

ARTICLE OPEN



All-weather, natural silent speech recognition via machine-learning-assisted tattoo-like electronics

Youhua Wang^{1,2,5}, Tianyi Tang^{1,2,5}, Yin Xu^{3,5}, Yunzhao Bai^{1,2}, Lang Yin^{1,2}, Guang Li⁴, Hongmiao Zhang³, Huicong Liu³ and YongAn Huang^{1,2}

The internal availability of silent speech serves as a translator for people with aphasia and keeps human–machine/human interactions working under various disturbances. This paper develops a silent speech strategy to achieve all-weather, natural interactions. The strategy requires few usage specialized skills like sign language but accurately transfers high-capacity information in complicated and changeable daily environments. In the strategy, the tattoo-like electronics imperceptibly attached on facial skin record high-quality bio-data of various silent speech, and the machine-learning algorithm deployed on the cloud recognizes accurately the silent speech and reduces the weight of the wireless acquisition module. A series of experiments show that the silent speech recognition system (SSRS) can enduringly comply with large deformation (~45%) of faces by virtue of the electricity-preferred tattoo-like electrodes and recognize up to 110 words covering daily vocabularies with a high average accuracy of 92.64% simply by use of small-sample machine learning. We successfully apply the SSRS to 1-day routine life, including daily greeting, running, dining, manipulating industrial robots in deafening noise, and expressing in darkness, which shows great promotion in real-world applications.

npj Flexible Electronics (2021)5:20; <https://doi.org/10.1038/s41528-021-00119-7>

INTRODUCTION

Silent speech can offer people with aphasia an alternative communication way. More importantly, compared to voice interactions or visual interactions, human–machine interactions using silent speech are versatile enough to work in all-weather surroundings, such as obscured, dynamic, quiet, dark, and noisy. Speaking is owned from babyhood, thus silent speech requires less specialized learning and carries more information than most silent alternatives (from typing to sign language to Morse code). Human brains manipulate the voices by neural signals, and therefore it is effective to learn about human intentions by recognition of surface electromyographic (sEMG) signals on faces. Natural silent speech in daily and working life hinges on high-fidelity sEMG acquisition, accuracy classification, and imperceptible wearable devices.

sEMG signals are ubiquitously distributed on skins and have significant spatiotemporal variability¹. The diversity of sEMG even occurs in the same actions¹. Such complexity of sEMG motivates researchers to develop various classifiers, such as supporting vector machine (SVM)², deep learning^{3,4}, and machine learning^{5,6}, to construct the mapping relations between facial sEMG and silent speech. Silent speech recognition based on EMG can be traced back to the mid-1980s. Sugie⁷ of Japan and Morse⁸ of the United States published their research almost at the same time. Sugie used three-channel electrodes to classify five Japanese vowels, while Morse successfully separated two English words with 97% accuracy. In the past two decades, the number of classified words has been increasing. In 2003, Jorgensen et al. recognized six independent words with 92% accuracy⁹. In 2008, Lee further expanded the number of words to 60, using the hidden Markov model to achieve 87.07% accuracy¹⁰. In 2018, Meltzner et al. recognized >1200 phrases generated from a 2200-word vocabulary with high accuracy

of 91.1%¹¹. In 2020, Wang et al. used the bidirectional long short-term memory to recognize ten words, and the accuracy reached 90%¹². Although the silent speech recognition of sEMG has made great progress in recent years, most of these works use non-flexible electrodes and sampling equipment with a high sampling rate and precision, they are only verified in the laboratory environment, and their long-term performance is not evaluated. Machine learning and deep learning are commonly used in previous research. Deep learning needs to collect a large number of training data with labels, which is very tired and monotonous for people. Compared with deep learning, machine learning has better performance in the case of small sample size and multi-classification. The processing speed is faster, which is more suitable for real-time recognition. Furthermore, it is a trend to reduce the complexity of both the acquisition side and application side by deploying algorithms on the cloud, which is of great importance for wearable devices^{13–15}.

Human faces have complex features, such as geometrically nondevelopable surfaces, softness, dynamical behaviors, and large deformation (~45%)¹⁶. However, current inherently planar and rigid electrodes, including wet gels (Ag/AgCl electrodes)¹⁷, invasive silicon needle electrodes¹⁸, and bulk metal electrodes¹⁹, cannot comply with skin textures, forming unstable gaps between skin and electrodes and correspondingly reducing signal-to-noise ratio. A commercial solution is to employ large-area and strong-adhesion materials (foams, nonwovens, etc.) to wrap electrodes; however, the auxiliary materials severely constrain the movements of muscles and cause uncomfortable experience. Users cannot normally express intentions when a mass of conventional electrodes is attached on faces. The emergence of lightweight, bendable, stretchable tattoo-like electronics shifts the paradigm of the conventional wearable field and show great prospect in

¹State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, China. ²Flexible Electronics Research Center, Huazhong University of Science and Technology, Wuhan, China. ³School of Mechanical and Electric Engineering, Jiangsu Provincial Key Laboratory of Advanced Robotics, Soochow University, Suzhou, China. ⁴State Key Laboratory of Industrial Control Technology, Institute of Cyber Systems and Control, Zhejiang University, Hangzhou, China. ⁵These authors contributed equally: Youhua Wang, Tianyi Tang, Yin Xu. ✉email: zhanghongmiao@suda.edu.cn; hcliu078@suda.edu.cn; yahuang@hust.edu.cn

clinical diagnosis, personal healthcare monitoring, and human-machine interaction^{1,20–25}. The mechanical performance of tattoo-like electronics similar to human skin renders the devices seamlessly conformal with the morphology of skin. The softness and conformability of tattoo-like electronics not only extend the effective contact area of skin-device interfaces, facilitating the accurate transmission of bio-signals from human bodies to external devices, but also achieve imperceptible wearing. Currently, few researchers except us apply the tattoo-like electrodes to the acquisition of silent speech sEMG signals^{1,26}. However, our previous works take a simple try of recording several words, which evidently cannot be extensively implemented in practice.

The proposed strategy in this paper fuses tattoo-like electrodes, wireless data acquisition (DAQ) modulus, and machine-learning algorithm into one all-weather silent speech recognition system (SSRS). The tattoo-like electrodes made up of ultrathin filamentary serpentine keep laminated on facial skins even under long-term, large deformation. The wireless DAQ modulus is a reusable wearable device, serving as real-time bio-data transmission from tattoo-like electrodes to machine learning. The machine-learning algorithm, suitable for multi-label classification of small samples, is deployed on the cloud and used for the accuracy recognition of 110 daily words. To show the applicability of SSRS, we apply SSRS to various scenarios close to daily life.

RESULTS AND DISCUSSION

Design of the SSRS

Figure 1 illustrates the schematics of the all-weather, natural SSRS, which not only helps people naturally communicate in their daily lives but also benefits the users by silently interacting in all-weather conditions. Compared with sign language, with our SSRS system, users do not need a professional training. As shown in Fig. 1a, the SSRS includes four parts: four-channel tattoo-like electronics, a wireless DAQ module, a server-based machine-learning algorithm, and a terminal display of silent speech recognition. Without the use of large sEMG acquisition devices, the user only needs to wear the tattoo-like electronics properly assisted with an ear-mounted wireless DAQ module to capture, process, and transmit the four-channel sEMG signals. The users' real-time sEMG signals are transmitted to a cloud server with powerful computing power and are online classified through the model trained by the machine-learning algorithm. By a Bluetooth connection, a mobile terminal is used to display the recognized speech information and play the audio. The advantages for all-whether, natural use of the SSRS come from the user's long-time wearing, portable device, stable, and high-rate recognition in a variety of scenarios, such as greeting, exercise, repast, work, and dark scenes. Besides, our SSRS uses natural speech, lowering the training cost, and therefore is user-friendly for beginners.

Different speaking is generated through the coordination among facial and neck muscles such that the placements of electrodes are of critical importance. Four pairs of tattoo-like electrodes are selectively attached on the muscles with significant sEMG signals as one silently speaks, including levator anguli oris (LAO), depressor anguli oris (DAO), buccinators (BUC), and anterior belly of digastric (ABD), to elevate the accuracy of silent speech recognition. Each channel includes one reference electrode and one working electrode. To guarantee high-fidelity delivery of sEMG through skin-electrode interfaces, the tattoo-like epidermal electrodes are designed to be only 1.2 μm thick and integrated within a skin-like 3 M Tegaderm patch (Young's module ~ 7 KPa, 47 μm), which are able to perfectly conform with topologies of skins. The electrodes are further patterned to be filamentary serpentine to improve elastic stretchability. Specifically, the width, the ribbon-width-to-arc-radius ratio, and the arc angle of filamentary serpentine are 500 μm , 0.32, and 20°, respectively. According to the mechanics

theory of serpentine ribbons²⁷, the design can simply reach 4% elastic stretchability, much less than the deformation of facial skins. We introduce a so-called "electricity-preferred" method to enable the tattoo-like electrodes adequate to tough stretch, which will be described in "Wearable characterizations of tattoo-like electrodes." The overall size of one electrode is about 18 mm \times 32 mm. The tattoo-like electrodes are prepared by the low-cost but high-efficiency "Cut and Paste" methods^{24,28} and the processes are described in the "Methods" section.

Figure 1b shows the picture of a user wearing the tattoo-like electronics and an ear-mounted wireless DAQ module²⁹. As shown in Supplementary Fig. 1, the method of low-temperature alloy welding effectively increases the strength of the connection, which ensures all-weather use without damage. The specific operation of connection is described in "Methods." The block diagram in Fig. 1c summarizes the system architecture and overall wireless operation procedures³⁰. The wireless DAQ module has four signal collection channels and each channel is connected with a working electrode and a reference electrode of tattoo-like electronics. The sEMG signal from each channel is processed by an instrumentation amplifier and an analog filter. Then a micro control unit and a Bluetooth transmission unit are employed to convert and transmit the four-channel signals, simultaneously. The DAQ module amplifies the high-fidelity sEMG signals by 1000 times and then extracts the effective signals, which carry speech information through 10–500 Hz band-pass filtering³¹. A 10-bit analog-to-digital converter of the micro control unit operating at a sampling frequency of 500 Hz digitizes the signals collected from each channel. Under the tests of the recognition rate under different sampling frequencies in Supplementary Table 1, the low sampling frequency of 500 Hz can not only meet the requirements of a high recognition rate but also reduce the processing time and power consumption of the SSRS. The Bluetooth transmission unit uses the fifth-generation Bluetooth protocol, which allows continuous data transmission at a rate of up to 256 kb/s^{32,33}. With the help of a Bluetooth receiver, the mobile terminal receives the recognition information and performs a proper interaction in daily applications. The recognition of silent speech is achieved by training facial sEMG signals with the linear discriminant analysis (LDA) algorithm, which will be described in "sEMG-based silent speech recognition by machine learning." Figure 1d shows the confusion matrix of recognition results of the proposed frequently used 110 words in daily life, which are divided into 13 categories. The high recognition rate of 92.64% of the SSRS can fully meet the users' daily communication requirements.

Wearable characterizations of tattoo-like electrodes

The mechanical mismatch of the skin-electrode interface constrains the natural deformation of human skins, thus causing an uncomfortable wearing experience. Figure 2a and Supplementary Fig. 2 compares the mechanical constrain of tattoo-like electrodes and commercial gel electrodes to human skins under large deformation. It is obvious that, no matter how a human face deforms extremely, including opening mouth, inflating cheeks, and twitching mouth toward left/right, the ultrathin tattoo-like electrodes comply with deformed skins while the gel electrodes constrain the deforming movements of skins. The strong driving forces between skin-gel electrode interfaces not only decrease the wearability but also delaminate the interfaces. Figure 2b displays the conformability of tattoo-like electrodes on skin textures at different scales. It is evident that tattoo-like electrodes are able to perfectly match both coarse and fine skin textures. The overlapped curves in Supplementary Fig. 3 evidently indicate that, due to the excellent conformability of tattoo-like electrodes, the skin-electrode interface has robust electrical performance even after suffering large deformation. Figure 2c exhibits the micro-optical photographs of stretchability of skin-electrode interfaces. Soft silicone rubbers

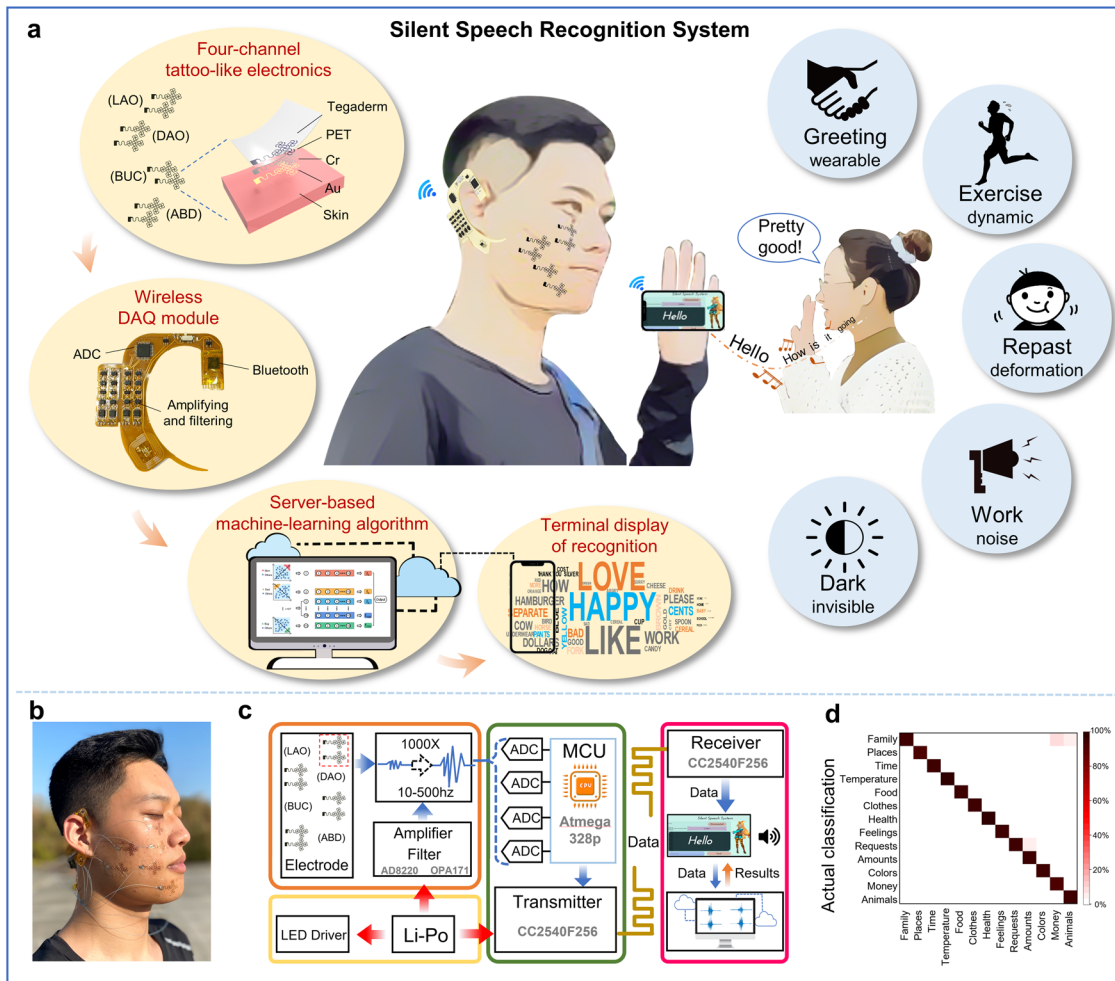


Fig. 1 Design of silent speech recognition system. **a** Schematic illustration of an all-weather, natural SSRS, including four-channel tattoo-like electronics, the wireless DAQ module, the server-based machine-learning algorithm, and the terminal display of recognition, with adaptability in various scenarios. The identifiable drawing is fully consented by the written consent. **b** The photograph of a participant wearing the SSRS. The identifiable photograph is fully consented by the written consent. **c** Functional block diagram of the wireless DAQ module. **d** Confusion matrix of recognition results of the frequently used 110 words. The words are from 13 categories.

(Young's modulus is ~ 0.35 MPa, close to that of human skin) were used to mimic human skin, and the mimic skin laminated with an ultrathin tattoo-like electrode was stretched by 30%, equal to the elastic limit of human skin. The comparison results in Fig. 2c show that the skin–electrode interface is still intact after tensed. Motion artifact has a critical impact on the signal-to-noise ratio. The robust conformability on complex skin textures in Fig. 2b and under large deformation in Fig. 2c has the ability to suppress the motion artifact of various silent voices.

Though researches have shown that the elastic limit of human skin is about 30%^{25,34,35} and that tensing limit without pain is about 20%^{36,37}, the tension ability of human faces reaches up to 45%¹⁶. Figure 2d, e present the strain contours of ultrathin tattoo-like electronics axially extended with 45% applied strain in horizontal and vertical directions. The maximum principal strains are, respectively, 4.1% (horizontal direction) and 1.8% (vertical direction), both beyond the yield limit of nano-film gold (0.3%)³⁷. However, for physiological electrodes, we pay more attention to electrical conductivity. Figure 2f plots the change of electrical resistivity of nano-film gold in the 100 nm Au/10 nm Cr/1.1 μ m polyethylene terephthalate (PET) composite with respect to applied strain under the uniaxial tension (the inserted schematics in Fig. 2f). The change of electrical resistivity is simply about 5% when tensed to 2% while it sharply rises to $\sim 30\%$ when tensed to 4%. Supplementary Fig. 4 clearly shows that the parts beyond 2%

in Fig. 2d are at inner crests of serpentine structures. The parts $<2\%$ and those beyond 2% can be equivalent to parallel circuits, seen in the schematics in Fig. S4. According to the electrical resistance change rule in a parallel circuit, the parts beyond 4% cannot bring remarkable influence on the whole resistance of the ultrathin tattoo-like electronics. Therefore, it is still claimed that the ultrathin tattoo-like electrodes are effective structures. Such thought, giving priority to the electrical performance, enables the structural design and fabrication of electrodes simple, which is called the “electricity-preferred” method.

The practical applications of silent speech demand long-term wearing performance, and the background noises and the skin–electrode contact impedance directly determine whether the deserved silent speech are collected or not, therefore we tested both electrical parameters of ultrathin tattoo-like electrodes during a 10-h wearing period. Commercial gel electrodes (3 M) were used as the gold standard to study the long-term electrical performances. Two pairs of ultrathin tattoo-like electrodes and gel electrodes were closely attached on the subject's forearm and the distance between two tattoo-like electrodes, or gel electrodes, were set to 7 cm. The subject was required to begin to run for half an hour at the eighth hour. The results are illustrated in Fig. 2g. The gel electrodes have robust noise and impedance during the whole measurement while the noise and impedance of tattoo-like electrodes gradually degrade. The background noise and the

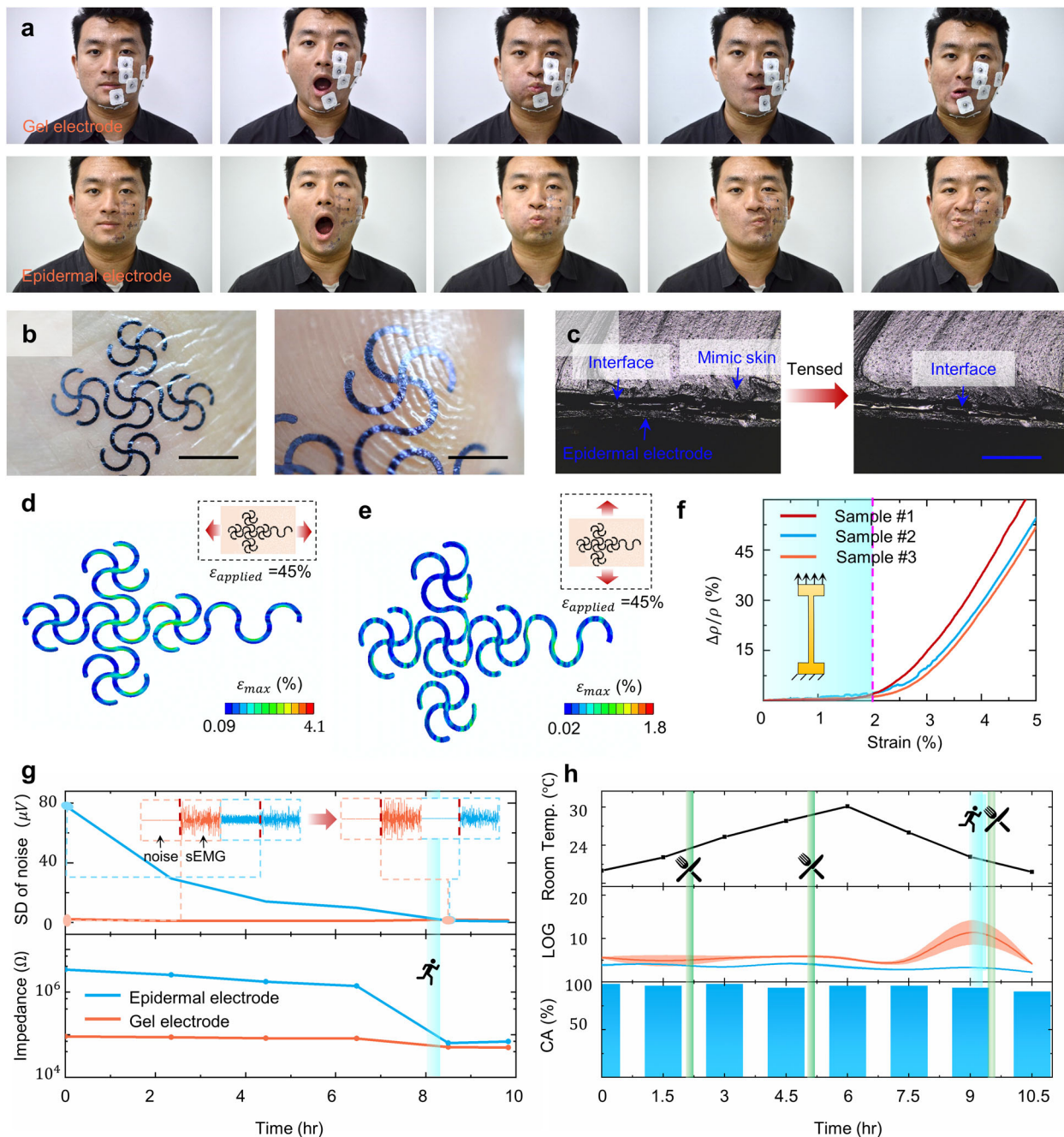


Fig. 2 Characterizations of tattoo-like electrodes. **a** The wearability of tattoo-like electrodes and gel electrodes when attached on the subject's face. The identifiable photographs are fully consented by the written consents. **b** The tattoo-like electrodes conform with skin textures at different scales. The scale bars on the left and right panel are 6 and 2.5 mm, respectively. **c** The skin–electrode interface before and after being stretched by 30%. The scale bar is 150 μm . **d, e** The strain distributions of tattoo-like electrodes under horizontal and vertical tensing with 45%. **f** The resistivity changes of gold ribbons with respect to strain. **g** The long-term measurement to investigate the change of background noise and impedance. The standard deviation (SD) characterizes the strength of background noises. **h** Long-term measurement of log detector (LOG) and classification accuracy (CA).

impedance highly depend on the skin–electrode interface and skin properties³⁸. The chloride ions contained in gel electrodes freely permeate through the stratum corneum, significantly suppressing noise and impedance. After running, both electrical parameters of tattoo-like electrodes sharply go down and the noise is even weaker than that of gel electrodes. It is mainly because, during the test, the Tegaderm film prevents the evaporation of sweat, and the sweat goes through the stratum corneum and finally accumulates at the skin–electrode interface, dramatically reducing the noise and impedance. Now that the

long-term wear can affect the background noise, it is reasonable to consider the effect of daily usages on sEMG. Thus, we study the effect of complicated daily activities on signal features and classification accuracy. Tattoo-like electrodes and gel electrodes were, respectively, attached on the left and right faces. The subject was required to dine at the 2.3th hour, 5th hour, and 9.5th hour and run for 20 min at the 9th hour. Additionally, the room temperature gradually rises to 30.1 from 20 $^{\circ}\text{C}$ in 6 h and then declines to 19.8 $^{\circ}\text{C}$ in 4.5 h, shown in the upper panel in Fig. 2h. The middle panel in Fig. 2h plots the log detector (LOG) of the first

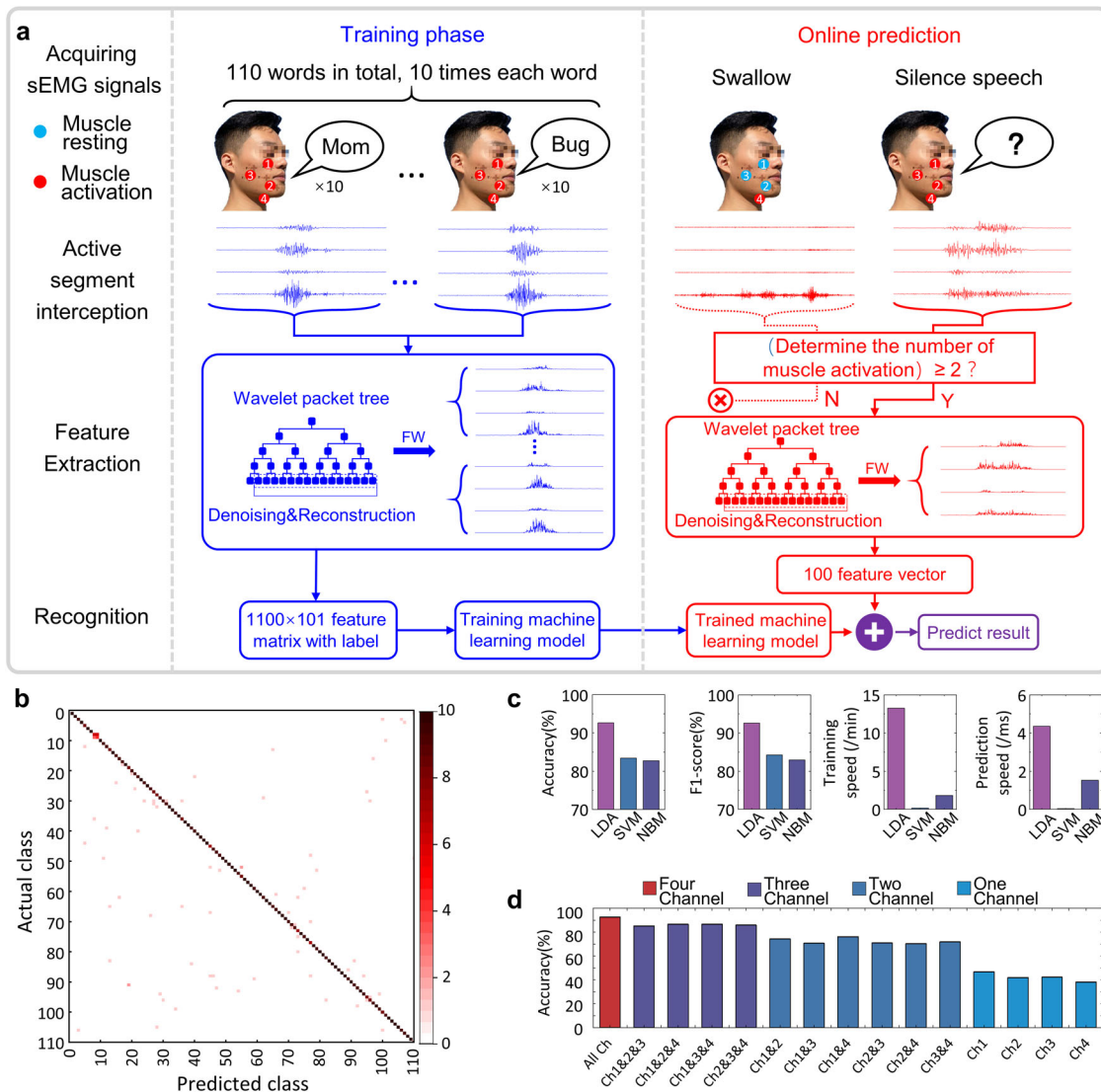


Fig. 3 Flow chart and the evaluation of silent speech recognition system. **a** Recognition flow chart of training phase (left) and online prediction (right). The identifiable photographs are fully consented by the written consents. **b** Confusion matrix of recognition results of 110 ASL words. **c** Prediction performance of different classifiers LDA, SVM, and NBM. **d** Accuracy rate from multiple channels to a single channel.

channel with respect to wearing time. The results show that the signal feature captured by tattoo-like electrodes are immune to room temperature change, running, and dining during the long-term test while that of gel electrodes keep fluctuating. The bottom panel in Fig. 2h shows that silently saying “Hello, nice to meet you” at different times can achieve the high classification accuracy of 90% and the average classification accuracy is 95%. In conclusion, the long-term experiments in Fig. 2g, h offer the silent speech the probability in extensive applications.

sEMG-based silent speech recognition by machine learning

To keep the normal communications going smoothly, a collection of 110 words and phrases in American Sign Language (ASL) covering the words frequently used in daily life is selected for recognition (Supplementary Table 2)³⁹. The collection is divided into 13 categories, including time, place, emotion, health, etc. According to the previous researches, eight muscles were selected (Supplementary Fig. 5)^{9,10,26,40}. However, the more channels are used, the more power consumption of the wireless DAQ modulus the transmitted data demand, thus it is necessary to pick out the optimal combination. The classification accuracy of different

channel combinations from one muscle to eight muscles was calculated. The results in Supplementary Fig. 6 clearly show that the mean classification accuracy gradually increases and approaches 92.1% along with the growth of channels. The number of channels is finally selected as four. Furthermore, the combination with the highest classification accuracy among the four channels is selected (Supplementary Fig. 7), specifically, LAO, DAO, BUC, and ABD.

The high-quality sEMG for the recognition is immediately recorded by the flexible tattoo-like electronics and transmitted through the wireless DAQ module to the cloud server in real time once participants perform silent speech tasks. The proposed recognition procedure shown in Fig. 3a is deployed on the cloud server and comprises the active segment interception, the training phase (left panel in Fig. 3a), and online prediction (right panel in Fig. 3a).

The active segment interception plays an essential role in distinguishing between the silent speech-related sEMG and non-silent speech-related sEMG (swallowing, blinking, etc.). According to our experience and the previous research⁴¹, the sEMG absolute amplitude threshold and the number threshold of facial muscles activated by the silent speech are set to 50 μV and 2, respectively. The active segment interception extracts the sEMG of 800 ms

before and 1200 ms after the moment when the sEMG signals achieve beyond both thresholds above.

Due to the significant discrimination of the ultrathin devices and the welding spots in terms of thickness (1.2 and about 300 μm , respectively), there is a huge difference in bending stiffness, which may easily cause motion artifacts. The baseline wandering of signals is unavoidable though the violent shaking of connection between tattoo-like electronics and wires is suppressed by the adhesive Tegaderm. To remove the baseline wandering, a 4-level wavelet packet with a soft threshold is used to decompose the extracted signals and reconstruct signals with the node coefficients from the 2nd to 16th node in the 4th layer⁴². Fifteen relative wavelet packet energy⁴³ as frequency-domain features are extracted from 15 nodes, respectively. Then the denoised signals are treated by full-wave rectification and ten time-domain features are extracted from the rectified signals. The definitions of all features are listed in Supplementary Table 3^{44–47}. Considering that there are four channels, a silent speech word corresponds to a feature vector composed of 100 features. In the training phase, the silent speech users are required to speak 110 words and repeat 10 times each word. With an additional vector of 101 labels, the dimensions of the feature matrix reach up to 1100×101 . The feature matrix is input into an LDA model for training and tenfold cross-validation is used to evaluate the training effectiveness. To speed up the recognition, one vs rest is selected as the multiple pattern recognition. In the online prediction, a vector of 100 features is input into the well-trained LDA model to predict the silent speech users' intentions.

For offline recognition, the average classification accuracy of LDA reaches up to 92.64% in the case of 110 words (Fig. 3b). Supplementary Fig. 8 shows the classification accuracy of each word. SVM and naive Bayesian model (NBM) are the other two machine-learning methods used extensively and compared with LDA in four aspects: classification accuracy, F1-score, training speed, and prediction speed. Only a small amount of data to be collected is of great importance for users to avoid monotony and fatigue. The recognition of 110 words with 10 samples is the typical few-shot classification. The comparison results in Fig. 3c clearly show that LDA is superior to SVM and NBM in whatever aspect of performances. In conclusion, LDA with high classification accuracy and high prediction speed renders silent speech users to naturally communicate.

Considering the contamination from eye blinking to the facial sEMG, the influence of electrooculogram (EOG) on SSRS is discussed. Only channel 1 (LAO) closest to the eye is affected by the EOG. The maximum amplitude of EOG is 30 μV , which is less than the threshold of the muscle activity detection (50 μV). Although the muscle activity segment detection will not misjudge the EOG signal as the silent speech-related signal, sometimes speaking and blinking happen at the same time. When such situation occurs, the EOG signal can be eliminated by preprocessing. When preprocessing the raw signal, the first node in the fourth layer decomposed by wavelet packet is not involved in reconstructing the signal. This means that the original signal is filtered by a 15 Hz high-pass filter (the frequency range of EOG is 0–12 Hz⁴⁸). After preprocessing, the maximum amplitude of EOG is <12 μV (Supplementary Fig. 9). The effect of EOG signal on SSRS can be ignored.

Considering possible extreme conditions during the long-term usage, such as wire disconnecting or electrode damaging, SSRS may lose some sEMG channels. All the possible scenarios from four channels (normal state) to only one channel are tested to examine the robustness of LDA (Fig. 3d). When three channels are in good condition, the average classification accuracy can reach >85%. When two channels work, the average classification accuracy can reach >70%. Even when merely one channel remains intact, the average classification accuracy can reach 42.27%, which is much higher than the random recognition (0.91%) of one word. Therefore, our SSRS has promising applications in extreme conditions.

All-weather demonstration of the SSRS

Figure 4a exhibits five typical scenarios that a user often experiences in daily life, including greeting, exercise, repast, working in a noisy environment, and communicating in darkness. The model training and signal recognition of the SSRS are based on cloud servers with powerful computing capabilities, which allow users to only need a mobile phone with basic communication functions. The popularization of the fifth-generation communication technology has great potential to further reduce the delay of SSRS and bring users a more natural interactive experience. With the help of SSRS, users can not only communicate point to point but also express their intentions point to net. Some excellent capabilities and advantages of the SSRS are demonstrated in detail as below.

Figure 4bi demonstrates a typical greeting scene in which the user needs to communicate with people in his/her life. Figure 4bii shows the real-time sEMG signals collected from the four channels of the SSRS, as long as the user silently pronounces the words "Hello," "Morning," "Thanks," "Goodbye," and so on. The characteristics of different words can be easily identified from different channels in real time. It is proved that the sEMG of facial muscles carries enough speech information. As shown in Supplementary Video 1, the subject is able to communicate naturally with his friend in silent speech with the help of SSRS. The confusion matrix in Fig. 4biii indicates that the recognition rate of eight words in the greeting scene is 95%. The SSRS is able to meet the user's natural communication in three aspects. First, the wearable flexible printed circuit and the wireless connection of the SSRS provide more convenience, which greatly extends the activity range of the users. Second, the ultrathin tattoo-like electronics can collect high-fidelity facial sEMG and can be worn all day long. Finally, the LDA algorithm achieves a high recognition rate of 92.64% in 110 classifications. This is more than enough for the user to naturally communicate in daily life without any sign-language training.

The proposed ultrathin and super-conformal tattoo-like electronics ensure the stability of acquisition of the sEMG signal, even if the user experiences strenuous exercise activities. Figure 4ci demonstrates a typical exercise scene. The comparative experiments on background noise and recognition rate at different running speeds were carried out. We selected five common words ("Home," "Work," "School," "Store," "Church"), and tested the recognition rates in four motion states, i.e., resting (0 m/s), walking (1 m/s), jogging (3 m/s), and running (5 m/s). Each word was repeated ten times by subject. In Fig. 4cii, the recognition rate at resting, walking, and jogging state maintain as high as $\geq 96\%$ and at running state it is up to 86%. The average recognition rate at four states is 96% in Fig. 4ciii, which proves the excellent stability of SSRS. Supplementary Video 2 shows the exercise scene where SSRS can still recognize words correctly when the subject is jogging. It is also verified that the SSRS does not get affected by the user's body shaking and hence has great potential to replace touch control to operate smart devices.

In addition, to maintain a high recognition rate in dynamic conditions, the all-weather SSRS has good tolerance to mouth deformation and muscle fatigue. Figure 4di displays the scene of a user repasting in a restaurant with the help of SSRS (in Supplementary Video 3). According to the statistics, users chew about 400–600 times during a meal. Therefore, the subject was asked to repeat mouth movement actions 0–200 times, and four kinds of mouth deformations each time were performed, as seen in Fig. 4dii. The recognition rate of five words related to food, such as "Pizza," "Milk," "Hamburger," "Hotdog," "Egg," remains consistently as high as $\geq 96\%$, as shown in Fig. 4diii, and the total recognition rate of five different repeat times is 98%.

Compared with Automatic Speech Recognition (ASR), the SSRS has a good capability of tolerating sound especially in noisy or quiet-required environments, such as workplaces and public places. Figure 4ei shows a noisy industry environment, which

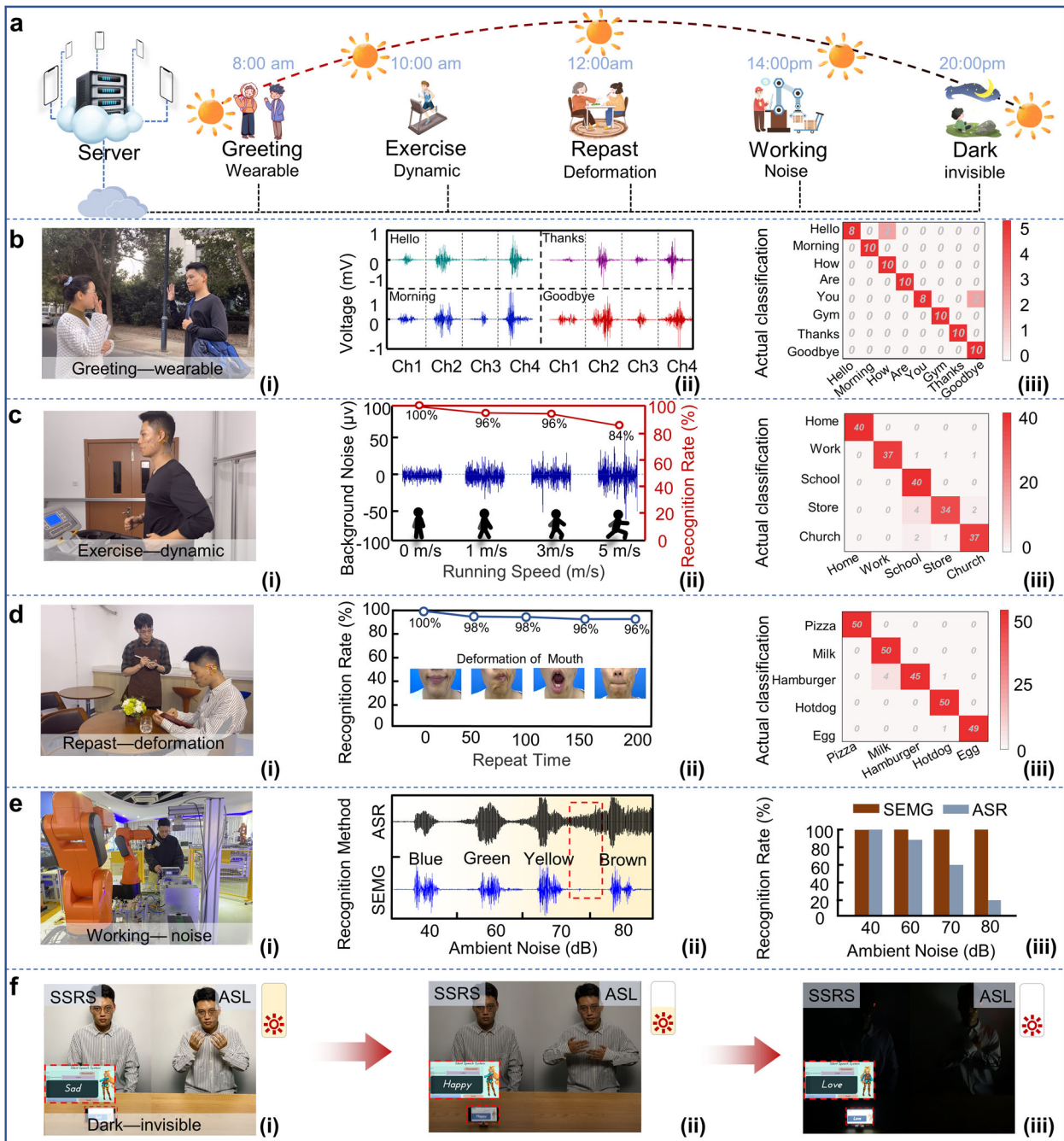


Fig. 4 All-weather demonstration of the SSRS. **a** Five typical scenarios experienced in daily life. **b** Wearable and natural communication: (i) The scene of greeting. (ii) Four-channel sEMG of four representative words. (iii) The recognition rates of eight words in the greeting scene. **c** All-weather use in dynamic condition: (i) The scene of exercise. (ii) The recognition rates and background noises of five words related to locations under four different running speeds. (iii) The recognition rates of five words related to locations in four different exercise states. **d** All-weather use in the large deformation condition: (i) The scene of the repast. (ii) The recognition rates of five words related to food under different repeat times, each time contains four kinds of mouth deformation. (iii) The recognition rates of five words related to food. **e** Adaptability in the noisy environment. (i) The scene of the noisy working environment. (ii) The signals of ASR and sEMG. (iii) The comparison of the recognition rates of two recognition methods under four different ambient noises. **f** Adaptability in the darkened environment. i–iii The comparison of recognition effect of Silent Speech Recognition (left) and American Sign Language (right) in the darkened environment. The identifiable photographs in **b–f** are fully consented by the written consents.

the user may experience at work. The comparative experiments on the performance of SSRS (left) and ASR (right) in a noisy environment were carried out in Supplementary Video 4. As four words related to colors are tested in Fig. 4eii, the noise of ASR decreases with the increase of ambient decibels, while the noise of sEMG remains unchanged (the details seen in the

dashed box). Figure 4eiii depicts the comparison of the recognition rate of SSRS and ASR under different decibels of ambient noise. When the ambient noise reaches 80 dB, the recognition rate of ASR is dropped down to only 20%, while the recognition rate of SSRS remains as 100%, as shown in Supplementary Fig. 10. It can be seen from the above comparison that SSRS has great potential

to be an effective interface of human–machine interaction. People can easily control the equipment through SSRS in the noisy working environment (Supplementary Fig. 11).

One alternative to overcome darkness is by means of smart gloves. In the past few years, some smart gloves have been able to effectively recognize sign language. In 2019, Sundaram et al. proposed a scalable tactile glove, which realized the classification of 8 gestures, and the recognition rate was 89.4%⁴⁹. In 2020, Zhou et al. demonstrated a recognition rate of up to 98.63% by using machine-learning-assisted stretchable sensor arrays to analyze 660 acquired sign language hand gestures⁵⁰. However, sensor-based gesture recognition is different from SSRS. It takes a lot of time to master sign language, which becomes an obstacle to assisting pronunciation. Compared with the sign language, the all-weather SRSS can be used naturally without any technical threshold or any influence of brightness. Figure 4f compares the recognition effects of SSRS (left) and ASL (right) in a gradually darkening environment. The subject expresses “Happy,” “Sad,” “Sorry,” “Angry,” and “Love” through SSRS and ASL, respectively. As the light becomes darker, the ASL cannot be identified anymore as seen in Supplementary Video 5, while the SSRS still works well. Therefore, the SSRS would be a better choice than ASL for users in the future.

We have successfully proposed a silent speech strategy by designing an all-weather SSRS, realizing natural silent speech recognition. The ultrathin tattoo-like electronics are able to be conformal with various skin textures and the simple but effective electricity-preferred design method renders filamentary serpentine bear ~45% extension of facial skins. Long-term attachment not only decreases the interface impedance but also maintains the features of sEMG. The wireless DAQ module bridges the tattoo-like electronics with LDA algorithm. The LDA algorithm is deployed on the cloud to lightweight the wireless DAQ module and achieves a high recognition rate of 110 words. The 1-day routine life demonstrates the competence for future all-weather, natural silent speech, including exchanging of communication in people with aphasia, communication while keeping quiet, and human–machine interactions free from surrounding disturbance.

METHODS

Manufacturing processes of the tattoo-like electronics

The fabrication processes started with the lamination of an ultrathin PET film (1.1- μm thickness) on the wetted water transfer paper (Huizhou Yibite Technology, China). The composite substrate of PET and water transfer paper was baked in an oven (ZK-6050A, Wuhan Aopusen Test Equipment Inc., China) at 50–60 °C for ~1 h and subsequently at 100–110 °C for ~2 h for adequate drying. Ten-nm-thick chromium (Cr) and 100-nm-thick gold (Au) were deposited on PET. Then the film was cut by a programmable mechanical cutter (CE6000-40, GRAPHTEC, Japan) to a designed pattern. A tweezer was used to carefully remove the unnecessary part of the pattern on the re-wetted water transfer paper. Then the patterned film was flipped over using the thermally released tape (TRT) (REVALPHA, Nitto, Japan). The TRT was deactivated on a hotplate at ~130 °C for 3 min, followed by sticking to the 3M Tegaderm. Finally, the deactivated TRT was removed to get the tattoo-like electronics.

Method of connecting electrodes and wireless DAQ module

The processed electrode was placed on the platform with the Tegaderm layer facing up. Then we used low-temperature welding to realize the connection between the pad and the wire. The electrode was peeled off and folded in half along the pad and then used low-temperature alloy to weld the wire on the pad. When the temperature dropped to room temperature, we used another Tegaderm to fix the connection between the pad and the wire.

Design of wireless DAQ module

The wireless DAQ module used were AD8220, OPA171, Atmega328p, and CC2540F256. AD8220 and OPA171 were used to amplify the original sEMG signal 1000 times. Atmega328p with a 10-bit precision was used for analog-to-digital conversion and the sampling frequency of the microprocessor was set to 500 Hz. CC2540F256 was used to send and receive data.

Experimental process of EMG signal acquisition

The reference electrode needs to be placed on electrically neutral tissue⁵¹; the position of the posterior mastoid closest to the acquisition device is selected as the reference electrode placement position. Another 8 electrodes were attached to the designated 4 muscles, with every 2 electrodes targeting 1 muscle, and the distance between the 2 electrodes was set to 2 cm. Before applying the electrode, the target locations were cleaned with clean water. The wireless DAQ module and the electrodes were connected by wires, and the wireless circuit module was hung on the subject’s ear. The subject was instructed to read each word silently ten times. During the experimental sessions, the subject was asked to avoid swallowing, coughing, and other facial movements unrelated to silent reading.

System environment and parameters of SSRS

The SSRS was built in Windows 10 environment. The LDA algorithm in the machine-learning toolbox of MATLAB 2019b was used in SSRS. In real-time recognition, the time window length was 2000 ms and the sliding window was 200 ms. The sampling frequency was 500 Hz.

Ethical information for studies involving human subjects

All experiments involving human subjects were conducted in compliance with the guidelines of Institutional Review Board and were reviewed and approved by the Ethics Committee of Soochow University (Approval Number: SUDA20210608A01). All participants for the studies were fully voluntary and submitted the informed consents. The SSRS is located on the silent speech users’ faces and thus the necessary but limited identifiable images have to be used. All identifiable information was totally consented by the user.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

CODE AVAILABILITY

The custom code and mathematical algorithm that support the findings of this study are available at <https://doi.org/10.5281/zenodo.4925493>. The most recent version of this code can be found at https://github.com/xsjzbx/paper_All-weather-natural-silent-speech-recognition-via-ML-assisted-tattoo-like-electronic.

Received: 8 April 2021; Accepted: 30 June 2021;

Published online: 13 August 2021

REFERENCES

- Wang, Y. H. et al. Electrically compensated, tattoo-like electrodes for epidermal electrophysiology at scale. *Sci. Adv.* **6**, eabd0996 (2020).
- Cai, S. et al. SVM-based classification of sEMG signals for upper-limb self-rehabilitation training. *Front. Neurobot.* **13**, 31 (2019).
- Cote-Allard, U. et al. Interpreting deep learning features for myoelectric control: a comparison with handcrafted features. *Front. Bioeng. Biotechnol.* **8**, 158 (2020).
- Orjuela-Canon, A. D., Ruiz-Olaya, A. F. & Forero, L. Deep neural network for EMG signal classification of wrist position: Preliminary results. In *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)* 5 pp. (IEEE, 2017).
- Jaramillo-Yáñez, A., Benalcázar, M. E. & Mena-Maldonado, E. Real-time hand gesture recognition using surface electromyography and machine learning: a systematic literature review. *Sensors* **20**, 2467 (2020).
- Jiang, Y. et al. Shoulder muscle activation pattern recognition based on sEMG and machine learning algorithms. *Comput. Biol. Med.* **197**, 105721 (2020).
- Sugie, N. & Tsunoda, K. A speech prosthesis employing a speech synthesizer - vowel discrimination from perioral muscle activities and vowel production. *IEEE Trans. Biomed. Eng.* **32**, 485–490 (1985).
- Morse, M. S. & Obrien, E. M. Research summary of a scheme to ascertain the availability of speech information in the myoelectric signals of neck and head muscles using surface electrodes. *Comput. Biol. Med.* **16**, 399–410 (1986).
- Jorgensen, C., Lee, D. D. & Agabon, S. Sub Auditory Speech Recognition Based on EMG Signals. In *International Joint Conference on Neural Networks 2003* 3128–3133 (Institute of Electrical and Electronics Engineers Inc., 2003)
- Lee, K. S. EMG-based speech recognition using hidden Markov models with global control variables. *IEEE Trans. Biomed. Eng.* **55**, 930–940 (2008).

11. Meltzner, G. S. et al. Development of sEMG sensors and algorithms for silent speech recognition. *J. Neural Eng.* **15**, 046031 (2018).
12. Wang, Y. et al. Silent speech decoding using spectrogram features based on neuromuscular activities. *Brain Sci.* **10**, 442 (2020).
13. Molina-Molina, A. et al. Validation of mDurance, a wearable surface electromyography system for muscle activity assessment. *Front. Physiol.* **11**, 606287 (2020).
14. Peng, Y. H., Wang, X. J., Guo, L., Wang, Y. C. & Deng, Q. X. An efficient network coding-based fault-tolerant mechanism in WBAN for smart healthcare monitoring systems. *Appl. Sci.* **7**, 18 (2017).
15. Mehmood, G., Khan, M. Z., Abbas, S., Faisal, M. & Rahman, H. U. An energy-efficient and cooperative fault-tolerant communication approach for wireless body area network. *IEEE Access* **8**, 69134–69147 (2020).
16. Hsu, V. M., Wes, A. M., Tahiri, Y., Cornman-Homonoff, J. & Percec, I. Quantified facial soft-tissue strain in animation measured by real-time dynamic 3-dimensional imaging. *Plast. Reconstr. Surg. Glob. Open* **2**, e211 (2014).
17. Bracken, D. J., Ornelas, G., Coleman, T. P. & Weissbrod, P. A. High-density surface electromyography: a visualization method of laryngeal muscle activity. *Laryngoscope* **129**, 2347–2353 (2019).
18. Kim, S. et al. Integrated wireless neural interface based on the Utah electrode array. *Biomed. Microdevices* **11**, 453–466 (2009).
19. Liao, L. D., Wang, I. J., Chen, S. F., Chang, J. Y. & Lin, C. T. Design, fabrication and experimental validation of a novel dry-contact sensor for measuring electroencephalography signals without skin preparation. *Sensors* **11**, 5819–5834 (2011).
20. Kim, D. H. et al. Epidermal electronics. *Science* **333**, 838–843 (2011).
21. Kim, Y. et al. A bioinspired flexible organic artificial afferent nerve. *Science* **360**, 998–1003 (2018).
22. Yang, J. C. et al. Electronic skin: recent progress and future prospects for skin-attachable devices for health monitoring, robotics, and prosthetics. *Adv. Mater.* **31**, e1904765 (2019).
23. Son, D. et al. An integrated self-healable electronic skin system fabricated via dynamic reconstruction of a nanostructured conducting network. *Nat. Nanotechnol.* **13**, 1057–1065 (2018).
24. Zhou, Y. et al. Multichannel noninvasive human-machine interface via stretchable μm thick sEMG patches for robot manipulation. *J. Micromech. Microeng.* **28**, 014005 (2018).
25. Wang, Y. et al. Low-cost, μm -thick, tape-free electronic tattoo sensors with minimized motion and sweat artifacts. *npj Flex. Electron.* **2**, 6 (2018).
26. Liu, H. C. et al. An epidermal sEMG tattoo-like patch as a new human-machine interface for patients with loss of voice. *Microsyst. Nanoeng.* **6**, 16 (2020).
27. Widlund, T., Yang, S. X., Hsu, Y. Y. & Lu, N. S. Stretchability and compliance of freestanding serpentine-shaped ribbons. *Int. J. Solids Struct.* **51**, 4026–4037 (2014).
28. Yang, X. et al. “Cut-and-paste” method for the rapid prototyping of soft electronics. *Sci. China Technol. Sci.* **62**, 199–208 (2019).
29. Vesa, E. P. & Ilie, B. Equipment for SEMG signals acquisition and processing. In *International Conference on Advancements of Medicine and Health Care through Technology, MEDITECH 2014* 187–192 (Springer Verlag, 2014).
30. Ferreira, J. M. & Lima, C. Distributed system for acquisition and processing the sEMG signal. In *1st International Conference on Health Informatics, ICHI 2013* 335–338 (Springer, 2013).
31. Alemu, M., Kumar, D. K. & Bradley, A. Time-frequency analysis of SEMG—with special consideration to the interelectrode spacing. *IEEE Trans. Neural Syst. Rehabil. Eng.* **11**, 341–345 (2003).
32. Sheikh, M. U., Badihi, B., Ruttik, K. & Jantti, R. Adaptive physical layer selection for bluetooth 5: measurements and simulations. *Wirel. Commun. Mob. Comput.* **2021**, 1–10 (2021).
33. Sun, D. Z., Sun, L. & Yang, Y. On secure simple pairing in bluetooth standard v5.0-Part II: Privacy analysis and enhancement for low energy. *Sensors* **19**, 3259 (2019).
34. Lee, K. et al. Mechano-acoustic sensing of physiological processes and body motions via a soft wireless device placed at the suprasternal notch. *Nat. Biomed. Eng.* **4**, 148–158 (2020).
35. Koh, A. et al. A soft, wearable microfluidic device for the capture, storage, and colorimetric sensing of sweat. *Sci. Transl. Med.* **8**, 366ra165 (2016).
36. Hyoyoung, J. et al. NFC-enabled, tattoo-like stretchable biosensor manufactured by cut-and-paste method. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 4094–4097 (IEEE, 2017).
37. Tian, L. et al. Large-area MRI-compatible epidermal electronic interfaces for prosthetic control and cognitive monitoring. *Nat. Biomed. Eng.* **3**, 194–205 (2019).
38. Huigen, E., Peper, A. & Grimbergen, C. A. Investigation into the origin of the noise of surface electrodes. *Med. Biol. Eng. Comput.* **40**, 332–338 (2002).
39. Vicars, W. First 100 signs: American Sign Language (ASL). <http://www.lifeprint.com/asl101/pages-layout/concepts.htm> (2002).
40. Meltzner, G. S. et al. Speech recognition for vocalized and subvocal modes of production using surface EMG signals from the neck and face. In *INTER-SPEECH 2008 - 9th Annual Conference of the International Speech Communication Association* 2667–2670 (International Speech Communication Association, 2008).
41. Hooda, N., Das, R. & Kumar, N. Fusion of EEG and EMG signals for classification of unilateral foot movements. *Biomed. Signal Process. Control* **60**, 101990 (2020).
42. Mithun, P., Pandey, P. C., Sebastian, T., Mishra, P. & Pandey, V. K. A wavelet based technique for suppression of EMG noise and motion artifact in ambulatory ECG. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2011**, 7087–7090 (2011).
43. Xiao et al. Classification of surface EMG signal using relative wavelet packet energy. *Comput. Methods Prog. Biomed.* **79**, 189–195 (2005).
44. Pancholi, S. & Joshi, A. M. Electromyography-based hand gesture recognition system for upper limb amputees. *Electron. Lett.* **3**, 1–4 (2019).
45. Yikang, Y. et al. A Multi-Gestures Recognition System Based on Less sEMG Sensors. In *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)* 105–110 (IEEE, 2019).
46. Too, J., Abdullah, A. R. & Saad, N. M. Classification of hand movements based on discrete wavelet transform and enhanced feature extraction. *Int. J. Adv. Comput. Sci. Appl.* **10**, 83–89 (2019).
47. Savur, C. & Sahin, F. Real-Time American Sign Language Recognition System Using Surface EMG Signal. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* 497–502 (IEEE, 2015).
48. Halder, S. et al. Online artifact removal for brain-computer interfaces using support vector machines and blind source separation. *Comput. Intell. Neurosci.* **2007**, 82069–82069 (2007).
49. Sundaram, S. et al. Learning the signatures of the human grasp using a scalable tactile glove. *Nature* **569**, 698–702 (2019).
50. Zhou, Z. et al. Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. *Nat. Electron.* **3**, 571–578 (2020).
51. Luca, C. D. Surface electromyography: detection and recording. *DelSys Incorporated* **10**, 1–10 (2002).

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (grant nos. 51925503, U1713218) and the Program for HUST Academic Frontier Youth Team.

AUTHOR CONTRIBUTIONS

The design, preparation, and characterizations of tattoo-like electronics were completed by Y.W., Y.B. and L.Y.; the typical scenarios of SSRS were carried out by T.T. and Y.X.; the cloud-based machine-learning algorithm was developed by Y.X., T.T. and G.L.; Y.A.H., H.L., H.Z., Y.W., T.T. and Y.X. contributed to the writing of the manuscript; Y.A.H., H.L. and H.Z. supervised the overall research.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41528-021-00119-7>.

Correspondence and requests for materials should be addressed to H.Z., H.L. or Y.H.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.