatial dimensions of the input and reduce computation. In the end, the fully connected layers connect every neuron from the previous layers to the current layers and lead to the output layers that produce the final predictions. Jin et al. developed an MXene-based AT and harnessed CNN to accomplish the categorization task of distinguishing between long and short vowels (Figure 12) [12]. Owing to the deep structure of the neural network, a large amount of data is required. In this work, a total of 1500 data was adopted, including 750 long vowels and 750 short vowels. Among them, 1050 data were randomly selected as the training data set, and the others were used as the testing data set. After about 200 epochs of training, the result of accuracy for long vowels and short vowels reached 83.6% and 88.9% respectively.

4.3. Relief

Relief is a feature selection algorithm employed in machine learning and data mining. Especially beneficial when dealing with datasets containing numerous features, Relief aims to recognize the most crucial features for a predictive model. As an algorithm capable of identifying the most relevant features, Relief can collaborate with other feature extraction algorithms such as CNN and others, to identify valuable features, and reduce data dimensions. Yang et al. utilizing an integrated machine learning model proposed a graphene-based intelligent wearable AT for speech recognition and interaction [13]. In this work, they take advantage of the groundbreaking architectures within the realm of CNN, AlexNet for feature extraction, and introduce an improved AlexNet model. Furthermore, they choose Relief for feature selection and SVM as a classifier. As shown in Figure 13, the improved AlexNet extracts 10 features through the five convolution layers and other

layers, then the Relief sorts the most important features for the SVM to classify. Compared with other models, including single AlexNet and another ensemble model (improved AlexNet + SVM), the model composed of improved AlexNet, Relief, and SVM attains significant enhancement in classification and time cost (Figure 14a). The result of accuracy can reach more than 90% in the task of recognizing daily words.

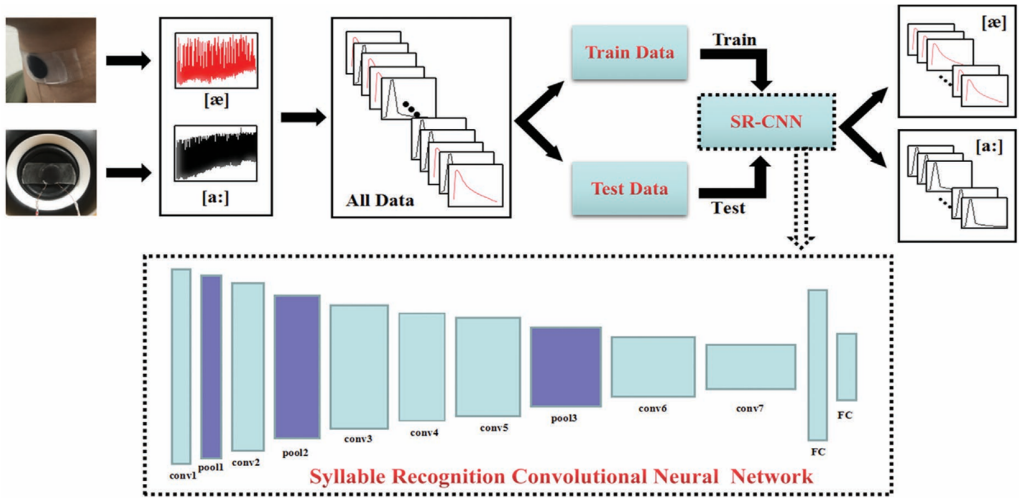

Figure 12. Experimental flow chart and the structure of overall classification by SR-CNN [12]. The SR-CNN is composed of seven convolution layers, three pooling layers, and two fully connected layers.

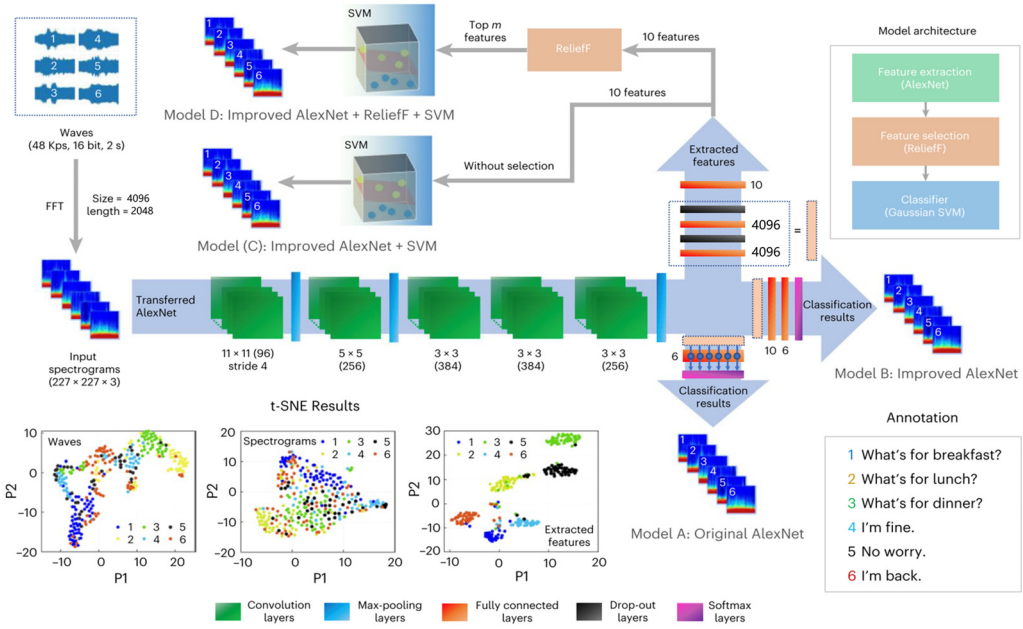

Figure 13. Structure of different integrated models [13]. Model A is the original AlexNet, model B is the improved model, model C is a combination model of two artificial algorithms, improved AlexNet and SVM, and model D is a combination of three artificial algorithms, improved AlexNet, Relief, and SVM.

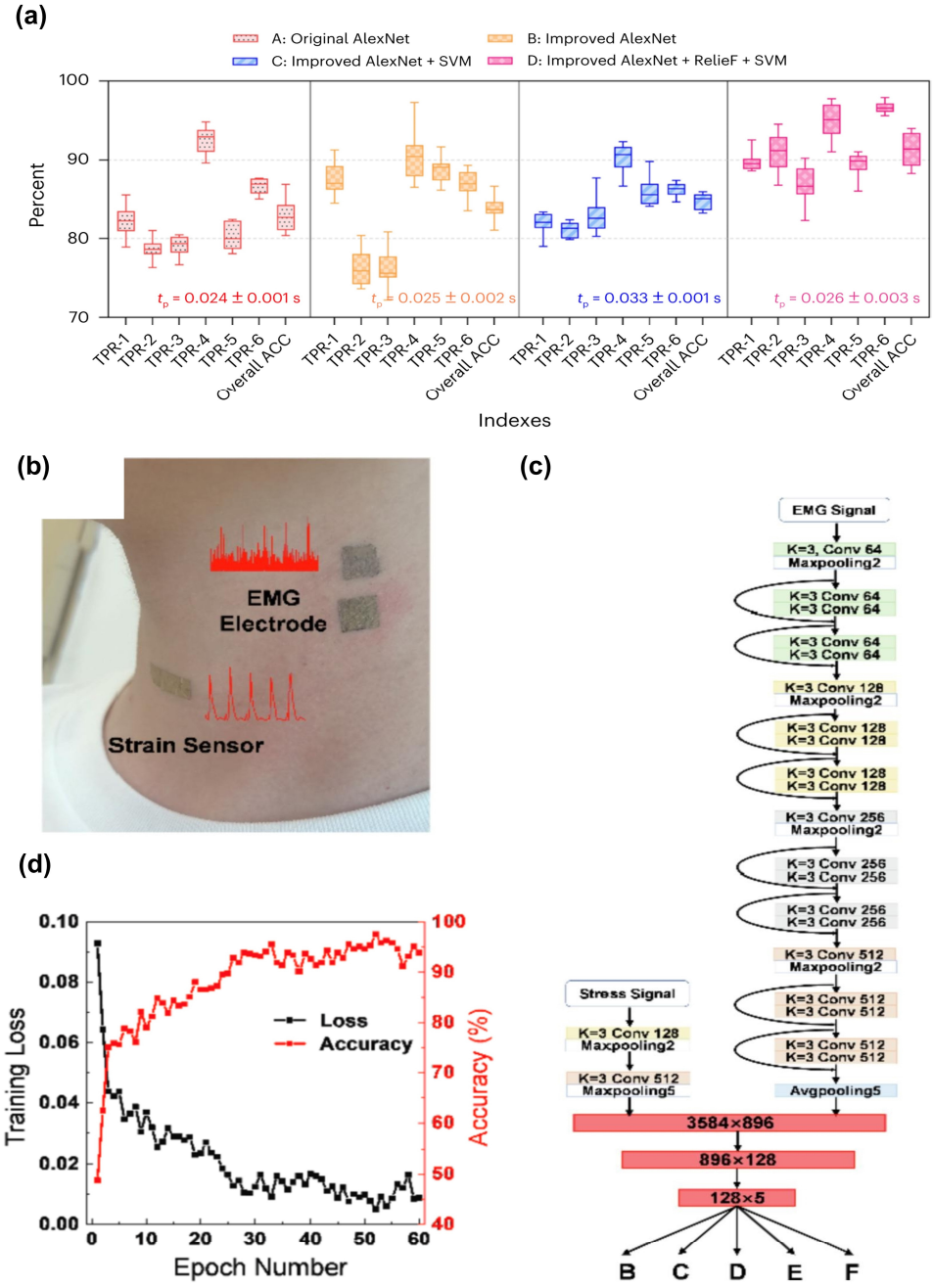

Figure 14. Post-processing and recognition of the detected signals. (**a**) Comparison of the improved AlexNet model with the original AlexNet. ACC, accuracy; tp, time for prediction; TPR, true positive rate [13]. (**b**) The illustration of Au/PU nanomesh strain sensor and Au nanomesh EMG electrodes. (**c**) The SCNN algorithm consists of ResNet18 for the EMG signal and two-layer CNN for the stress signal. (**d**) The training loss and classification accuracy for the SCNN model [15].

In addition to employing machine learning algorithms, mixed-modality is also harnessed in the signals acquisition and process, a mixed-modality signals can capture different aspects of information, resulting in enhanced accuracy and performance compared to using a single modality in isolation [76]. Qiao et al. applied the Au nanomesh as the physiological electrodes to detect EMG signal, while leveraging the Au/PU nanomesh as the strain sensor in the throat (Figure 14b). Furthermore, they introduced a synergetic CNN algorithm (Figure 14c) consisting of a modified CNN to analyze the EMG signals and a two-layer CNN to analyze the stress-strain, aiming at distinguishing voice signals. The result of the accuracy can reach as high as 98.9% (Figure 14d) [15].

In summary, the speech recognition function serves as the bridge connecting the sound sensor component of AT to the sound emitter component. Nowadays, the advancement of machine learning algorithms, including CNN, AlexNet, and other artificial algorithms, has significantly improved recognition accuracy and expanded the language corpus. This expansion has broadened the application landscape of AT.

5. AT Serving as a Sound Sensor and Emitter

This paper has introduced the three components of AT in early sections, namely, the sound-sensing part, the sound-emitting part, and the speech recognition part. However, it should be noted that the AT is not comprised of a single component; rather, it is a combination of the three parts.

Wei et al. developed a device that integrates both sound sensing and sound emission capabilities with speech recognition functions. They devised a system for sound sensing utilizing a custom-made circuit board (Figure 15a,b) and performed feature extraction in the time domain based on changes in resistance (Figure 15c), then they connected the AT to the microcontroller which transforms the changes of the resistances into different voltages. Variances in voltage will result in different sounds in the emitter section of the AT. Consequently, if the tester executes strong movements, there will be a significant voltage variation, causing the sound emitter to say "OK". Conversely, if the tester's movements are weak, the sound emitter will state "NO" [80].

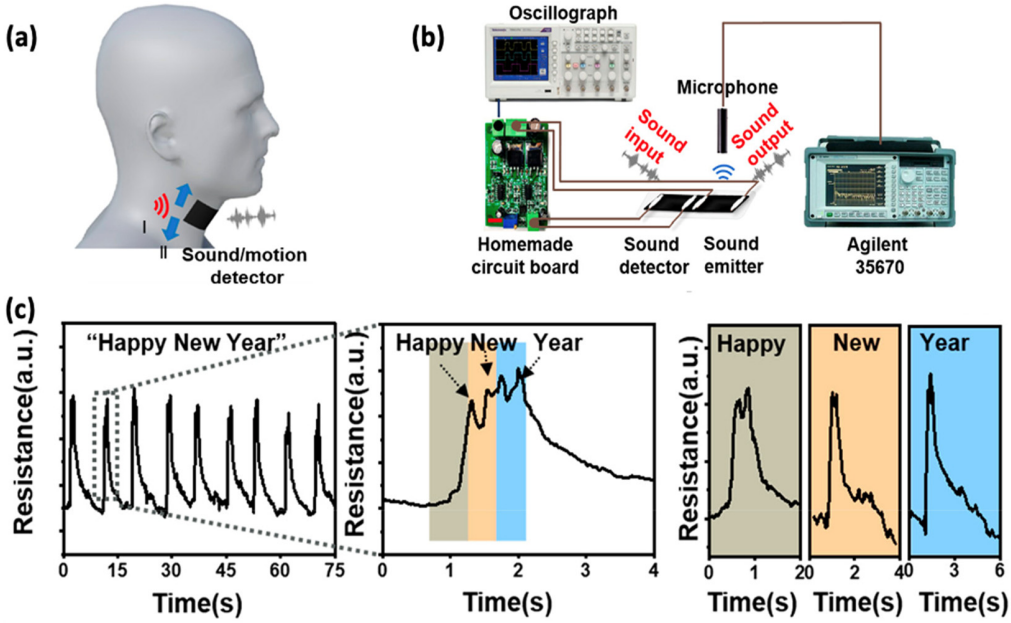

Figure 15. AT is a combination of sound detection, emission, and recognition. (**a**) The AT can serve as a sound and motion sensor. (**b**) The sound detection system. The sound detection device is connected to the circuit board and displays resistance. (**c**) The resistance response to the sound "Happy New Year" [80].

Tao et al. utilizing the microcontroller also developed a device comprising a sound receiver and sound emitter with a voice recognition function in the time domain. Figure 16a shows the workflows of the recognition process. The microcontroller will initially detect the amplitude and the duration of the voice by capturing the resistance of the graphene AT until either the amplitude or time reaches the thresholds. Afterward, the digital function generator will be applied to the graphene AT for 3 s. In their device, different amplitude and last time will lead to the activation of different digital function generators, producing varied volumes and frequencies [58].

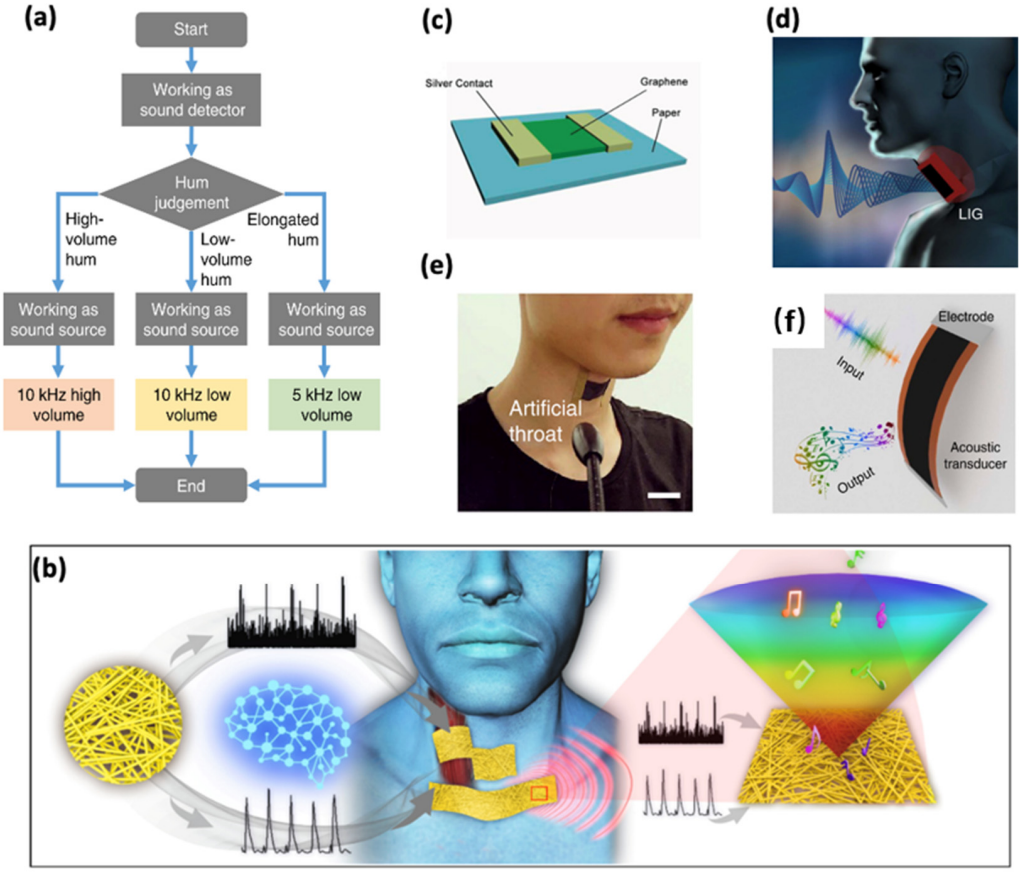

Figure 16. The AT can serve as a sound sensor and emitter with a speech recognition function. (**a**) The working procedure of the artificial throat [58]. (**b**) The composition of the AT based on Au/PVA and Au/PU nanomesh [15]. (**c**) Schematic view of a sound emitter using graphene as the emission component [88]. (**d**) The AT can detect the movement of the throat and emit sound. (**e**) The tester wearing the graphene AT. Scale bar, 1 cm [58]. (**f**) The AT serves as the sound emitter and sound sensor simultaneously [58].

Qiao et al. further proposed AT combined with sound sensors, sound emitters, and speech recognition, as previously mentioned [15]. Utilizing machine learning algorithms for speech recognition, individual English letters, such as 'B', 'C', 'D', 'E', and 'F', can be discerned through the combination of EMG and strain sensor, Au/PVA nanomesh, in the neck muscle. Following the classification by the algorithms, the sound sensor is then repurposed as a sound emitter, producing the corresponding letter at an intensity of 78 dB.

With larger datasets and more sophisticated algorithms, as introduced in the paper previously, Yang et al. proposed an enhanced and more intelligent AT [13]. Their innovation extends the scope from identifying individual letters to recognizing complete sentences. Common everyday language sentences like 'I'm back', 'I'm fine', 'What's for breakfast', 'What's for lunch', and 'What's for dinner' can be accurately identified, achieving a high correct rate in patients with a laryngectomy. Following the classification of these sentences by artificial intelligence algorithms, the AT can speak the corresponding sounds at an approximate intensity of 60 dB. Additionally, the device exhibits robust performance, effectively recognizing sentences in subtle sounds or noisy environments. Table 1 compares various devices with sound emitting and detecting functions.

Table 1. The device with sound emitting and sound detecting functions.

Material	Substrate	Principle of Sound Emitting	Signal of Sound Detecting	Algorithm	Accuracy	Reference
Graphene	PI	TA effect	Vibration	None	None	[58]
Graphene	PI	TA effect	Vibration	None	None	[80]
Au and Au/PU	PU	TA effect	Vibration and EMG	Synergetic GNN	98.9%	[15]
Graphene	PI	TA effect	Vibration and EMG	CNN	>88.14%	[13]
MXene	Parylene and so on	TA effect	Unable to detect	None	None	[20]
MXene	PDMS	Unable to emit	Vibration	CR-CNN	>70%	[12]
CNT	None	TA effect	Unable to detect	None	None	[16]
CNT	Zn	Unable to emit	Vibration	None	None	[125]

In conclusion, AT is a combination of the sound sensor and sound emitter with a speech recognition function. Each part performs a specific function in speech recognition and sound emitting. In terms of sound sensors, a variety of materials, such as graphene and Au/Pu nanomesh, have been widely used in sound harvesting, which can effectively capture the physiological signals and vibration signals of the human body (Figure 16b) [15]. Then, artificial intelligence algorithms such as SVM, CNN, and Relief are applied to infer physiological signals from the voice signals, recognizing distinctive features. Following this, by utilizing thermoacoustic materials such as graphene film (Figure 16c) [88], the collected information will be output as audio signals. The AT commonly used is often a device that combines all of the three functions in one unit (Figure 16d–f) [58].

6. Challenge and Prospect

The recent research developments of intelligent flexible AT with sound emitting, detecting, and recognizing abilities are demonstrated above. However, many challenges still need to be overcome.

In the post-process and artificial algorithms, the current artificial algorithms used for AT are less general, and smaller in datasets with a restricted capability to recognize content. On the one hand, the existing databases are still small, much less than the size of databases in image recognition such as the ImageNet database. In the ImageNet database, there are millions of annotated images, totaling approximately 14 million images with 20 thousand various kinds of objects captured from various angles, perspectives, and environments. However, most of the databases for AT have been built only by the researchers themselves, and the databases are limited in terms of the daily language they can cover, making it difficult to cover other aspects of life. In Yang's model [13], the dataset size is less than ten thousand, with fewer than ten categories for each classification task. In Jin's model [12], there are only 1900 elements in their dataset with only two categories in the classification task. Consequently, the existing trained models are less general, and limited to some basic phrases and sentences. On the other hand, the researchers lack the willingness to upload the self-built databases. In consequence, the existing public database for AT is extremely rare, lacking a substantial foundation for training large-scale AT speech recognition models.

In terms of hardware, the flexibility of the current AT circuit is still deficient. In the existing soft wearable instruments, the circuit is always flexible and combined with the sensors. In 2023, Yoo et al., in Rogers' group, proposed a wireless sensing system for physiological monitoring that integrated the circuit and sensors into a single combination [126]. Similarly, in 2023, Shinjae Kwon et al. also developed a sleep monitoring system that combined the circuits and sensors into a unified assembly [127]. However, current AT often requires pairing with external power supplies, microcontrollers, etc., which hampers their portability and impedes their widespread adoption and practical use. The cooperation

relationship between sensors, circuits, and microcontrollers also should be optimized. Furthermore, to enhance stability and achieve a higher signal-to-noise ratio, it is necessary to incorporate facilities and devices designed for shielding against external interference. The ideal solution is a system where the sensors, circuits, microprocessors, etc. are flexible and as small as possible.

In addition to the two aspects mentioned earlier, the existing AT also falls short of meeting the requirements of portable medical products. The current trend in the development of portable medical products is toward multifunctionality, catering to a wide range of operating scenarios. The pursuit of multifunctionality not only enhances efficiency but also reduces costs and resource consumption, expanding the scope of applications. However, the functions of the existing AT are relatively homogeneous and require further improvement to align with the evolving direction of medical device development. In the near future, it is crucial to expand the application scope of the AT. Firstly, we can leverage the AT's capability to acquire EMG signals for sleep monitoring to diagnose certain diseases. Then, we can enhance our algorithms to enable language translation functionalities, allowing the AT to assist individuals with communication barriers due to language differences, not just limited to these people with larynx diseases.

Additionally, there is another issue that requires attention. There are relatively limited experimental studies on the AT in clinical applications, which leads to a lack of experimental data to support the clinical safety of the AT. In Yang's study [13], patients wore the AT for a short duration during testing. However, the effects of prolonged wear and environmental factors such as temperature and humidity on the functionality of the AT remain unclear. Additionally, the study only involved one patient as a tester, which may limit the generalizability of the findings. More participants are needed for a more comprehensive and reliable assessment. In the studies conducted by Jin [12], the evaluation of AT was limited to healthy individuals, lacking data from patients and clinical practice settings. Although the intelligent AT technology theoretically has the potential in medical applications, more experimental data is needed to assess its feasibility and safety in practical medical settings.

In conclusion, a comprehensive review of the AT has been demonstrated which consists of the detection, emitting sound, and algorithms for speech recognition. The sensor for detecting sound can be divided into capacitive, piezoelectric, electromagnetic, and piezoresistive. Then some devices for emitting sound, including the TA effect are discussed. The algorithms utilized by AT for speech recognition also has been analyzed carefully. Finally, we state the challenge and outlook of this AT. Compared with the conventional electrolarynx, the AT with flexible material and adhesive to the skin very well is more portable and easier to use for mute people and is superior in other aspects.

Author Contributions: Investigation, J.F., Z.D., C.L. (Chuting Liu), L.S., X.L. and M.P.; Resources, S.P. and H.L.; Writing—Original Draft Preparation, J.F. and Z.D.; Writing—Review & Editing, Y.Q., C.L. (Chang Liu), J.L. and J.W.; Visualization, J.F. and Z.D.; Supervision, Y.Q. and J.Z.; Project Administration, Y.Q. and J.Z.; Funding Acquisition, Y.Q. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Natural Science Foundation of China (No. 62201624; 32000939; 21775168; 22174167; 51861145202; U20A20168), Shenzhen Science and Technology Program (RCBS20221008093310024), Shenzhen Research Funding Program (JCYJ20190807160401657; JCYJ201908073000608), the Open Research Fund Program of Beijing National Research Center for Information Science and Technology (BR2023KF02010). The authors are also thankful for the support from Key Laboratory of Sensing Technology and Biomedical Instruments of Guangdong Province (No. 2020B1212060077).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siegel, R.L.; Miller, K.D.; Wagle, N.S.; Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 17–48. [CrossRef] [PubMed]
2. Ferlay, J.; Colombet, M.; Soerjomataram, I.; Mathers, C.; Parkin, D.M.; Piñeros, M.; Znaor, A.; Bray, F. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* **2019**, *144*, 1941–1953. [CrossRef]
3. Siegel, R.L.; Miller, K.D.; Goding Sauer, A.; Fedewa, S.A.; Butterly, L.F.; Anderson, J.C.; Cercek, A.; Smith, R.A.; Jemal, A. Colorectal cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 145–164. [CrossRef]
4. da Silva, A.P.; Feliciano, T.; Freitas, S.V.; Esteves, S.; e Sousa, C.A. Quality of life in patients submitted to total laryngectomy. *J. Voice Off. J. Voice Found.* **2015**, *29*, 382–388. [CrossRef]
5. Tang, C.G.; Sinclair, C.F. Voice restoration after total laryngectomy. *Otolaryngol. Clin. N. Am.* **2015**, *48*, 687–702. [CrossRef]
6. Liu, H.; Ng, M.L. Electrolarynx in voice rehabilitation. *Auris Nasus Larynx* **2007**, *34*, 327–332. [CrossRef]
7. Barney, H.; Haworth, F.; Dunn, H. An experimental transistorized artificial larynx. *Bell Syst. Tech. J.* **1959**, *38*, 1337–1356. [CrossRef]
8. Kaye, R.; Tang, C.G.; Sinclair, C.F. The electrolarynx: Voice restoration after total laryngectomy. *Med. Devices Evid. Res.* **2017**, *10*, 133–140. [CrossRef]
9. Isshiki, N.; Tanabe, M. Acoustic and aerodynamic study of a superior electrolarynx speaker. *Folia Phoniatr. Logop.* **1972**, *24*, 65–76. [CrossRef]
10. Wu, L.; Wan, C.; Wang, S.; Wan, M. Improvement of Electrolaryngeal Speech Quality Using a Supraglottal Voice Source With Compensation of Vocal Tract Characteristics. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 1965–1974.
11. Qiao, Y.; Gou, G.; Wu, F.; Jian, J.; Li, X.; Hirtz, T.; Zhao, Y.; Zhi, Y.; Wang, F.; Tian, H.; et al. Graphene-Based Thermoacoustic Sound Source. *ACS Nano* **2020**, *14*, 3779–3804. [CrossRef]
12. Jin, Y.; Wen, B.; Gu, Z.; Jiang, X.; Shu, X.; Zeng, Z.; Zhang, Y.; Guo, Z.; Chen, Y.; Zheng, T.; et al. Deep-Learning-Enabled MXene-Based Artificial Throat: Toward Sound Detection and Speech Recognition. *Adv. Mater. Technol.* **2020**, *5*, 2000262. [CrossRef]
13. Yang, Q.; Jin, W.; Zhang, Q.; Wei, Y.; Guo, Z.; Li, X.; Yang, Y.; Luo, Q.; Tian, H.; Ren, T.-L. Mixed-modality speech recognition and interaction using a wearable artificial throat. *Nat. Mach. Intell.* **2023**, *5*, 169–180. [CrossRef]
14. Abd Almisreb, A.; Jamil, N.; Din, N.M. Utilizing AlexNet deep transfer learning for ear recognition. In Proceedings of the 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), Kota Kinabalu, Malaysia, 26–28 March 2018; IEEE: New York, NY, USA, 2018; pp. 1–5.
15. Qiao, Y.; Gou, G.; Shuai, H.; Han, F.; Liu, H.; Tang, H.; Li, X.; Jian, J.; Wei, Y.; Li, Y.; et al. Electromyogram-strain synergetic intelligent artificial throat. *Chem. Eng. J.* **2022**, *449*, 137741. [CrossRef]
16. Romanov, S.A.; Aliev, A.E.; Fine, B.V.; Anisimov, A.S.; Nasibulin, A.G. Highly efficient thermophones based on freestanding single-walled carbon nanotube films. *Nanoscale Horiz.* **2019**, *4*, 1158–1163. [CrossRef]
17. Lang, C.; Fang, J.; Shao, H.; Ding, X.; Lin, T. High-sensitivity acoustic sensors from nanofibre webs. *Nat. Commun.* **2016**, *7*, 11108. [CrossRef]
18. Zhao, Y.; Gao, S.; Zhang, X.; Huo, W.; Xu, H.; Chen, C.; Li, J.; Xu, K.; Huang, X. Fully flexible electromagnetic vibration sensors with annular field confinement origami magnetic membranes. *Adv. Funct. Mater.* **2020**, *30*, 2001553. [CrossRef]
19. Ma, Y.; Yue, Y.; Zhang, H.; Cheng, F.; Zhao, W.; Rao, J.; Luo, S.; Wang, J.; Jiang, X.; Liu, Z.; et al. 3D Synergistical MXene/Reduced Graphene Oxide Aerogel for a Piezoresistive Sensor. *ACS Nano* **2018**, *12*, 3209–3216. [CrossRef]
20. Kim, J.; Jung, G.; Jung, S.; Bae, M.H.; Yeom, J.; Park, J.; Lee, Y.; Kim, Y.R.; Kang, D.h.; Oh, J.H.; et al. Shape-Configurable MXene-Based Thermoacoustic Loudspeakers with Tunable Sound Directivity. *Adv. Mater.* **2023**, *35*, 2306637. [CrossRef]
21. Sujatha, C. Fundamentals of Acoustics. In *Vibration, Acoustics and Strain Measurement: Theory and Experiments*; Sujatha, C., Ed.; Springer International Publishing: Cham, Switzerland, 2023; pp. 161–217.
22. Beranek, L.; Mellow, T. *Acoustics: Sound Fields, Transducers and Vibration*; Academic Press: Cambridge, MA, USA, 2019.
23. Peters, R. *Acoustics and Noise Control*; Routledge: Abingdon, UK, 2013.
24. Schmitz, T.L.; Smith, K.S. Two degree of freedom forced vibration. In *Mechanical Vibrations: Modeling and Measurement*; Springer: New York, NY, USA, 2012; pp. 167–198.
25. Tohyama, M. *Sound in the Time Domain*; Springer: Singapore, 2018.
26. Sivian, L.; White, S. On minimum audible sound fields. *J. Acoust. Soc. Am.* **1933**, *4*, 288–321. [CrossRef]
27. Lee, C.-Y.; Lee, G.-B. Humidity sensors: A review. *Sens. Lett.* **2005**, *3*, 1–15. [CrossRef]
28. Hong-Tao, S.; Ming-Tang, W.; Ping, L.; Xi, Y. Porosity control of humidity-sensitive ceramics and theoretical model of humidity-sensitive characteristics. *Sens. Actuators* **1989**, *19*, 61–70. [CrossRef]
29. Reddy, A.; Narakathu, B.; Atashbar, M.; Rebros, M.; Rebrosova, E.; Joyce, M. Fully printed flexible humidity sensor. *Procedia Eng.* **2011**, *25*, 120–123. [CrossRef]
30. Miles, R.N. A compliant capacitive sensor for acoustics: Avoiding electrostatic forces at high bias voltages. *IEEE Sens. J.* **2018**, *18*, 5691–5698. [CrossRef]
31. Zawawi, S.A.; Hamzah, A.A.; Majlis, B.Y.; Mohd-Yasin, F. A review of MEMS capacitive microphones. *Micromachines* **2020**, *11*, 484. [CrossRef]
32. Jung, Y.H.; Hong, S.K.; Wang, H.S.; Han, J.H.; Pham, T.X.; Park, H.; Kim, J.; Kang, S.; Yoo, C.D.; Lee, K.J. Flexible piezoelectric acoustic sensors and machine learning for speech processing. *Adv. Mater.* **2020**, *32*, 1904020. [CrossRef]

33. Lee, S.; Kim, J.; Yun, I.; Bae, G.Y.; Kim, D.; Park, S.; Yi, I.-M.; Moon, W.; Chung, Y.; Cho, K. An ultrathin conformable vibration-responsive electronic skin for quantitative vocal recognition. *Nat. Commun.* **2019**, *10*, 2468. [CrossRef]
34. Broadhurst, M.; Davis, G. Physical basis for piezoelectricity in PVDF. *Ferroelectrics* **1984**, *60*, 3–13. [CrossRef]
35. Cauda, V.; Stassi, S.; Bejtka, K.; Canavese, G. Nanoconfinement: An effective way to enhance PVDF piezoelectric properties. *ACS Appl. Mater. Interfaces* **2013**, *5*, 6430–6437. [CrossRef]
36. Wang, Y.; Zheng, J.; Ren, G.; Zhang, P.; Xu, C. A flexible piezoelectric force sensor based on PVDF fabrics. *Smart Mater. Struct.* **2011**, *20*, 045009. [CrossRef]
37. Lim, J.; Kim, H.S. Effects of SWCNT/PVDF composite web behavior on acoustic piezoelectric property. *Sens. Actuators A Phys.* **2021**, *330*, 112840. [CrossRef]
38. Kang, M.-G.; Jung, W.-S.; Kang, C.-Y.; Yoon, S.-J. Recent Progress on PZT Based Piezoelectric Energy Harvesting Technologies. *Actuators* **2016**, *5*, 5. [CrossRef]
39. Jain, A.; KJ, P.; Sharma, A.K.; Jain, A.; PN, R. Dielectric and piezoelectric properties of PVDF/PZT composites: A review. *Polym. Eng. Sci.* **2015**, *55*, 1589–1616. [CrossRef]
40. Venkatragavaraj, E.; Satish, B.; Vinod, P.; Vijaya, M. Piezoelectric properties of ferroelectric PZT-polymer composites. *J. Phys. D Appl. Phys.* **2001**, *34*, 487. [CrossRef]
41. Le, A.T.; Ahmadipour, M.; Pung, S.-Y. A review on ZnO-based piezoelectric nanogenerators: Synthesis, characterization techniques, performance enhancement and applications. *J. Alloys Compd.* **2020**, *844*, 156172. [CrossRef]
42. Gullapalli, H.; Vemuru, V.S.; Kumar, A.; Botello-Mendez, A.; Vajtai, R.; Terrones, M.; Nagarajaiah, S.; Ajayan, P.M. Flexible piezoelectric ZnO–paper nanocomposite strain sensor. *Small* **2010**, *6*, 1641–1646. [CrossRef]
43. Zhang, C.; Wang, X.; Chen, W.; Yang, J. An analysis of the extension of a ZnO piezoelectric semiconductor nanofiber under an axial force. *Smart Mater. Struct.* **2017**, *26*, 025030. [CrossRef]
44. Wang, Z.L. Nanogenerators, self-powered systems, blue energy, piezotronics and piezo-phototronics–a recall on the original thoughts for coining these fields. *Nano Energy* **2018**, *54*, 477–483. [CrossRef]
45. Qi, S.; Oudich, M.; Li, Y.; Assouar, B. Acoustic energy harvesting based on a planar acoustic metamaterial. *Appl. Phys. Lett.* **2016**, *108*, 263501. [CrossRef]
46. Lin, L.; Jing, Q.; Zhang, Y.; Hu, Y.; Wang, S.; Bando, Y.; Han, R.P.; Wang, Z.L. An elastic-spring-substrated nanogenerator as an active sensor for self-powered balance. *Energy Environ. Sci.* **2013**, *6*, 1164–1169. [CrossRef]
47. Cui, N.; Gu, L.; Liu, J.; Bai, S.; Qiu, J.; Fu, J.; Kou, X.; Liu, H.; Qin, Y.; Wang, Z.L. High performance sound driven triboelectric nanogenerator for harvesting noise energy. *Nano Energy* **2015**, *15*, 321–328. [CrossRef]
48. Shao, H.; Wang, H.; Cao, Y.; Ding, X.; Bai, R.; Chang, H.; Fang, J.; Jin, X.; Wang, W.; Lin, T. Single-layer piezoelectric nanofiber membrane with substantially enhanced noise-to-electricity conversion from endogenous triboelectricity. *Nano Energy* **2021**, *89*, 106427. [CrossRef]
49. Yang, B.; Lee, C.; Xiang, W.; Xie, J.; He, J.H.; Kotlanka, R.K.; Low, S.P.; Feng, H. Electromagnetic energy harvesting from vibrations of multiple frequencies. *J. Micromechanics Microengineering* **2009**, *19*, 035001. [CrossRef]
50. Liu, H.; Qian, Y.; Lee, C. A multi-frequency vibration-based MEMS electromagnetic energy harvesting device. *Sens. Actuators A Phys.* **2013**, *204*, 37–43. [CrossRef]
51. Horng, R.-H.; Chen, K.-F.; Tsai, Y.-C.; Suen, C.-Y.; Chang, C.-C. Fabrication of a dual-planar-coil dynamic microphone by MEMS techniques. *J. Micromechanics Microengineering* **2010**, *20*, 065004. [CrossRef]
52. Li, Y.; Qi, Z.; Yang, J.; Zhou, M.; Zhang, X.; Ling, W.; Zhang, Y.; Wu, Z.; Wang, H.; Ning, B.; et al. Origami NdFeB flexible magnetic membranes with enhanced magnetism and programmable sequences of polarities. *Adv. Funct. Mater.* **2019**, *29*, 1904977. [CrossRef]
53. Barlian, A.A.; Park, W.T.; Mallon, J.R.; Rastegar, A.J.; Pruitt, B.L. Review: Semiconductor Piezoresistance for Microsystems. *Proc. IEEE* **2009**, *97*, 513–552. [CrossRef]
54. Fiorillo, A.; Critello, C.; Pullano, S. Theory, technology and applications of piezoresistive sensors: A review. *Sens. Actuators A Phys.* **2018**, *281*, 156–175. [CrossRef]
55. Irani, F.S.; Shafaghi, A.H.; Tasdelen, M.C.; Delipinar, T.; Kaya, C.E.; Yapici, G.G.; Yapici, M.K. Graphene as a piezoresistive material in strain sensing applications. *Micromachines* **2022**, *13*, 119. [CrossRef]
56. Stassi, S.; Cauda, V.; Canavese, G.; Pirri, C.F. Flexible tactile sensing based on piezoresistive composites: A review. *Sensors* **2014**, *14*, 5296–5332. [CrossRef]
57. Lin, Z.; Duan, S.; Liu, M.; Dang, C.; Qian, S.; Zhang, L.; Wang, H.; Yan, W.; Zhu, M. Insights into Materials, Physics and Applications in Flexible and Wearable Acoustic Sensing Technology. *Adv. Mater.* **2023**, 2306880. [CrossRef] [PubMed]
58. Tao, L.-Q.; Tian, H.; Liu, Y.; Ju, Z.-Y.; Pang, Y.; Chen, Y.-Q.; Wang, D.-Y.; Tian, X.-G.; Yan, J.-C.; Deng, N.-Q.; et al. An intelligent artificial throat with sound-sensing ability based on laser induced graphene. *Nat. Commun.* **2017**, *8*, 14579. [CrossRef] [PubMed]
59. Wang, Y.; Yang, T.; Lao, J.; Zhang, R.; Zhang, Y.; Zhu, M.; Li, X.; Zang, X.; Wang, K.; Yu, W.; et al. Ultra-sensitive graphene strain sensor for sound signal acquisition and recognition. *Nano Res.* **2015**, *8*, 1627–1636. [CrossRef]
60. Yamada, T.; Hayamizu, Y.; Yamamoto, Y.; Yomogida, Y.; Izadi-Najafabadi, A.; Futaba, D.N.; Hata, K. A stretchable carbon nanotube strain sensor for human-motion detection. *Nat. Nanotechnol.* **2011**, *6*, 296–301. [CrossRef]
61. Liu, Z.; Qi, D.; Guo, P.; Liu, Y.; Zhu, B.; Yang, H.; Liu, Y.; Li, B.; Zhang, C.; Yu, J.; et al. Thickness-gradient films for high gauge factor stretchable strain sensors. *Adv. Mater.* **2015**, *27*, 6230–6237. [CrossRef]

62. Hata, K.; Futaba, D.N.; Mizuno, K.; Namai, T.; Yumura, M.; Iijima, S. Water-assisted highly efficient synthesis of impurity-free single-walled carbon nanotubes. *Science* **2004**, *306*, 1362–1364. [CrossRef]
63. Yue, Y.; Liu, N.; Ma, Y.; Wang, S.; Liu, W.; Luo, C.; Zhang, H.; Cheng, F.; Rao, J.; Hu, X.; et al. Highly Self-Healable 3D Microsupercapacitor with MXene-Graphene Composite Aerogel. *ACS Nano* **2018**, *12*, 4224–4232. [CrossRef]
64. Li, P.; Shi, W.; Liu, W.; Chen, Y.; Xu, X.; Ye, S.; Yin, R.; Zhang, L.; Xu, L.; Cao, X. Fabrication of high-performance MXene-based all-solid-state flexible microsupercapacitor based on a facile scratch method. *Nanotechnology* **2018**, *29*, 445401. [CrossRef]
65. Wang, Y.; Yue, Y.; Cheng, F.; Cheng, Y.; Ge, B.; Liu, N.; Gao, Y. Ti3C2T x MXene-based flexible piezoresistive physical sensors. *ACS Nano* **2022**, *16*, 1734–1758. [CrossRef]
66. Cheng, Y.; Ma, Y.; Li, L.; Zhu, M.; Yue, Y.; Liu, W.; Wang, L.; Jia, S.; Li, C.; Qi, T.; et al. Bioinspired microspines for a high-performance spray Ti3C2T x MXene-based piezoresistive sensor. *Acs Nano* **2020**, *14*, 2145–2155. [CrossRef]
67. Gong, S.; Schwalb, W.; Wang, Y.; Chen, Y.; Tang, Y.; Si, J.; Shirinzadeh, B.; Cheng, W. A wearable and highly sensitive pressure sensor with ultrathin gold nanowires. *Nat. Commun.* **2014**, *5*, 3132. [CrossRef]
68. Denby, B.; Schultz, T.; Honda, K.; Hueber, T.; Gilbert, J.M.; Brumberg, J.S. Silent speech interfaces. *Speech Commun.* **2010**, *52*, 270–287. [CrossRef]
69. Janke, M.; Wand, M.; Schultz, T. A spectral mapping method for EMG-based recognition of silent speech. In Proceedings of the International Workshop on Bio-Inspired Human-Machine Interfaces and Healthcare Applications, Valencia, Spain, 20–23 January 2010; SciTePress: Setúbal, Portugal, 2010; pp. 22–31.
70. Wand, M.; Janke, M.; Schultz, T. Tackling Speaking Mode Varieties in EMG-Based Speech Recognition. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 2515–2526. [CrossRef]
71. Janke, M.; Diener, L. EMG-to-Speech: Direct Generation of Speech From Facial Electromyographic Signals. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 2375–2385. [CrossRef]
72. Qiao, Y.; Luo, J.; Cui, T.; Liu, H.; Tang, H.; Zeng, Y.; Liu, C.; Li, Y.; Jian, J.; Wu, J.; et al. Soft Electronics for Health Monitoring Assisted by Machine Learning. *Nano-Micro Lett.* **2023**, *15*, 66. [CrossRef]
73. Schultz, T.; Wand, M. Modeling coarticulation in EMG-based continuous speech recognition. *Speech Commun.* **2010**, *52*, 341–353. [CrossRef]
74. Liu, H.; Dong, W.; Li, Y.; Li, F.; Geng, J.; Zhu, M.; Chen, T.; Zhang, H.; Sun, L.; Lee, C. An epidermal sEMG tattoo-like patch as a new human–machine interface for patients with loss of voice. *Microsyst. Nanoeng.* **2020**, *6*, 16. [CrossRef]
75. Ngiam, J.; Khosla, A.; Kim, M.; Nam, J.; Lee, H.; Ng, A.Y. Multimodal deep learning. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), Bellevue, WA, USA, 28 June–2 July 2011; pp. 689–696.
76. Baltrušaitis, T.; Ahuja, C.; Morency, L.P. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 423–443. [CrossRef]
77. Tian, H.; Li, X.; Wei, Y.; Ji, S.; Yang, Q.; Gou, G.-Y.; Wang, X.; Wu, F.; Jian, J.; Guo, H.; et al. Bioinspired dual-channel speech recognition using graphene-based electromyographic and mechanical sensors. *Cell Rep. Phys. Sci.* **2022**, *32*, 101075. [CrossRef]
78. Kumar, P.; Saini, R.; Roy, P.P.; Sahu, P.K.; Dogra, D.P. Envisioned speech recognition using EEG sensors. *Pers. Ubiquitous Comput.* **2018**, *22*, 185–199. [CrossRef]
79. Porbadnigk, A.; Wester, M.; Calliess, J.; Schultz, T. EEG-based speech recognition-impact of temporal effects. In Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing, Porto, Portugal, 14–17 January 2009; SciTePress: Setúbal, Portugal, 2009; pp. 376–381.
80. Wei, Y.; Qiao, Y.; Jiang, G.; Wang, Y.; Wang, F.; Li, M.; Zhao, Y.; Tian, Y.; Gou, G.; Tan, S.; et al. A Wearable Skinlike Ultra-Sensitive Artificial Graphene Throat. *ACS Nano* **2019**, *13*, 8639–8647. [CrossRef]
81. Shinoda, H.; Nakajima, T.; Ueno, K.; Koshida, N. Thermally induced ultrasonic emission from porous silicon. *Nature* **1999**, *400*, 853–855. [CrossRef]
82. Fuchs, A.K.; Hagmuller, M.; Kubin, G. The New Bionic Electro-Larynx Speech System. *IEEE J. Sel. Top. Signal Process.* **2016**, *10*, 952–961. [CrossRef]
83. Xiao, L.; Chen, Z.; Feng, C.; Liu, L.; Bai, Z.-Q.; Wang, Y.; Qian, L.; Zhang, Y.; Li, Q.; Jiang, K.; et al. Flexible, Stretchable, Transparent Carbon Nanotube Thin Film Loudspeakers. *Nano Lett.* **2008**, *8*, 4539–4545. [CrossRef]
84. Arnold, H.D.; Crandall, I.B. The Thermophone as a Precision Source of Sound. *Phys. Rev.* **1917**, *10*, 22–38. [CrossRef]
85. Hantanasirisakul, K.; Gogotsi, Y. Electronic and Optical Properties of 2D Transition Metal Carbides and Nitrides (MXenes). *Adv. Mater.* **2018**, *30*, 1804779. [CrossRef]
86. Hu, H.; Wang, D.; Wang, Z. Solution for acoustic field of thermo-acoustic emission from arbitrary source. *AIP Adv.* **2014**, *4*, 107114. [CrossRef]
87. Vesterinen, V.; Niskanen, A.O.; Hassel, J.; Helistö, P. Fundamental Efficiency of Nanothermophones: Modeling and Experiments. *Nano Lett.* **2010**, *10*, 5020–5024. [CrossRef]
88. Tian, H.; Ren, T.-L.; Xie, D.; Wang, Y.-F.; Zhou, C.-J.; Feng, T.-T.; Fu, D.; Yang, Y.; Peng, P.-G.; Wang, L.-G.; et al. Graphene-on-Paper Sound Source Devices. *ACS Nano* **2011**, *5*, 4878–4885. [CrossRef]
89. Xie, Q.-Y.; Ju, Z.-Y.; Tian, H.; Xue, Q.-T.; Chen, Y.-Q.; Tao, L.-Q.; Mohammad, M.A.; Zhang, X.-Y.; Yang, Y.; Ren, T.-L. A point acoustic device based on aluminum nanowires. *Nanoscale* **2016**, *8*, 5516–5525. [CrossRef]
90. Tian, H.; Li, C.; Mohammad, M.A.; Cui, Y.-L.; Mi, W.-T.; Yang, Y.; Xie, D.; Ren, T.-L. Graphene Earphones: Entertainment for Both Humans and Animals. *ACS Nano* **2014**, *8*, 5883–5890. [CrossRef]

91. Suk, J.W.; Kirk, K.; Hao, Y.; Hall, N.A.; Ruoff, R.S. Thermoacoustic Sound Generation from Monolayer Graphene for Transparent and Flexible Sound Sources. *Adv. Mater.* **2012**, *24*, 6342–6347. [CrossRef]
92. Heath, M.S.; Horsell, D.W. Multi-frequency sound production and mixing in graphene. *Sci. Rep.* **2017**, *7*, 1363. [CrossRef]
93. Gou, G.-Y.; Jin, M.L.; Lee, B.-J.; Tian, H.; Wu, F.; Li, Y.-T.; Ju, Z.-Y.; Jian, J.-M.; Geng, X.-S.; Ren, J.; et al. Flexible two-dimensional Ti_3C_2 MXene films as thermoacoustic devices. *ACS Nano* **2019**, *13*, 12613–12620. [CrossRef]
94. Aliev, A.E.; Gartstein, Y.N.; Baughman, R.H. Increasing the efficiency of thermoacoustic carbon nanotube sound projectors. *Nanotechnology* **2013**, *24*, 235501. [CrossRef]
95. Zhou, Z.; Wang, J.; Rong, D.; Tong, Z.; Xu, X.; Lim, C. Design and characteristic analysis of CNT thin film thermoacoustic transducer spherical array panel for low intensity focused ultrasound. *J. Therm. Stress.* **2021**, *44*, 582–596. [CrossRef]
96. Passeri, D.; Sassi, U.; Bettucci, A.; Tamburri, E.; Toschi, F.; Orlanducci, S.; Terranova, M.L.; Rossi, M. Thermoacoustic emission from carbon nanotubes imaged by atomic force microscopy. *Adv. Funct. Mater.* **2012**, *22*, 2956–2963. [CrossRef]
97. Aliev, A.E.; Codoluto, D.; Baughman, R.H.; Ovalle-Robles, R.; Inoue, K.; Romanov, S.A.; Nasibulin, A.G.; Kumar, P.; Priya, S.; Mayo, N.K.; et al. Thermoacoustic sound projector: Exceeding the fundamental efficiency of carbon nanotubes. *Nanotechnology* **2018**, *29*, 325704. [CrossRef]
98. Wang, K.; Yap, L.W.; Gong, S.; Wang, R.; Wang, S.J.; Cheng, W. Nanowire-Based Soft Wearable Human–Machine Interfaces for Future Virtual and Augmented Reality Applications. *Adv. Funct. Mater.* **2021**, *31*, 2008347. [CrossRef]
99. Mason, B.J.; Chang, S.-W.; Chen, J.; Cronin, S.B.; Bushmaker, A.W. Thermoacoustic Transduction in Individual Suspended Carbon Nanotubes. *ACS Nano* **2015**, *9*, 5372–5376. [CrossRef]
100. Tian, H.; Xie, D.; Yang, Y.; Ren, T.-L.; Lin, Y.-X.; Chen, Y.; Wang, Y.-F.; Zhou, C.-J.; Peng, P.-G.; Wang, L.-G.; et al. Flexible, ultrathin, and transparent sound-emitting devices using silver nanowires film. *Appl. Phys. Lett.* **2011**, *99*, 253507. [CrossRef]
101. Dutta, R.; Albee, B.; Van Der Veer, W.E.; Harville, T.; Donovan, K.C.; Papamoschou, D.; Penner, R.M. Gold Nanowire Thermophones. *J. Phys. Chem. C* **2014**, *118*, 29101–29107. [CrossRef]
102. Naguib, M.; Kurtoglu, M.; Presser, V.; Lu, J.; Niu, J.; Heon, M.; Hultman, L.; Gogotsi, Y.; Barsoum, M.W. Two-Dimensional Nanocrystals Produced by Exfoliation of Ti_3AlC_2. *Adv. Mater.* **2011**, *23*, 4248–4253. [CrossRef]
103. Richard, B.; Shahana, C.; Vivek, R.; M., A.R.; Rasheed, P.A. Acoustics Platform Meet MXenes–A New Paradigm Shift in the Palette of Biomedical Applications. *Nanoscale* **2023**, *15*, 18156–18172. [CrossRef]
104. Altan, A.; Namvari, M. Multifunctional, flexible, and mechanically robust polyimide-MXene nanocomposites: A review. *2D Mater.* **2023**, *10*, 042001. [CrossRef]
105. Niu, G.; Zhang, M.; Wu, B.; Zhuang, Y.; Ramachandran, R.; Zhao, C.; Wang, F. Nanocomposites of pre-oxidized $Ti_3C_2T_x$ MXene and SnO_2 nanosheets for highly sensitive and stable formaldehyde gas sensor. *Ceram. Int.* **2023**, *49*, 2583–2590. [CrossRef]
106. Daschewski, M.; Boehm, R.; Prager, J.; Kreutzbruck, M.; Harrer, A. Physics of thermo-acoustic sound generation. *J. Appl. Phys.* **2013**, *114*, 114903. [CrossRef]
107. Sofiah, A.G.N.; Samykano, M.; Kadirgama, K.; Mohan, R.V.; Lah, N.A.C. Metallic nanowires: Mechanical properties–Theory and experiment. *Appl. Mater. Today* **2018**, *11*, 320–337. [CrossRef]
108. Jiu, J.; Suganuma, K. Metallic nanowires and their application. *IEEE Trans. Compon. Packag. Manuf. Technol.* **2016**, *6*, 1733–1751. [CrossRef]
109. Untiedt, C.; Rubio, G.; Vieira, S.; Agraït, N. Fabrication and characterization of metallic nanowires. *Phys. Rev. B* **1997**, *56*, 2154. [CrossRef]
110. Bobinger, M.; La Torraca, P.; Mock, J.; Becherer, M.; Cattani, L.; Angeli, D.; Larcher, L.; Lugli, P. Solution-Processing of Copper Nanowires for Transparent Heaters and Thermo-Acoustic Loudspeakers. *IEEE Trans. Nanotechnol.* **2018**, *17*, 940–947. [CrossRef]
111. Mubeen, N.; Shahina, A.; Khan, A.N.; Vinoth, G. Combining spectral features of standard and throat microphones for speaker identification. In Proceedings of the 2012 International Conference on Recent Trends in Information Technology, Chennai, India, 19–21 April 2012; IEEE: New York, NY, USA, 2012; pp. 119–122.
112. Sahidullah, M.; Hautamäki, R.G.; Thomsen, D.A.L.; Kinnunen, T.; Tan, Z.-H.; Hautamäki, V.; Parts, R.; Pitkänen, M. Robust speaker recognition with combined use of acoustic and throat microphone speech. *Proc. Interspeech* **2016**, 1720–1724. [CrossRef]
113. Rastgoo, R.; Kiani, K.; Escalera, S. Sign language recognition: A deep survey. *Expert Syst. Appl.* **2021**, *164*, 113794. [CrossRef]
114. Fang, H.; Li, S.; Wang, D.; Bao, Z.; Xu, Y.; Jiang, W.; Deng, J.; Lin, K.; Xiao, Z.; Li, X.; et al. Decoding throat-language using flexibility sensors with machine learning. *Sens. Actuators A Phys.* **2023**, *352*, 114192. [CrossRef]
115. Wang, G.; Liu, T.; Sun, X.-C.; Li, P.; Xu, Y.-S.; Hua, J.-G.; Yu, Y.-H.; Li, S.-X.; Dai, Y.-Z.; Song, X.-Y.; et al. Flexible pressure sensor based on PVDF nanofiber. *Sens. Actuators A Phys.* **2018**, *280*, 319–325. [CrossRef]
116. Shuai, X.; Zhu, P.; Zeng, W.; Hu, Y.; Liang, X.; Zhang, Y.; Sun, R.; Wong, C.-p. Highly sensitive flexible pressure sensor based on silver nanowires-embedded polydimethylsiloxane electrode with microarray structure. *ACS Appl. Mater. Interfaces* **2017**, *9*, 26314–26324. [CrossRef]
117. Cui, P.; Athey, S. Stable learning establishes some common ground between causal inference and machine learning. *Nat. Mach. Intell.* **2022**, *4*, 110–115. [CrossRef]
118. Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep learning--based text classification: A comprehensive review. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–40. [CrossRef]
119. Zhang, Z.; Geiger, J.; Pohjalainen, J.; Mousa, A.E.-D.; Jin, W.; Schuller, B. Deep learning for environmentally robust speech recognition: An overview of recent developments. *ACM Trans. Intell. Syst. Technol. (TIST)* **2018**, *9*, 1–28. [CrossRef]

120. Rios, A.L.G.; Li, Z.; Xu, G.; Alonso, A.D.; Trajković, L. Detecting network anomalies and intrusions in communication networks. In Proceedings of the 2019 IEEE 23rd International Conference on Intelligent Engineering Systems (INES), Gödöllő, Hungary, 25–27 April 2019; IEEE: New York, NY, USA, 2019; pp. 29–34.
121. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]
122. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef]
123. Qiu, S.; Wang, J.; Tang, C.; Du, D. Comparison of ELM, RF, and SVM on E-nose and E-tongue to trace the quality status of mandarin (*Citrus unshiu* Marc.). *J. Food Eng.* **2015**, *166*, 193–203. [CrossRef]
124. Wang, J.; Zhang, L.; Cao, J.J.; Han, D. NBWELM: Naive Bayesian based weighted extreme learning machine. *Int. J. Mach. Learn. Cybern.* **2018**, *9*, 21–35. [CrossRef]
125. Chen, S.; Luo, J.; Wang, X.; Li, Q.; Zhou, L.; Liu, C.; Feng, C. Fabrication and Piezoresistive/Piezoelectric Sensing Characteristics of Carbon Nanotube/PVA/Nano-ZnO Flexible Composite. *Sci. Rep.* **2020**, *10*, 8895. [CrossRef]
126. Yoo, J.-Y.; Oh, S.; Shalish, W.; Maeng, W.-Y.; Cerier, E.; Jeanne, E.; Chung, M.-K.; Lv, S.; Wu, Y.; Yoo, S.; et al. Wireless broadband acousto-mechanical sensing system for continuous physiological monitoring. *Nat. Med.* **2023**, *29*, 3137–3148. [CrossRef]
127. Kwon, S.; Kim, H.S.; Kwon, K.; Kim, H.; Kim, Y.S.; Lee, S.H.; Kwon, Y.-T.; Jeong, J.-W.; Trotti, L.M.; Duarte, A.; et al. At-home wireless sleep monitoring patches for the clinical assessment of sleep quality and sleep apnea. *Sci. Adv.* **2023**, *9*, eadg9671. [CrossRef]