ADVANCED MATERIALS
www.advmat.de

Flexible Piezoelectric Acoustic Sensors and Machine Learning for Speech Processing

*Young Hoon Jung, Seong Kwang Hong, Hee Seung Wang, Jae Hyun Han, Trung Xuan Pham, Hyunsin Park, Junyeong Kim, Sunghun Kang, Chang D. Yoo, and Keon Jae Lee**

Flexible piezoelectric acoustic sensors have been developed to generate multiple sound signals with high sensitivity, shifting the paradigm of future voice technologies. Speech recognition based on advanced acoustic sensors and optimized machine learning software will play an innovative interface for artificial intelligence (AI) services. Collaboration and novel approaches between both smart sensors and speech algorithms should be attempted to realize a hyperconnected society, which can offer personalized services such as biometric authentication, AI secretaries, and home appliances. Here, representative developments in speech recognition are reviewed in terms of flexible piezoelectric materials, self-powered sensors, machine learning algorithms, and speaker recognition.

1. Introduction

The most intuitive human–machine interface is speech recognition,[1–8] which can replace conventional touch-based devices for hyper-connected society.[9–15] In addition, voice user interface (VUI) has attracted significant attention as the core technology of internet of things (IoT) artificial intelligence (AI) due to its exceptionally convenient and bilateral communication.[3,16–28] Smart acoustic sensors can be applied to speaker recognition, biometrics, personalized AI secretary, and smart home appliances.[29–45]

Speech recognition system involves two main parts: i) acoustic hardware sensors,[46–54] ii) speech recognition software.[55–60] Acoustic sensor detects sound pressure of human utterance and coverts original analog speech sound into electrical data.[61–68] The general speech signals have 60 dB of magnitude over audible frequency range (20 Hz to 20 kHz).[69,70] The most of voice energy is distributed in frequency range of 100–4000 Hz, which includes the surrounding noises.[71–74] The acoustic sensors should have enough sensitivity to detect the speech signal at least 60 dB over voice frequency range. The frequency domain signals are converted by using fast Fourier transform and short time Fourier transform. The converted signals are used to provide the training and test data for speech recognition software. The optimized machine learning algorithms are necessary to improve the speech recognition under the noise condition by filtering the surrounding sound such as typing, clock tick, car horn, and engine.

The condenser type microphones is widely commercialized acoustic sensor to detect sound signal by using the difference of capacitance between two conducting diaphragms.[61,67,69,75–81] This capacitive sensors belong to the nonresonance type that exhibits the flat frequency response. The uniform sensitivity over voice frequency range enables the easy processing of speech recognition with single channel. However, the condenser-type sensor has demerits such as insufficient sensitivity, limited recognition distance, high power consumption and the unstable circuit of large amplification.[3,82–87] The piezoresistive sensor detects the sound by sensing the resistance change depending on the movement of membrane.[88] Similar to capacitive microphone, the piezoresistive sensors should be powered to detect the sound, which requires high power consumption.[89] Furthermore, the noise signal can be generated at high temperature condition, since parameters of piezoresistance are dependent on temperature.[90,91] The triboelectric acoustic sensors detect the voice signal, utilizing the generated electrostatic charges of membrane by sound pressure.[51,53] This type is self-powered operation, like piezoelectric acoustic sensor, converting mechanical deformation into electrical potential. The electrostatic charges can be severely affected by humidity and temperature, due to the nature electrostatic phenomenon.[92–95] As one of the piezoelectric type, the surface acoustic wave (SAW) sensors consist of two bulky piezoelectric materials based rigid transducers to respond to high frequency signals above MHz.[96,97] In addition, the delay between the transmitted and received signal can restrict the real-time sensing. Therefore, properties of SAW sensors are not appropriate for frequency range of speech recognition.[98]

Recently, flexible piezoelectric acoustic sensors, mimicking the basilar membrane of human cochlear,[3,67,99–103] are spotlighted as a promising candidate for improving the sensitivity and recognition rate. The self-powered flexible acoustic sensor exhibits higher sensitivity at human utterance compared to

Y. H. Jung, Dr. S. K. Hong, H. S. Wang, J. H. Han, Prof. K. J. Lee
Department of Materials Science and Engineering
Korea Advanced Institute of Science and Technology (KAIST)
291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
E-mail: keonlee@kaist.ac.kr
T. X. Pham, Dr. H. Park, J. Kim, S. Kang, Prof. C. D. Yoo
Department of Electrical Engineering
Korea Advanced Institute of Science and Technology (KAIST)
291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adma.201904020.

DOI: 10.1002/adma.201904020

the capacitive and piezoresistive sensor, because the device is fabricated considering the material and mechanical design for superior piezoelectric coefficient and multiple resonant vibrations, respectively. The flexible acoustic sensors of highly sensitive piezoelectric membrane can consistently exhibit high sensitivity without degradation over humidity and heat, due to the durability of inorganic thin film materials. The thickness and shape of piezoelectric membrane should be considered to design resonance frequencies. Thin membrane can intensively vibrate at low voice frequencies, while thick membrane respond to high frequency. The flexible trapezoidal shape of piezoelectric membrane can generate the exceptional resonant vibration according to voice frequency band, which can detect minute sound under the 60 dB from far distance, compared to other type microphones.[3,66,67,104,105] The flexible piezoelectric membrane can produce multiple data set of frequency components depending on channel width, due to more than twice IDE channels. Multichannel signals can obtain more than twice voice information for speech processing, like the mechanism of human cochlear with more than 10 000 outer hair cells. Since the piezoelectric acoustic sensors acquire abundant speech information from multichannel signals, it has advantages of machine learning training based on the sufficient data, and useful signal selection from multichannel input. By adopting optimized machine learning algorithms with weighted value, the intentional selection of discrete information can filter noise signal and enhance speech recognition. The flexible piezoelectric acoustic sensors were proven to exhibit 97.5% accuracy for speaker recognition, due to high sensitivity of –76 dB under white noise condition, and abundant data set of seven channels.[3] **Table 1** shows the main characteristics in different acoustic sensor types.

Speech recognition processing transforms the voice information into a binary digital format for input data of machine learning algorithms.[55,106–108] Researches of speech processing have been developed from the perspectives of collection, manipulation, and storage of the incoming voice data.[109–114] For decades, the classical machine learning based algorithms such as Gaussian mixture models (GMMs), hidden Markov models (HMMs), and support vector machine (SVM) have been used widely for speech processing. Recent advances in deep learning have significantly improved the performance of speech processing tasks, surpassing the performance of classical machine learning algorithms. However, these technologies have still suffered from a low recognition rate due to hardware issue of low sensitivity and limited voice information.[115–117] Future voice technologies should be focused on synergistic collaboration between smart acoustic sensor and AI algorithm to overcome the fundamental weakness of speech recognition.

Here, we introduce brief overview of flexible piezoelectric acoustic sensors and speech processing. In particular, we concentrate on recent progress in bioinspired and self-powered inorganic acoustic sensors, and machine learning algorithms for speaker recognition. This review is classified into five main categories: i) Piezoelectric effect and materials, ii) self-powered flexible sensors, iii) flexible piezoelectric acoustic sensors, iv) machine learning algorithms for speech processing, and v) speaker recognition. **Figure 1** illustrates machine learning based flexible piezoelectric acoustic sensor and its applications as voice user interface platform.

Young Hoon Jung received his B.S. degree in materials science and engineering (MSE) at Korea University in 2018. He is currently working toward his M.S. at KAIST under the supervision of Prof. Keon Jae Lee. His doctoral research interests include piezoelectric energy harvesting, flexible acoustic sensors, and laser–material interaction.

Chang D. Yoo received his B.S. degree in engineering and applied science from the California Institute of Technology, the M.S. degree in electrical engineering from Cornell University and his Ph.D. degree in electrical engineering from MIT. His current research interest is machine learning, signal processing, computer vision and audio processing. He is a Member of Tau Beta Pi and Sigma Xi. He is director of Video Turing Test Research Center and also the AI Fairness Research Center.

Keon Jae Lee received his Ph.D. in materials science and engineering (MSE) at the University of Illinois, Urbana-Champaign (UIUC). During his Ph.D. at UIUC, he was involved in the first co-invention of "flexible single-crystalline inorganic electronics", using top-down semiconductors and soft lithographic transfer. Since 2009, he has been a professor in MSE at KAIST. His current research topics are self-powered flexible electronic systems including self-powered sensors/energy harvester, micro LEDs, neuromorphic memory/large-scale integration (LSI) and laser material interaction for in vivo biomedical applications.

2. Piezoelectric Effect and Materials

Piezoelectric effect that produces electric charge by the deformation is the well-established technology for generating electricity from tiny mechanical force.[118–122] In addition, the electrical output signals can be utilized to detect the motions for sensing the physical or chemical reactions.[123–127]

Table 1. Summary of main characteristics in different acoustic sensor types.

	Capacitive/MEMS type	Piezoresistive type	Triboelectric type	Piezoelectric type
Measurement type	Capacitance	Voltage (Resistance)	Voltage	Voltage
External power	Required	Required	Self-powered	Self-powered
Device type	Nonresonance	Resonance	Resonance	Resonance
Sensitivity (open circuit)	Low	Low	High	High
Data sets	Single signal	Single signal	Single/multisignal	Multisignal
Frequency response	Flat	Sharp peak	Sharp peaks	Tunable for voice frequency band

Figure 2a-i,ii shows wurtzite and perovskite-structured crystals of piezoelectric materials. Under mechanical strain on these piezoelectric crystals, the displacement of positive and negative electric charge leads to an electric polarization inside the piezoelectric structures.[118–120,128] The created net charges (Q) can be expressed by $Q = A \cdot E \cdot S \cdot d_{ij}$, where A is the surface area of crystals, E is the elastic modulus, S is the strain, and d_{ij} is the piezoelectric charge coefficient; i and j denotes the polarization and strain direction of 3D coordinate system, respectively.[128–131] The interdigitated electrodes (IDEs) based device is d_{33} mode case, while d_{31} is applied in generator with metal–insulator–metal (MIM) structure. The piezoelectric properties can vary depending on materials as well as the structure. By utilizing the vapor–liquid–solid method, deposition process or molecular beam epitaxy, wurtzite piezoelectric materials such as ZnO and ZnS can be easily synthesized in different structures of nanowires (NWs), nanotubes, and films for energy conversion, piezotronics and piezophotonic devices; however, their applications are limited due to low piezoelectricity.[126–145] The piezoelectric materials of perovskite structure such as PZT and PMN-PT have been widely commercialized for energy generators, actuators and sensors, due to excellent d_{33} piezoelectric properties.[128–130,146–149]

Flexible piezoelectric harvesters on plastics have attracted enormous interest since they can generate electrical energy not only by vertical pressures but also by tiny bending including wind, sound, and biomechanical movements.[150–153] Yang et al.[133] fabricated parallel ZnO NW on a flexible substrate for piezoelectric nanogenerators to overcome disadvantages of vertical NWs.[138,139] As shown in Figure 2b, the piezoelectric NWs were laterally aligned on plastics to increase output voltage and mechanical robustness. The stable condition was preserved at 22 stretching/releasing cycles per minute for 45 min, due to the package of insulating flexible film. Deformation-induced displacement of electric charges in wurtzite crystal, producing electrical piezoelectric potential distributions to charge the wire. While the bending motions induced the piezoelectric potential to charge the wire, the releasing motions caused electrons to flow back in the opposite direction. This procedure generated positive and negative open-circuit voltage of 20 and –50 mV, respectively, according to periodic bending and releasing with a strain of 0.05–0.1%. In addition, the series connection of laterally aligned ZnO wires enhanced the open-circuit voltage.

Park et al.[150] reported the flexible $BaTiO_3$ energy harvester by transferring high temperature annealed perovskite piezoelectric thin films onto a plastic substrate via soft lithographic printing method. The fabrication process was conducted by three main steps: i) sol–gel coating and high temperature crystallization of lead-free $BaTiO_3$ on Pt/Si sacrificial substrate, ii) formation of the freestanding MIM piezoelectric layer by removing the mother substrate, and iii) transfer of the MIM structures. Figure 2c displays the fabricated flexible $BaTiO_3$ thin

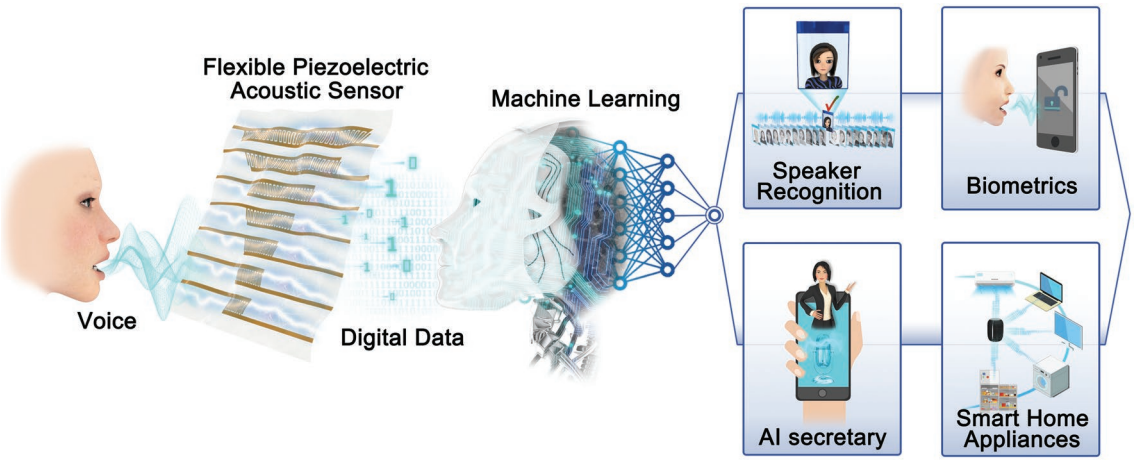

Figure 1. Schematic illustration of the promising applications of voice user interface platform. Flexible piezoelectric acoustic sensors covert utterance to electrical multi signals by vibrating in response to the speaker's voice, which can provide digitalized data for preprocessing. The data are trained using machine learning-based model and language information is extracted from the speech. This process will support the evolution from touch to sound-operated electronics system.

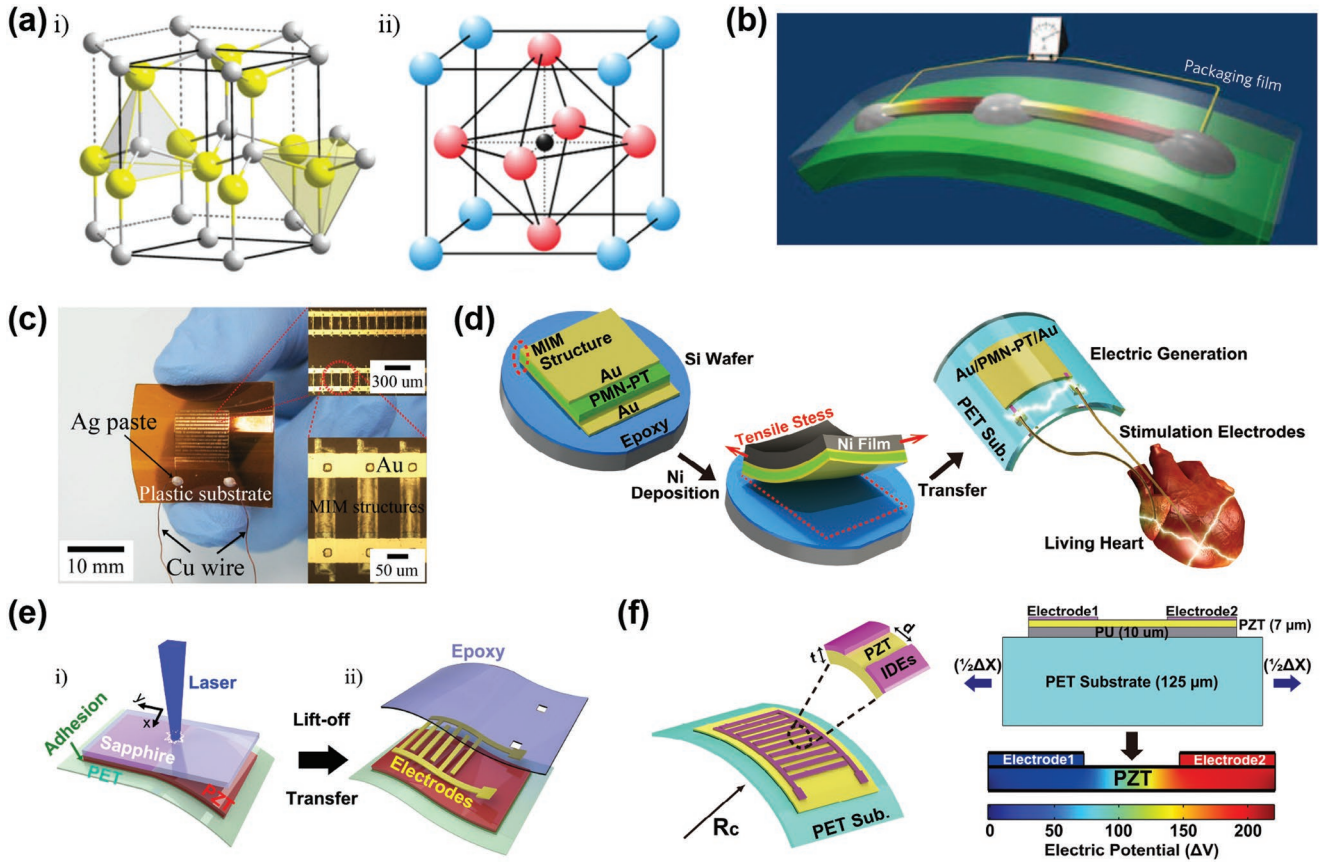

Figure 2. a) Atomic structures: i) a wurtzite structure and ii) a perovskite structure. a) Reproduced with permission.[128] Copyright 2014, Royal Society of Chemistry. b) Design of a piezoelectric single-wire generator on a flexible substrate and two generators connected in series. Reproduced with permission.[133] Copyright 2009, Springer Nature. c) A flexible $BaTiO_3$ nanogerator supported on a plastic substrate with the copper wires connected. Reproduced with permission.[150] Copyright 2011, American Chemical Society. d) Schematic illustration of the fabrication procedure and biomedical application of flexible PMN-PT piezoelectric harvester. Reproduced with permission.[129] Copyright 2014, Wiley-VCH. e) Depiction of transfer process for flexible PZT thin film via LLO method. Reproduced with permission.[147] Copyright 2014, Wiley-VCH. f) The simulation of piezopotential distribution inside the PZT thick film for IDE-type device. Reproduced with permission.[149] Copyright 2016, Wiley-VCH.

film nanogenerator to measure the electrical output properties. A magnified optical image shows that IDEs were connected with the MIM structure. The bent $BaTiO_3$ thin film produced output voltage between neighboring electrodes by creating electrical charges in the MIM structure. The poling process aligned randomly distributed charge dipoles in one direction by applying high electric fields of 100 kV cm^{-1} at high temperature of 140 °C.[154–156] After the piezoelectric coefficient of d_{33} increased to 105 pC N^{-1} by the poling process, the self-powered flexible $BaTiO_3$ generator exhibited short-circuit current of 12 nA and open-circuit voltage of 0.35 V at maximum strain of 0.55%. This work was significant because flexible perovskite piezoelectric thin film was successfully demonstrated on plastic substrates by adopting high temperature annealing process before transfer.

Single-crystalline thin films with exceptionally high piezoelectric properties improved the capability of MIM-based flexible energy devices for converting mechanical strain into electrical signal.[157,158] Hwang et al.[129] demonstrated flexible single crystalline PMN-PT piezoelectric nanogenerator to directly stimulate a heart of a living rat without any external power sources. Figure 2d presents the schematic of experimental procedures to fabricate a self-powered pacemaker. The detailed fabrication procedures are as follows: i) single crystal structure of PMN-PT ingot was grown from melt near the morphotropic phase boundary (MPB). The piezoelectric charge constant can be obtained up to 2500 pC N^{-1} at the composition of MPB.[159–162] The electrode deposited PMN-PT plate was thinned to a layer of 8 μm, and poled along the[98] direction after deposition of Au top electrode. ii) the piezoelectric thin film was peeled off from the mother substrate by utilizing electroplated thick Ni film stressor. Due to the directional stress mismatch between Ni layer and MIM piezoelectric thin film, uniform and spontaneous exfoliation was achieved without any wrinkling and cracking on active device materials. iii) piezoelectric PMN-PT film was transferred on PET substrates with UV-cured polyurethane, providing sufficient flexibility to respond slight bending motions. The piezoelectric pacemaker generator responded to the minute movement of flexible substrates and subsequently generated high output. The output voltage and current of flexible PMN-PT nanogenerator were measured up to 8.2 V and 145 μA, respectively, upon periodic

bending and releasing motions with 0.62% strain. The stability test showed mechanical durability without degradation during 30 000 bending/releasing cycles.

The electrode structures have been modified and optimized to enhance output voltage depending on current and voltage driven applications. While MIM energy harvesters can generate high current, IDE-based piezoelectric nanogenerator can produce sufficient output voltage.[129,149] Park et al.[147] demonstrated highly efficient IDE type piezoelectric harvester on plastic substrates using inorganic-based laser lift-off (ILLO) technique as illustrated in Figure 2e. The PZT thin film was separated from mother substrates by irradiating XeCl excimer laser on the transparent sapphire wafer. Epoxy was coated to protect the thin film nanogenerator after IDEs were deposited on the flexible PZT membrane. The commonly used equation of piezoelectric output voltage can be described by $V_{3j} = \sigma_{xx}\, g_{3j} L_j$, where σ_{xx} is mechanical stress, g_{3j} is the piezoelectric voltage constant ($g_{3j} = d_{3j}/\varepsilon^T$, ε^T is the permittivity), and L_j is the distance between electrodes. The IDE-type self-powered generator operates in d_{33} mode, which can have L_3 longer than a few hundred micrometers. The piezoelectric voltage constant of g_{33} is usually twice larger than g_{31} constant. Therefore, high open-circuit voltage of 200 V was achieved at bending radius of 1.61 cm (corresponding to tensile strain of 0.386%) from the 100 μm gap across the electrodes of flexible PZT device. This is one of the highest output voltage among flexible piezoelectric materials compared to a few volts of previous researches.[127,129,148] After 10 000 bending iterations with strain of 0.153% to 0.386%, the peak voltage values were maintained without physical cracks. This work showed that output voltage can be generated by slight vibration such as pulse, wind and voice.

Hwang et al.[149] developed flexible PZT nanogenertaor by utilizing the aerosol deposition (AD) method. The AD process can provide instantaneously deposited piezoelectric thick films with high quality. The piezoelectric coefficient of the films increases with the thickness, however, the layers cannot efficiently respond to the vibration due to the lack of flexibility. After optimization of trade-off between piezoelectric coefficient and flexibility, 7 μm thick AD PZT film of one-step deposition enabled high output voltage compared to multiple process of sol–gel coating and annealing. The efficient performance was attributed to high piezoelectric coefficient d_{33} and g_{33} of 406 pC N^{-1} and 49.5 mV m N^{-1}, respectively. Figure 2f depicts the simulation model of the flexible piezoelectric energy harvester. The simple model assumed that piezoelectric film was deformed under bending radius of 21 mm, which elongated PET substrates with a displacement (ΔX) of 150 μm. The maximum piezopotential difference (ΔV) based on finite-element analysis (FEA) was calculated to 370 V, that matched with measured high open-circuit voltage of 200 V at bending motions by fingers. The output voltage remained stably constant during 115 000 bending/releasing cycles at 0.3% strain.

3. Self-Powered Flexible Sensors

Flexible piezoelectric sensors have the advantage of self-powered operation without the external energy supply.[3,62–68,123–127] Self-powered sensor is an important technology for the future IoT era of always-on condition because human cannot manage the energy issue of millions of sensors.[149,163–165] Lin et al.[166] integrated piezoelectric nanogenerator with the mechanical component of weight measurement by growing the textured ZnO NW film on the elastic spring. **Figure 3**a–i depicts the structure of the spring-substrated nanogenerator (SNG). The fabricated device was composed of Ag electrode, metal spring, and ZnO NWs passivated with PMMA as a buffer layer. The weight measurement mode of the self-powered SNG is described in Figure 3a-ii. The tiny plates were placed on the active sensor to load the heavy object, producing the piezopotential across the ZnO layer. Figure 3a-iii shows linear relationship between the loading weight and electrical outputs. The piezoelectric sensor produced output voltage of 0.23 V by applying 15.2 N. The output decayed by only 3–4% after continuous 80 000 working cycles. The equivalent weight of the object was calculated with the output current and voltage, exhibiting the sensitivity of 2.8 nA kg^{-1} and 45 mV kg^{-1}, respectively. Furthermore, this research validated the effectiveness of the SNG self-powered sensor by varying displacements and vibration condition. The spring of 50 mm was deformed up to 35 mm, showing the sensitivity of 7 mV mm^{-1}. The oscillated output current was induced by free vibration, which indicated the availability of the SNG as vibration sensor.

In addition, flexible piezoelectric self-powered sensors have been applied to detect chemical substances.[127,167,168] Niu et al.[169] improved the oxygen sensor device based on ZnO piezotronic effect. Figure 3b-i illustrates the experimental setup and developed piezoelectric oxygen sensor on flexible substrates. The ZnO NW was transferred onto plastic substrates to form the metal–semiconductor–metal structure by fixing both ends with silver paste. After the piezoelectric chemical sensor was attached on the bracket inside gas chamber, positioner was moved to induce the strain in substrates. Figure 3b-ii shows the performance of oxygen sensor with the current under the bias of 1 V. As the oxygen pressure rose from 16 to 700 Torr, the current dropped from 899 to 401 nA, due to the oxygen adsorption. Tensile strain improved the sensitivity in terms of the relative current change. The relative current change was enhanced from −55.4% to −75.4% as the strain rose from 0% to 0.2%, due to the potential distribution inside ZnO NW. Figure 3b-iii illustrates the simple mechanism of enhanced sensitivity of self-powered oxygen chemical sensor. When tensile strain created the negative piezopotential across the ZnO NW, Schottky barrier height increased. Therefore, the movements of electrons responding to oxygen pressure were restricted, and sensitivity of relative current variation was enhanced. This improved detection efficiency at 0.2% strain indicated that minute movement can be appropriate input for flexible piezoelectric sensor.

The real-time biomedical information can be also monitored using self-powered piezoelectric sensors attached on rugged human skin. Park et al.[170] fabricated a self-powered health-care sensor to detect blood vessel movements in situ. As shown in Figure 3c-i,iii, the piezoelectric PZT thin film was transferred onto an ultrathin flexible substrate via ILLO method, allowing the conformal contact on the epidermis to monitor the biomechanical vibrations by human pulse. Characteristics of

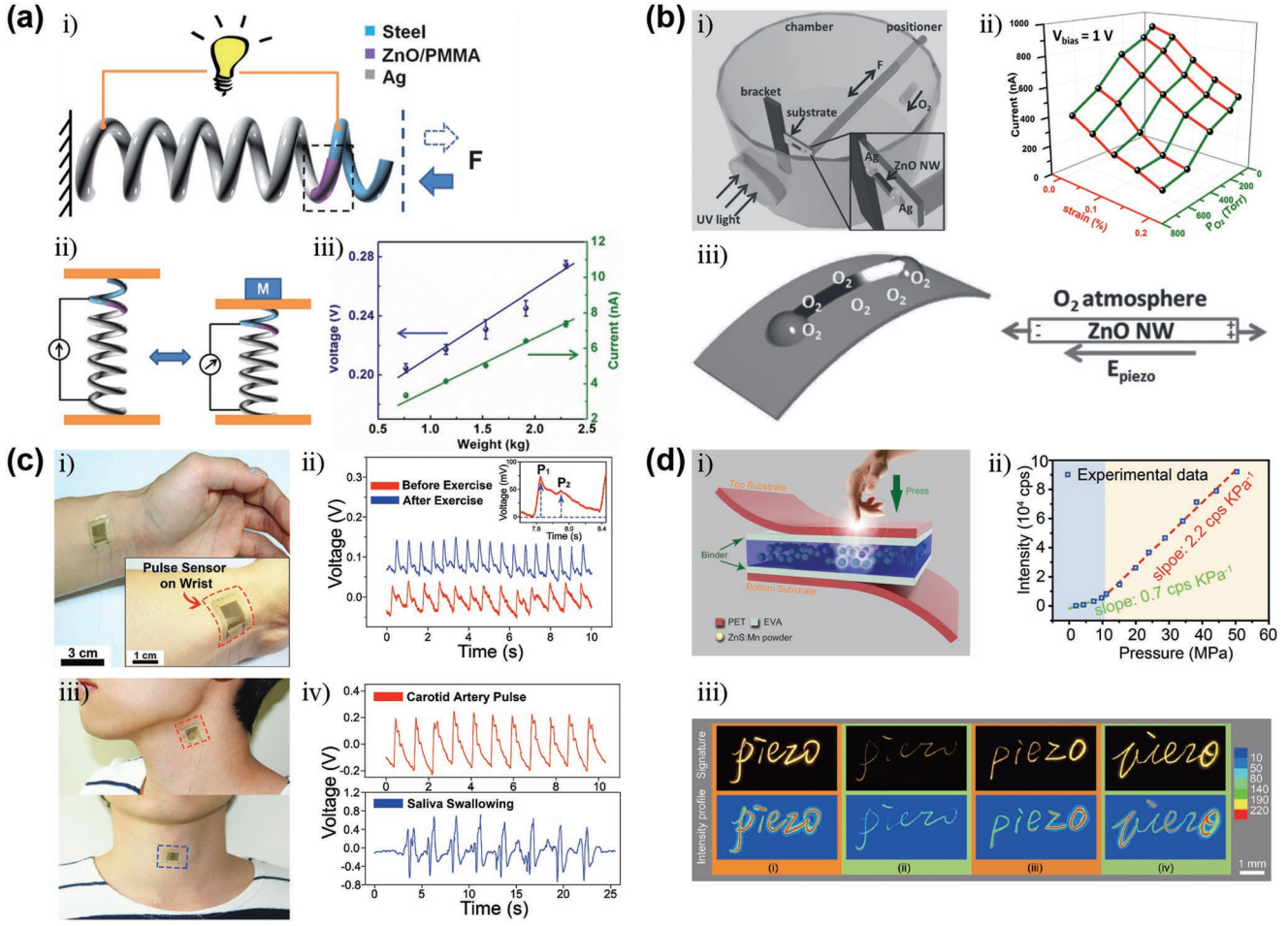

Figure 3. a-i) Schematic structure of nanogenerator based on ZnO nanowires grown on the surface of the elastic wire. ii) The working mechanism of the self-powered sensor for weight measurement. iii) Linear relationship between the electrical output and loaded weight on the nanogenerator. a) Reproduced with permission.[166] Copyright 2013, Royal Society of Chemistry. b-i) Schematic of the experiment setup for analyzing the piezotronic effect in a ZnO NW oxygen sensor. ii) 3D graph showing the output current of the ZnO NW-based oxygen sensor under different strains and oxygen pressures at a bias of 1 V. iii) Simplified structures of ZnO NW oxygen sensor under vacuum with no strain, only in an oxygen atmosphere, and in both oxygen atmosphere and with tensile strain. b) Reproduced with permission.[169] Copyright 2013, Wiley-VCH. c-i) Optical image of a self-powered pulse sensor conformally contacted on wrist utilizing a biocompatible liquid bandage. The inset shows the deformed device due to blood vessel movement. ii) Radial artery pulse detected by the piezoelectric pulse sensor, showing different output voltages in response to heart rates before and after physical exercise. The inset represents the magnified voltage by the radial artery pulse before exercise, clearly indicating pulse pressure (*P*1) and late systolic augmentation (*P*2). iii) Photograph of self-powered pulse sensor conformally attached on a carotid artery spot (top) and the middle of the throat (bottom). iv) The generated piezoelectric voltage according to carotid arterial pressure (top) and saliva swallowing actions (bottom). c) Reproduced with permission.[170] Copyright 2017, Wiley-VCH. d-i) Schematic depiction of pressure sensor matrix (PSM) devices. ii) Integral intensity over 525–685 nm under different pressures. iii) PSM devices showing the recorded signing habits of four signees. d) Reproduced with permission.[173] Copyright 2015, Wiley-VCH.

the self-powered health-care sensor, such as the sensitivity to external pressure, the response to the audible frequency and the durability against mechanical damage, were examined before direct pulse detection. The exerted pressure of 25 kPa produced the output voltage of 1.85 V, exhibiting the pressure sensitivity of 0.018 kPa^{-1}. The peak-to-peak voltage (V_{pp}) of 100 mV was generated from the 240 Hz sound of 80 dB. The dynamic pushing test verified the stable operation of the sensor during 5000 cycles under the pressure of 20 kPa. The flexible piezoelectric pulse sensor was successfully adhered on a human wrist by biocompatible liquid bandage. The inset is the optical image of the stably deformed pulse sensor that effectively responded to tiny movement of radial artery. Figure 3c-ii shows the real-time pulse of a healthy male before (red line) and after (blue line) physical exercise. The self-powered pulse sensor exhibited the averaged output V_{pp} of 65 mV before exercise and 81.5 mV after exercise, respectively. The result indicated that the rate of heartbeats increased to supply the consistent amount of oxygen to muscles. The enlarged radial arterial signal (inset) represents the characteristic artery waveforms with important physiological information including myocardial infarction and artery disease. Flexible piezoelectric sensor was also attached on a human neck to obtain biomedical data, such as carotid artery pulse and muscle movements. Figure 3c-iv presents the generated voltage V_{pp} of 400 mV and 1000 mV from the carotid arterial pressure and saliva swallowing actions, respectively.

Some researchers have attempted to utilize flexible piezoelectric sensors as human-machine interface.[171,172] Wang et al.[173] developed flexible self-powered pressure sensor matrix (PSM) to record the signature and signing habits by employing piezophotonic effect of ZnS:Mn particles (ZMPs). Figure 3d-i illustrates the sandwiched PSM device composed of two polymeric substrates and middle layer of ZMPs. Ethylene-vinyl acetate copolymer (EVA), bonding ZMPs on transparent PET films, protected the photon emitting layer and transmitted the light. The emitted light intensity increased as the loaded force changed from 0.8 to 10 N. Figure 3d-ii shows the integral light intensity depending on pressures. The self-powered sensors exhibited linearly increasing light intensity with sensitivity of 2.2 cps kPa^{-1} over the range of 10–50 MPa. After pressures generated the piezopotential within ZnS, nonradioactive recombination occurred to excite the electrons of Mn^{2+} ion. The yellow visible light was emitted when the electrons fell back to the original state. The intensity was stably preserved while periodic pressure of 15 MPa was applied for 10 000 cycles. Figure 3d-iii illustrates the securer signature collecting system demonstrated with flexible PSM device. Signatures "piezo" were handwritten by four different signees, showing obvious differences in writing habits. The ZMP-based PSM device enabled the precise mapping for planar pressure distributions, which can be further applied in identity verification.

4. Flexible Piezoelectric Acoustic Sensors

4.1. Flexible Organic Piezoelectric Acoustic Devices

Shintaku et al.[62] developed multichannel flexible PVDF film acoustic sensor with selective frequency response. **Figure 4**a displays that the electrodes deposited PVDF layer was bonded on a substrate with trapezoidal hole to mimic the passive basilar membrane. The width of film varied along the longitudinal direction, changing resonant frequencies according to the position. The electrodes were named from Ch. 1 to Ch. 24 as the width increased. Self-powered piezoelectric acoustic device was adhered on the stainless substrate filled with silicone oil to model an in vivo environment. Figure 4b shows the experimental data of vibrational motion (solid line) and output voltage (broken line). The peaks of both amplitudes corresponded to each other, showing identical frequency dependence. The resonant frequencies were measured as 3.64, 2.32, and 1.88 kHz at Ch. 6, Ch.12, and Ch.18, respectively. The output voltage was not sensitive to detect minute sound under 60 dB, although the polymer piezoelectric acoustic sensor exhibited the selective response to frequencies.

Park et al.[63] demonstrated flexible self-powered MIM structured PVDF acoustic sensor by depositing top and bottom metal electrodes on the surface of the piezoelectric thin film.

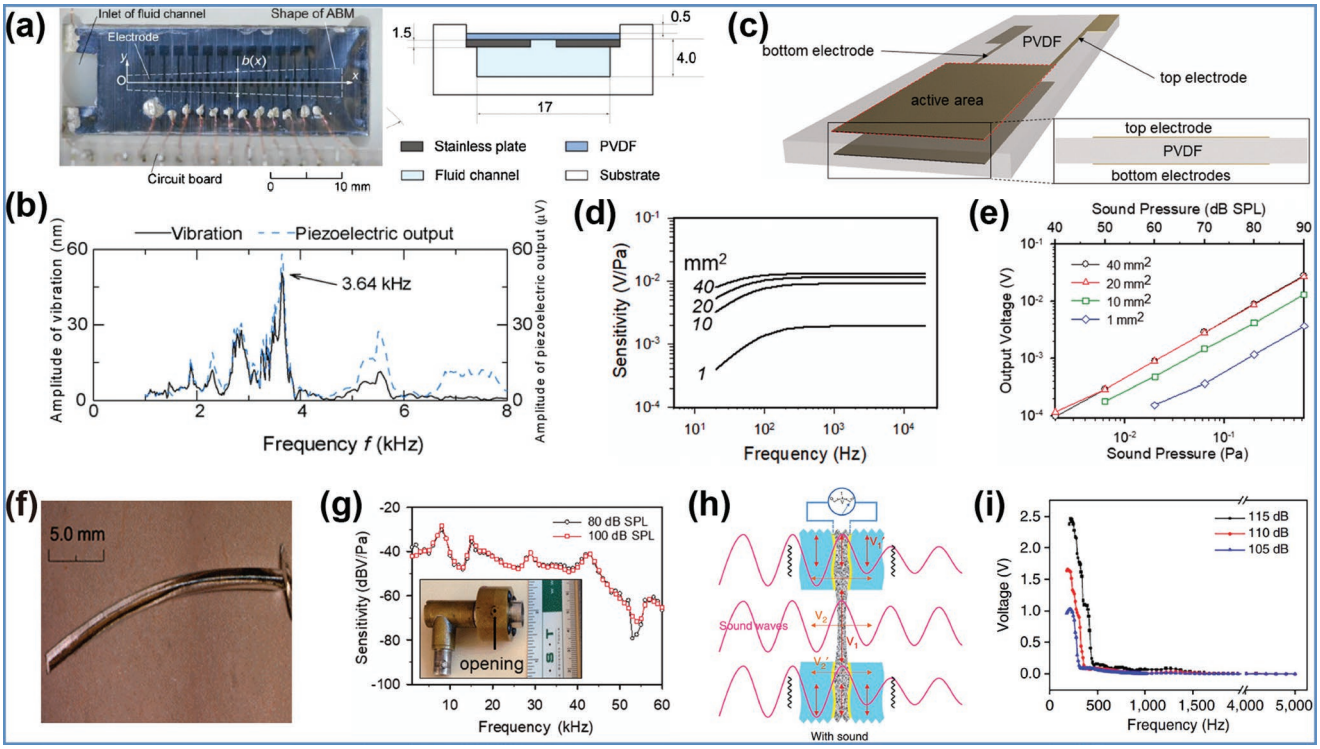

Figure 4. a) Structure of PVDF-based piezoelectric acoustic sensor. b) Resultant vibration amplitude and output voltage from Ch. 6, Ch. 12, and Ch. 18 in silicone oil of 1.75×10^{-3} Pa s. a,b) Reproduced with permission.[62] Copyright 2010, Elsevier. c) Schematic of acoustic device with electrodes patterned on top and bottom surface of PVDF film. d) Plot of expected sensitivity (V Pa^{-1}) of PVDF sensors with different active areas (40, 20, 10, and 1 mm^2), when amplified with a PARC amplifier with a gain of 1000. e) Output voltage of PVDF device according to different sound pressure. f) Photograph of the PVDF sensor encapsulated in PDMS cylinder. g) The sensitivity of PVDF acoustic device inside an air pressure chamber. Sound pressure was delivered at two levels, 80 and 100 dB SPL. Inset shows air pressure chamber. c–g) Reproduced with permission.[63] Copyright 2018, SAGE Publications Inc. h) The mechanism of piezoelectric conversion for the nanofiber acoustic sensors. i) Voltage outputs as a function of sound frequency. h,i) Reproduced under the terms of the Creative Commons Attribution 4.0 International License(http://creativecommons.org/licenses/by/4.0/).[64] Copyright 2016, Springer Nature.

Figure 4c illustrates the developed structure with electrode-overlapped area where the surface charges were created to generate the piezoelectric voltage. The width was longer than 100 μm to prevent fracture and open-circuit between electrodes. An amplifier (gain of 1000) was connected to the piezoelectric PVDF acoustic device to enhance the output voltage. The input voltage to the amplifier (V_{IN}) is expressed as following equation:

$$V_{IN} = V_{PVDF}\ \frac{Z_{I.AMP}}{Z_{I.AMP} + Z_{O.PVDF}} = V_{PVDF}\ \frac{\left(j\omega C_I + 1/R_I\right)^{-1}}{\left(j\omega C_I + 1/R_I\right)^{-1} + \left(j\omega C_{PVDF}\right)^{-1}} \quad (1)$$

V_{PVDF} is the PVDF-generated voltage. $Z_{O.PVDF}$ and $Z_{I.AMP}$ are the impedance of the acoustic sensor, and amplifier, respectively. Figure 4d shows the calculated sensitivity for different active areas. The expected output voltage increased as the active areas got larger from 1 to 40 mm^2, due to the decrease in impedance of the piezoelectric PVDF acoustic sensor. However, the sensitivity was lowered by saturated $Z_{I.AMP}$ as the frequency decreased. Figure 4e displays the measured voltage at each sound pressure level by taking the average output over the frequency range between 10 and 20 kHz. The piezoelectric sensors exhibited the sensitivity of 45, 43, 21, and 6 Pa^{-1} for active areas of 40, 20, 10, and 1 mm^2, respectively. To further prevent the shorting of electrodes in the fluid, the flexible piezoelectric acoustic sensor was encapsulated with an insulating material as shown in Figure 4f. The entire area of PVDF (the width of 0.5 mm and the length of 15 mm) was covered with metal electrodes to maximize the output voltage. The injection molding process was performed to last a lifetime of this piezoelectric acoustic device by using PDMS cylinder with a diameter of 0.6 mm. Figure 4g presents the sensitivity of the flexible acoustic sensor inserted into air pressure chamber. The PDMS-encapsulated self-powered sensor was stimulated by the sound in the range of 1–60 kHz at 80 and 100 dB SPL (decibel sound pressure level). The responsive voltage showed many peaks and valleys, due to the resonance of the PDMS tube.

Piezoelectric polymer acoustic devices have required the complex circuits to compensate the relatively low sensitivity.[63] Lang et al.[64] demonstrated the highly sensitive flexible piezoelectric nanofiber acoustic sensors by electrospinning the PVDF polymer. Figure 4h illustrates the structure and the mechanism of the piezoelectric sound sensing device. The electrospun piezoelectric nanofiber layer was contacted on the surface of electrode-coated PET films. The transparent plastic substrate was cut to create the circular hole where sensor directly received the acoustic waves. When the sound hits the acoustic sensor, the sound absorption caused the vibration of nanofiber web and PET sheets. The exposed through hole area vibrated more intensively in the in-plane (V_1) and cross-plane directions (V_2) compared to plastic film-covered nanofibers. While the cross-plane vibrations generated the piezoelectric charge, in-plane vibrations induced wave propagation through the whole device. This asymmetric vibration improved the sensitivity. Flexible PVDF acoustic sensor exhibited peak output voltage at 220 Hz as plotted in Figure 4i. The measured voltage fluctuated to nearly zero over the frequency range of 400–1500 Hz, after the voltage showed the abrupt decrease above 220 Hz. Although the piezoelectric acoustic sensor exhibited intrinsically high output voltage without amplification, the limited frequency coverage is insufficient to extract the human voice information.

4.2. Flexible Inorganic Piezoelectric Acoustic sensors

Inorganic-based flexible piezoelectric acoustic sensors have been spotlighted as a promising candidate for enhancing sensitivity and speech recognition, due to its superior inherent piezoelectric properties.[174] Jang et al.[65] developed a piezoelectric artificial basilar membrane (ABM) to detect sound with selective frequencies by mimicking the tonotopy of the cochlear. **Figure 5**a shows the piezoelectric aluminum nitride (AlN) cantilever array consisted of eight channels with different length from 600 to 1350 μm. The piezoelectric acoustic sensor detected each resonance frequency depending on its beam dimensions. Figure 5b describes the vibration pattern and displacement of piezoelectric acoustic cantilever unit at resonance frequency collected by the laser Doppler vibrometer (LDV), indicating the constrained oscillation of beam structure. The inset demonstrates that the resonance frequency, piezoelectric output, and displacement increased with inversely proportional relation to the cantilever length when the channel number decreased. Figure 5c represents the frequency response of each cantilever channel, exhibiting high quality factor (Q-factor) at resonance frequencies and thus discrete piezoelectric voltage signal over audible frequency range. The piezoelectric AlN acoustic sensor showed 0.354–1.22 mV Pa^{-1} of sensitivity in response to sound of the frequency range from 2.92 to 12.6 kHz.

Lee et al.[66] demonstrated an inorganic piezoelectric acoustic nanosensor (iPANS) by transferring PZT thin film onto trapezoidal silicone-based membrane (SM) to mimic biomimetic basilar membrane (BM). The conceptual image for organ of Corti and flexible piezoelectric acoustic sensor is described as shown in the Figure 5d, representing that the iPANS as artificial hair cells could produce piezoelectric potential by minute sound wave. The three iPANS were adhered on different location of an artificial trapezoidal silicone BM to accomplish a similar mechanism of the hair cells. Figure 5e provides an optical image of the SM with iPANS array attached at the apex, intermediate, and base region, respectively. While low frequency acoustic signal vibrates SM at the apex, high frequency sound induces the resonance in the base region. As presented in Figure 5f, the vibrational motion of the artificial BM (red plot) with short width corresponded to piezoelectric output voltage of the iPANS (blue plot) placed on the base. The flexible PZT thin film converted the vibration displacement (7.6 nm) into piezoelectric voltage (59.7 μV) at the resonance frequency of 1000 Hz. Furthermore, other iPANS array at apex and intermediate regions also successfully distinguished the low and middle frequency of 500 and 600 Hz within the targeted voice frequency range, respectively.

Han et al.[67] reported a multichannel flexible piezoelectric acoustic sensor (f-PAS) with outstanding sensitivity and the tunable multifrequency band by mimicking the curved BM structure. Multiple sensing channels were integrated on the piezoelectric PZT thin film to respond to tiny vibration over the voice frequency band. The experiments were performed

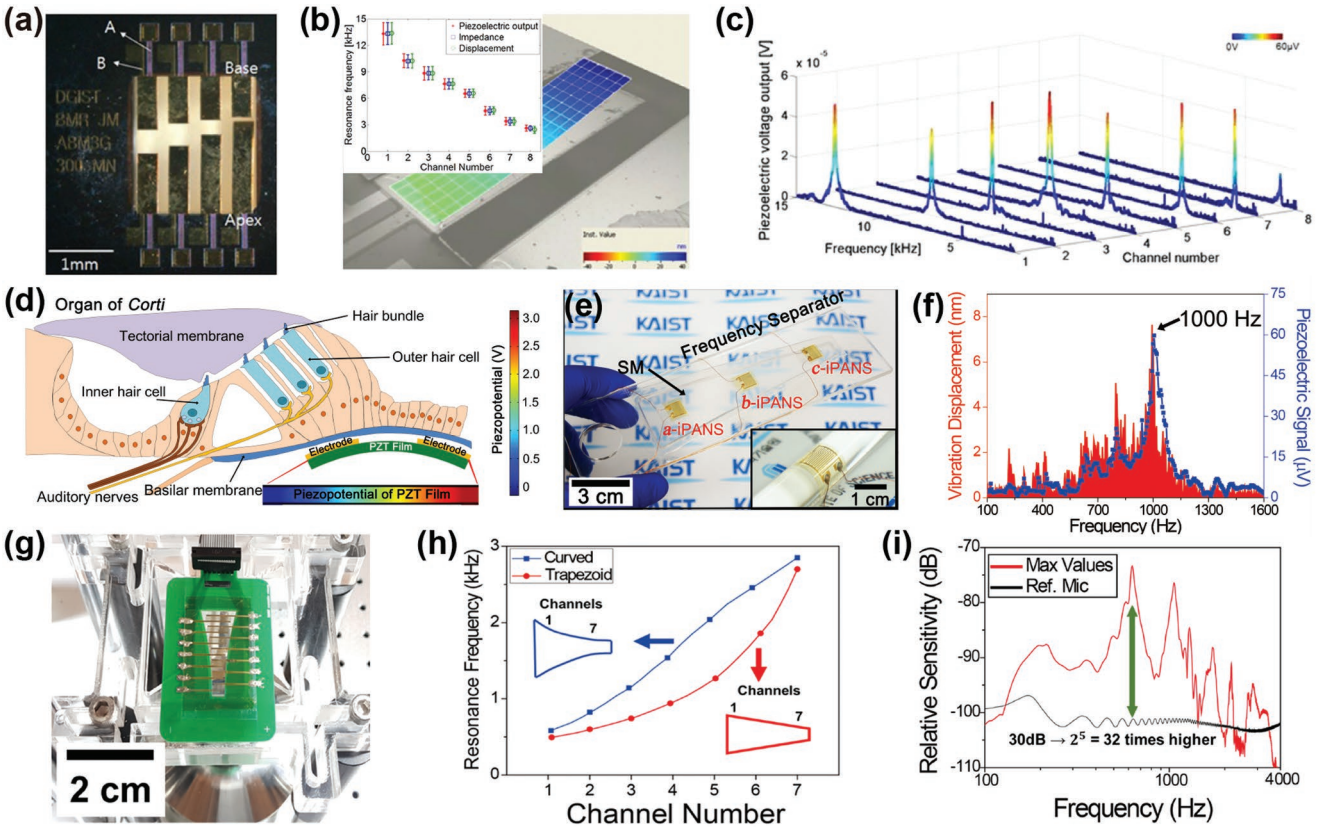

Figure 5. a) Optical image of the fabricated ABM with eight cantilevers by using aluminum nitride (AlN) as the active piezoelectric layer of MIM structure. b) A deformed cantilever by the resonant vibration utilizing a scanning laser Doppler vibrometer. c) Piezoelectric outputs for all cantilevers of an ABM. a–c) Reproduced under the terms of the Creative Commons Attribution 4.0 International License(http://creativecommons.org/licenses/by/4.0/).[65] Copyright 2016, Springer Nature. d) Conceptual illustration of organ of Corti in mammalian cochlea. The PZT film bent upwards generated piezopotential of ≈3 V when a tensile stress applied parallel to the surface of PZT film. e) Optical image of the frequency separator with iPANS. The inset shows a single iPANS unit conformally attached on cylinder. f) A vibration displacement and a piezoelectric signal generated by c-iPANS in the frequency range of 100–1600 Hz. A distribution of the vibration displacement and the piezoelectric signal were closely overlapped all over the frequency bandwidth. d–f) Reproduced with permission.[66] Copyright 2014, Wiley-VCH. g) A photograph of the multichannel f-PAS fixed on a PCB with a curvilinear trapezoidal hole. h) Comparison of calculated resonance frequencies of the curvilinear f-PAS according to the channel number. i) Relative sensitivity of the f-PAS and reference microphone over a voice frequency range from 100 to 4000 Hz. g–i) Reproduced with permission.[67] Copyright 2018, Elsevier.

to measure the performance of self-powered f-PAS attached on the printed circuit board (PCB) with a curvilinear shape hole as shown in Figure 5g. The sensor characteristics of voltage signals from the resonance of multichannels covered the entire voice frequency band from 100 to 4000 Hz with high sensitivity. FEA simulation was conducted to analysis the multiresonance in the curved f-PAS, which showed sound wave propagation similar in the BM of the cochlea. The first, second, and third mode of the resonance were sequentially generated at the region of base, intermediate, and apex areas in membrane, corresponding to resonance frequency of the 650, 1080, and 1440 Hz, respectively. The frequency of each resonant mode was selectively detected by multichannel where the oscillatory motion occurred. The resonance frequency of the flexible piezoelectric membrane is described as following equation

$$f_r = \frac{\omega}{2\pi} = C\frac{t}{l^2}\sqrt{\frac{E}{\rho}} \quad (2)$$

ω are angular frequency of the trapezoidal membrane. C, t, E, l, and ρ indicate the capacitance, thickness, Young's modulus, width, and density of the film, respectively. Figure 5h represents the linearly increasing resonant frequencies of the multichannel acoustic sensor, obtained from the quadratic decrease of the width length in the curvilinear membrane. Figure 5i illustrates the relative sensitivity of reference microphone and the f-PAS in the frequency range of 100–4000 Hz. By taking the maximum voltage signals among seven channels, the tunable multiresonant frequency band was achieved, which showed 32 times higher sensitivity (30 decibels difference) at 650 Hz compared to a condenser type reference microphone. **Table 2** presents the performance and challenge of the flexible piezoelectric acoustic sensors in Section 4.

5. Machine Learning Algorithms for Speech Processing

Machine learning (ML) is the algorithms and statistical models that help the computer to perform specific tasks without explicit instructions.[175–177] In machine learning, data is often split into training, validation and test data. Given a training set with N

Table 2. Summary of main parameters in piezoelectric acoustic sensors.

	[Ref. [62]]	[Ref. [63]]	[Ref. [64]]	[Ref. [65]]	[Ref. [66]]	[Ref. [67]]
Material	PVDF film	PVDF film	PVDF nanoweb	AlN film	PZT film	PZT film
Thickness of material [μm]	40	50	40	0.5	2	1
Device structure	Microbeam array + trapezoidal membrane	MIM + cylinder-shape encapsulation	MIM + substrate with trough hole	Cantilevers with different length	IDE + trapezoidal membrane	IDE + curvilinear trapezoidal membrane
The number of channel	24	1	1	8	3	7
Resonance frequency[a] [kHz]	1.4–4.9	<8	0.22	2.6–13.3	0.5–1	0.65–3
Size[b] [mm^2]	799	7.5	1200	7.05	300	<450
Sensitivity[c] [mV Pa^{-1}]	0.0253[d]	<0.045[e]	266[f] (@ 0.22 kHz) 8.9[g] (@ 1 kHz)	1.67 (@ 7.62 kHz)	[h]	>45[i] (@ 0.65 kHz)
Opportunities	Multiple resonance bands	Small size	High sensitivity	Multiple resonance bands Broad frequency response Small dimension	Multiple resonance bands Thin film	High sensitivity Multiple resonance bands Thin film
Challenges	Low sensitivity Large dimension	Low sensitivity One resonance band Thick film	One resonance band Narrow frequency response Low sensitivity at high frequency Large dimension	Low sensitivity	Narrow frequency response Large size	Large size

[a]The ranges were described in cases of multiresonance, since every resonance frequency was not mentioned; [b]Size was based on the area of piezoelectric materials; [c]Sensitivity = $V/P = V/(P_0 \times 10^{L_p/20})$, P_0 is the reference sound pressure of 0.00002 Pa and L_p is the SPL in decibel.[64] The sensitivity was calculated by using the peak voltage under monochromatic sound wave input; [d]The measurement frequency of peak voltage was not mentioned in the paper; [e]The sensitivity was converted by excluding the gain of 1000; [f]The value can be different by the condition of dB SPL. (60 mV Pa^{-1} was calculated by using the output voltage of 1.2 mV at 60 dB); [g]The sensitivity decreased at the frequency above 500 Hz, since the output voltage fluctuated to almost zero in the frequency range of 400–1500 Hz; [h]In this paper, the output voltage under monochromatic sound wave was not described; [i]The sensitivity at 650 Hz was predicted based on 1 kHz condition, since the output voltage under monochromatic sound input of 650 Hz was not measured.

samples $\mathbf{X} = \{x_1, x_2, \cdots, x_N\}$ and its label set $\mathbf{Y} = \{y_1, y_2, \cdots, y_N\}$, the machine learning algorithm can learn a function $f(\boldsymbol{X})$ to generate the output vector $\mathbf{Y}$ from the input $\mathbf{X}$. After function $f(\boldsymbol{X})$ is formulated in the training phase, the trained model can predict the labels for the new data examples. Validation data are used for model selection with the parameter that is estimated by training data. The validation loss/error is the loss/error that can be obtained by using the trained model for the validation data. The performance of the algorithm is evaluated on the test data. Machine learning has exhibited its marvelous capability of signal processing for image, video, and audio data.[178–183]

Speech processing has been developed to analyze the speech signals by computer for bilateral communication.[184] Recent researches in automatic speech recognition (ASR) systems allowed the smooth interaction between machine and human.[185,186] The ML techniques open avenues for performing various speech processing tasks such as speaker identification, speaker localization, language identification, speech translation, and speech recognition.[187–189] There are major machine learning methods for speech processing that have been discussed below including: the standard methods (GMM, SVM, HMM), and the deep learning approaches (DNN, CNN, RNN). The input features of these machine learning algorithms are the extracted data from the waveform speech, including mel-frequency cepstral coefficients (MFCC), spectrogram, constant-Q transforms, and auditory filter bank.[190–193] A spectrogram can be generated by bank of band-pass filters, Fourier transform, and wavelet transform. STFT is Fourier transforms of short segments of a long-time signal. Auditory filter bank is set of parallel band pass filters that are designed to mimic the frequency resolution of human hearing.

Gaussian mixture models (GMM) demonstrated the great performance for speaker recognition without extensive data, which allowed GMM one of the best choices in modeling speech data for decades.[194–196] GMM-based algorithm was adequate in showing the performance difference between the piezoelectric acoustic device and the commercialized microphone. GMM is a probabilistic model that represents normally distributed subpopulations among the overall population. **Figure 6**a illustrates the GMM model that consists of the mixture weights, the means and variances of Gaussians. Given GMM with K components, the k^{th} component is represented by Gaussian mean μ_k and variance σ_k^2. The mixture component weight ϕ_k is constrained with $\sum_{k=1}^{K} \phi_k = 1$. The probability density function is expressed by the following equation

$$p(x) = \sum_{k=1}^{K} \phi_k \, N\left(x|\mu_k, \sigma_k^2\right), N\left(x|\mu_k, \sigma_k^2\right) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{\left\{\frac{(x-\mu_k)^2}{2\sigma_k^2}\right\}} \quad (3)$$

The optimal parameters for Gaussian distributions are determined by expectation maximization (EM), which is an iterative algorithm for estimating maximum likelihood on the data.[197] For mixture models, EM consists of two main steps: E-step and M-step. Given a joint distribution $p(x, z; \theta)$ of the observed and

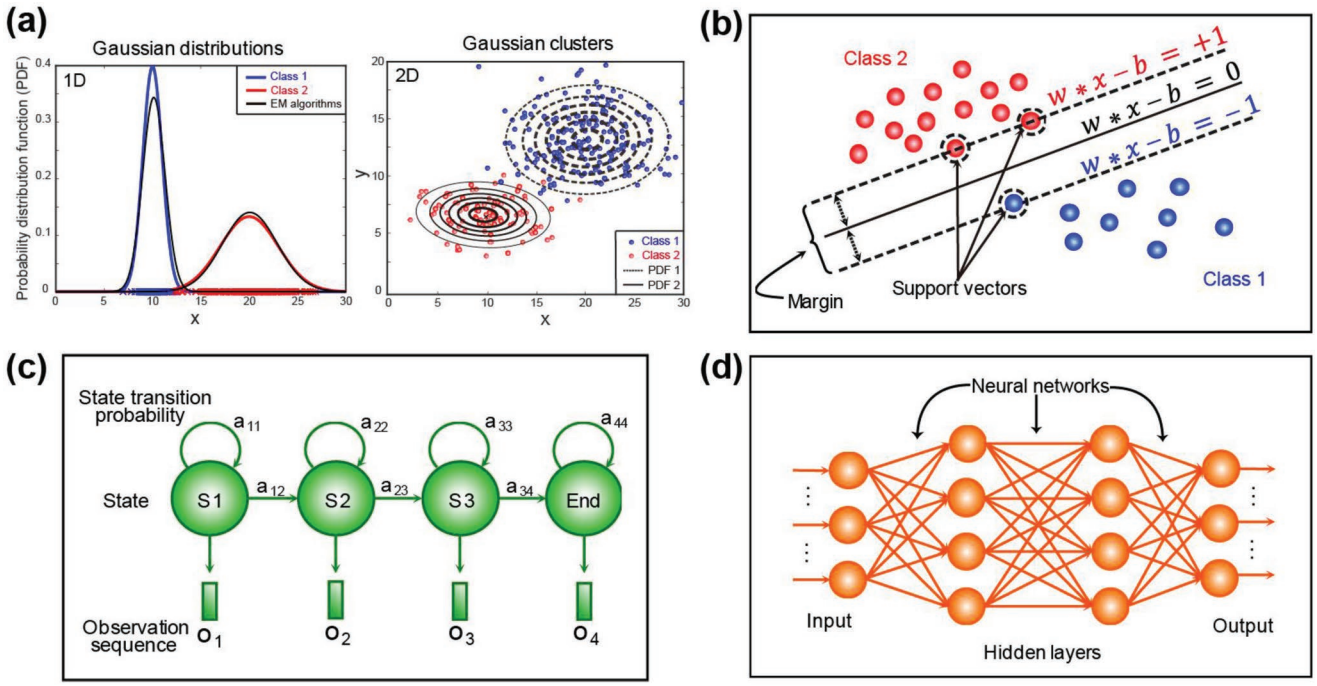

Figure 6. a) A simple GMM binary classifier. There are two classes represented by blue and red data points. In 1D plot, the original data were located in a horizontal line, showing their estimated probability distribution functions (PDFs) with EM algorithm (black line). The blue and red lines represent the exact PDFs. In 2D plot, the estimated PDFs were drawn by the contour plots and the original data of two classes were shown in the blue and red circles. b) Illustration of the SVM for binary classification. The black line represents the boundary that maximizes the margin between two classes. There are three support vectors that determine the boundary (data points within the dashed circles). c) HMM model with four states: S_1, S_2, S_3, and end. The state transition probability between state i and j is represented by a_{ij}. The sequence of the observation (o_1, o_2, o_3 and o_4) was generated by four states of HMM. d) A general neural network that is the premise for deep learning. This artificial neural network models the human brain. The model contains input, output layers, and many hidden layers which are connected to construct a large mesh network, however, there is no connection within the neurons of layer.

latent variables x and z, the maximum of the likelihood $p(x|\theta)$ is required. The parameters θ^0 is initialized from the beginning. The first step is expectation (E-step) that evaluate $p(z|x;\ \theta^{n-1})$ by computing the EM auxiliary function Q. The EM auxiliary function is determined by the expectation of the probability density function (PDF), $p(x,\ z;\ \theta)$

$$Q\left(\theta;\theta^{n-1}\right)=E_{Z|X;\theta^{n-1}}\left[\log p\left(x,z;\theta\right)\right] \tag{4}$$

where $n > 0$ denotes the iteration index, θ is the unknown parameters vector that contains μ_k, σ_k and ϕ_k, x is the observed data, and z is the random variable referred to as missing data. The second step is maximization (M-step) that maximizes the expectation computed from the E step with respect to the model parameters and then, update the model parameters μ_k, σ_k and ϕ_k:

$$\theta^{(n)}=\arg\max\left[Q\left(\theta;\theta^{n-1}\right)\right] \tag{5}$$

The M-step helps to re-estimate parameters based on the auxiliary function Q. The two above steps are repeated until the convergence condition satisfied. GMM+SVM model has been applied effectively for speaker verification task.[198] In this system, input to GMM are the MFCC features which were extracted from the speech data. GMM produces GMMs scores for these features and then trains an SVM classifier to map those score values to posterior probabilities. The final classification for speakers is determined by selecting the maximum probability.

SVM is an effective binary-classifier. Given two binary classes {−1, 1}, the decision boundary for two classes can be defined as follows

$$y\left(x\right)=W^T\ \phi\left(x\right)+b \tag{6}$$

where b is the bias, W is the model parameter, and $\phi(x)$ is the feature-space transformation. Assuming the training set is linearly separable with N samples, the desired solution should yield the best generalization among the many solutions of the form (6). The best generalization can lead to highly accurate performance on unseen data that was not used for training. The SVM approaches the solution with the concept of margin that is the perpendicular distance between the decision boundary ($y = 0$) and the closest point as shown in Figure 6b. A set of data points in dotted circles are support vectors helping SVM to locate the decision boundary. In addition, SVM obtains better generalization by applying the structural risk minimization (SRM) that can select an optimal model for finite dataset.[199] Although SVM is the state-of-the-art classifier, it requires a space complexity of $O(n^2)$, where n is the training set size to find the optimal solution of SVM problem.[200] SVM is an effective and powerful algorithm in the voice activity detection (VAD) problem.[201–203] VAD is the task that identifies whether speech is present or absent in

the audio data. The MFCC feature is extracted from the input audio. These MFCC feature vectors have been used as the input for training a binary classifier SVM with two classes of speech and nonspeech. The trained SVM classifier produces the scores to obtain the labels for test utterance frame. For speech recognition, the combination of SVM and HMM is necessary. HMM is used to represent the temporal dynamics of the input speech and align those features into states. An SVM classifier in each state of HMM is trained to classify the speech frames into single phonemes. These outputs are used for estimating the emission probabilities in the HMM decoder.

Hidden Markov models (HMMs) are based on the assumption that the very short segment of a voice signal (milliseconds) could be approximated as a stationary process.[204] HMMs are the important models for processing the sequential speech data. Figure 6c shows that the observation sequence was generated by four states with state transition probability. In HMMs, there are hidden states, each state generates the observation with emission probability. The Markov chain provides the joint distribution of random sequences. In speech recognition system, given the acoustic data X, HMM finds the most likely sequence of words as follows

$$\hat{w} = \arg\max p(w \mid \mathbf{X}) \tag{7}$$

where $\hat{w}$ is the desired output sequence, $\mathbf{X}$ is the feature vector extracted from the input waveform, w is the set of all possible sequences, and p denotes the probability. The processes of HMM include evaluation, decoding, and training. The evaluation phase computes the probability of observation sequences with the forward algorithm. The state sequence that maximizes the observation probability is found by Viterbi algorithm to decode the speech to text.[205] Finally, the most likely words are determined using the Baum-Welch (Forward–Backward) algorithm to customize the parameters to maximize the probability of observed sequences. In the GMM/HMM, GMM is used for emission (observation) probability at each hidden state of an HMM. The GMM/HMM-based method is the standard in the acoustic models.[206]

Deep learning can learn complex representations to provide high performance for tasks, while classical algorithms are based on hand-crafted features. Furthermore, deep learning based architectures exceeded the human performance in many tasks by using the mechanism of artificial neural networks, as illustrated in Figure 6d.[178,183,207,208] Multilayer perceptron (MLP) is a kind of neural network having input, hidden and output layer where each node in each layer (except the input layer) uses an activation function such as ReLU, sigmoid, and tanh.[209] MLP with more hidden layers can be considered as the deep neural networks (DNNs). MLP was used to approximate the maximum of a posterior probability.[210] This idea is to combine MLP and the standard HMM to utilize the advantage of MLP on the discriminative capability and the ability to estimate the posterior. A sequential MLP (SMLP) was introduced to apply a sequence of MLPs to recognize words.[211] Although the performance of recognizing words is increased by a sequence of MLPs, the difficulties in learning process of MLPs should be solved such as the heavy computational load, longer time for training, and greater liability of overfitting for the MLP with deeper architecture.[212] DNN can replace GMM method in speaker recognition by modeling the distribution of features extracted from the speech data.

Convolutional neural network (CNN) is a neural network that contains modules with convolutional layers and pooling layers.[213] This kind of architecture is more efficient in terms of memory and complexity compared to a conventional neural network, sharing the weights that led to avoid overfitting.[214] In the speech processing tasks such as speaker recognition and speech to text, CNN takes MFCC features (2D) as the input, and treats these features similar to image for classification. A typical CNN contains a series of the convolutional layers calculating feature maps via the corresponding learnable kernels. The max-pooling layers are behind convolutional layers, which perform down-sampling the feature maps by selecting the maximum value in the subregion that have values as input for pooling layers. The maximum value is the output from max-pooling layers. Using the subsampling layer, the model not only mitigates the computational cost but also obtains the robustness to the variation from the local pattern. The CNN network is trained to learn local speech features from MFCC input. The trained CNN produces the local speech feature vectors and these outputs are fed into RNN. This RNN is capable of learning history and future contextual features from the input (temporal dynamic of speech). A fully connected softmax layer (FC) is then applied for the RNN's output. This FC layer produces the probability for the given input speech. The network CNN/RNN is trained end-to-end using the stochastic gradient descent (SGD) optimizer with momentum. Recent studies showed that advancements in CNN enabled the replacement of GMM/HMM-based ASR system.[215]

Recurrent neural network (RNN) is a powerful deep learning architecture for sequential data like speech or text. The corresponding hidden state is obtained from the previous state, which can provide the output by using the input at a timestep. Given input $\mathbf{X} = \{x_1, x_2, \cdots, x_N\}$, RNN determines the hidden vector of sequence $H = (h_1, h_2, \cdots, h_T)$ and the output $\mathbf{Y} = (\gamma_1, \gamma_2, \cdots, \gamma_T)$ by the formula as follows

$$h_t = \mathrm{f}\left(W_{\mathrm{xh}} x_t + W_{\mathrm{hh}} h_{t-1} + b_{\mathrm{h}}\right), \gamma = W_{\mathrm{hy}}\ h_t + b_{\mathrm{y}} \tag{8}$$

b and $f(.)$ are the bias and the activation function.[216] W_{xh}, W_{hh}, and W_{hy} indicate weights of input-hidden, hidden-hidden, and hidden-output, respectively. RNN can learn and process the sequential speech data to predict what will be coming next by extracting relevant information. Graves et al.[216] achieved the best score in TIMIT phoneme recognition by using RNNs to train end-to-end. The speech data were processed to extract auditory filter-bank to form the input to the RNN model for phoneme recognition. RNNs suffer from vanishing gradients problem in training. The various studies have proposed to tackle that problem, such as long short-term memory (LSTM) or gate recurrent units (GRU).[217–219]

6. Speaker Recognition

Speaker recognition combined with artificial intelligence has attracted considerable attentions in the fields of biometric

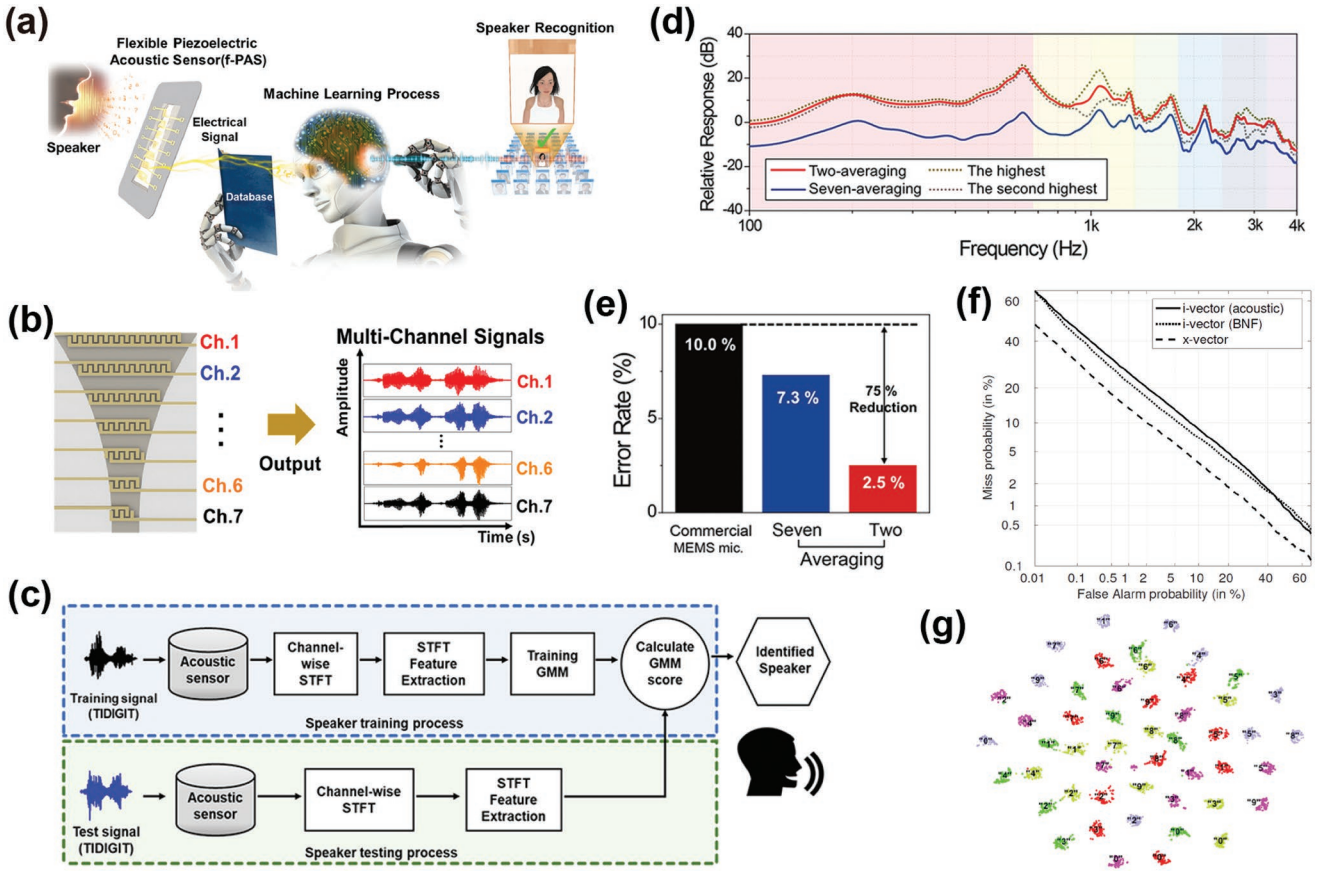

Figure 7. a) Overall illustration of the speaker recognition system based on machine learning. The f-PAS produces multi signals while the multichannel piezoelectric film vibrates according to the speaker's voice. The GMM algorithm-based training and test procedure are demonstrated by using the multi signals. Speaker recognition is performed by this process. b) Schematic depiction of flexible piezoelectric acoustic device with its output voice signal from each channel. c) The flow chart of speaker training and test process using the TIDIGITS dataset (90% of TIDIGITS used for training data, and 10% for testing data). d) Relative response of two-averaged (red), seven-averaged (blue), the highest and the second highest outputs over the voice range of 100 Hz to 4 kHz. All signals were normalized to 0 dB at 1 kHz. e) Comparison of recognition error rate of commercial phone and the f-PAS (seven-averaged and two-averaged) in the mixture number of 30. a–e) Reproduced with permission.[3] Copyright 2018, Elsevier. f) Comparison of detection error tradeoff curve among acoustic i-vector, phonetic bottleneck i-vector, and x-vector. Produced with permission.[228] Copyright 2018, IEEE. g) Visualized speaker embeddings with t-SNE. Five speaker embeddings were projected on two axes. Reproduced with permission.[229] Copyright 2018, IEEE.

security, bilateral AI communication, and IoT appliances.[16,17,21,22,26,27,42–45] Innovative personalized AI services such as AI secretary, authentication and security can be designed by analyzing the data of individual speaker such as preference and habit, depending on user voice.[220–222] Advanced acoustic sensors with optimized machine learning algorithms should be required for accurate speaker recognition.

Han et al.[3] demonstrated a speaker recognition utilizing flexible piezoelectric acoustic sensor (f-PAS) and machine learning process as illustrated in **Figure 7**a. A flexible piezoelectric thin film was fabricated by ILLO process to transfer PZT thin film onto plastic substrates. To cover entire speech frequency range of human utterance, seven channels of IDEs were printed on the piezoelectric thin film using photolithography technique.[223–225] This seven channel f-PAS can produce the multiple electrical signal with abundant voice information in a single speech. The Gaussian mixture modeling (GMM)-based machine learning algorithm was applied to store the frequency characteristics of multi voice signal generated by f-PAS in the training database. Finally, the random speeches were analyzed with the trained database to evaluate speaker recognition. Figure 7b illustrates the multichannel structure and electrical waveform produced by input sound pressure, respectively. This seven channel f-PAS was designed in the form of a piezoelectric thin film of concaved trapezoidal shape by mimicking the active BM within the human cochlear. Self-powered flexible acoustic sensor successfully expressed the output waveforms as a function of time, showing similar shape with standard speech of TIDIGIT. As shown in Figure 7c, the generated speech data from the multichannel f-PAS was processed through machine learning algorithm to recognize speaker. The GMM-based speaker process algorithm utilized the standard TIDIGITS dataset (77 speeches per each person, 40 speakers, and a total of 3080 sound data). Among these 3080 sound data, 90% (2800 sound data) were used for speaker training process and the other 10% (280 sound data) were utilized for testing process.

After the training and testing sound data were compared with short-time Fourier transform (STFT) features, the speaker was identified by calculating STFT absolute value.

Figure 7d shows the relative responses of f-PAS when the white noise was swept from 100 to 4000 Hz under the 94 dB sound pressure level. The red line was obtained by averaging the most and the second highest value among the seven channel output according to the relative frequency, while the blue line was acquired by averaging all the electrical output signals. The speaker recognition rate of f-PAS and a commercialized condenser type microphone were compared according to the number of mixture. The speaker recognition rates of the two-averaging f-PAS and reference microphone sharply increased until the 10 mixtures, and saturated to show 97.5% and 90% speaker recognition rate, respectively. As the number of mixture increased, the speaker recognition rate of the f-PAS exceeded the commercialized microphone since the output of f-PAS contained more sound information of Gaussian profiles rather than reference microphone.

Figure 7e presents the speaker recognition error rate of f-PAS and the reference microphone in the mixture number of 30. The two-averaged error rate was only 2.5%, showing the 75% reduction in comparison with that of the commercialized microphone. This outstanding speaker recognition rate was attributed to sufficient voice information of multichannel sensor outputs, and the highly sensitive performance of f-PAS. This work compared the recognition rate between the f-PAS and the commercial device, using speech processing of 40 speakers/3080 utterances. Other researches of triboelectric and ferroelectret acoustic sensors showed sensor sensitivity and simple voice verification.[52,53,226]

Most of the speaker recognition researches have been studied based on commercial microphones.[227–229] Snyder et al.[228] improved the speaker recognition by using DNN embedding architecture and commercial microphones. For speaker recognition, the x-vectors extracted from DNN were used, like i-vectors. Figure 7f shows the comparison of recognition result among acoustic i-vector, phonetic bottleneck i-vector, and x-vector. Acoustic i-vector is the GMM-based traditional i-vector system, however, phonetic bottleneck i-vector is based on ASR DNN acoustic model. The x-vector DNN required only speaker labels for training, although transcribed data are necessary for other systems. The DNN embedding x-vector system exhibited low EER of 4.16% after adding augmented microphone speech.

Novoselov et al.[229] analyzed the performance of CNN based speaker recognition system. The deep neural network extractors were trained in multitask mode with voice database recorded by the Android-based mobile phone, iOS-based mobile phone, and web-camera. In addition, the network was trained to discriminate between speakers and digits simultaneously. The speaker pronouncing a particular digit was assigned an individual class, which increased total amount of classes corresponding to the number of neurons. The deep CNN-based system showed the embedding discriminative capability as illustrated in Figure 7g. Randomly selected five speaker embeddings were visualized on two principal axes. The speaker and text discriminative embeddings enabled not only automatic validation of correctness but also usage of multi languages (English and Russian). Furthermore, the CNN system with multitask learning surpassed the classical algorithms to achieve 2.85% EER. Since the error rate of algorithms changes depending on the quality of recorded speech data,[230] speaker recognition rate can be enhanced by integrating the f-PAS with advanced machine learning algorithms.

7. Conclusion

We have summarized the recent progress of self-powered flexible piezoelectric acoustic sensors and machine learning algorithms.[3,62–67,175–219] Self-powered devices on plastics were developed using various piezoelectric materials such as ZnO NW array, PVDF nano fiber web, and perovskite thin films.[64,68,133,231,232] Highly sensitive flexible inorganic acoustic sensors covered entire voice frequency band (100–4000 Hz) depending on multichannel resonances of bioinspired trapezoidal membrane. In addition, multiple electrical signals from self-powered acoustic sensors were detected upon single utterance, which provided abundant speech information.[67] Classical machine learning techniques of speech processing were applied in simple recognition with TIMIT database.[233,234] The mechanism of artificial neural networks was utilized in machine learning to exceed the human performance in speech processing by deep learning. Speaker recognition of multichannel flexible piezoelectric acoustic sensors was demonstrated by utilizing GMM-based machine learning algorithms.[3] The f-PAS is expected to be improved by further development such as miniaturization, broadening the frequency band, filtering the surrounding noises, modularization, and efficient deep learning algorithms. Synergistic collaboration between advanced acoustic sensors and optimized speech recognition algorithm will open the new AI services such as biometric authentication and personalized IoT services.

Acknowledgements

Y.H.J., S.K.H., and H.S.W. contributed equally to this work. This work was supported by the Wearable Platform Materials Technology Center (WMC) funded by the National Research Foundation of Korea (NRF) Grant by the Korean Government (MSIT) (No. 2016R1A5A1009926), and Convergent Technology R&D Program for Human Augmentation through the National Research Foundation of Korea (NRF) funded by Ministry of Science and ICT (NRF-2020M3C1B8081519). This study was also supported by the National Research Foundation (NRF) of Korea (Grant No. NRF-2017R1A2A1A18071765), and Nano Material Technology Development Program (NRF-2016M3A7B4905621) through the NRF funded by the Ministry of Science, ICT and Future Planning.

Note: The acknowledgements section was corrected on September 1, 2020, after initial publication online.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

acoustic sensors, flexible piezoelectrics, machine learning algorithm, speech processing

Received: June 25, 2019
Revised: August 28, 2019
Published online: October 16, 2019

[1] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, M. Ekelid, *Science* **1995**, *270*, 303.
[2] E. Formisano, F. De Martino, M. Bonte, R. Goebel, *Science* **2008**, *322*, 970.
[3] J. H. Han, K. M. Bae, S. K. Hong, H. Park, J. H. Kwak, H. S. Wang, D. J. Joe, J. H. Park, Y. H. Jung, S. Hur, C. D. Yoo, K. J. Lee, *Nano Energy* **2018**, *53*, 658.
[4] L. Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions*, Cambridge University Press, Cambridge, UK **2006**.
[5] D. W. Massaro, in *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed., (Ed: J. D. Wright), Elsevier, London, UK **2015**, pp. 235–242.
[6] L. Millward, in *Understanding Occupational and Organizational Psychology* (Eds: M. Carmichael, S. Price, F. Pedroletti, F. Bryant), SAGE, Thousand Oaks, CA, USA **2005**, pp. 337–381.
[7] A. J. Maren, C. T. Harston, R. M. Pap, *Handbook of Neural Computing Applications*, Academic Press, San Diego, CA, USA **2016**.
[8] R. F. Lyon, *Human and Machine Hearing: Extracting Meaning from Sound*, Cambridge University Press, Cambridge, UK **2017**.
[9] S. Cho, S. Kang, A. Pandya, R. Shanker, Z. Khan, Y. Lee, J. Park, S. L. Craig, H. Ko, *ACS Nano* **2017**, *11*, 4346.
[10] T. Y. Choi, B. U. Hwang, B. Y. Kim, T. Q. Trung, Y. H. Nam, D. N. Kim, K. Eom, N. E. Lee, *ACS Appl. Mater. Interfaces* **2017**, *9*, 18022.
[11] J. Lee, P. Lee, H. Lee, D. Lee, S. S. Lee, S. H. Ko, *Nanoscale* **2012**, *4*, 6408.
[12] J. K. Song, D. Son, J. Kim, Y. J. Yoo, G. J. Lee, L. Wang, M. K. Choi, J. Yang, M. Lee, K. Do, J. H. Koo, N. Lu, J. H. Kim, T. Hyeon, Y. M. Song, D. H. Kim, *Adv. Funct. Mater.* **2017**, *27*, 1605286.
[13] W. Bauer, M. Hämmerle, S. Schlund, C. Vocke, *Procedia Manuf.* **2015**, *3*, 417.
[14] A. J. Choi, in *Proc. - 2014 IEEE Asian Solid-State Circuits Conf. A-SSCC 2014* (Ed: W. Z. Chen), IEEE, Piscataway, NJ, USA **2015**, pp. 5–8.
[15] A. Yarali, *IOT: Platforms, Connectivity, Applications and Services*, Nova Science Publisher, New York, NY **2018**.
[16] R. Dasgupta, *Voice User Interface Design*, Apress, Berkeley, CA, USA **2018**.
[17] I. Lee, K. Lee, *Bus. Horiz.* **2015**, *58*, 431.
[18] J. A. Stankovic, *IEEE Internet Things J.* **2014**, *1*, 3.
[19] G. B. Cogan, T. Thesen, C. Carlson, W. Doyle, O. Devinsky, B. Pesaran, *Nature* **2014**, *507*, 94.
[20] J. Li, L. Zhang, *J. Neurosci. Methods* **2010**, *193*, 373.
[21] T. Mandel, in *The Elements of User Interface Design Part 1 : Foundations of User Interface Design* (Ed: T. Hudson), Wiley, New York **1997**, pp. 1–6.
[22] A. Breen, H. H. Bui, R. Crouch, K. Farrell, F. Faubel, R. Gemello, W. F. Ganong, T. Haulick, R. M. Kaplan, C. L. Ortiz, P. F. Patel-Schneider, H. Quast, A. Ratnaparkhi, V. Sejnoha, J. Shen, P. Stubley, P. Van Mulbregt, in *Interactive Displays: Natural Human-Interface Technologies* (Ed: A. K. Bhowmik), Wiley, New York **2014**, pp. 107–163.
[23] Z. Ghahramani, *Nature* **2015**, *521*, 452.
[24] G. Brewka, *Knowl. Eng. Rev.* **2009**, *11*, 78.
[25] D. J. Ward, D. J. C. MacKay, *Nature* **2002**, *418*, 838.
[26] A. Bassi, M. Bauer, M. Fiedler, T. Kramp, R. van Kranenburg, S. Lange, S. Meissner, *Enabling Things to Talk: Designing IoT Solutions with the IoT Architectural Reference Model*, Springer, Berlin, Germany **2013**.
[27] A. Lele, in *Disruptive Technologies for the Militaries and Security*, (Eds: R. J. Howlett, L. C. Jain), Springer, Singapore **2019**, pp. 187–195.
[28] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, *Future Gener. Comput. Syst.* **2013**, *29*, 1645.
[29] M. Liu, X. Pu, C. Jiang, T. Liu, X. Huang, L. Chen, C. Du, J. Sun, W. Hu, Z. L. Wang, *Adv. Mater.* **2017**, *29*, 1703700.
[30] J. Cramer, F. Vogt, K. S. Booksh, in *Comprehensive Chemometrics* (Eds: S. D. Brown, R. Tauler, B. Walezak), Elsevier, Oxford, UK **2010**, pp. 357–376.
[31] M. Robertson, *Nature* **1976**, *262*, 435.
[32] D. C. Parkes, M. P. Wellman, *Science* **2015**, *349*, 267.
[33] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, H. J. W. L. Aerts, *Nat. Rev. Cancer* **2018**, *18*, 500.
[34] E. J. Topol, *Nat. Med.* **2019**, *25*, 44.
[35] K. H. Yu, A. L. Beam, I. S. Kohane, *Nat. Biomed. Eng.* **2018**, *2*, 719.
[36] A. Zanella, N. Bui, A. Castellani, L. Vangelista, M. Zorzi, *IEEE Internet Things J.* **2014**, *1*, 22.
[37] S. Davidoff, M. K. Lee, C. Yiu, J. Zimmerman, A. K. Dey, in *UbiComp 2006: Ubiquitous Computing* (Eds: P. Dourish, A. Friday), Springer, Berlin, Germany **2006**, pp. 19–34.
[38] D. S. Hecht, L. Hu, G. Irvin, *Adv. Mater.* **2011**, *23*, 1482.
[39] A. Doğan, E. Uzgur, in *Piezoelectric and Acoustic Materials for Transducer Applications* (Eds: A. Safari, E. K. Akdogan), Springer, Boston, MA, USA **2008**, pp. 341–371.
[40] C. Lang, J. Fang, H. Shao, H. Wang, G. Yan, X. Ding, T. Lin, *Nano Energy* **2017**, *35*, 146.
[41] G. Dahl, A.-R. Mohamed, G. Hinton, in *23th Int. Conf. on Neural Information Processing Systems*, Vol. *1* (Eds: J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, A. Culotta), Curran Associate Inc., Red Hook, NY, USA **2010**, p. 469.
[42] H. Beigi, *Fundamentals of Speaker Recognition*, Springer, Boston, MA, USA **2011**.
[43] T. Kinnunen, H. Li, *Speech Commun.* **2010**, *52*, 12.
[44] A. K. Jain, A. Ross, S. Prabhakar, *IEEE Trans. Circuits Syst. Video Technol.* **2004**, *14*, 4.
[45] A. K. Jain, P. Flynn, A. A. Ross, *Handbook of Biometrics*, Springer, Boston, MA, USA **2008**.
[46] S. Gong, W. Schwalb, Y. Wang, Y. Chen, Y. Tang, J. Si, B. Shirinzadeh, W. Cheng, *Nat. Commun.* **2014**, *5*, 3132.
[47] R. Fleury, D. Sounas, A. Alù, *Nat. Commun.* **2015**, *6*, 5905.
[48] J. Yang, J. Chen, Y. Liu, W. Yang, Y. Su, Z. L. Wang, *ACS Nano* **2014**, *8*, 2649.
[49] B. Park, J. Kim, D. Kang, C. Jeong, K. S. Kim, J. U. Kim, P. J. Yoo, T. Kim, *Adv. Mater.* **2016**, *28*, 8068.
[50] L. Q. Tao, K. N. Zhang, H. Tian, Y. Liu, D. Y. Wang, Y. Q. Chen, Y. Yang, T. L. Ren, *ACS Nano* **2017**, *11*, 8790.
[51] J. Yang, J. Chen, Y. Su, Q. Jing, Z. Li, F. Yi, X. Wen, Z. Wang, Z. L. Wang, *Adv. Mater.* **2015**, *27*, 1316.
[52] W. Li, D. Torres, R. Díaz, Z. Wang, C. Wu, C. Wang, Z. Lin Wang, N. Sepúlveda, *Nat. Commun.* **2017**, *8*, 15310.
[53] H. Guo, X. Pu, J. Chen, Y. Meng, M.-H. Yeh, G. Liu, Q. Tang, B. Chen, D. Liu, S. Qi, C. Wu, C. Hu, J. Wang, Z. L. Wang, *Sci. Rob.* **2018**, *3*, eaat2516.
[54] T. Yamada, Y. Hayamizu, Y. Yamamoto, Y. Yomogida, A. Izadi-Najafabadi, D. N. Futaba, K. Hata, *Nat. Nanotechnol.* **2011**, *6*, 296.
[55] J. Hirschberg, C. D. Manning, *Science* **2015**, *349*, 261.
[56] N. Jones, *Nature* **2014**, *505*, 146.
[57] R. J. Zatorre, A. C. Evans, E. Meyer, A. Gjedde, *Science* **1992**, *256*, 846.
[58] S. A. Fulop, in *Speech Spectrum Analysis* (Ed: C. Baumann), Springer, Berlin/Heidelberg, Germany **2011**, pp. 41–68.
[59] M. J. Mardini, A. Mehta, in *PACS: A Guide to the Digital Revolution* 2nd ed., (Eds: K. J. Dreyer, D. S. Hirschorn, J. H. Thrall, A. Mehta), Springer-Verlag, New York **2006**, pp. 467–482.
[60] A. van Berlo, K. Hallenborg, J. M. C. Rodríguez, D. I. Tapia, P. Novais, *Ambient Intelligence - Software and Applications*, Springer, Heidelberg, Germany **2012**.
[61] H. Ding, X. Shu, Y. Jin, T. Fan, H. Zhang, *Nanoscale* **2019**, *11*, 5839.

[62] H. Shintaku, T. Nakagawa, D. Kitagawa, H. Tanujaya, S. Kawano, J. Ito, *Sens. Actuators, A.* **2010**, *158*, 183.
[63] S. Park, X. Guan, Y. Kim, F. (Pete) X. Creighton, E. Wei, I. Kymissis, H. H. Nakajima, E. S. Olson, *Trends Hear.* **2018**, *22*, 1.
[64] C. Lang, J. Fang, H. Shao, X. Ding, T. Lin, *Nat. Commun.* **2016**, *7*, 11108.
[65] J. Jang, J. Lee, S. Woo, D. J. Sly, L. J. Campbell, J. H. Cho, S. J. O'Leary, M. H. Park, S. Han, J. W. Choi, J. Hun Jang, H. Choi, *Sci. Rep.* **2015**, *5*, 12447.
[66] H. S. Lee, J. Chung, G. T. Hwang, C. K. Jeong, Y. Jung, J. H. Kwak, H. Kang, M. Byun, W. D. Kim, S. Hur, S. H. Oh, K. J. Lee, *Adv. Funct. Mater.* **2014**, *24*, 6914.
[67] J. H. Han, J. H. Kwak, D. J. Joe, S. K. Hong, H. S. Wang, J. H. Park, S. Hur, K. J. Lee, *Nano Energy* **2018**, *53*, 198.
[68] S. Wang, T. Zhang, K. Li, S. Ma, M. Chen, P. Lu, L. Wei, *Adv. Electron. Mater.* **2017**, *3*, 1600449.
[69] J. W. Weigold, T. J. Brosnihan, J. Bergeron, X. Zhang, in *Proc. IEEE Int. Conf. Micro Electro Mechanical Systems* (Eds: T. Akin, R. Zengerle), IEEE, Piscataway, NJ, USA **2006**, pp. 86–89.
[70] B. Hagerman, *Scand. Audiol.* **1982**, *11*, 79.
[71] J. M. Festen, R. Plomp, *J. Acoust. Soc. Am.* **1990**, *88*, 1725.
[72] M. W. Skinner, *Ann. Otol. Rhinol. Laryngol.* **2003**, *112*, 4.
[73] B. W. Y. Hornsby, T. A. Ricketts, *J. Acoust. Soc. Am.* **2006**, *119*, 1752.
[74] B. B. Monson, E. J. Hunter, A. J. Lotto, B. H. Story, *Front. Psychol.* **2014**, *5*, 587.
[75] N. Mohamad, *Engineering* **2010**, *2*, 762.
[76] C. H. Je, J. Lee, W. S. Yang, J. Kim, Y. H. Cho, *J. Micromech. Microeng.* **2013**, *23*, 055018.
[77] W. Sui, W. Zhang, K. Song, C. H. Cheng, Y. K. Lee, in *TRANSDUCERS 2017 - 19th Int. Conf. Solid-State Sensors, Actuators Microsystems*, Vol. *1* (Ed: W. Fang), IEEE, Piscataway, NJ, USA **2017**, p. 946.
[78] X. Li, R. Lin, H. Kek, J. Miao, Q. Zou, *Sens. Actuators, A* **2001**, *92*, 257.
[79] R. Kressmann, M. Klaiber, G. Hess, *Sens. Actuators, A* **2002**, *100*, 301.
[80] Y. Gao, J. Song, S. Li, C. Elowsky, Y. Zhou, S. Ducharme, Y. M. Chen, Q. Zhou, L. Tan, *Nat. Commun.* **2016**, *7*, 12316.
[81] M. N. Niu, E. S. Kim, *J. Microelectromech. Syst.* **2003**, *12*, 892.
[82] D. J. Lipomi, M. Vosgueritchian, B. C. K. Tee, S. L. Hellstrom, J. A. Lee, C. H. Fox, Z. Bao, *Nat. Nanotechnol.* **2011**, *6*, 788.
[83] L. Viry, A. Levi, M. Totaro, A. Mondini, V. Mattoli, B. Mazzolai, L. Beccai, *Adv. Mater.* **2014**, *26*, 2659.
[84] X. Wang, T. Li, J. Adams, J. Yang, *J. Mater. Chem. A* **2013**, *1*, 3580.
[85] J. Y. Sun, C. Keplinger, G. M. Whitesides, Z. Suo, *Adv. Mater.* **2014**, *26*, 7608.
[86] W. Bracke, P. Merken, R. Puers, C. Van Hoof, *Sens. Actuators, A* **2005**, *125*, 25.
[87] S. Walser, C. Siegel, M. Winter, G. Feiertag, M. Loibl, A. Leidl, *Sens. Actuators, A* **2016**, *247*, 663.
[88] A. Passian, P. G. Evans, V. K. Varma, T. L. Ferrell, T. Thundat, *Rev. Sci. Instrum.* **2003**, *74*, 1031.
[89] S. Kon, K. Oldham, R. Horowitz, *Proc. SPIE* **2007**, *2007*, 65292V.
[90] S. Luo, T. Liu, *Adv. Mater.* **2013**, *25*, 5650.
[91] J. C. Suhling, R. C. Jaeger, *IEEE Sens. J.* **2001**, *1*, 14.
[92] Z. L. Wang, T. Jiang, L. Xu, *Nano Energy* **2017**, *39*, 9.
[93] V. Nguyen, R. Yang, *Nano Energy* **2013**, *2*, 604.
[94] X. Wen, Y. Su, Y. Yang, H. Zhang, Z. L. Wang, *Nano Energy* **2014**, *4*, 150.
[95] Y. Zhang, Z. Yang, K. Song, X. Pang, B. Shangguan, *Friction* **2013**, *1*, 259.
[96] K. Hashimoto, *Surface Acoustic Wave Devices in Telecommunications*, Springer, Berlin/Heidelberg, Germany **2000**.
[97] S. Shiokawa, J. Kondoh, *Jpn. J. Appl. Phys.* **2004**, *43*, 2799.
[98] D. B. Go, M. Z. Atashbar, Z. Ramshani, H. C. Chang, *Anal. Methods* **2017**, *9*, 4112.
[99] G. Von Békésy, *Nature* **1970**, *225*, 1207.
[100] A. F. Huxley, *Nature* **1969**, *221*, 935.
[101] D. K. Chan, A. J. Hudspeth, *Nat. Neurosci.* **2005**, *8*, 149.
[102] A. J. Hudspeth, *Nature* **1989**, *341*, 397.
[103] T. D. Griffiths, S. Uppenkamp, I. Johnsrude, O. Josephs, R. D. Patterson, *Nat. Neurosci.* **2001**, *4*, 633.
[104] P. Belin, R. J. Zatorre, P. Lafallie, P. Ahad, B. Pike, *Nature* **2000**, *403*, 309.
[105] T. D. Nguyen, S. Mao, Y. W. Yeh, P. K. Purohit, M. C. McAlpine, *Adv. Mater.* **2013**, *25*, 946.
[106] J. Li, L. Deng, R. Haeb-Umbach, Y. Gong, in *Robust Automatic Speech Recognition* (Eds: J. Li, L. Deng, R. Haeb-Umbach, Y. Gong), Academic Press, San Diego, CA, USA **2015**, pp. 9–40.
[107] D. Jurafsky, J. H. Martin, in *Speech and Language Processing* (Eds: S. Russell, P. Norvig), Pearson Education, Upper Saddle River, NJ, USA **2008**, p. 988.
[108] T. Hain, P. N. Garner, in *Multimodal Signal Processing* (Eds: S. Renals, H. Bourlard, J. Carletta, A. Popescu-Belis), Cambridge University Press, Cambridge, UK **2012**, pp. 56–83.
[109] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, B. Kingsbury, T. Sainath, *IEEE Signal Process. Mag.* **2012**, *29*, 82.
[110] G. E. Dahl, D. Yu, L. Deng, A. Acero, *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 30.
[111] A. R. Mohamed, G. E. Dahl, G. Hinton, *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 14.
[112] A. M. Darcy, A. K. Louie, L. W. Roberts, *J. Am. Med. Assoc.* **2016**, *315*, 551.
[113] T. N. Sainath, A. R. Mohamed, B. Kingsbury, B. Ramabhadran, in *ICASSP, IEEE Int. Conf. on Acoustics, Speech and Signal Processing - Proc.* (Eds: M. Adams, V. Zhao), IEEE, Piscataway, NJ, USA **2013**, pp. 8614–8618.
[114] T. Mikolov, A. Deoras, D. Povey, L. Burget, J. Černocký, in *2011 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2011, Proc.* (Eds: T. Hain, K. Yu), IEEE, Piscataway, NJ, USA **2011**, pp. 196–201.
[115] X. Wang, Y. Gu, Z. Xiong, Z. Cui, T. Zhang, *Adv. Mater.* **2014**, *26*, 1336.
[116] J. Park, Y. Lee, J. Hong, M. Ha, Y. Do Jung, H. Lim, S. Y. Kim, H. Ko, *ACS Nano* **2014**, *8*, 4689.
[117] M. Jian, K. Xia, Q. Wang, Z. Yin, H. Wang, C. Wang, H. Xie, M. Zhang, Y. Zhang, *Adv. Funct. Mater.* **2017**, *27*, 1606066.
[118] D. Damjanovic, *Rep. Prog. Phys.* **1998**, *61*, 1267.
[119] J. Yang, *An Introduction to the Theory of Piezoelectricity*, Springer, Boston, MA, USA **2006**.
[120] H. Fu, R. E. Cohen, *Nature* **2000**, *403*, 281.
[121] Z. Wang, J. Hu, A. P. Suryavanshi, K. Yum, M. F. Yu, *Nano Lett.* **2007**, *7*, 2966.
[122] B. Dudem, D. H. Kim, L. K. Bharat, J. S. Yu, *Appl. Energy* **2018**, *230*, 865.
[123] G. T. Hwang, M. Byun, C. K. Jeong, K. J. Lee, *Adv. Healthcare Mater.* **2015**, *4*, 646.
[124] C. Dagdeviren, Y. Shi, P. Joe, R. Ghaffari, G. Balooch, K. Usgaonkar, O. Gur, P. L. Tran, J. R. Crosby, M. Meyer, Y. Su, R. C. Webb, A. S. Tedesco, M. J. Slepian, Y. Huang, J. A. Rogers, *Nat. Mater.* **2015**, *14*, 728.
[125] Q. Sun, W. Seung, B. J. Kim, S. Seo, S. W. Kim, J. H. Cho, *Adv. Mater.* **2015**, *27*, 3411.
[126] Z. Chen, Z. Wang, X. Li, Y. Lin, N. Luo, M. Long, N. Zhao, J. Bin Xu, *ACS Nano* **2017**, *11*, 4507.
[127] R. Yu, C. Pan, J. Chen, G. Zhu, Z. L. Wang, *Adv. Funct. Mater.* **2013**, *23*, 5868.
[128] J. Zhang, C. Wang, C. Bowen, *Nanoscale* **2014**, *6*, 13314.

[129] G. T. Hwang, H. Park, J. H. Lee, S. Oh, K. Il Park, M. Byun, H. Park, G. Ahn, C. K. Jeong, K. No, H. Kwon, S. G. Lee, B. Joung, K. J. Lee, *Adv. Mater.* **2014**, *26*, 4880.
[130] C. R. Bowen, H. A. Kim, P. M. Weaver, S. Dunn, *Energy Environ. Sci.* **2014**, *7*, 25.
[131] H. Cui, R. Hensleigh, D. Yao, D. Maurya, P. Kumar, M. G. Kang, S. Priya, X. (R.) Zheng, *Nat. Mater.* **2019**, *18*, 234.
[132] Z. L. Wang, J. Song, *Science* **2006**, *312*, 242.
[133] R. Yang, Y. Qin, L. Dai, Z. L. Wang, *Nat. Nanotechnol.* **2009**, *4*, 34.
[134] W. Liu, A. Zhang, Y. Zhang, Z. Lin Wang, *Nano Energy* **2015**, *14*, 355.
[135] Z. L. Wang, *Adv. Mater.* **2012**, *24*, 4632.
[136] D. H. Kim, B. Dudem, J. S. Yu, *ACS Sustainable Chem. Eng.* **2018**, *6*, 8525.
[137] Y. H. Ko, G. Nagaraju, S. H. Lee, J. S. Yu, *ACS Appl. Mater. Interfaces* **2014**, *6*, 6631.
[138] Y. Qin, X. Wang, Z. L. Wang, *Nature* **2008**, *451*, 809.
[139] S. Xu, Y. Qin, C. Xu, Y. Wei, R. Yang, Z. L. Wang, *Nat. Nanotechnol.* **2010**, *5*, 366.
[140] H. D. Espinosa, R. A. Bernal, M. Minary-Jolandan, *Adv. Mater.* **2012**, *24*, 4656.
[141] Z. R. Tian, J. A. Voigt, J. Liu, B. Mckenzie, M. J. Mcdermott, M. A. Rodriguez, H. Konishi, H. Xu, *Nat. Mater.* **2003**, *2*, 821.
[142] X. Wang, J. Zhou, J. Song, J. Liu, N. Xu, Z. L. Wang, *Nano Lett.* **2006**, *6*, 2768.
[143] J. Song, J. Zhou, Z. L. Wang, *Nano Lett.* **2006**, *6*, 1656.
[144] Z. L. Wang, *Adv. Mater.* **2007**, *19*, 889.
[145] W. Wu, L. Wang, Y. Li, F. Zhang, L. Lin, S. Niu, D. Chenet, X. Zhang, Y. Hao, T. F. Heinz, J. Hone, Z. L. Wang, *Nature* **2014**, *514*, 470.
[146] R. E. Cohen, *Nature* **1992**, *358*, 136.
[147] K. Il Park, J. H. Son, G. T. Hwang, C. K. Jeong, J. Ryu, M. Koo, I. Choi, S. H. Lee, M. Byun, Z. L. Wang, K. J. Lee, *Adv. Mater.* **2014**, *26*, 2514.
[148] J. F. Scott, *Science* **2007**, *315*, 954.
[149] G.-T. Hwang, V. Annapureddy, J. H. Han, D. J. Joe, C. Baek, D. Y. Park, D. H. Kim, J. H. Park, C. K. Jeong, K.-I. Park, J.-J. Choi, D. K. Kim, J. Ryu, K. J. Lee, *Adv. Energy Mater.* **2016**, *6*, 1600237.
[150] K. Il Park, S. Xu, Y. Liu, G. T. Hwang, S. J. L. Kang, Z. L. Wang, K. J. Lee, *Nano Lett.* **2010**, *10*, 4939.
[151] J. H. Lee, K. Y. Lee, B. Kumar, N. T. Tien, N. E. Lee, S. W. Kim, *Energy Environ. Sci.* **2013**, *6*, 169.
[152] A. I. Kingon, S. Srinivasan, *Nat. Mater.* **2005**, *4*, 233.
[153] J. Briscoe, S. Dunn, *Nano Energy* **2015**, *14*, 15.
[154] V. Nagarajan, A. Roytburd, A. Stanishevsky, S. Prasertchoung, T. Zhao, L. Chen, J. Melngailis, O. Auciello, R. Ramesh, *Nat. Mater.* **2003**, *2*, 43.
[155] C. L. Jia, S. B. Mi, K. Urban, I. Vrejoiu, M. Alexe, D. Hesse, *Nat. Mater.* **2008**, *7*, 57.
[156] T. Sluka, A. K. Tagantsev, D. Damjanovic, M. Gureev, N. Setter, *Nat. Commun.* **2012**, *3*, 748.
[157] G. T. Hwang, Y. Kim, J. H. Lee, S. Oh, C. K. Jeong, D. Y. Park, J. Ryu, H. Kwon, S. G. Lee, B. Joung, D. Kim, K. J. Lee, *Energy Environ. Sci.* **2015**, *8*, 2677.
[158] G. T. Hwang, J. Yang, S. H. Yang, H. Y. Lee, M. Lee, D. Y. Park, J. H. Han, S. J. Lee, C. K. Jeong, J. Kim, K. Il Park, K. J. Lee, *Adv. Energy Mater.* **2015**, *5*, 1500051.
[159] D. Damjanovic, *J. Am. Ceram. Soc.* **2005**, *88*, 2663.
[160] M. Ahart, M. Somayazulu, R. E. Cohen, P. Ganesh, P. Dera, H. K. Mao, R. J. Hemley, Y. Ren, P. Liermann, Z. Wu, *Nature* **2008**, *451*, 545.
[161] F. Li, D. Lin, Z. Chen, Z. Cheng, J. Wang, C. Li, Z. Xu, Q. Huang, X. Liao, L. Q. Chen, T. R. Shrout, S. Zhang, *Nat. Mater.* **2018**, *17*, 349.
[162] Y. Liu, H. Aziguli, B. Zhang, W. Xu, W. Lu, J. Bernholc, Q. Wang, *Nature* **2018**, *562*, 96.
[163] Z. L. Wang, *Adv. Mater.* **2012**, *24*, 280.
[164] X. He, Y. Zi, H. Yu, S. L. Zhang, J. Wang, W. Ding, H. Zou, W. Zhang, C. Lu, Z. L. Wang, *Nano Energy* **2017**, *39*, 328.
[165] F. Narita, M. Fox, *Adv. Eng. Mater.* **2018**, *20*, 1700743.
[166] L. Lin, Q. Jing, Y. Zhang, Y. Hu, S. Wang, Y. Bando, R. P. S. Han, Z. L. Wang, *Energy Environ. Sci.* **2013**, *6*, 1164.
[167] R. Yu, C. Pan, Z. L. Wang, *Energy Environ. Sci.* **2013**, *6*, 494.
[168] R. Zhou, G. Hu, R. Yu, C. Pan, Z. L. Wang, *Nano Energy* **2015**, *12*, 588.
[169] S. Niu, Y. Hu, X. Wen, Y. Zhou, F. Zhang, L. Lin, S. Wang, Z. L. Wang, *Adv. Mater.* **2013**, *25*, 3701.
[170] D. Y. Park, D. J. Joe, D. H. Kim, H. Park, J. H. Han, C. K. Jeong, H. Park, J. G. Park, B. Joung, K. J. Lee, *Adv. Mater.* **2017**, *29*, 1702308.
[171] S. Lim, D. Son, J. Kim, Y. B. Lee, J. K. Song, S. Choi, D. J. Lee, J. H. Kim, M. Lee, T. Hyeon, D. H. Kim, *Adv. Funct. Mater.* **2015**, *25*, 375.
[172] L. Persano, C. Dagdeviren, Y. Su, Y. Zhang, S. Girardo, D. Pisignano, Y. Huang, J. A. Rogers, *Nat. Commun.* **2013**, *4*, 1633.
[173] X. Wang, H. Zhang, R. Yu, L. Dong, D. Peng, A. Zhang, Y. Zhang, H. Liu, C. Pan, Z. L. Wang, *Adv. Mater.* **2015**, *27*, 2324.
[174] Y. M. You, W. Q. Liao, D. Zhao, H. Y. Ye, Y. Zhang, Q. Zhou, X. Niu, J. Wang, P. F. Li, D. W. Fu, Z. Wang, S. Gao, K. Yang, J. M. Liu, J. Li, Y. Yan, R. G. Xiong, *Science* **2017**, *357*, 306.
[175] M. I. Jordan, T. M. Mitchell, *Science* **2015**, *349*, 255.
[176] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, *Nature* **2015**, *518*, 529.
[177] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, *Nature* **2016**, *529*, 484.
[178] M. Tan, Q. V. Le, in *36th Int. Conf. on Machine Learning*, Vol. *97* (Eds: K. Chaudhuri, R. Salakhutdinov), PMLR, **2019**, p. 6105.
[179] Y. Huang, Y. Cheng, D. Chen, H. Lee, J. Ngiam, Q. V. Le, Z. Chen, arXiv preprint **2018**, arXiv:1811.06965.
[180] K. Simonyan, A. Zisserman, in *27th Int. Conf. on Neural Information Processing Systems*, Vol. *1* (Eds: Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger), NIPS, San Diego, CA, USA **2014**, p. 568.
[181] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, F. F. Li, in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, (Eds: E. Mortensen, S. Fidler), IEEE, Piscataway, NJ, USA **2014**, p. 1725.
[182] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, in *9th ISCA Speech Synthesis Workshop* (Eds: A. Bonafonte, K. Prahallad), ISCA, Baixas, France **2016**, p. 125.
[183] Y. Wang, R. J. Skerry-Ryan, D. Stanton, Y. Wu, R. J. Weiss, N. Jaitly, Z. Yang, Y. Xiao, Z. Chen, S. Bengio, Q. Le, in *Proc. Annual Conf. Int. Speech Communication Association INTERSPEECH* (Ed: F. Lacerda), ISCA, Baixas, France **2017**, p. 4006.
[184] P. A. Abhang, B. W. Gawali, S. C. Mehrotra, *Introduction to EEG- and Speech-Based Emotion Recognition*, Academic Press, San Diego, CA, USA **2016**, pp. 1–17.
[185] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, J. Chen, J. Chen, Z. Chen, M. Chrzanowski, A. Coates, G. Diamos, K. Ding, N. Du, E. Elsen, J. Engel, W. Fang, L. Fan, C. Fougner, L. Gao, C. Gong, A. N. Hannun, T. Han, L. V. Johannes, B. Jiang, C. Ju, B. Jun, P. Legresley, L. Lin, J. Liu, Y. Liu, W. Li, X. Li, D. Ma,

S. Narang, A. Ng, S. Ozair, Y. Peng, R. Prenger, S. Qian, Z. Quan, J. Raiman, V. Rao, S. Satheesh, D. Seetapun, S. Sengupta, K. Srinet, A. Sriram, H. Tang, L. Tang, C. Wang, J. Wang, K. Wang, Y. Wang, Z. Wang, Z. Wang, S. Wu, L. Wei, B. Xiao, W. Xie, Y. Xie, D. Yogatama, B. Yuan, J. Zhan, Z. Zhu, in *33rd Int. Conf. on Machine Learning ICML 2016*, Vol. *1* (Eds: M. F. Balcan, K. Q. Weinberger), PMLR **2016**, p. 312.

[186] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, C. Wellekens, *Speech Commun.* **2007**, *49*, 763.

[187] G. E. Hinton, S. Osindero, Y.-W. Teh, *Neural Comput.* **2006**, *18*, 1527.

[188] R. Salakhutdinov, A. Mnih, G. Hinton, in *24th Int. Conf. on Machine Learning* (Ed: Z. Ghahramani), ACM, New York **2008**, p. 791.

[189] G. Hinton, *Scholarpedia* **2009**, *4*, 5947.

[190] J. H. Connolly, E. A. Edmonds, J. J. Guzy, S. R. Johnson, A. Woodcock, *Int. J. Man-Mach. Stud.* **1986**, *24*, 611.

[191] B. Yang, in *ICASSP, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing - Proc.* (Eds: T. Adali, B. Rao), IEEE, Piscataway, NJ, USA **2008**, p. 3541.

[192] J. C. Brown, *J. Acoust. Soc. Am.* **1991**, *89*, 425.

[193] M. Slaney, *An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank*, Apple Technical Report **1993**, *35*, 1.

[194] A. N. Jadhav, N. V. Dharwadkar, *Int. J. Mod. Educ. Comput. Sci.* **2018**, *10*, 19.

[195] P. K. Nayana, D. Mathew, A. Thomas, *Procedia Comput. Sci.* **2017**, *115*, 47.

[196] A. Maurya, D. Kumar, R. K. Agarwal, *Procedia Comput. Sci.* **2018**, *125*, 880.

[197] R. L. Streit, *Poisson Point Processes: Imaging, Tracking, and Sensing*, Springer, Boston, MA, USA **2010**.

[198] S. Fine, J. Navrátil, R. A. Gopinath, in *ICASSP, IEEE Int. Conf. Acoustics, Speech, and Signal Processing - Proc.* Vol. *1*, (Ed: B. Farhang-Boroujeny), IEEE, Piscataway, NJ, USA **2001**, pp. 417–420.

[199] R. Solera-Ureña, J. Padrell-Sendra, D. Martín-Iglesias, A. Gallardo-Antolín, C. Peláez-Moreno, F. Díaz-de-María, in *Progress in Nonlinear Speech Processing* (Eds: Y. Stylianou, M. Faundez-Zanuy, A. Eposito), Springer, Berlin, Germany **2007**, pp. 190–216.

[200] A. Bordes, S. Ertekin, J. Weston, L. Bottou, *J. Mach. Learn. Res.* **2005**, *6*, 1579.

[201] J. Dey, M. S. B. Hossain, M. A. Haque, in *ICECE 2018 - 10th Int. Conf. on Electrical and Computer Engineering* (Eds: C. Shahnaz, F. M. Mohammedy, Y. Arafat, A. Zubair, Md. S. Islam, M. A. A. Shohel, D. J. Paul, A. Subhana), IEEE, Piscataway, NJ, USA **2019**, pp. 297–300.

[202] M. Baig, S. Masud, M. Awais, in *2006 Int. Symp. on Intelligent Signal Processing and Communications* (Ed: H. Yasukawa), IEEE, Piscataway, NJ, USA **2007**, pp. 319–322.

[203] T. Kinnunen, E. Chernenko, *Int. Conf. Speech* **2007**, *2*, 555.

[204] L. R. Rabiner, B. H. Juang, *IEEE ASSP Mag.* **1986**, *3*, 4.

[205] A. J. Viterbi, *IEEE Trans. Inf. Theory* **1967**, *13*, 260.

[206] P. Swietojanski, A. Ghoshal, S. Renals, in *ICASSP, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing - Proc.* (Eds: M. Adams, V. Zhao), IEEE, Piscataway, NJ, USA **2013**, p. 6744.

[207] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (Eds: E. Mortensen, W. Brendel), IEEE, Piscataway, NJ, USA **2018**, p. 8759.

[208] N. Rusk, *Nat. Methods* **2016**, *13*, 35.

[209] F. Rosenblatt, *Psychol. Rev.* **1958**, *65*, 386.

[210] H. Bourlard, N. Morgan, *Adv. Neural Inf. Process. Syst.* **1990**, *2*, 186.

[211] W. Y. Chen, S. H. Chen, C. J. Lin, *Neural Networks* **1996**, *9*, 655.

[212] C. Cai, Y. Xu, D. Ke, K. Su, *J. Rob.* **2015**, *2015*, 1.

[213] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, *Neural Comput.* **1989**, *1*, 541.

[214] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, T. Chen, *Pattern Recognit.* **2018**, *77*, 354.

[215] W. Song, J. Cai, CS224N Project, https://cs224d.stanford.edu/reports/SongWilliam.pdf (accessed: September 2019).

[216] A. Graves, A. R. Mohamed, G. Hinton, in *ICASSP, IEEE Int. Conf. on Acoustics, Speech and Signal Processing - Proc.* (Eds: M. Adams, V. Zhao), IEEE, Piscataway, NJ, USA **2013**, p. 6645.

[217] S. Hochreiter, J. Schmidhuber, *Neural Comput.* **1997**, *9*, 1735.

[218] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, in *EMNLP 2014 - 2014 Proc. Conf. on Empirical Methods in Natural Language Processing* (Ed: Y. Marton), ACL, Stroudsburg, PA, USA **2014**, p. 1724.

[219] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, in *27th Int. Conf. on Neural Information Processing Systems*, Vol. *1* (Eds: Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger), NIPS, San Diego, CA, USA **2014**, p. 1.

[220] M. H. Huang, R. T. Rust, *J. Serv. Res.* **2018**, *21*, 155.

[221] A. V. Anttiroiko, P. Valkama, S. J. Bailey, *AI Soc.* **2014**, *29*, 323.

[222] L. Xiao, X. Wan, X. Lu, Y. Zhang, D. Wu, *IEEE Signal Process. Mag.* **2018**, *35*, 41.

[223] W. I. Park, B. K. You, B. H. Mun, H. K. Seo, J. Y. Lee, S. Hosaka, Y. Yin, C. A. Ross, K. J. Lee, Y. S. Jung, *ACS Nano* **2013**, *7*, 2651.

[224] B. K. You, W. I. Park, J. M. Kim, K. Il Park, H. K. Seo, J. Y. Lee, Y. S. Jung, K. J. Lee, *ACS Nano* **2014**, *8*, 9492.

[225] H. S. Wang, C. K. Jeong, M. H. Seo, D. J. Joe, J. H. Han, J. B. Yoon, K. J. Lee, *Nano Energy* **2017**, *35*, 415.

[226] M. Jin, C. D. Yoo, in *Behav. Biometrics Hum. Identif. Intell. Appl.* (Eds: L. Wang, X. Geng), IGI Global, Hershey, USA **2009**, pp. 264–289.

[227] M. Senoussaoui, P. Kenny, N. Dehak, P. Dumouchel, in *Proc. Odyssey Speaker and Language Recogntion Workshop* (Eds: W. Hess, M. Cooke), ISCA, Baixas, France **2010**, pp. 28–33.

[228] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, S. Khudanpur, in *ICASSP, IEEE Int. Conf. on Acoustics Speech and Signal Processing - Proc.* (Eds: M. Hayes, H. Ko), IEEE, Piscataway, NJ, USA **2018**, pp. 5329–5333.

[229] S. Novoselov, O. Kudashev, V. Shchemelinin, I. Kremnev, G. Lavrentyeva, in *ICASSP, IEEE Int. Conf. on Acoustics Speech and Signal Processing - Proc.* (Eds: M. Hayes, H. Ko), IEEE, Piscataway, NJ, USA **2018**, p. 5334.

[230] F. Bimbot, J. F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-García, D. Petrovska-Delacrétaz, D. A. Reynolds, *EURASIP J. Appl. Signal Process.* **2004**, *2004*, 430.

[231] T. Jeon, H. M. Jin, S. H. Lee, J. M. Lee, H. Il Park, M. K. Kim, K. J. Lee, B. Shin, S. O. Kim, *ACS Nano* **2016**, *10*, 7907.

[232] C. K. Jeong, J. H. Han, H. Palneedi, H. Park, G. T. Hwang, B. Joung, S. G. Kim, H. J. Shin, I. S. Kang, J. Ryu, K. J. Lee, *APL Mater.* **2017**, *5*, 074102.

[233] F. Sha, L. K. Saul, in *IEEE Int. Conf. on Acoustics Speech and Signal Processing - Proc.*, Vol. *1* (Ed: N. Thomas), IEEE, Piscataway, NJ, USA **2006**, p. 265.

[234] J. Ming, F. J. Smith, in *ICASSP, IEEE Int. Conf. on Acoustics Speech and Signal Processing - Proc.* (Ed: A. Acero), IEEE, Piscataway, NJ, USA **1998**, p. 409.