Encoding of multi-modal emotional information via personalized skin-integrated wireless facial interface

Jin Pyo Lee[1,2], Hanhyeok Jang[1], Yeonwoo Jang [1], Hyeonseo Song[1], Suwoo Lee[1], Pooi See Lee [2] ✉ & Jiyun Kim [1,3] ✉

Human affects such as emotions, moods, feelings are increasingly being considered as key parameter to enhance the interaction of human with diverse machines and systems. However, their intrinsically abstract and ambiguous nature make it challenging to accurately extract and exploit the emotional information. Here, we develop a multi-modal human emotion recognition system which can efficiently utilize comprehensive emotional information by combining verbal and non-verbal expression data. This system is composed of personalized skin-integrated facial interface (PSiFI) system that is self-powered, facile, stretchable, transparent, featuring a first bidirectional triboelectric strain and vibration sensor enabling us to sense and combine the verbal and non-verbal expression data for the first time. It is fully integrated with a data processing circuit for wireless data transfer allowing real-time emotion recognition to be performed. With the help of machine learning, various human emotion recognition tasks are done accurately in real time even while wearing mask and demonstrated digital concierge application in VR environment.

The utilization of human affects, encompassing emotions, moods, and feelings, is increasingly recognized as a crucial factor in improving the interaction between humans and diverse machines and systems[1–3]. Consequently, there is a growing expectation that technologies capable of detecting and recognizing emotions will contribute to advancements across multiple domains, including HMI device[4–6], robotics[7–9], marketing[10–12], healthcare[13–15], education[16–18], etc. By discerning personal preferences and delivering immersive interaction experiences, these technologies have the potential to offer more user-friendly and customized services. Nonetheless, decoding and encoding emotional information poses significant challenges due to the inherent abstraction, complexity, and personalized nature of emotions[19,20]. To overcome these challenges, the successful utilization of comprehensive emotional information necessitates the extraction of meaningful patterns through the detection and processing of combined data from multiple modalities, such as speech, facial expression, gesture, and various physiological signals (e.g., temperature, electrodermal activity)[21–23]. Encoding these extracted patterns into interaction parameters tailored for specific applications also becomes essential.

Conventional approaches for recognizing emotional information from humans often rely on analyzing images of facial expressions[24–26] or speech of verbal expression[27–29]. However, these methods are frequently impeded by environmental factors such as lighting conditions, noise interference, and physical obstructions. As an alternative, text analysis techniques[30–32] have been explored for emotion detection, utilizing vast amounts of information available on diverse social media platforms. However, this approach presents challenges due to the diverse ambiguities and new terminologies being introduced, which further complicates the accurate detection of emotions from the text.

[1]School of Material Science and Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, South Korea. [2]School of Materials Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore. [3]Center for Multidimensional Programmable Matter, Ulsan National Institute of Science and Technology, Ulsan 44919, South Korea. ✉e-mail: pslee@ntu.edu.sg; jiyunkim@unist.ac.kr

To overcome these limitations, sensing devices capable of capturing changes in physiological signals, including EEG[33–35], EMG[36–38], ECG[39–41] and GSR[42–44] have been employed to collect more accurate and reliable data. These devices can establish correlations between these signals and human emotions irrespective of environmental factors, but the requirement of bulky equipment limits their application to everyday communication scenarios.

In recent studies, flexible skin-integrated devices have shown the possibility of providing real-time detection and recognition of emotional information through various modalities such as facial expressions, speech, text, hand gestures, physiological signals, etc.[45–56]. Specifically, a resistive strain sensor has been employed to directly detect facial strain deformations that occur during facial expressions[46,47,51,52]. This approach offers simplicity by using thin and soft skin-integrated electrode interfaces for current flow, allowing for wearable or portable applications. However, an additional power source, low working frequency range, and extra components for the signal conversion cause simple modality only limited to one-to-one correlation that imposes constraints on the range of applications such as healthcare, VR where complementary information is needed to approximate natural interaction, and user experience can be enhanced by multiple ways of inputs. Furthermore, most existing studies have primarily focused on recognizing and exploiting human emotions, intentions or commands using the single-modal data that can have weaknesses in specific context, thus limiting the use of higher-level and comprehensive emotional contexts[45,48–50,53–56]. On the other hand, to overcome the drawbacks of each modality for a more resilient system, multi-modal emotion recognition was conducted to draw embedded high-level information by using the combined knowledge from all the accessible data sensing[57–59]. Consequently, to effectively and precisely encode emotional information, an advanced format of the skin-integrated device necessitates improved wearability seamlessly integrating with individuals, while possessing multi-modal sensing capabilities to process and extract higher-level of information. Also, this personalized device, capable of real-time collection of reliable and accurate multi-modal data regardless of external environmental factors, should be accompanied by the corresponding classification technique to encode the gathered data into personalized feedback parameters for target applications.

Here, we proposed a human emotion recognition system in an attempt to utilize complex emotional states with our personalized skin-integrated facial interface (PSiFI) offering simultaneous detection and integration of facial expression and vocal speech. The PSiFI incorporates a personalized facial mask that is self-powered, easily applicable, stretchable, transparent, capable of wireless communication, and highly customized to conformally fit into an individual's face curvatures based on 3D face reconstruction. These features enhance the device's usability and reliability in capturing and analyzing emotional cues, facilitating the real-time detection of multi-modal sensing signals derived from facial strains and vocal vibrations. To encode the combinatorial sensing signals into personalized feedback parameters, we employ a convolutional neural network (CNN)-based classification technique that rapidly adapts to an individual's context through transfer learning. In the context of human emotion recognition, we specifically focus on facial expression and vocal speech as the chosen multi-modal data, considering their convenience for data collection and classification based on prior research findings.

The PSiFI device is basically comprised of strain and vibration sensing units based on triboelectrification to detect facial strain for facial expression and vocal vibration for speech recognition, respectively. The incorporation of a triboelectric nanogenerator (TENG) enables the sensor device to possess self-powering capabilities while offering a broad range of design possibilities in terms of materials and architectures[60,61], thus fulfilling the requirements of personalized and multi-modal sensing devices. The sensing units are made of PDMS film as a dielectric layer and

PEDOT:PSS coated PDMS film as an electrode layer prepared by the semi-curing method which enables the film to exhibit good transparency with decent electrical conductivity. Furthermore, we demonstrated real-time emotion recognition with data processing circuit for wireless data transfer and real-time classification based on rapidly adapting convolution neural network (CNN) model with the help of transfer learning using data augmentation methods. Last, we demonstrated digital concierge application as an exciting possibility in virtual reality (VR) environment via human machine interfaces (HMIs) with our PSiFI. The digital concierge recognizes a user's intention and interactively offers helpful services depending on the user's affectivity. Our work presents a promising way to help to consistently collect data regarding emotional speech with barrier-free communication and can pave the way toward acceleration of digital transformation.

Results

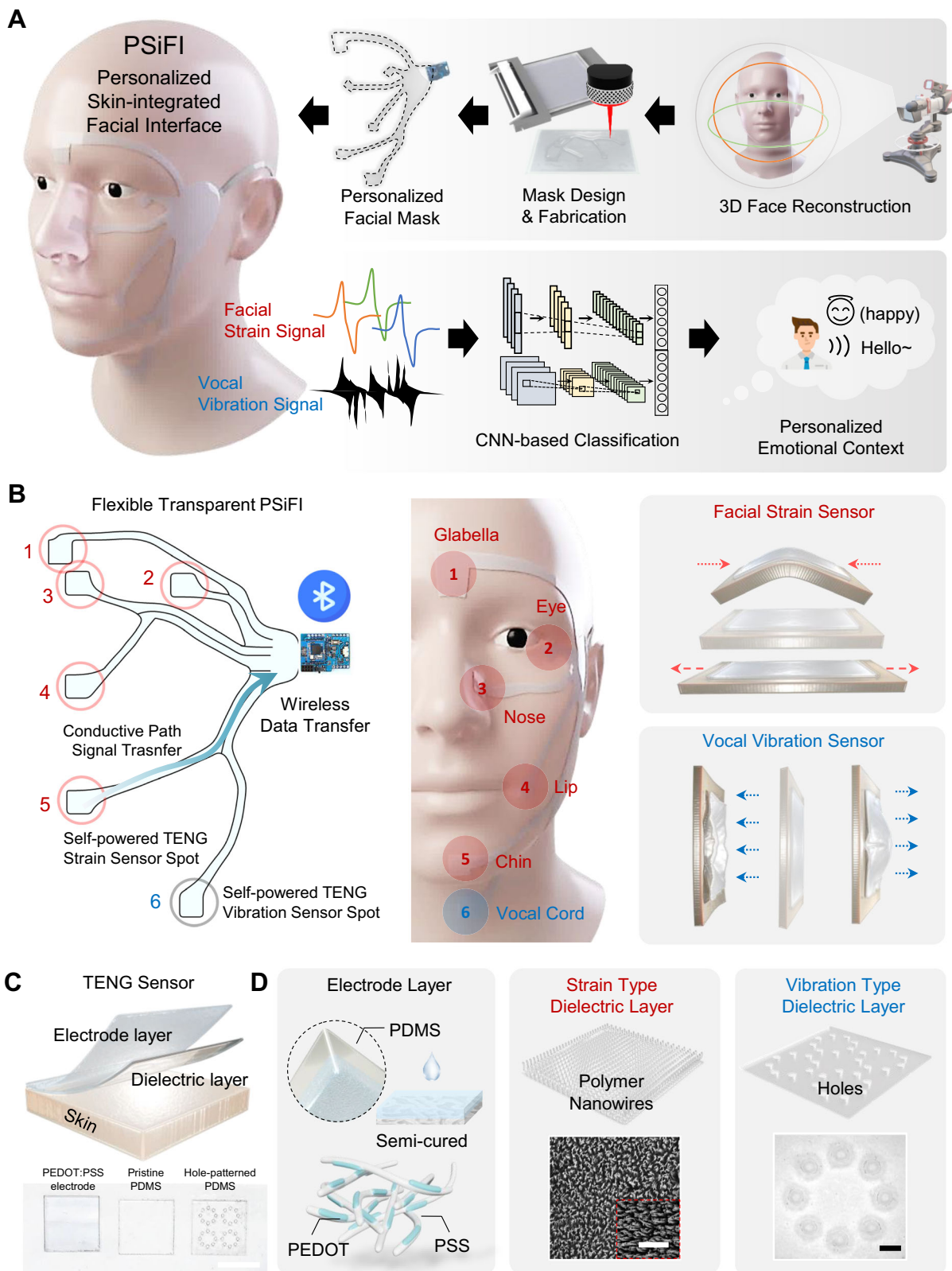### Personalized skin-integrated facial interface (PSiFI) system

We devised personalized skin-integrated facial interface (PSiFI) system consisting of multimodal triboelectric sensors (TES), data processing circuit for wireless data transfer and deep-learned classifier. Figure 1A illustrates the schematics of overall process for human emotion recognition with PSiFI from fabrication to classification task. As for making personalized device, we brought in 3D face reconstruction process by collecting 3D data of user's appearance from scanned photos and converting the data to digital models. This process allowed us to fabricate personalized device fitted in well with various user faces and successfully secure individual user data for accurate recognition task. (Supplementary Fig. 1). Subsequently, we utilized both verbal/non-verbal expression information detected from multimodal sensors and classified human emotions in real-time using transfer learning applied convolution neural network (CNN).

As shown in Fig.1B, the emotional information based on verbal/non-verbal expression in the form of digital signals was sent to be the PSiFI mask and wirelessly transferred with data processing circuit. To effectively detect the signals for the emotional information, the PSiFI was integrated with multi-modal TES to capture facial skin strains and vocal cord vibrations by detecting electrical signals from glabella, eye, nose, lip, chin and vocal cord selected as representative regions based on previous studies regarding facial muscle activation patterns during facial expression[62–64].

Figure 1C provides the schematic and real image of the TES consisting of simple two-layer structure where PEDOT:PSS-coated polydimethylsiloxane (PDMS) and nanostructured PDMS were used as stretchable electrode and dielectric layer respectively so that our TES are based on single electrode mode in principle. Figure 1D shows schematics of the PEDOT:PSS-coated PDMS and dielectric layers for each strain and vibration type. The PEDOT:PSS-coated PDMS was fabricated by semi-cured process[65,66] where coating is conducted before full-curing of the elastomer (Supplementary Movie 1). Our stretchable electrode based on the semi-curing process was characterized and showed better performance when it compared to conventional surface treated electrode in terms of optical, mechanical, and electrical aspects. (Supplementary Fig. 2) As shown in scanning electron microscope (SEM) image in Fig.1D, for the dielectric layers we fabricated, nano surface engineering was introduced by inductively coupled plasma reactive ion etching process (ICP-RIE) to improve triboelectric performance by enhancing specific surface area. (Supplementary Fig. 3) Additionally, the dielectric layer for the vibration sensing was perforated like the acoustic holes which enhance vibrate the volume of air inside (Supplementary Movie 2).

Working mechanism and characterization of the strain sensing unit

Converting facial skin strain during facial expression into distinct electrical signals and sending the data as non-verbal information to the

A PSiFI — Personalized Skin-integrated Facial Interface

Personalized Facial Mask — Mask Design & Fabrication — 3D Face Reconstruction

Facial Strain Signal / Vocal Vibration Signal → CNN-based Classification → Personalized Emotional Context — (happy) Hello~

B Flexible Transparent PSiFI

1, 3, 2, 4, 5 — Conductive Path Signal Transfer — Wireless Data Transfer — Self-powered TENG Strain Sensor Spot — 6 Self-powered TENG Vibration Sensor Spot

Glabella 1 — Eye 2 — Nose 3 — Lip 4 — Chin 5 — Vocal Cord 6

Facial Strain Sensor

Vocal Vibration Sensor

C TENG Sensor — Electrode layer — Dielectric layer — Skin

PEDOT:PSS electrode — Pristine PDMS — Hole-patterned PDMS

D Electrode Layer — PDMS — Semi-cured — PEDOT — PSS

Strain Type Dielectric Layer — Polymer Nanowires

Vibration Type Dielectric Layer — Holes

circuit system is the function of our strain sensing unit. As depicted schematically in Fig. 2A, the strain sensing unit was fabricated with the nanostructured PDMS for its high effective contact area as a dielectric layer and PEDOT:PSS embedded PDMS as an electrode layer to make TES with the single electrode structure for simple configuration to be facilitated as wearable sensors. These two layers were separated by double sided tapes applied to both ends of the layers as a spacer to be

consistently generate a series of electrical signals during the operation cycle. Besides, all the parts in the sensing units are made of stretchable and skin-friendly viable materials and can be prepared through scalable fabrication processes (for the details see the "Methods" section and Supplementary Fig. 4). These characteristics of the materials used in the strain sensing unit allow our strain sensor to retain relatively good electrical conductivity even under stretching in the range of

Fig. 1 | The system overview with PSiFI. A Schematic illustration of personalized skin-integrated facial interfaces (PSiFI) including triboelectric sensors (TES), data processing circuit for wireless communication and deep-learned classifier for facial expression and voice recognition. **B** Schemes showing 2d layout for the PSiFI in the form of wearable mask and depicting two different types of TES in terms of sensory stimulus such as facial strain and vocal vibration. **C** Schematic diagram of the TES which consists of simple two-layer structure such as electrode layer and dielectric layer and photograph of the TES components, respectively. Scale bar: 1 cm.

D Schematics demonstrating fabricated components for our TES. As for the electrode layer, PEDOT:PSS based electrode was made via semi-curing process. (left). As for the dielectric layer, it was designed differently considering sensing stimuli such as strain and vibration to achieve optimal sensing performance. The inset in center showing SEM image for nanostructured surface of strain type dielectric layer and in right showing photograph for punched holes as acoustic holes of vibration type dielectric layer. Scale bar: 2 μm and 1 mm.

facial skin strain during facial expression and guarantee robustness of the sensing unit. As schematically shown in Fig. 2B, an electrical potential builds up due to the difference between triboelectric series based on different affinity for electrons, which the PDMS played a triboelectrically negative material by receiving electrons and the PEDOT:PSS based stretchable electrode played a triboelectrically positive material by donating electrons in TES. On top of that, our strain sensing unit makes the contact area changes when stretched and achieved even buckled states so that it can detect bidirectional strain motion among the triboelectric based strain sensors for the first time, according to our knowledge. Correspondingly, the generated output signals of our strain sensing unit during the buckle-stretch cycle were shown in Fig. 2C. The comprehensive working mechanism of the bidirectional strain sensor for each mode was demonstrated in Supplementary Fig. 5.

To characterize the strain sensing unit in terms of mechanical and electrical properties, a linear motor was employed to exert a cyclic force on the sensing unit as shown in Fig. 2D. Figure 2E and F provides our strain sensing unit sensitivity measurement in a strain range from 0% to 100% by buckling and stretching, respectively. The sensitivity was derived from $S = \Delta V/\Delta\varepsilon$ where ΔV is the relative potential change and ε is the strain. As for the buckling strain, linearity of the electrical responses and a sensitivity of 5 mV was obtained in a strain range up to 50% despite non-linear region occurred beyond the strain due to anomalous shape change. The signals in the non-linear region were differentiated with the difference in the width of time as shown in Supplementary Fig. 6. As for the stretching strain, an acceptable linearity and sensitivity of 3 mV was obtained in wide strain range up to 90%. We measured the response time of the strain sensing unit to evaluate the performance of the unit as it can be executed real-time classification tasks. As shown in Fig. 2G, there is no apparent latency time between the stretching force and corresponding the output voltage so that we can make sure the sensing unit can detect the sensing in real time. The stretch–release of one cycle (Fig. 2G, inset) exhibits a response time of below 20 ms. Therefore, compared with other strain sensors, our strain sensing unit has an advantage because of its high sensitivity in bi-direction, fast-response time and high stretchability, which can ensure an accurate sensing of the facial expression via converted electrical signals in real time.

We also measured the output voltage at constant strain of 40% depending on the working frequencies ranging from 0.5 to 3 Hz and confirmed that our strain sensing unit can show reliable performance regardless of the frequencies as shown in Fig. 2H. When it comes to long-term use in practical application, the mechanical stability of our sensing unit also can be considered as important property. As demonstrated in Fig. 2I, apparent output voltages changes were not observed for the strain sensing unit after 3000 continuous working cycles under 40% strain. It is noteworthy that the 40% strain change is way beyond the requirement for most facial skin strain during facial expression demonstrations[45,67].

Working mechanism and characterization of the vocal sensing unit

Our vocal sensing unit has a function of capturing vocal vibrations on the vocal cord during verbal expression and sending the data as verbal information to the circuit system. As shown in Fig. 3A, the vocal sensing unit was fabricated with the holes patterned PDMS as dielectric layer and PEDOT:PSS embedded PDMS as an electrode layer to make TES. The holes were introduced into the vocal sensing unit as acoustic holes which not only act as communicating vessels to ventilate an air between two contact surfaces to the ambient air, which results in enhanced flat frequency response but also reduce the stiffness by improving the movement of the rim of diaphragms[68–70] (Supplementary Fig. 7 and Table S1). To be configured into TES, like the strain sensing unit, the dielectric and electrode layer were separated by double-sided tapes applied to both ends of the layers as a spacer for consistent operations during working cycles. The inset to Fig. 3A provides an enlarged view of the vocal sensing unit capturing vocal vibrations on vocal cord. As schematically depicted in Fig. 3B, an electrical potential builds up due to triboelectric series difference based on an electron affinity. Figure 3C provides the schematic drawing showing hole pattern configuration applied in vocal vibration sensor to see how the pattern influence the output and SEM images of the holes.

We measured output voltage signals of the vibration sensing units with different open ratios (ORs) considered the proportion of area perforated with acoustic holes in the whole area on the frequency response of the devices as shown in Fig. 3D. The frequency ranges we tested encompass the fundamental frequency of typical adult men and women ranging from 100 to 150 Hz (Fig. 3D, blue) and from 200 to 250 Hz (Fig. 3D, red), respectively[71]. The results indicate that the vibration sensing unit with OR value of 10 exhibited best output voltage performance and the wideset bandwidth of flat frequency response. This experimental observation is originated from a trade-off between the deflection of dielectric layer and the effective contact area. Larger OR leads to a larger deflection of the dielectric diaphragm and thus a higher electric output. However, increased OR will reduce the effective contact area for triboelectrification, and thus a lower electrical output. Accordingly, an optimized value of OR is needed for maximization of the electrical output. Figure 3E provides measured data plots of output voltage signals per each different OR at the testing frequency of 100 Hz.

As shown in Fig. 3F and G, the output voltage of the vibration sensing unit was affected by structural parameters such as the support thickness and number of holes. As the support thickness is increased, the gap between the triboelectric layers is larger so that the effective contact area can be reduced thus the generated triboelectric output signals is decreased. On the other hand, the larger number of holes with the same OR condition makes the diaphragms deflect more vigorously, thus enhancing the triboelectric output performance. These experiments were carried out at the testing frequency of 100 Hz. Lastly, as shown in Fig. 3H, we measured the output voltage between the vibration sensing unit with and without holes as a function of input vibration acceleration in the ranging from 0.1 to 1.0 g at the same testing frequency of 100 Hz. Both sensing units have a uniform sensitivity obtained from dividing the measured output voltage by the vibration acceleration. As for the sensitivity, the hole-patterned vibration sensing units exhibits 5.78 V/g around 2.8 times larger than that of the pristine vibration sensing unit.
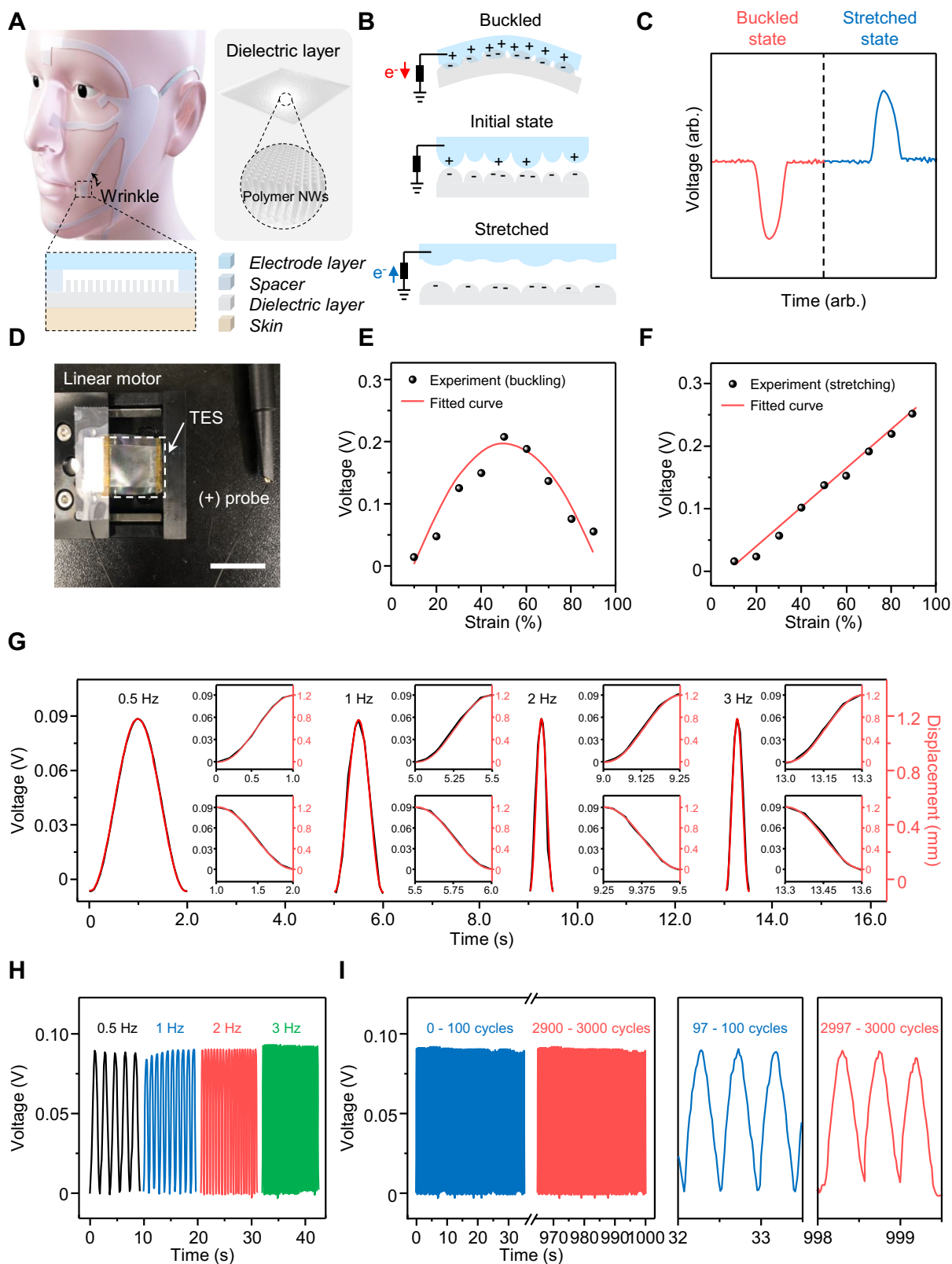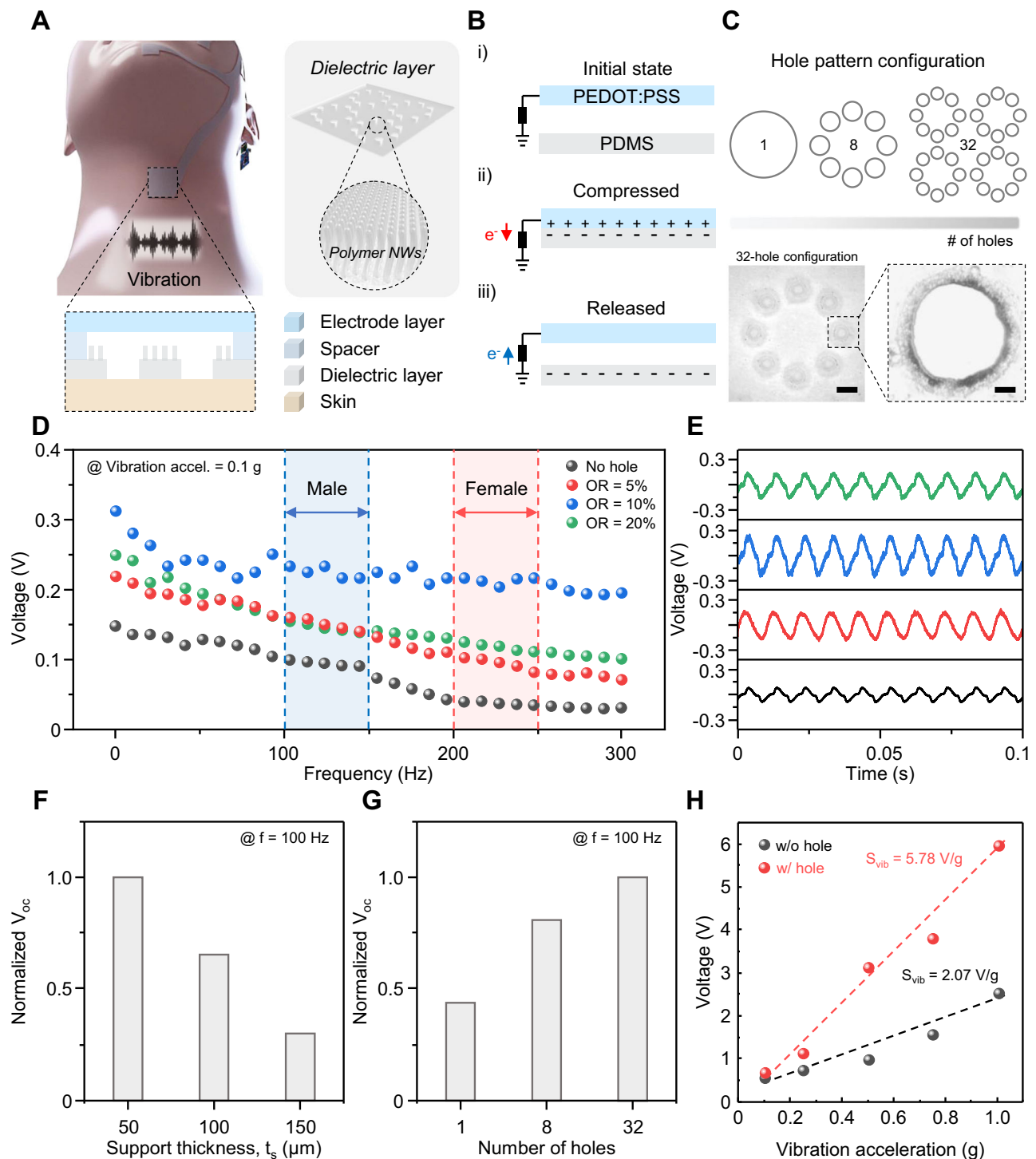
Fig. 2 | Working mechanism and characterization of the strain sensing unit.
A Schematic illustration of the strain sensing unit. Inset: enlarged view of the sensing unit detecting facial strain. **B** Electrical potential distribution of the strain sensing unit under buckled and stretched state. **C** Output electrical signals of the strain sensing unit during the buckle-stretch cycle. **D** Real image of experimental set-up for output measurements. Scale bar: 1 cm. **E** and **F** Sensitivity measurement during buckling (**E**) and stretching of the sensing unit (**F**). **G** Response time measurement with various frequencies. Insets: enlarged views of the loading and unloading processes in one cycle. **H** Generated voltage signals of the sensing unit with various frequencies at a constant strain of 40%. **I** Mechanical durability test for up to 3000 continuous working cycles and enlarged views of different operation cycles, respectively.

Fig. 3 | Working mechanism and characterization of the vibration sensing unit.
A Schematic illustration of the vibration sensing unit. Inset: enlarged view of the sensing unit detecting vocal-cord vibration. **B** Electrical potential distribution of the sensing unit during working cycle. **C** Schematic of hole pattern configuration applied in vocal vibration sensor and SEM images of the holes in 32-hole configuration. Scale bar: 2 mm (inset: magnified view showing an acoustic hole. Scale bar: 400 μm). **D** Frequency response data (V_{oc} as a function of acoustic frequency) for the vibration sensing unit with different open ratios (ORs) of 5, 10 and 20. The vocal cord frequency ranges for male and female are colored blue and red, respectively. **E** Measured data plots of output voltage signals per each different OR at the testing frequency of 100 Hz. **F**, **G** Effects of support thickness and number of holes on vibration sensitivity at working frequency of 100 Hz. For each graph, PDMS used as diaphragm material, acoustic holes were patterned on the diaphragm, and the structural parameters were fixed as follows unless otherwise specified: diaphragm thickness of 50 μm, support thickness of 50 μm and an array of 32 holes. The error bars indicate the s.d. of the normalized V_{oc} at the measured frequency of 100 Hz. **H** Comparison of measured output voltage between the vibration sensing unit with and without holes.

Wireless data processing process and machine learning based real time classification

Figure 4A and B provides real images of the whole PSiFI mask and the participant wearing the PSiFI mask properly laminated onto the participant's face, which made it look transparent and comfortable enough to be worn for long time and communicate well without interrupting expressions that can be caused by a colored device. As schematically depicted in Fig. 4C, our wireless data acquisition and transfer process was carried out from the data collection of the skin-integrated facial mask by the several centimeter size of circuit board as a signal transmitter powered by a tiny portable battery to wirelessly transmitted data received by the main board as the receiver connected to the laptop for storing data used to be datasets for the machined learning.

Figure 4D provides collected triboelectric signal patterns from each modal sensor such as lip, eye, glabella, nose, chin (for strain sensing unit) and vocal cord (for vibration sensing unit). As for the acquired signals from the strain sensing units, distinct patterns were exhibited in accordance with the different facial expressions such as happiness, surprise, disgust, anger and sadness that the participant expressed. As for the signals from the vocal sensing unit, each signals for different speech from the syllable such as "A", "B", "C" to the simple sentence such as "I love you" clearly exhibited its own distinct patterns and were further transformed by fast Fourier transformation (FFT) which converts data from time domain to frequency domain to find remarkable patterns in frequency domain so that the pattern recognition performed well. We conducted separate training for the vocal and strain signals as the interdependence between verbal and non-verbal expressions appears to be relatively insignificant when compared to the distinct and concurrent measurements of the multi-modal inputs (Supplementary Fig. 8).

When it comes to machine learning, we applied the CNN algorithm as an example of algorithm for classification. Specifically, we utilized one-dimensional CNN to classify the facial expressions and two-dimensional CNN for speech classification, respectively (Supplementary Fig. 9 and Table S2). Generally, the more datasets our classifier trains, the better performance it shows. However, it is not viable and time consuming to test the sensor integrated wearable mask to many people in practical terms. The facial muscle movements, vocal cord vibration and sensor values corresponding to the verbal/non-verbal expressions of the new users would be different from those of the previous users since every human has its own characteristics. We therefore need to adapt to a network which can be trained with even small amounts of datasets and tuned with the new datasets from the new users.

Figure 4E provides schematic diagrams showing the overall process from data achieving pre-trained model trained with enhanced accuracy by introducing data augmentation technique (Supplementary Fig. 10 and Table S3) to fine-tunned network for personalization by exploiting pre-trained parameters called as transfer learning, which enables the network to be trained in reduced time and effectively adapt to new user's datasets so that it made the real time classification possible. In detail, a participant repeated, respectively, verbal and non-verbal expression 20 times to demonstrate reliability for a total acquisition of 100 recognition signal patterns per each expression. 70 patterns of total were randomly selected from the acquired signals to serve as the training set which are subsequently augmented 8-fold based on different methods (Jittering, Scaling, Time-warping, Magnitude-warping) for effective learning, and the remaining 30 signals were assigned as the test set. Furthermore, according to the previous report, it was found that the movement and activation patterns of facial muscles during facial expressions was not dissimilar depending on the individuals[62-64]. Based on this fact, we anticipate that the network can get used to adapt to new expressions from new users by rapidly training the corresponding learning data. As for the transfer learning,

after the initial participant had firstly trained with the classifier by the above-mentioned training method, the following participants were wearing with the PSiFI device and able to fast train with the classifier by only repeating 10 times each on both expressions, which successfully allow the real-time classification to be demonstrated. When it comes to practical application, compared with other classification methods based on various kinds of video camera and microphone, our PSiFI mask is free from environmental restrictions such as the location, obstruction, and time. As shown in Fig. 4F, the real-time classification result for combined verbal/nonverbal expressions without any restriction exhibited very high accuracy of 93.3% and even the decent accuracy of 80.0% was achieved despite carrying out the classification with obstruction such as wearing a facial mask (Supplementary Movie 3).

Digital concierge application in VR environment

As for the application with the PSiFI, we brought in VR environment which allows individuals to experiment with how their emotions could influence and can be expressed and implemented into specific situations in the virtual world[72-74]. This in turn can deepen communications in VR environment by engaging with human emotions. In this sense, we selected digital concierge application that can be enriched with emotional information in terms of practical use and usability. The digital concierge is likely to be anticipated that it can provide user-oriented services which improve quality of user's life by promoting user's experience. Herein, for the first time, we demonstrated the application which offers a digital concierge service operated with our PSiFI based on HMI in VR environment of Unity software as shown in Fig. 5.

Figure 5A provides conceptual schematic showing how human and machine can interact smartly with personalized emotional context by wearing the PSiFI. To realize this, we demonstrate VR-based digital concierge application via HMI with our PSiFI as the overall process was shown in Fig. 5B. Specifically, the digital concierge system was operated based on conversation between the user's avatar and randomly generated avatar who serves as the virtual concierge. Additionally, we built the digital concierge to provide various application services from smart home to entertainment by taking into account the situations which take place very probably in real life.

Figure 5C provides three different scenarios demonstrating smart home, office, and entertainment application in Unity space (Supplementary Movie 4; for details, see the "Methods" secton). As for the first scenario for smart home application, the digital concierge accessed the user's mood of sadness and recommend some playlist from website to relieve the mood despite of user's simple word. As for the second scenario for office application, the digital concierge was able to check if the user understands contents of presentation and pop out new window showing content interpretation that helps to promote user's understanding. As for the last scenario for entertainment application, the digital concierge identifies user's reaction to the movie trailer and curates user-friendly contents in accordance with user's reaction. The applications with our PSiFI-based HMI and built-in VR space can be greatly diversified with learning and adapting new data regarding verbal and non-verbal expressions from new users so that we strongly anticipate our highly personalized PSiFI platform contributes to various practical applications such as education, marketing, and advertisements that can be enriched with emotional information.

Discussion

In this work, we proposed a machine-learning assisted PSiFI for wearable human emotion recognition system. The PSiFI was made of PDMS-based dielectric and stretchable conductor layers that are highly transparent and comfortable as possible to wear in real life. By endowing our PSiFI with multi-modality to detect simultaneously both facial and vocal expressions using self-powered triboelectric-based sensing units, we can acquire better emotional information regardless
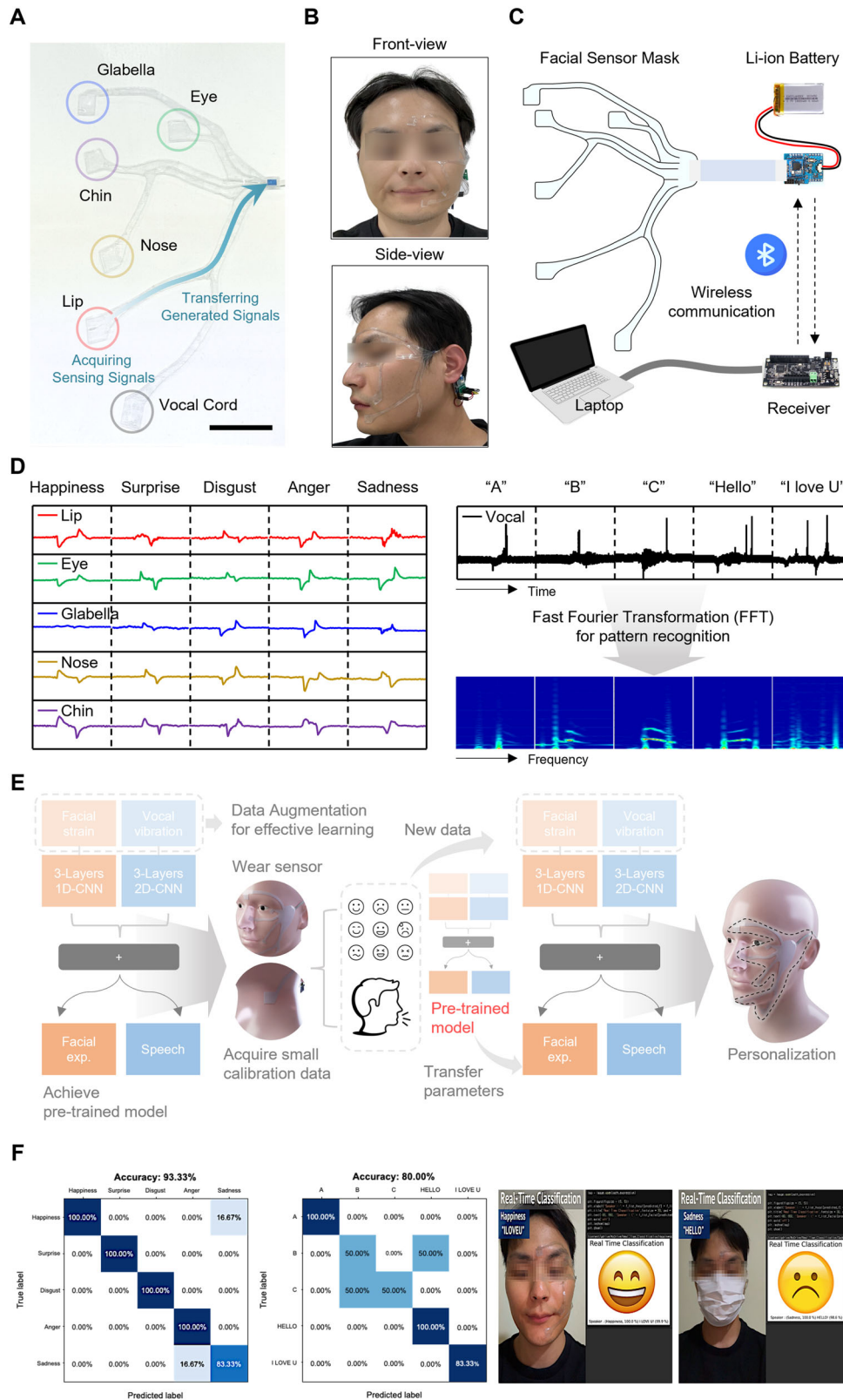
Fig. 4 | Real-time emotional speech acquisition. A Photograph showing the multimodality of the PSiFI attached to active units such as glabellar, eye, nose, lip, chin, and vocal cord for simultaneous verbal/non-verbal data collection. Scale bar: 2 cm. **B** Real images of front (top) and side view (bottom) of the participant wearing the PSiFI. **C** Schematic diagrams of the wireless emotional speech classifying system including PSiFI, signal processing board for wireless data transfer. **D** Facial strain and vocal vibration signals were collected from the skin-integrated interface. **E** The processes of learning algorithm architecture implemented in our classification system where machine learning methods such as data augmentation and transfer learning were applied to efficiently reduce training time for the real-time classification. **F** Comparison of confusion matrix (left) and captured images (right) in real-time classification between without and with an obstacle such as a mask.
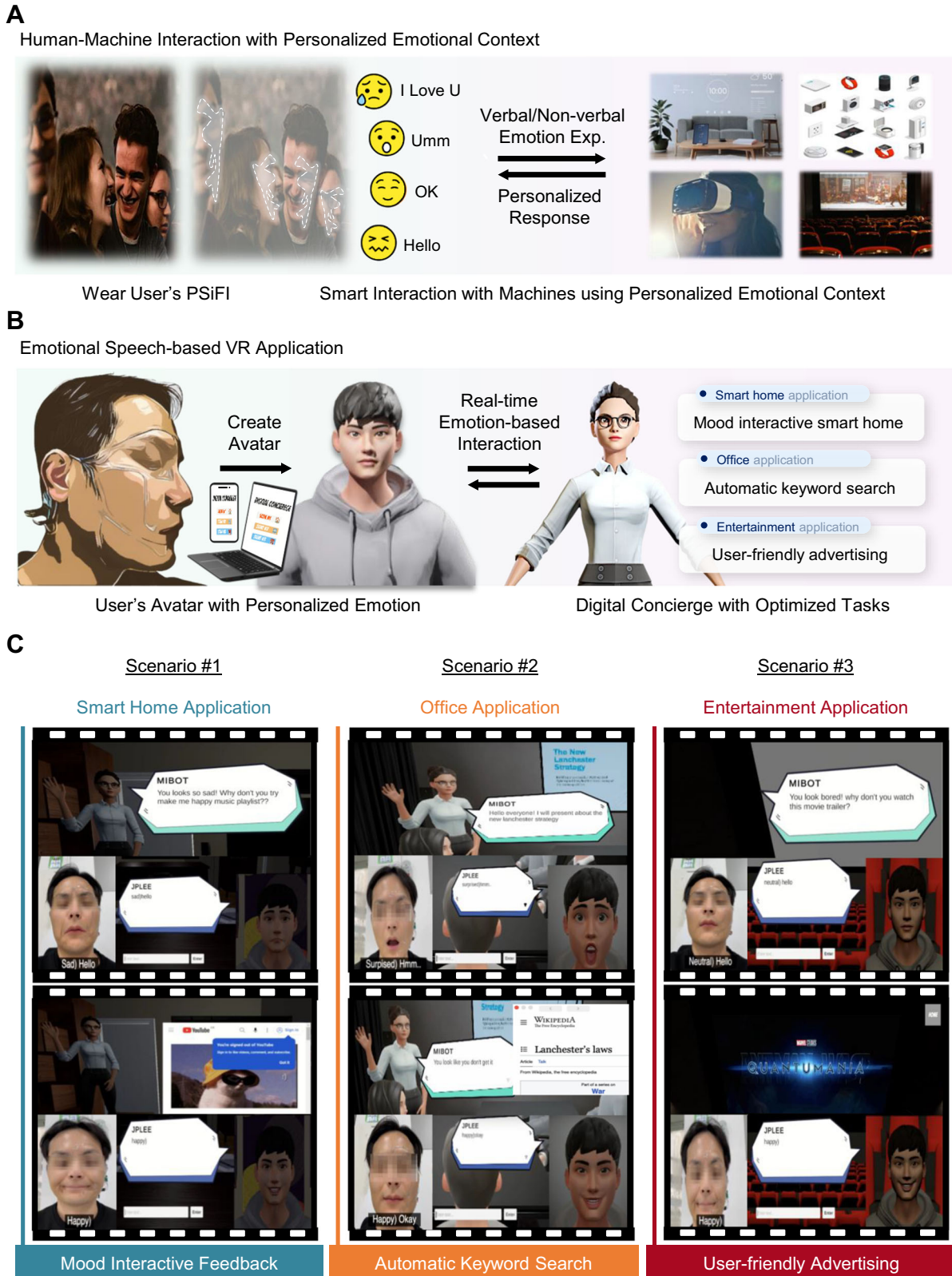
Fig. 5 | The demonstration for digital concierge based on the emotional speech classifying system in VR environment. A Conceptual illustration of human machine interaction with personalized emotional context achieved by wearing user's PSiFI. **B** Schematic diagram of the way the user interacts with the digital concierge providing various helpful services. **C** The corresponding captured images of three different scenarios as tasks (such as mood interactive feedback, automatic keyword search and user-friendly advertising) of digital concierge which likely take place in various places such as home, office and theater in VR environment of Unity software.

of external factors such as time, place, and obstacles. Furthermore, we realized wireless data communication for real-time human emotion recognition with the help of designed data-processing circuit unit and the rapid adapting learning model and achieved acceptable standard in terms of test accuracy even with the barrier such as mask. Finally, we first demonstrated digital concierge application in VR environment capable of responding to user's intention based on the user's emotional speech information. We believe that the PSiFI could assist and accelerate the active usage of emotions for digital transformation in the near future.

Methods

Materials
PDMS was purchased from Dow corning which consists of elastomer base and curing set (10:1 wt/wt). Aqueous dispersions of PEDOT:PSS solution (>3%), ethylene glycol (99.8%), and Au nanoparticles (Au NPs) (~100 nm) dispersion in deionized water (DI) was purchased from Sigma-Aldrich. Acetone (99.5%) and isopropyl alcohol (IPA) (99.5%) were purchased from Samchun Chemical.

Preparation of conductive dispersion and stretchable conductor
An aqueous solution of PEDOT:PSS was firstly filtered through a 0.45 mm nylon syringe filter. Next, 5 wt% DMSO was added to the solution, and it was then mixed with 50 wt% IPA solvent by vigorously stirring at room temperature for half an hour. Subsequently, the base monomer and curing agent were mixed with a weight ratio of 10:1 at room temperature and then, placed into the vacuum desiccator to degas the PDMS mixture. After 40 min, 1 mL of mixture was spread in the form of a continuous layer onto the cleaned Kapton film as a substrate using a micrometer adjustable film applicator, and allowed to solidify into an amorphous free-standing film by heating on an oven at 90 °C for 5 min. The prepared conductive dispersion was subsequently coated on the PDMS to anchor the conductive polymers within the PDMS matrix before the film is fully solidified.

Fabrication of nanowire-based surface modification of dielectric film
Nanowires on the surface of the PDMS film were formed by using inductively coupled plasma (ICP) reactive ion etching. The dielectric films with a thickness of 50 μm were first cleaned subsequently by Acetone, IPA and DI, then blown dry with nitrogen gas. In the etching process, Au NPs were prepared by vortex mixer for homogeneous distribution and deposited by drop-casting. After 30 min of drying in oven at 80 °C, the Au NPs were coated on the dielectric surface as a nano-patterned mask. Subsequently, a mixed gas including Ar, O_2, and CF_4 was introduced in the ICP chamber, with a corresponding flow rate of 15.0, 10.0, and 30.0 sccm, respectively. The dielectric films were etched for 300 s to obtain a nanowire structure on the surface. One power source of 400 W was used to yield a large density of plasma, while another 100 W was used to accelerate the plasma ions.

Fabrication of hole-patterned dielectric films
Arrays of circular acoustic holes with various shapes and distributions were fabricated and punched through the PDMS film (thickness 100 μm) using laser-cutting technology (Universal Laser Systems Inc.). The diameter of the smallest hole is 500 μm, which is close to the line-width limitation of the laser cutting on a plate surface.

Fabrication of self-powered sensing units
As for the strain sensing unit, the prepared stretchable conductor was cut in the size of 1 cm × 1 cm. Next, a flat flexible cable (FFC) was attached with the double-sided medical silicone tape (3M 2476P, 3M Co., Ltd) for electrical connection (Supplementary Fig. 11). Then, the surface modified dielectric film (thickness 50 μm) was subsequently placed on the layer and used as space-charge carrying layer.

As for the vibration sensing unit, the prepared stretchable conductor was cut in the size of 1 cm × 1 cm. Next, the FFC was attached with the double-sided medical tape for electrical connection like in the strain sensing unit. Then, the 50 μm-thick surface modified and hole patterned PDMS film as dielectric layer was sequentially applied on the layer and used as diaphragm deflecting with the vocal vibration.

Characterization and measurement
The morphologies and thickness of the PEDOT:PSS embedded stretchable conductor and the nano-patterned dielectrics were investigated by using a Nano 230 field-emission scanning electron microscope (FEI, USA) at an accelerating voltage of 10 kV. Optical transmission measurements of the stretchable conductors were performed on ultraviolet–visible spectrophotometer (Cary 5000, Agilent) from 400 to 800 nm. The sheet resistances (Rs) of the stretchable conductors were measured using the four-point van der Pauw method with collinear probes (0.5 cm spacing) connected to a four-point probing system (CMT2000N, AIT). For the electrical measurement of the strain sensor unit, an external shear force was applied by a commercial linear mechanical motor (X-LSM 100b, Zaber Technologies) and a programmable electrometer (Keithley model 6514) was used to measure the open-circuit voltage and short-circuit current. For the vibration sensor unit, a Digital Phosphor Oscilloscope (DPO 3052, Tektronix) was used to measure the electrical output signals at the sampling rate of 2.5 GS/s. For the multi-channel sensing system, a DAQ system (PCIe-6351, NI) was used to simultaneously measure electrical output signals of multi-channel sensor units.

Attachment of the device on the skin
To mount the sensor device completely onto the facial and neck skin, we applied a bio-compatible, ultrathin, and transparent medical tape (Tegaderm TM Film 1622W, 3M) over the edge of the sensor and the metal lines connected to the interface circuit. The medical tape is developed and widely utilized for skin-friendly adhesive solution. Therefore, there was no skin irritation or itch during several hours of wearing. The test was exempted from IRB in accordance with the approval by UNIST IRB Committee. The authors affirm that human research participants provided informed consent prior to inclusion in this study and for publication of the images in Figs. 4 and 5.

Machine learning for emotion recognition
For the pre-training, a total acquisition of 100 recognition signal patterns per each expression were collected from a participant repeating 20 times each on both verbal and non-verbal expressions, respectively. 70 patterns of total were randomly selected as training set, further augmented 8-fold based on different augmentation methods (Jittering, Scaling, Time-warping, Magnitude-warping), and the remaining 30 signals were assigned as the test set. After pre-processing step for the datasets such as trimming in accordance with input size of the neural network and converting to image by FFT, the 1D-CNN and 2D-CNN were applied for non-verbal expression and verbal-expression training. With this pre-trained classifier, a new user can rapidly customize the classifier with its own data by repeating 10 times each on both expressions, known as transfer learning, the real-time classification was successfully demonstrated.

Demonstration of the application
The three-dimensional (3D) VR environment that the user saw was provided by Unity3D on a computer, the facial strain and vocal vibration sensing data were sent to Unity3D through wireless serial communication from Buleinno2, and the interaction between PSiFI and the computer was done by PySerial package in python. We built VR-based digital concierge scenario comprising of environmental assets and generated avatars as follows. The virtual environments assets such as home, office, and theater were downloaded at Unity Asset Store. The

avatars used in the VR environments were simply created from individual photo using readyplayer.me website. In demonstration, the generated avatar proceeded the scenario based on the real-time information transmitted from PSiFI and got adaptive responses from the avatar called MIBOT virtually created for digital concierge.

Reporting summary

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

Data availability

The data that support the plots within this paper and other finding of the study are present in the paper and/or the Supplementary Information. The original datasets for human emotion recognition are available from https://github.com/MATTER-INTEL-LAB/PSIFI.git.

Code availability

All codes used for implementation of the data augmentation and classification are available from https://github.com/MATTER-INTEL-LAB/PSIFI.git.

References

1. Rahman, M. M., Poddar, A., Alam, M. G. R. & Dey, S. K. Affective state recognition through EEG signals feature level fusion and ensemble classifier. Preprint at https://doi.org/10.48550/arXiv.2102.07127 (2021).
2. Niklander, S. & Niklander, G. Combining sentimental and content analysis for recognizing and interpreting human affects. in *HCI International 2017—Posters' Extended Abstracts* (ed. Stephanidis, C.) 465–468 (Springer International Publishing, 2017).
3. Torres, E. P., Torres, E. A., Hernández-Álvarez, M., Yoo, S. G. & EEG-Based, B. C. I. Emotion recognition: a survey. *Sensors* **20**, 5083 (2020).
4. Palaniswamy, S. & Suchitra, A. Robust pose & illumination invariant emotion recognition from facial images using deep learning for human–machine interface. In *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)* 1–6 (2019).
5. Thirunavukkarasu, G. S., Abdi, H. & Mohajer, N. A smart HMI for driving safety using emotion prediction of EEG signals. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* 004148–004153 (2016).
6. Huo, F., Zhao, Y., Chai, C. & Fang, F. A user experience map design method based on emotional quantification of in-vehicle HMI. *Humanit. Sci. Soc. Commun.* **10**, 1–10 (2023).
7. Breazeal, C. Emotion and sociable humanoid robots. *Int. J. Hum.–Comput. Stud.* **59**, 119–155 (2003).
8. Stock-Homburg, R. Survey of emotions in human–robot interactions: perspectives from robotic psychology on 20 years of research. *Int. J. Soc. Robot.* **14**, 389–411 (2022).
9. Chuah, S. H.-W. & Yu, J. The future of service: The power of emotion in human-robot interaction. *J. Retail. Consum. Serv.* **61**, 102551 (2021).
10. Consoli, D. A new concept of marketing: the emotional marketing. *BRAND Broad Res. Account. Negot. Distrib.* **1**, 52–59 (2010).
11. Bagozzi, R. P., Gopinath, M. & Nyer, P. U. The role. *Emot. Mark. J. Acad. Mark. Sci.* **27**, 184–206 (1999).
12. Yung, R., Khoo-Lattimore, C. & Potter, L. E. Virtual reality and tourism marketing: conceptualizing a framework on presence, emotion, and intention. *Curr. Issues Tour.* **24**, 1505–1525 (2021).
13. Hasnul, M. A., Aziz, N. A. A., Alelyani, S., Mohana, M. & Aziz, A. A. Electrocardiogram-based emotion recognition systems and their applications in healthcare—a review. *Sensors* **21**, 5015 (2021).
14. Dhuheir, M. et al. Emotion recognition for healthcare surveillance systems using neural networks: a survey. Preprint at https://doi.org/10.48550/arXiv.2107.05989 (2021).
15. Jiménez-Herrera, M. F. et al. Emotions and feelings in critical and emergency caring situations: a qualitative study. *BMC Nurs.* **19**, 60 (2020).
16. Schutz, P. A., Hong, J. Y., Cross, D. I. & Osbon, J. N. Reflections on investigating emotion in educational activity settings. *Educ. Psychol. Rev.* **18**, 343–360 (2006).
17. Tyng, C. M., Amin, H. U., Saad, M. N. M. & Malik, A. S. The influences of emotion on learning and memory. *Front. Psychol.* **8**, 1454 (2017).
18. Li, L., Gow, A. D. I. & Zhou, J. The role of positive emotions in education: a neuroscience perspective. *Mind Brain Educ.* **14**, 220–234 (2020).
19. Ben-Ze'Ev, A. *The Subtlety of Emotions* (MIT Press, 2001).
20. Lane, R. D. & Pollermann, B. Z. Complexity of emotion representations. in *The Wisdom in Feeling: Psychological Processes in Emotional Intelligence* 271–293 (The Guilford Press, 2002).
21. Boehner, K., DePaula, R., Dourish, P. & Sengers, P. How emotion is made and measured. *Int. J. Hum.–Comput. Stud.* **65**, 275–291 (2007).
22. Mauss, I. B. & Robinson, M. D. Measures of emotion: a review. *Cogn. Emot.* **23**, 209–237 (2009).
23. Meiselman, H. L. *Emotion Measurement* (Woodhead Publishing, 2016).
24. Ioannou, S. V. et al. Emotion recognition through facial expression analysis based on a neurofuzzy network. *Neural Netw.* **18**, 423–435 (2005).
25. Tarnowski, P., Kołodziej, M., Majkowski, A. & Rak, R. J. *Emotion recognition using facial expressions. Procedia Comput. Sci.* **108**, 1175–1184 (2017).
26. Abdat, F., Maaoui, C. & Pruski, A. Human–computer interaction using emotion recognition from facial expression. In *2011 UKSim 5th European Symposium on Computer Modeling and Simulation* (ed Sterritt, R.) 196–201 (IEEE computer society, 2011).
27. Akçay, M. B. & Oğuz, K. Speech emotion recognition: emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Commun.* **116**, 56–76 (2020).
28. Issa, D., Fatih Demirci, M. & Yazici, A. Speech emotion recognition with deep convolutional neural networks. *Biomed. Signal Process. Control* **59**, 101894 (2020).
29. Lech, M., Stolar, M., Best, C. & Bolia, R. Real-time speech emotion recognition using a pre-trained image classification network: effects of bandwidth reduction and companding. *Front. Comput. Sci.* **2**, 14 (2020).
30. Nandwani, P. & Verma, R. A review on sentiment analysis and emotion detection from text. *Soc. Netw. Anal. Min.* **11**, 81 (2021).
31. Acheampong, F. A., Wenyu, C. & Nunoo-Mensah, H. Text-based emotion detection: advances, challenges, and opportunities. *Eng. Rep.* **2**, e12189 (2020).
32. Alm, C. O., Roth, D. & Sproat, R. Emotions from text: machine learning for text-based emotion prediction. In *Proc. Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* 579–586 (Association for Computational Linguistics, 2005).
33. Murugappan, M., Ramachandran, N. & Sazali, Y. Classification of human emotion from EEG using discrete wavelet transform. *J. Biomedical Science and Engineering* **3**, 390–396 (2010).
34. Gannouni, S., Aledaily, A., Belwafi, K. & Aboalsamh, H. Emotion detection using electroencephalography signals and a zero-time windowing-based epoch estimation and relevant electrode identification. *Sci. Rep.* **11**, 7071 (2021).
35. Jenke, R., Peer, A. & Buss, M. Feature Extraction and Selection for Emotion Recognition from EEG. *IEEE Trans. Affect. Comput.* **5**, 327–339 (2014).
36. Balconi, M., Bortolotti, A. & Gonzaga, L. Emotional face recognition, EMG response, and medial prefrontal activity in empathic behaviour. *Neurosci. Res.* **71**, 251–259 (2011).

37. Künecke, J., Hildebrandt, A., Recio, G., Sommer, W. & Wilhelm, O. Facial EMG responses to emotional expressions are related to emotion perception ability. *PLoS ONE* **9**, e84053 (2014).

38. Kulke, L., Feyerabend, D. & Schacht, A. A comparison of the affectiva imotions facial expression analysis software with EMG for identifying facial expressions of emotion. *Front. Psychol.* **11**, 329 (2020).

39. Brás, S., Ferreira, J. H. T., Soares, S. C. & Pinho, A. J. Biometric and emotion identification: an ECG compression based method. *Front. Psychol.* **9**, 467 (2018).

40. Selvaraj, J., Murugappan, M., Wan, K. & Yaacob, S. Classification of emotional states from electrocardiogram signals: a non-linear approach based on hurst. *Biomed. Eng. OnLine* **12**, 44 (2013).

41. Agrafioti, F., Hatzinakos, D. & Anderson, A. K. ECG pattern analysis for emotion detection. *IEEE Trans. Affect. Comput.* **3**, 102–115 (2012).

42. Goshvarpour, A., Abbasi, A. & Goshvarpour, A. An accurate emotion recognition system using ECG and GSR signals and matching pursuit method. *Biomed. J.* **40**, 355–368 (2017).

43. Dutta, S., Mishra, B. K., Mitra, A. & Chakraborty, A. An analysis of emotion recognition based on GSR signal. *ECS Trans.* **107**, 12535 (2022).

44. Wu, G., Liu, G. & Hao, M. The analysis of emotion recognition from GSR based on PSO. In *2010 International Symposium on Intelligence Information Processing and Trusted Computing*. (ed Sterritt, R.) 360–363 (IEEE computer society, 2010).

45. Wang, Y. et al. A durable nanomesh on-skin strain gauge for natural skin motion monitoring with minimum mechanical constraints. *Sci. Adv.* **6**, eabb7043 (2020).

46. Roh, E., Hwang, B.-U., Kim, D., Kim, B.-Y. & Lee, N.-E. Stretchable, transparent, ultrasensitive, and patchable strain sensor for human–machine interfaces comprising a nanohybrid of carbon nanotubes and conductive elastomers. *ACS Nano* **9**, 6252–6261 (2015).

47. Su, M. et al. Nanoparticle based curve arrays for multirecognition flexible electronics. *Adv. Mater.* **28**, 1369–1374 (2016).

48. Yoon, S., Sim, J. K. & Cho, Y.-H. A flexible and wearable human stress monitoring patch. *Sci. Rep.* **6**, 23468 (2016).

49. Jeong, Y. R. et al. A skin-attachable, stretchable integrated system based on liquid GaInSn for wireless human motion monitoring with multi-site sensing capabilities. *NPG Asia Mater.* **9**, e443–e443 (2017).

50. Hua, Q. et al. Skin-inspired highly stretchable and conformable matrix networks for multifunctional sensing. *Nat. Commun.* **9**, 244 (2018).

51. Ramli, N. A., Nordin, A. N. & Zainul Azlan, N. Development of low cost screen-printed piezoresistive strain sensor for facial expressions recognition systems. *Microelectron. Eng.* **234**, 111440 (2020).

52. Sun, T. et al. Decoding of facial strains via conformable piezoelectric interfaces. *Nat. Biomed. Eng.* **4**, 954–972 (2020).

53. Wang, M. et al. Gesture recognition using a bioinspired learning architecture that integrates visual data with somatosensory data from stretchable sensors. *Nat. Electron.* **3**, 563–570 (2020).

54. Zhou, Z. et al. Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. *Nat. Electron.* **3**, 571–578 (2020).

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

56. Zhuang, M. et al. Highly robust and wearable facial expression recognition via deep-learning-assisted, soft epidermal electronics. *Research* **2021**, 2021/9759601 (2021).

57. Zheng, W.-L., Dong, B.-N. & Lu, B.-L. *Multimodal emotion recognition using EEG and eye tracking data*. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (ed Melley, D.) 5040–5043 (IEEE express conference publishing, Chicago, IL, USA, 2014).

58. Schirmer, A. & Adolphs, R. Emotion perception from face, voice, and touch: comparisons and convergence. *Trends Cogn. Sci.* **21**, 216–228 (2017).

59. Ahmed, N., Aghbari, Z. A. & Girija, S. A systematic survey on multimodal emotion recognition using learning algorithms. *Intell. Syst. Appl.* **17**, 200171 (2023).

60. Zhang, R. & Olin, H. Material choices for triboelectric nanogenerators: a critical review. *EcoMat* **2**, e12062 (2020).

61. Kim, W.-G. et al. Triboelectric nanogenerator: structure, mechanism, and applications. *ACS Nano* **15**, 258–287 (2021).

62. Schumann, N. P., Bongers, K., Guntinas-Lichius, O. & Scholle, H. C. Facial muscle activation patterns in healthy male humans: a multichannel surface EMG study. *J. Neurosci. Methods* **187**, 120–128 (2010).

63. Lee, J.-G. et al. Quantitative anatomical analysis of facial expression using a 3D motion capture system: application to cosmetic surgery and facial recognition technology: quantitative anatomical analysis of facial expression. *Clin. Anat.* **28**, 735–744 (2015).

64. Zarins, U. *Anatomy of Facial Expression* (Exonicus Incorporated, 2018).

65. Kim, K. N. et al. Surface dipole enhanced instantaneous charge pair generation in triboelectric nanogenerator. *Nano Energy* **26**, 360–370 (2016).

66. Lee, J. P. et al. Boosting the energy conversion efficiency of a combined triboelectric nanogenerator-capacitor. *Nano Energy* **56**, 571–580 (2019).

67. Lu, Y. et al. Decoding lip language using triboelectric sensors with deep learning. *Nat. Commun.* **13**, 1401 (2022).

68. Yang, J. et al. Triboelectrification-based organic film nanogenerator for acoustic energy harvesting and self-powered active acoustic sensing. *ACS Nano* **8**, 2649–2657 (2014).

69. Yang, J. et al. Eardrum-inspired active sensors for self-powered cardiovascular system characterization and throat-attached anti-interference voice recognition. *Adv. Mater.* **27**, 1316–1326 (2015).

70. Lee, S. et al. An ultrathin conformable vibration-responsive electronic skin for quantitative vocal recognition. *Nat. Commun.* **10**, 2468 (2019).

71. Calvert, D. R. Clinical measurement of speech and voice, by Ronald J. Baken, PhD, 528 pp, paper, College-Hill Press, Boston, MA, 1987, $35.00. *Laryngoscope* **98**, 905–906 (1988).

72. Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y. & Mühlberger, A. The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Front. Psychol.* **6**, 26 (2015).

73. Allcoat, D. & Mühlenen, A. von. Learning in virtual reality: Effects on performance, emotion and engagement. *Res. Learn. Technol.* **26**, 2140 (2018).

74. Colombo, D., Díaz-García, A., Fernandez-Álvarez, J. & Botella, C. Virtual reality for the enhancement of emotion regulation. *Clin. Psychol. Psychother.* **28**, 519–537 (2021).

Acknowledgements

Author contributions

J.P.L. carried out and designed most of the experimental work and data analysis. H.J., H.S., and S.L. assisted in the materials processing and device fabrication. Y.J. assisted in the machine learning algorithms and analysis of the results. P.S.L. revised and improved the manuscript with technical comments. J.K. supervised the whole project and was the lead contact. All authors discussed and wrote and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information