Electromyogram-Based Lip-Reading via Unobtrusive Dry Electrodes and Machine Learning Methods

Penghao Dong, Yuanqing Song, Shangyouqiao Yu, Zimeng Zhang, Sandeep K. Mallipattu, Petar M. Djurić,* and Shanshan Yao*

Lip-reading provides an effective speech communication interface for people with voice disorders and for intuitive human–machine interactions. Existing systems are generally challenged by bulkiness, obtrusiveness, and poor robustness against environmental interferences. The lack of a truly natural and unobtrusive system for converting lip movements to speech precludes the continuous use and wide-scale deployment of such devices. Here, the design of a hardware–software architecture to capture, analyze, and interpret lip movements associated with either normal or silent speech is presented. The system can recognize different and similar visemes. It is robust in a noisy or dark environment. Self-adhesive, skin-conformable, and semi-transparent dry electrodes are developed to track high-fidelity speech-relevant electromyogram signals without impeding daily activities. The resulting skin-like sensors can form seamless contact with the curvilinear and dynamic surfaces of the skin, which is crucial for a high signal-to-noise ratio and minimal interference. Machine learning algorithms are employed to decode electromyogram signals and convert them to spoken words. Finally, the applications of the developed lip-reading system in augmented reality and medical service are demonstrated, which illustrate the great potential in immersive interaction and healthcare applications.

1. Introduction

Lip-reading or speech-reading techniques that can convert lip movements to speech have long been pursued to help people with voice disorders or hearing impairments.[1–3] These techniques facilitate communications and have the potential to replace human interpreters[4] in medical service applications. Lip-reading also provides an effective human–machine interface that uses natural speech for interactions with smart machines,[5,6] including robotics,[7] prosthetics,[8] computers,[9] and even augmented reality (AR)/virtual reality (VR) environments.[10,11] Silent speech interfaces, where no acoustic sound is produced, are often in demand when privacy or a quiet environment is desired, such as in public areas or hospitals. To realize lip-reading, camera-based visual solutions have been extensively explored to capture the visual features of lip movements.[12–16] In addition to external video tracking devices, visual lip-reading requires the users to be constantly within the line-of-sight of a camera, which may not be possible when the user is on the move and when the lighting conditions are unsatisfactory. Alternatively, ultrasound-based solutions exploit bulky ultrasonic imaging devices to build 2D lip and tongue images.[17,18] Another more portable ultrasound-based method is to emit inaudible signals using speakers and detect signals reflected by the moving lips using microphones.[19,20] This method is vulnerable to multipath interferences from users' body movements and objects in the surroundings. In vivo on-skin sensing technologies for tracking lip movements are emerging.[21] For example, permanent magnets attached to the tongue,[22,23] strain sensors placed on the face,[24,25] electromyography (EMG) electrodes placed on the face,[26–29] and an electroencephalogram (EEG) helmet worn around the head[30–32] have been utilized to capture lip movements and tongue positions. However, the magnets in the mouth are intrusive, and EEG signals suffer from noise when humans go astray mentally.[2]

Face-worn EMG sensing is an attractive method for lip-reading. During silent (subvocal) or normal (vocal) speech, lip movements result in subtle muscular activities, which can be reflected in the EMG signals collected near the mouth

P. Dong, S. Yu, Z. Zhang, S. Yao
Department of Mechanical Engineering
Stony Brook University
Stony Brook, NY 11794, USA
E-mail: shanshan.yao@stonybrook.edu
Y. Song, P. M. Djurić
Department of Electrical and Computer Engineering
Stony Brook University
Stony Brook, NY 11794, USA
E-mail: petar.djuric@stonybrook.edu
S. K. Mallipattu
Department of Medicine
Stony Brook University
Stony Brook, NY 11794, USA
S. K. Mallipattu
Renal Section
Northport VA Medical Center
Northport, NY 11768, USA

and throat.[33,34] EMG-based lip-reading has some advantages, including non-invasiveness, robustness in a noisy or dark environment, and high sensitivity for measuring subtle lip movements. However, there remains a gap for an intuitive and unobtrusive technology to track and interpret lip movements, both on the hardware (i.e., EMG sensing system) and software (i.e., data analysis) side.

In terms of EMG sensing, most existing systems employed commercial pre-gelled Ag/AgCl electrodes. These electrodes have limitations for long-term use, including bulkiness that adversely constrains the natural lip movements, skin irritation during long-term wear, and poor robustness owing to the dehydration of the gel.[35–37] More recently, skin-attachable gel-free dry electrodes have been developed to acquire EMG data for speech recognition.[38,39] The reported dry electrodes employed rigid metal films as conductive sensing materials, which have a large mismatch in the mechanical properties of the skin. In addition, these electrodes are not transparent on the face and rely on external tapes to adhere to the skin, which may preclude the continued use of such devices. In terms of signal processing, although researchers have used machine learning methods to decode EMG-speech relationships,[38,39] the main challenges lie in ambiguities at the word level. Various vocabularies with different phonemes[40] may belong to the same or similar viseme groups,[41] resulting in very similar muscular movements and EMG signals.[12,42,43] Namely, several phonemes produce lip movements that can be indistinguishable (also known as homophones), entailing that there is no one-to-one mapping between phonemes and visemes. Besides, the position of electrodes deserves further investigations to use fewer electrodes while still achieving higher accuracy.

To address the existing challenges for lip-reading, in this work, we present the design of a truly natural and robust EMG-based lip-reading system that can capture speech-relevant lip gestures and decode lip movements for speech. The system allows for a mechanically imperceptible, skin-friendly, visually unobtrusive platform for tracking and interpreting lip movements with minimal interference. The system is enabled by self-adhesive, semi-transparent, and conformable dry electrodes for high-quality EMG signal acquisition. The gel-free dry EMG electrode comprises silver nanowires (AgNWs) as a conductive interface to the skin surface, and D-sorbitol modified waterborne polyurethane (WPU) as a self-adhesive compliant substrate. Furthermore, ML algorithms (linear discriminant analysis [LDA] and support vector machines [SVM]) were implemented to decode the EMG signals. The combination of high-fidelity unobtrusive EMG sensing and ML-assisted speech recognition enables a highly robust non-acoustic speech communication interface. This speech interface is not affected by environmental conditions (lighting conditions and noise levels) and can be used no matter whether acoustic sounds are produced (in situations of normal and silent speech). Based on the EMG-based lip-reading interface, one application in the AR environment was demonstrated to convert lip movements into words in real time. The interpreted words were then used as commands to control the motion of AR characters by speech. Another application was developed to enable users to use non-acoustic speech interfaces to call for assistance, which facilitates communications with healthcare providers.

2. Result and Discussion

2.1. Overview of the Developed EMG-Based Lip-Reading System

Figure 1a presents the design of the EMG-based lip-reading system. As essential components of the hardware of the system, epidermal EMG sensing electrodes are developed to capture speech-relevant lip motions in a minimally invasive manner. The collected EMG data are analyzed by ML methods to decode lip movements for speech recognition. Eight-channel EMG electrodes are employed to obtain lip muscle activities from changes in biopotentials, as shown in Figure 1b. The monopolar configuration[44] for EMG sensing is adopted here. Eight electrodes are placed on the muscular regions relevant to speech articulation[27] as active sensing electrodes. A shared ground electrode is placed on the clavicle because the clavicle generates negligible muscular signals. Specifically, four electrodes are placed near the mouth to track muscle activities of the supralabial region (zygomaticus minor and zygomaticus major) and infralabial region (depressor anguli oris and mentalis).[45–48] Detection of the tongue position is indispensable to differentiate words that look exactly alike in lip movements when spoken.[27,46,49] Four extra EMG sensors are placed on the neck to capture tongue movements through muscles located in the ventromedial and submental regions.[45,46] Symmetrical equivalents of target locations across the craniocaudal axis are ignored to avoid feature repetition.

EMG sensing electrodes are key components of the lip-reading system and should be truly unobtrusive to encourage long-term use and wide-scale deployment. Ideal EMG electrodes should: 1) be hydrogel free (i.e., dry) to maintain signal quality over the long-term and reduce skin irritation; 2) can fully conform to the microscopically rough skin surface and therefore minimize the electrode-skin impedance, which is crucial for capturing subtle muscle activities with a high signal-to-noise ratio (SNR); 3) be stretchable to accommodate facial and neck skin deformations (up to ≈15% strain[33]); 4) be self-adhesive without using external tapes; 5) have good transparency to ensure visual unobtrusiveness.

As schematically illustrated in Figure 1c, the dry EMG electrode that satisfies these requirements are designed by optimizing material selections, electrode structures, and electrode thicknesses for conformable contact. Figure 1d shows the acquired EMG signal when the words "ship" and "sheep" were spoken. ML methods are used to classify the acquired eight-channel EMG signals and convert them to spoken words. The flowchart of EMG signal processing includes the data acquisition, de-noising of the signal, feature selection, classification, and the output of the recognized words. More details regarding the EMG electrodes and the ML-assisted signal processing are discussed below.

2.2. Design, Fabrication, and Characterization of EMG Sensing Electrodes

The EMG electrode includes the sensing and interconnection parts (Figure 1c). The sensing part is a bilayer structure including the conductive AgNW nanocomposites for acquiring
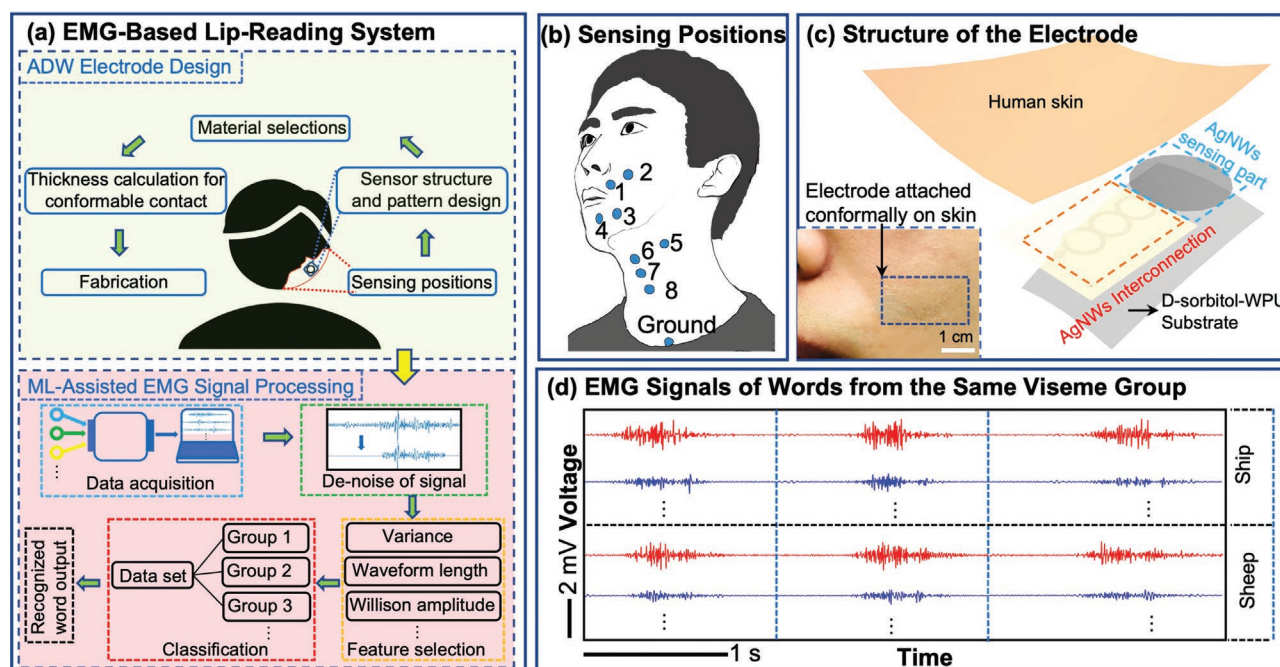
Figure 1. Overview of the developed EMG-based lip-reading system. a) Design of the hardware (top) and software (bottom) of the AgNWs-D-sorbitol-WPU (ADW) EMG-enabled lip-reading system comprising the optimization of EMG electrodes and the ML-assisted EMG signal processing. b) Schematics showing sensor positions on the skin near the mouth and on the neck. c) Exploded view of the EMG electrode and a picture of the electrode placed on the facial skin. d) Examples of EMG signals acquired by the developed EMG electrodes for one pair of words ("sheep" and "ship") from the same viseme group. Only two channels of EMG signals are shown as examples.

EMG signals. The self-adhesive D-sorbitol-WPU (D-W) is used as the substrate for embedding AgNWs and facilitating the attachment to the skin. AgNWs were employed as active sensing materials. Among various conductive materials, AgNWs are competitive candidates for stretchable transparent electrodes due to their exceptional conductivity, good mechanical compliance that can achieve reliable electromechanical performance, and long aspect ratio that can achieve good transmittance at a given conductivity.[50–52] As shown in **Figure 2**b, AgNWs make electrical contact with each other to form conductive pathways while maintaining decent transparency. The interconnection part has the same structure (AgNWs embedded in the D–W substrate). It is coated with an extra thin layer of self-adhesive silicone to isolate the interconnection from the skin electrically. The interconnection is patterned into the ring shape due to its good structural stretchability and thus a reduced resistance change during skin deformations.[53,54]

Figure 2a–e presents the fabrication process of the AgNW/D-W (ADW) electrodes. A shadow mask was first prepared and attached to a tape liner with low surface energy that makes it easy to transfer AgNWs to another substrate (Figure 2a). AgNW/ethanol solution was spray-coated over the mask to form the desired pattern (Figure 2b). After the mask was peeled off, a solution of D-sorbitol and WPU in N, N-Dimethylformamide (DMF) (DWD) was spin-coated on top of the AgNWs followed by evaporating the solvent (Figure 2c). AgNWs were then embedded just below the surface of the D–W substrate to improve stability during long-term wear. Afterward, the toner transfer paper with a water-soluble top coating (blue side) was attached to the formed ADW film to transfer it (Figure 2d). A

thin layer of self-adhesive silicone was then drop-casted onto the interconnection part (Figure 2e). The edges were cut off to get the final ADW electrodes. For skin attachment (Figure 2f–h), the ADW electrode was laminated to the skin. The toner transfer paper can be detached with the help of a wet cloth to dissolve the water-soluble layer. The specific dimensions of each part of the ADW electrode can be seen in Figure S2, Supporting Information.

In addition to high conductivity, stretchability, and transmittance, the conformable contact between the electrode and the free-form human skin surface[55,56] is also essential in electrode development. The conformable contact can help achieve low electrode-skin impedance and acquire high-fidelity EMG signals[50,57] during dynamic and subtle lip movements. As detailed in the Supporting Information, an energy-based analytical model was used to optimize the electrode for achieving conformable contact. The model was modified based on the literature[58–60] to extend the applicability for nonlinear elastic materials. For conformable contact, the work done by the adhesion force between the electrode and the skin should be equal to or larger than the sum of the bending energy of the electrode and the strain energy of the skin. Softer materials with higher adhesion force and smaller electrode thickness are beneficial for achieving conformable contact. The former two parameters can be tuned by changing the ratio between D-sorbitol and WPU. The critical electrode thickness to achieve the conformable contact can thus be determined at the specific ratio (see Supporting Information).

Figure 3a–c compares the stress–strain curves, adhesion forces tested on the pig skin, and transmittance for electrodes with

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com
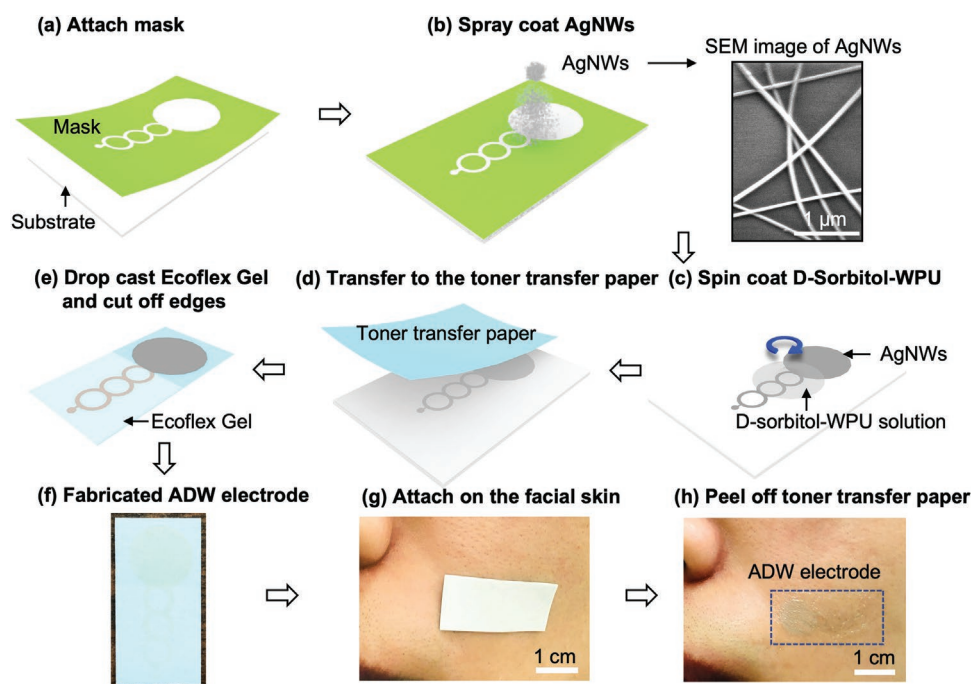
NANO · MICRO
small
www.small-journal.com

Figure 2. Schematic illustrations of the fabrication and attachment process of the developed ADW electrodes. a) Attach the mask to a substrate. b) Spray coating of AgNWs over the mask. An SEM image of spray-coated AgNW networks is shown. c) Spin coating the DWD solution on top of the AgNW networks to form thin AgNW/D-W composites. d) Transfer the prepared AgNW/D-W composites using the toner transfer paper. e) Drop-casting self-adhesive silicone onto the interconnection part and cut off excessive edges. f) Picture of a fabricated ADW electrode on the toner transfer paper. g,h) Attach the fabricated electrode to the skin followed by removing the toner transfer paper with water.

different ratios of D-sorbitol and WPU. As the ratio of D-sorbitol decreases, the transmittance increases, while the modulus and adhesion force of the electrode decrease. As shown in Table S2, Supporting Information, the critical thickness of electrodes made by AgNWs and pure WPU (without D-sorbitol) is around 7 μm, which means that electrodes with a thickness less than or equal to 7 μm can achieve a conformable contact. When adding 10 wt% or 20 wt% of D-sorbitol, the critical thickness jumps to 16–17 μm. 30 wt% of D-sorbitol results in a critical thickness of around 23 μm. To ensure high transmittance, low modulus, good self-adhesiveness, and relatively high critical thickness for easy fabrication, 10 wt% D-sorbitol was selected as the ratio for fabricating ADW electrodes. Figure 3d–e show SEM images of ADW electrodes with different thicknesses on the skin replica. The thick electrode (≈215 μm) shows apparent gaps with the skin replica (Figure 3d). The electrode with a thickness lower than the critical thickness (≈14 μm) conforms to the texture of the skin (Figure 3e).

As expected, good skin conformability leads to minimized electrode-skin impedances (Figure 3f), which is inducive to achieving high SNRs of EMG signals.[61] For pure AgNW-WPU (AW) electrodes, when the thickness dropped from over 200 μm to below 20 μm, the impedance drastically decreased because thinner electrodes possess better conformability. The impedance further decreased when adding 10 wt% D-sorbitol due to the increased skin adhesion, achieving a comparable impedance to commercial pre-gelled electrodes. Meanwhile, the ADW electrode showed similar performance with or without about 15% compressive and 15% tensile strains, as indicated in Figure 3g, illustrating good electromechanical stability to accommodate skin deformations. As shown in Figure 3h,i, when subject to

up to 15% strain, the resistance changes of ADW electrodes had a relatively large variation at the first three cycles of stretching, but the resistance change stabilized in the following stretching cycles. After the initial cycles of stretching and releasing, the ADW electrode exhibited a slight resistance variation (≈11%). Such variations would not affect the EMG sensing performance if the skin-electrode impedance is stable, which is much more important for biopotential sensing.[62]

The electrode can be detached from the skin using a tweezer or similar tools (Figure S3, Supporting Information). No water or other solvents were needed to help with the detachment process. There was no residue on the skin. The adhesion of the ADW electrode on the pig skin decreased slightly by around 13% after ten cycles of attachment/detachment, as shown in Figure S4, Supporting Information. The adhesion after ten cycles was still sufficient for reliable attachment on the skin. The SNRs of EMG signals during ten cycles of attachment/detachment are quite similar, varying between 16.06 and 17.26 dB (Figure S5, Supporting Information). The adhesion of the ADW electrode on the pig skin decreased by around 20% after 7 days of testing (Figure S6, Supporting Information). One attachment/detachment cycle was performed per day. The adhesion during seven days of testing was still sufficient for reliable attachment on the skin. The SNRs of EMG signals for each day are very similar, changing from 15.37 to 16.62 dB (Figure S7, Supporting Information). The results illustrate that the ADW electrode can maintain sufficient adhesion and EMG sensing properties during ten cycles of attachment/detachment and seven days of testing.

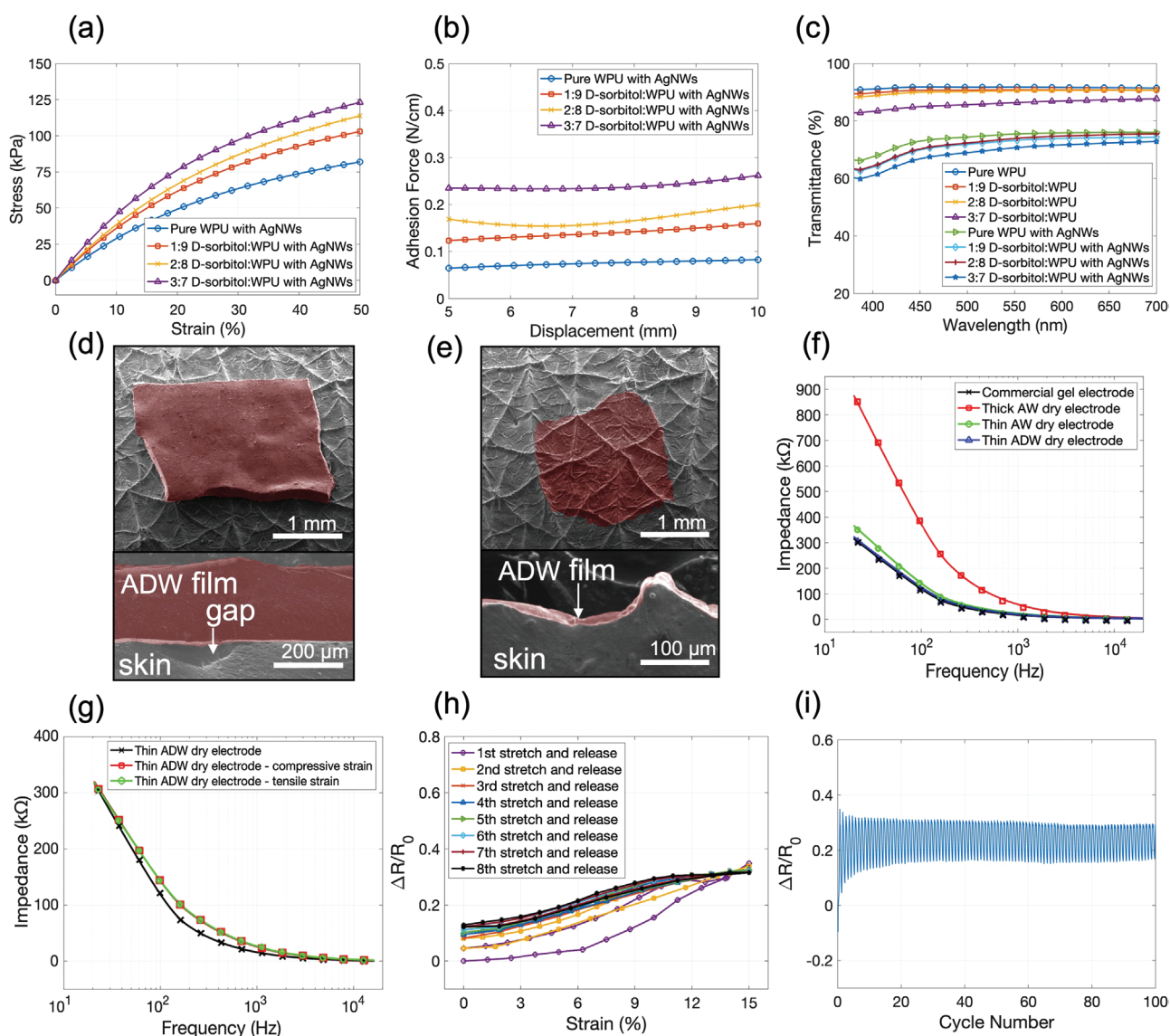The ADW electrode was found to be robust to changes in skin humidity. The EMG signals for the word "sheep"

Figure 3. Characterizations of ADW electrodes. a) Stress–strain curves, b) adhesion force to the pig skin, and c) transmittance of electrodes with different weight ratios between D-sorbitol and WPU. The top view (tilted 30°) and cross-sectional view of d) the thick ADW electrode (≈215 μm) and e) the thin ADW electrode (≈14 μm) on the skin replica. f) Comparisons of the electrode-skin impedance of commercial gel electrode, thick AgNWs-WPU (AW) electrode (around 200 μm in thickness), thin AW electrode, and thin ADW electrode. g) Comparisons of the electrode-skin impedance of the ADW electrode with and without strain. h) Relative resistance changes of the ADW electrode under tensile strain. i) Relative resistance changes of the ADW electrode under repeated stretching/releasing cycles. The ADW electrode was stretched to the strain of 15% in each cycle.

were acquired on the skin without any sweat and with sweat (Figure S8, Supporting Information). The subject was asked to jump for 5 min to generate sweat on the skin surface. The SNRs were 16.04 and 16.54 dB for the skin without any sweat and with sweat respectively, indicating no significant changes due to sweat. The adhesion was found to be robust in high skin humidity. Water or sweat are also known to be favorable to the skin-electrode signal transduction due to the improved conductivity channel.[35,63]

The electrode can be cleaned using DI water and ethanol. The AgNW sensing area is not suggested to be rubbed rigorously to avoid degradation in the conductivity. Complete cleaning of an electrode with excessive dust and skin debris

can be challenging due to the adhesion of the electrode. The electrode is reusable if it is protected from excessive dust in a good manner and the skin surface is properly cleaned before attachment. As our experiments have demonstrated, the electrode can be reusable for at least ten attachments/detachment cycles and over one week. After this time period, new electrodes may be used.

2.3. EMG Sensing Performance and Vocabulary Selection

To test the EMG sensing performance of the developed dry ADW electrodes, the EMG signal quality, evaluated by the SNR,
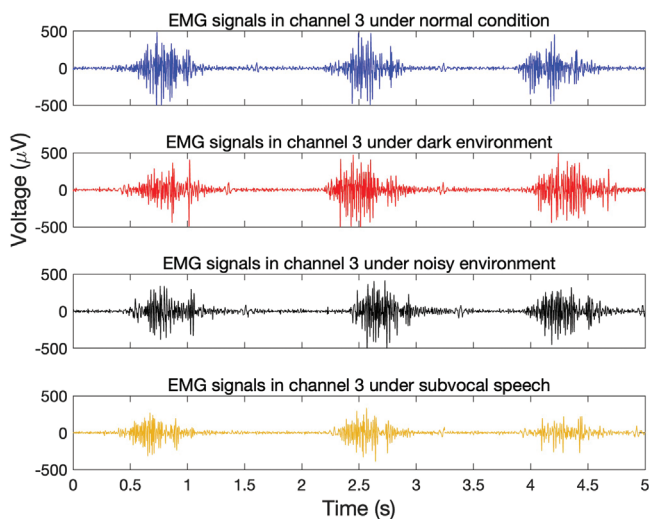
Figure 4. EMG signals (channel 3) for the word "ship" corresponding to normal conditions, dark environments, noisy environments, and silent speech.

was compared against the gold standard (commercial pre-gelled Ag/AgCl electrodes). Eight-channel EMG signals related to the lip movements for the word "ship" were captured using both types of electrodes, as given in Figures S9a and S10a, Supporting Information, respectively. The SNRs of the EMG signals obtained using ADW electrodes and pre-gelled electrodes in channel 3 were 14.83 and 11.85 dB, respectively. Conformable electrode-skin contact achieved by the optimized materials and electrode thickness guarantees the high SNR of the ADW electrodes. The final ADW electrodes are gel-free, conformable to the skin surface, stretchable, self-adhesive, and semi-transparent, allowing for mechanically and visually unobtrusive monitoring of EMG signals with high precision.

The robustness of EMG sensing was further evaluated in different situations, as summarized in **Figure 4**. For the silent speech, the subjects were asked to mouth their speech with facial and tongue movements but without producing any audible sound. The SNRs of EMG data collected by ADW electrodes at channel 3 under the normal condition, dark environment, noisy environment, and silent speech were found to be 14.83, 14.78, 14.92, and 13.02 dB, respectively. The results demonstrate the excellent adaptability of EMG sensing under various scenarios. The SNR decreased (≈1.8 dB) for silent speech because the muscular movement is slightly weaker. Table S3, Supporting Information shows the detailed SNRs of EMG data collected by ADW and pre-gelled electrodes from channels 1 to 8 under normal, noisy, and dark environments. Compared with visual and acoustic-based solutions, EMG-based lip-reading is not susceptible to environmental lighting conditions and noises.

To test the accuracy of the lip-reading system (discussed in the next section), a list of words covering all classes of visemes[64,65] (Figure S11, Supporting Information) was selected for EMG signal acquisition. Each word was repeated 50–60 times. There are 12 different visemes in the English language that represent distinct lip gestures unique to different phonemes.[66,67] A phoneme is a unit of sound that distinguishes

the pronunciation of the words,[68] while a viseme is a unit of visual speech[64] that includes phonemes with identical visual representations. Words with different phonemes may have very similar visemes. For instance, "sheep" and "ship" have different second phonemes, which are "/iː/" and "/ɪ/" respectively. But they share similar visual lip gestures and belong to the same viseme as shown in Figure S11, Supporting Information (viseme class "Spread"). This many-to-one mapping between phonemes and visemes poses significant challenges to lip-reading, especially for commonly used visual solutions.

2.4. Lip-Reading via Machine Learning Methods

In this work, the ML algorithms were trained by labeled EMG signals. There were eight electrodes attached to different positions of the volunteers' face and neck. Each electrode acquired an EMG signal when the volunteer spoke. The eight EMG signals were segmented (Figures S12 and S13, Supporting Information), where each segment contained a single word. Since the speaker would pause for a random time between words, we would extract the word portions of the recordings.

Six commonly used EMG features (Table S4, Supporting Information) from the time and frequency domains were employed. For each channel, the root means square, variance, mean, median frequency, zero crossing, and Wilson amplitude were computed. Therefore, for each word, there were 48 features. With the computed features, several ML methods for classification were trained, including the LDA and SVM classifiers. We obtained their performance via fivefold cross-validation, where the training and validation sets were selected by random sampling. Once the classifiers were trained with the four sets, we validated them with the remaining set. This process was repeated five times. The results of each classifier were summarized by finding its average validation accuracy.

Words with different pronunciations were first analyzed. Eleven words from 11 distinct viseme groups were selected (neglecting the "closed" viseme group) from Figure S11, Supporting Information: Bird, bite, boat, fan, choke, kay, pay, sheep, tea, thin, and way. **Figure 5**a and Figure S14a, Supporting Information show the confusion matrix (accuracy) of LDA and SVM classifiers for subject 1, respectively, where the vertical/horizontal axes represent true/predicted classes. The recognition accuracy of the LDA classifier was 97.6%, which has a better performance than that of SVM (accuracy of 94.4%). The receiver operating characteristic (ROC) curve related to the word "pay" (Figure 5c for LDA and Figure S14b, Supporting Information for SVM) also shows excellent performance when using both classifiers.

The vocabularies with similar pronunciations were then analyzed. Nine pairs of words were selected from the same viseme group (Figure S11, Supporting Information): Sea versus tea, bay versus pay, then versus thin, fan versus van, bite versus but, choke versus joke, boat versus book, sheep versus ship, and gay versus kay. The accuracies of LDA and SVM for subject 1 were 90.4% (Figure 5b) and 85.4% (Figure S15a, Supporting Information), respectively. The high accuracy and the ROC curve related to the word "pay" (Figure 5d and Figure S15b, Supporting Information) show that the proposed EMG-based

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com
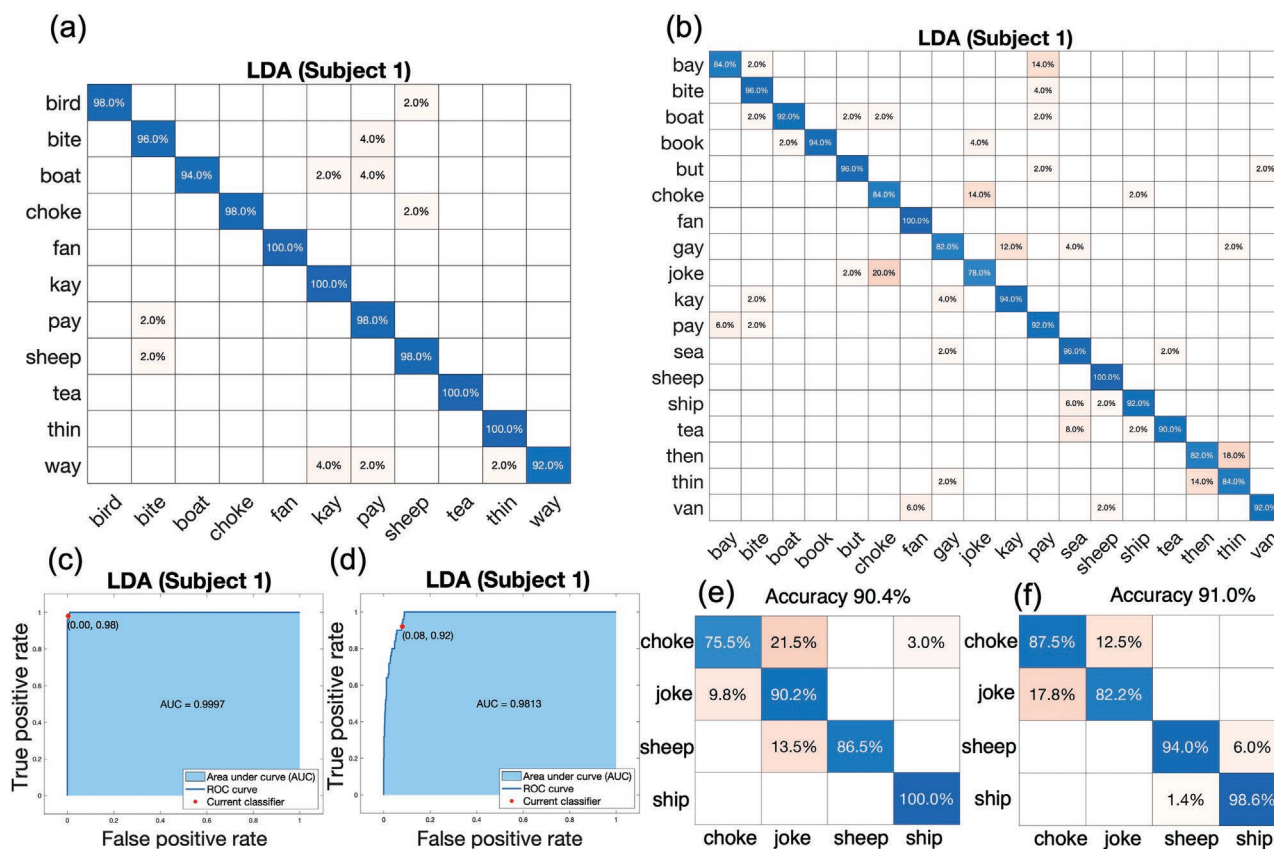
NANO · MICRO
small
www.small-journal.com

Figure 5. Recognition results of subject 1 enabled by the ML-assisted EMG signal processing. a) Confusion matrix for 11 words from different viseme groups when using the LDA classifier. b) Confusion matrix for nine pairs of words with similar pronunciations (from the same viseme group) when using the LDA classifier. c) The corresponding ROC curve related to the word "pay" in (a). d) The corresponding ROC curve related to the word "pay" in (b). e) Confusion matrix for subvocal words using the SVM classifier. f) Confusion matrix for vocal words using the SVM classifier. For all confusion matrices, the vertical/horizontal axes represent true/predicted classes.

lip-reading method has the capability to identify vocabularies with very similar pronunciations. To compare the performance between normal and silent EMG signals, the lip-reading results of silent and vocal EMG signals for two pairs of words from the same viseme groups (ship vs sheep, choke vs joke) were compared. The difference is minimal (0.6%), indicating the applicability of the proposed lip-reading to silent speech (Figure 5e,f). Results for subjects 2–4 can be seen in Figures S16–S21, Supporting Information. Overall, the LDA showed better performances than SVM among all the four subjects.

To suggest the optimal electrode position and the minimal number of electrodes, we analyzed EMG signals from different channel combinations. **Figure 6**a presents the results for the words with different pronunciations (from different viseme groups) related to subject 1, while Figure 6b for words with similar pronunciations (from the same viseme group) related to subject 1. Tables S5 and S6, Supporting Information show the recognition accuracy of each channel alone and the overall accuracy using the LDA classifier for all subjects. Overall, the EMG signals collected from position 4 have the best accuracy and signals collected from position 8 have the worst accuracy, which is applicable among all the four subjects. Besides, the electrodes on positions 2, 5, and 8 can be removed if we want to reduce the number of electrodes at the cost of sacrificing

some accuracy. The used ML classifiers are not computationally intensive. The ML classifiers complete the recognition in less than 0.1 s.

Table S7, Supporting Information shows the comparison of the proposed system to other systems using various sensing methods. All listed systems in Table S7, Supporting Information were portable or potentially portable. It should be noted that the mentioned systems in the literature did not show results for recognizing words with similar pronunciations (i.e., viseme groups). Thus, the recognition result of words from different viseme groups was used for comparison. The comparison reveals that our system can achieve a high speech recognition accuracy with portable and skin-conformable sensing electrodes that are mechanically and visually unobtrusive.

Both hardware and software contribute to improved accuracy in speech recognition. In terms of hardware, signals acquired from the developed nanowire electrodes have high SNRs due to the conformable contact with the skin surface and low skin-electrode impedance. The SNRs are comparable to and even slightly higher than SNRs of signals acquired from commercial pre-gelled electrodes (Table S3, Supporting Information). Higher SNRs are crucial for capturing subtle muscle activities with high precision and contribute to the accuracy of speech recognition.
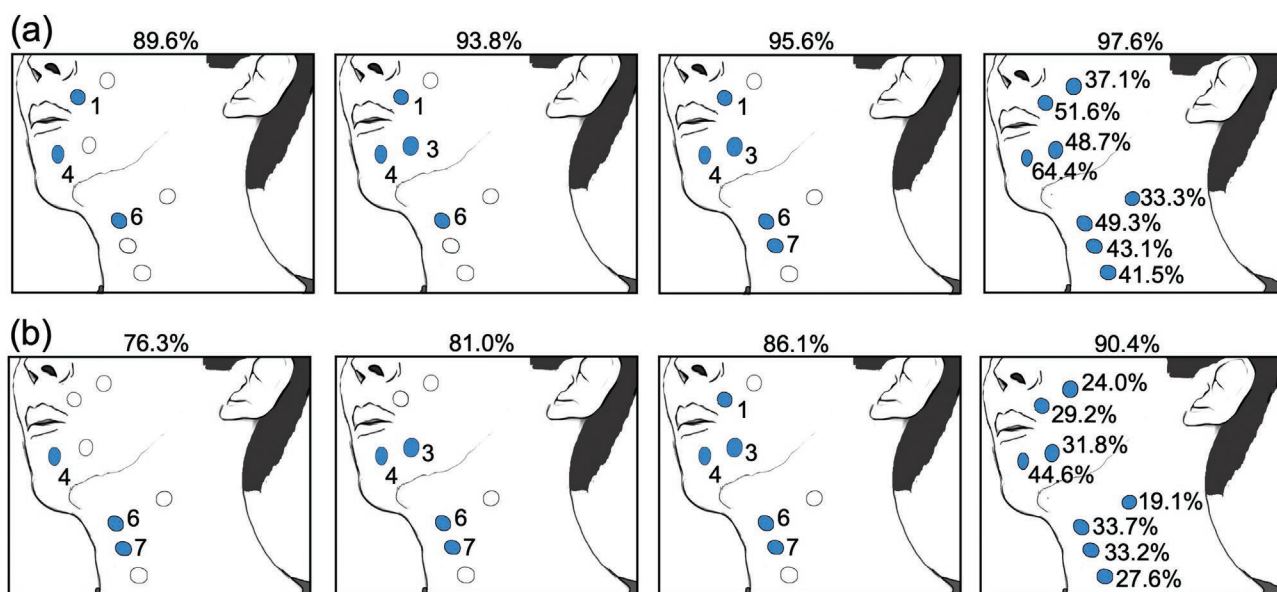
**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

NANO · MICRO

small

www.small-journal.com

Figure 6. Recognition results of subject 1 for different combinations of EMG channels when signals were processed by LDA. a) Accuracies of 11 words from different viseme groups. b) Accuracies of nine pairs of words from the same viseme group. The last configuration used all eight-channel. The accuracy of each channel alone is also indicated on the schematic.

In terms of software, ML-assisted conversion of EMG signals to speech improves accuracy. The selected six parameters provide distinguishable features for the ML algorithm to identify each word. Further, we tested several popular state-of-art algorithms and chose the best of them. We decided to work with six features instead of using the high sampling rate raw data. This choice speeds up the algorithms without sacrificing accuracy. In future work, we plan to search for other informative features and work with more sophisticated ML algorithms.

2.5. Applications in Augmented Reality Environment and Healthcare

Two demonstrations were developed to illustrate the potential applications of the developed EMG-based lip-reading system as speech communication interfaces (**Figures 7** and **8**). More details are provided in the experimental section. AR technologies have shown enormous commercial applications in training, entertainment, and marketing, to name a few. A proof-of-concept application was developed to demonstrate the EMG-based lip-reading system for interacting with the AR environment (Figure 8a and Video S1, Supporting Information). An AR software was employed on the iOS platform (Figure 7), where the EMG-based lip-reading system can directly control the movement of an AR character in real time with the help of a data transfer technique, the MQ telemetry transport (MQTT) method.[69] The ML lip-reading model was pre-trained ahead of time. For real-time applications, when mouth and neck muscles corresponding to different words (e.g., "run," "walk," "idle," and "dance") were formed and collected in the EMG data, the trained ML model interpreted the word. The system then sent the interpreted words as control commands to the AR character by MQTT. Simultaneously, the AR character could make cor-

responding movements according to the commands generated by speech-relevant lip movements. Unlike voice control (having noise issues) and video-based control (having lighting issues), this control interface leads to a more robust and interactive experience enabled by natural speech.

The EMG-based lip-reading speech interface can also be used in healthcare. Another software was developed on the iOS platform (Figure 7) to display the text information and play corresponding audio based on the interpreted words from collected EMG signals (Figure 8b and Video S2, Supporting Information). The EMG signals may be acquired from patients lacking voice abilities, or when the room is too dark/noisy, or when privacy is a concern. The lip-reading system can decode the EMG signals to speech and output the corresponding text and audio via the software in real time. Then, healthcare providers can see the displayed text and/or hear the sound. In the demonstration, "help," "water," "yes," and "no" were integrated into the system as examples.

3. Conclusion

In summary, a series of materials, electrode design, fabrication process, analytical calculation, and speech recognition algorithm was presented that realize unobtrusive monitoring and interpreting of speech-relevant EMG signals. Considering all the materials used, the cost of one ADW electrode is ≈$0.5. The cost can be reduced for mass production. The cost of a customized portable device for EMG data acquisition and analysis is estimated to be less than $50.

The presented lip-reading system has the following salient features that address challenges faced by other lip-reading systems, including bulkiness, obtrusiveness, and poor robustness: 1) The proposed sensors can be directly attached to the skin to
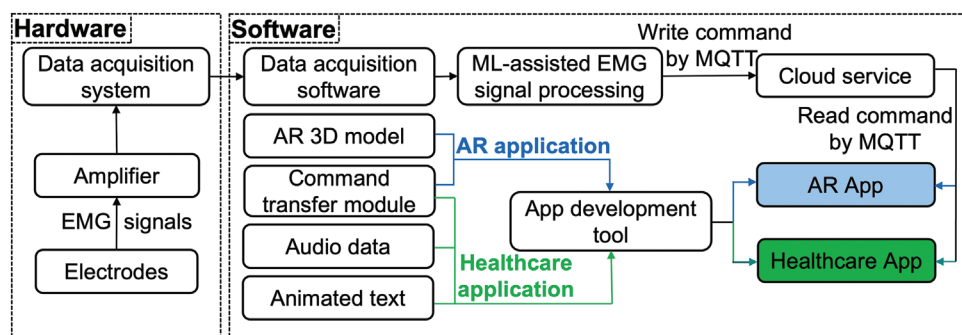
ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

NANO · MICRO
small
www.small-journal.com

Figure 7. Schematics of the application development process for the AR and healthcare applications shown in Figure 8. For hardware, the amplifier refers to the bioamplifier and the data acquisition system refers to the PowerLab instrument. For software, the data acquisition software refers to Lab-Chart. The ML-assisted EMG signal processing was implemented in MATLAB. The command transfer module refers to an open-source MQTT project M2MqttUnity. The cloud service was provided by Amazon Web Services. Unity 3D and Xcode were used as app development tools.

capture lip movements in vivo, eliminating complicated setups and line-of-sight constraints required for visual and acoustic-based methods. 2) Compared with visual and acoustic-based speech recognition, the EMG-based lip-reading system is more robust under noisy and dark environments and can recognize words from different or the same viseme groups via ML algorithms. 3) The sensors are compliant, lightweight, and can form seamless contact with the curvilinear skin surface under daily motions. The intimate contact and minimal interference are key to the accurate tracking of subtle lip movements for both normal and silent speech. 4) The sensors are gel-free, self-adhesive, and semi-transparent, allowing for unobtrusive monitoring while mitigating aesthetic concerns. To illustrate potential applications as a non-acoustic speech communication
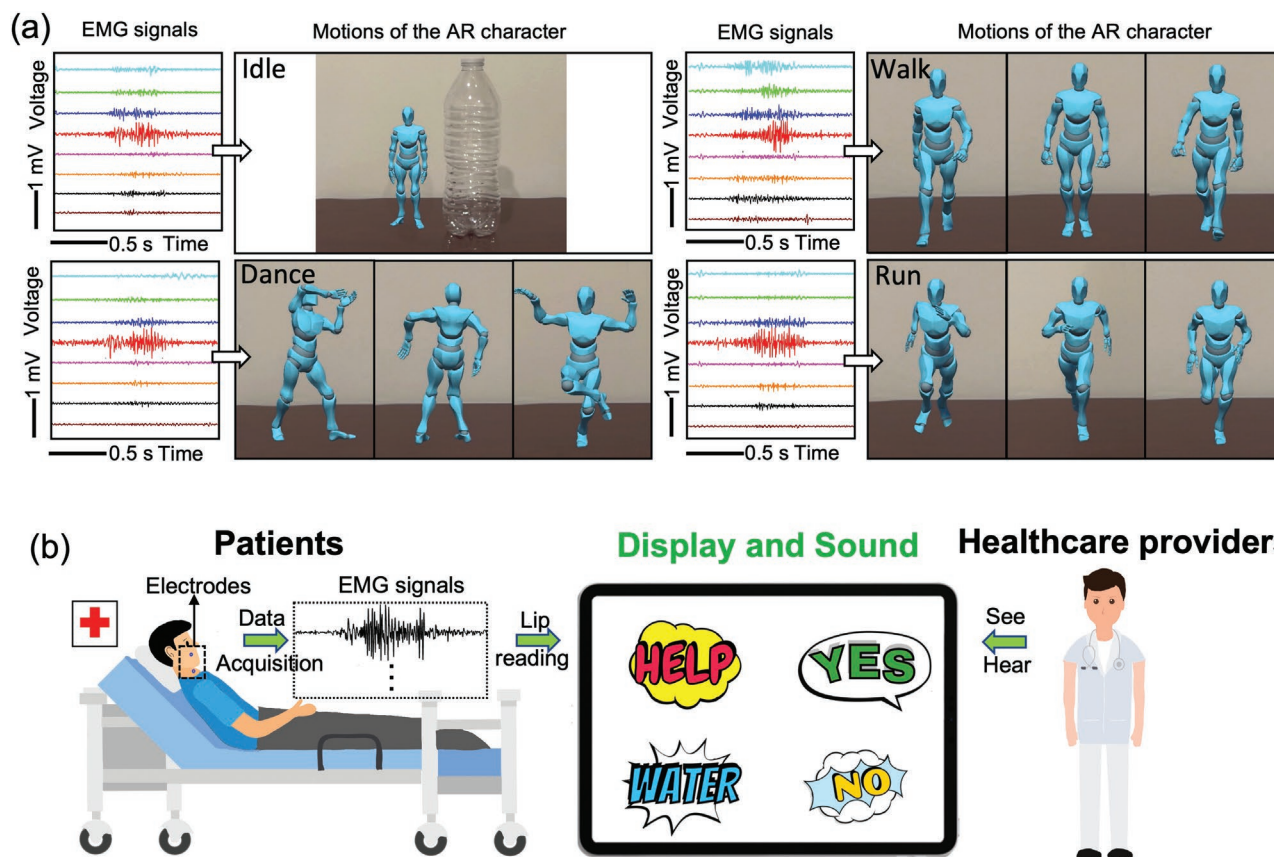

Figure 8. EMG-based lip-reading systems for AR and healthcare applications. Real-time EMG signals were captured using the EMG electrodes and speech was interpreted in real time. a) Applications in human-computer interfaces for AR environments. The interpreted speech (i.e., "idle," "dance," "walk," and "run") was used to control different motions of the AR character. b) Applications in healthcare applications. After lip-reading, animated texts corresponding to the interpreted speech (i.e., "help," "yes," "water," and "no") were displayed on the screen and corresponding sounds were played to call for assistance and to facilitate communications.

interface, two proof-of-concept prototypes for AR control and healthcare were developed. The EMG-based lip-reading enables speech (either silent or normal speech) recognition through muscular movements. It is expected to benefit individuals with speech and hearing disorders and lay the ground for the development of novel communication interfaces for prosthetics, AR/VR, gaming, and robotic control.

The presented system can be improved in several aspects for future work: 1) In terms of hardware, the proposed system should be converted to a portable version. The acquired signal can be transferred to a mobile device by Bluetooth for later signal processing. Multiple ADW electrodes should be integrated into one patch which is more convenient for users to attach and reduce position errors. 2) In terms of software, the proposed system should be improved to realize sentence-level speech recognition. The ML algorithm should identify each phoneme in a word and have the ability to construct new words using phonemes. Then the system should combine each recognized word to output the sentence. We are currently working on an advanced ML model in this regard.

4. Experimental Section

Materials: WPU (UD-410, 35 wt%) was provided by Bond Polymers International. D-sorbitol (98%), ethylene glycol (EG), polyvinylpyrrolidone (PVP, average Mw 360000), PVP (average Mw 55000), iron (III) chloride (FeCl$_3$), silver nitrate (AgNO$_3$), and *N,N*-dimethylformamide (DMF) were purchased from Sigma-Aldrich. The self-adhesive silicone was obtained from Smooth-On, Inc. All materials and reagents were used as received.

Synthesis of AgNWs: AgNWs were fabricated using a modified polyol reduction method. The PVP solution was prepared by dissolving 0.72 g PVP (Mw 55000) and 0.72 g PVP (Mw 350000) into 50 mL EG followed by magnetic stirring at 500 rpm at 100°C until fully dissolved. The AgNO$_3$ solution was prepared by dissolving 0.32 g AgNO$_3$ into 50 mL EG and stirring for 5 min at 500 rpm. Then, the PVP solution was heated to 150 °C in an oil bath. FeCl$_3$ solution (10 mL 600 μM in EG) was subsequently added to the PVP solution and stirred for 1 min at 500 rpm. The AgNO$_3$ solution was then added to the solution in the oil bath gradually and stirred for 1 min. The solution in the oil bath was allowed to sit in the oil bath at 150 °C. After 2.5 h, the solution was submerged in a room-temperature water bath to stop the reaction. Finally, AgNWs were purified with acetone and ethanol, respectively. The solution was centrifuged three times at 2000 rpm for 10 min to remove solvents, surfactants, and other impurities in the supernatant.

Preparation of EMG Electrodes: 5 g WPU solution was dried to get 1.75 g solid PU films. 0.194 g D-sorbitol was added to 0.5 g DI water and stirred for 5 min at 500 rpm. Afterward, 6 g DMF and 1.75 g solid PU were added to the D-sorbitol solution and stirred for 12 h at 800 rpm until completely dissolved to get the DWD solution. The pattern mask was cut by a mechanical plotter (Silhouette, Cameo 4) and laminated to a slippery tape liner attached to a glass substrate. AgNW/ethanol solution was spray-coated over the masks using the airbrush (Master Airbrush, Model G222). After peeling off the pattern mask, 0.8 g DWD solution was coated on the AgNWs using a spin coater (Laurell, WS-650Mz-23NPPB) at 3000 rpm for 1 min to get the ADW thin film. The toner transfer paper (Pulsar PRoFX) was then attached to the ADW thin film to transfer the film to the toner transfer paper. Copper lead wires were attached to the end of the ring-shaped interconnection part by the silver epoxy. Finally, the self-adhesive silicone and silicone thinner (Smooth-On, United States) were mixed with a ratio of 10:1 and then drop-casted on the interconnection part of the electrode.

Characterizations: To evaluate the electrode-skin impedance, two electrodes were attached to the forearm skin at a distance of 8 cm. The impedance between two electrodes was measured using the impedance analyzer (Keysight, E4990A). The morphologies of prepared ADW electrodes were obtained by the scanning electron microscope (Hitachi, S-4800). The transmittance of ADW electrodes was conducted using a UV–vis Spectroscopy (Thermal Scientific, Genesys 30) with wavelengths from 350 nm to 700 nm.

The stress–strain curves of ADW electrodes were measured using the materials testing systems (MTS, 858 Mini Bionix II) at a speed of 10 mm min⁻¹. The load cell of the tensile stage has a resolution of 0.001 N. The adhesion test of ADW electrodes was performed by the same materials testing system. Briefly, the rectangular ADW film of 2 × 1 cm was attached to the pig skin horizontally and subjected to a perpendicularly delaminated speed of 10 mm min⁻¹. The force exerted on the electrodes during the delamination process was measured by the load cell of the tensile stage. The adhesion force was calculated by dividing the maximum stable force by the film width.[70]

EMG Signal Acquisition and Processing: Eight-channel EMG signals were acquired by attaching 8 ADW electrodes as active sensing electrodes and one ADW electrode as the ground electrode. The position of 8 ADW electrodes attached to the face and neck is shown in Figure 1b. Meanwhile, the ground electrode was attached to the collarbone. All electrodes were linked to the Bioamplifier (ADInstruments, Octal Bio Amp ML138) and a data acquisition system (AD Instruments, PowerLab 8/30 ML 870). The software LabChart was used to collect all EMG signals at a sampling rate of 1 k s⁻¹. The signals were acquired with a bandpass filter between 10 to 500 Hz. Then, a bandpass digital filter with cutoff frequencies of 20 to 200 Hz was applied to further de-noise the signals. The SNR of the EMG signals was calculated using MATLAB and the following equation:[62]

$$SNR_{dB} = 10\log_{10}\left(\frac{A_{signal}}{A_{noise}}\right)^2 = 20\log_{10}\left(\frac{A_{signal}}{A_{noise}}\right) \tag{1}$$

where A_{signal} is the root mean square of the EMG signals, A_{noise} is the root mean square of the noise. Signals acquired when the face and neck remained stationary were considered noises. In the experiments, EMG signals in the noisy environment were acquired in an environment with a noise of around 85 dB produced by a speaker. When the ambient noise reaches 80 dB, the recognition rate of audio-based speech recognition will decrease drastically.[38] EMG signals in the dark environment were acquired when all lights were turned off. The normal environment refers to an environment with sufficient light and a noise level under 45 dB.

Applications in AR and Healthcare: Schematic illustrations of the developed applications in AR and healthcare are shown in Figure 7. EMG signals acquired from the ADW electrodes were collected using the data acquisition system (PowerLab) and data acquisition software (LabChart). A plugin[71] for LabChart was then used to transfer the EMG signal to MATLAB in real time for ML-assisted EMG signal processing. It should be noted that the lip-reading ML model was pre-trained in MATLAB. After the lip-reading ML model recognized the input EMG signals, interpreted words were sent to the cloud of amazon web services (AWS) using the MQTT method. Then, the developed software (app on the iOS platform as discussed below) received the command by MQTT and performed the corresponding actions. MQTT was selected here because it is a wireless machine-to-machine network protocol. The developed software can receive the commands in real time with low power consumption. In terms of software development, Unity 3D and Xcode were used as development tools. Unity 3D was used to code and integrate each module for the software. Xcode was used for transforming a Unity 3D project into an app on the iOS platform. The AR app integrated AR 3D animation models (downloaded from the Adobe Mixamo website with permission) and the MQTT command transfer module (open-source project M2MqttUnity[72]). The healthcare app integrated the MQTT command transfer module, audio data (word pronunciations available from the Google translate website), and animated texts (downloaded from the Flexclip website with permission).

Statistical Analysis: EMG signals were acquired from four subjects (including one female and three males, aged between 20 and 30 years

old). The authors had complied with all relevant ethical regulations. The Stony Brook University provided guidelines for study procedures. Informed consent was obtained from all participants. After data acquisition, the data were segmented into every single trial, and features were calculated. LDA/SVM training programs were applied to get the classification results. Figures 5 and 6 and Figures S14–S21, Supporting Information summarize the recognition results of all four subjects. The results of 11 words from different viseme groups using the LDA and the SVM were 94.8% ± 0.035 and 90.5% ± 0.061, respectively. The results for nine pairs of similar words were 86.8% ± 0.053 and 82.6% ± 0.056 when using the LDA and the SVM, respectively. The recognition accuracy was represented as mean ± standard derivation. Matlab was used as the software for the statistical analysis.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

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

[1] Y. Lu, H. Tian, J. Cheng, F. Zhu, B. Liu, S. Wei, L. Ji, Z. L. Wang, *Nat. Commun.* **2022**, *13*, 1401.

[2] W. Lee, J. J. Seong, B. Ozlu, B. S. Shim, A. Marakhimov, S. Lee, *Sensors* **2021**, *21*, 1399.

[3] G. S. Meltzner, J. T. Heaton, Y. Deng, G. De Luca, S. H. Roy, J. C. Kline, *J. Neural Eng.* **2018**, *15*, 046031.

[4] E. C. Meltzer, J. J. Gallagher, A. Suppes, J. J. Fins, *Crit. Care Med.* **2012**, *40*, 1529.

[5] Y. Lu, J. Yan, K. Gu, *Int. J. Pattern Recognit. Artif. Intell.* **2018**, *32*, 1856007.

[6] F. Karray, M. Alemzadeh, J. A. Saleh, M. N. Arab, *Int. J. Smart Sens. Intell. Syst.* **2008**, *1*, 137.

[7] S. A. Moubayed, G. Skantze, J. Beskow, *Int. J. Humanoid Rob.* **2013**, *10*, 1350005.

[8] V. Vincenti, E. Pasanisi, M. Guida, G. Di Trapani, M. Sanna, *Audiol. Neurotol.* **2008**, *13*, 273.

[9] G.-L. Zheng, M. Zhu, L. Feng, in *Proc. 2014 Seventh Int. Symp. on Computational Intelligence and Design*, IEEE, Hangzhou, China **2014**, p. 293, https://ieeexplore.ieee.org/abstract/document/7064194.

[10] Y. Zhang, Y.-C. Chen, H. Wang, X. Jin, in *Adjunct Proc. 2021 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing and Proc. 2021 ACM International Symp. on Wearable Computers*, ACM, Virtual USA **2021**, p. 580, https://dl.acm.org/doi/abs/10.1145/3460418.3480163.

[11] N. Kimura, M. Kono, J. Rekimoto, in *Proc. 2019 CHI Conf. on Human Factors in Computing Systems*, Glasgow, Scotland, UK **2019**, p. 146, https://www.mdpi.com/1424-8220/21/4/1399.

[12] Z. H. Zhou, G. Y. Zhao, X. P. Hong, M. Pietikainen, *Image Vision Comput.* **2014**, *32*, 590.

[13] Y. M. Assael, B. Shillingford, S. Whiteson, N. De Freitas, (Preprint) arXiv:1611.01599, v2, submitted: Dec **2016**.

[14] A. Fernandez-Lopez, O. Martinez, F. M. Sukno, in *Proc. 2017 12th Int. Conf. on Automatic Face and Gesture Recognition*, IEEE, Washington, DC **2017**, p. 208, https://ieeexplore.ieee.org/abstract/document/7961743.

[15] T. Afouras, J. S. Chung, A. Senior, O. Vinyals, A. Zisserman, *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *44*, 8717.

[16] S. Petridis, T. Stafylakis, P. C. Ma, G. Tzimiropoulos, M. Pantic, in *Proc. 2018 IEEE Workshop on Spoken Language Technology*, IEEE, Athens, Greece **2018**, p. 513, https://ieeexplore.ieee.org/abstract/document/8639643.

[17] B. Denby, M. Stone, in *Proc. the 2004 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, IEEE, Montreal, Canada **2004**, p. 685, https://ieeexplore.ieee.org/abstract/document/1326078.

[18] A. Jaumard-Hakoun, K. Xu, C. Leboullenger, P. Roussel-Ragot, B. Denby, in *ISCA Interspeech 2016*, ISCA, San Francisco, CA **2016**, p. 1467, https://hal.science/hal-01529630/.

[19] J. Luo, J. Z. Wang, N. Cheng, G. L. Jiang, J. Xiao, in *Proc. 2021 IEEE Spoken Language Technology Workshop*, IEEE, Shenzhen, China **2021**, p. 606, https://ieeexplore.ieee.org/abstract/document/9383622.

[20] J. Tan, C.-T. Nguyen, X. Wang, in *Proc. 2017 IEEE Conf. on Computer Communications*, IEEE, Atlanta, GA **2017**, p. 1, https://ieeexplore.ieee.org/abstract/document/8057099.

[21] W. Wu, H. Haick, *Adv. Mater.* **2018**, *30*, 1705024.

[22] A. Bedri, H. Sahni, P. Thukral, T. Starner, D. Byrd, P. Presti, G. Reyes, M. Ghovanloo, Z. Guo, *Computer* **2015**, *48*, 54.

[23] R. Hofe, S. R. Ell, M. J. Fagan, J. M. Gilbert, P. D. Green, R. K. Moore, S. I. Rybchenko, *Speech Commun.* **2013**, *55*, 22.

[24] Y. Kunimi, M. Ogata, H. Hiraki, M. Itagaki, S. Kanazawa, M. Mochimaru, in *Proc. Augmented Humans 2022*, ACM, Kashiwa, Japan **2022**, p. 26, https://dl.acm.org/doi/abs/10.1145/3519391.3519399.

[25] Y. Wang, S. Lee, T. Yokota, H. Wang, Z. Jiang, J. Wang, M. Koizumi, T. Someya, *Sci. Adv.* **2020**, *6*, abb7043.

[26] T. Schultz, M. Wand, *Speech Commun.* **2010**, *52*, 341.

[27] A. Kapur, S. Kapur, P. Maes, in *Proc. 23rd Int. Conf. on Intelligent User Interfaces*, ACM, Tokyo, Japan **2018**, p. 26, https://dl.acm.org/doi/abs/10.1145/3172944.3172977.

[28] C. Jorgensen, K. Binsted, in *Proc. 38th Annual Hawaii Int. Conf. on System Sciences*, IEEE, Big Island, HI **2005**, p. 294c, https://ieeexplore.ieee.org/abstract/document/1385845.

[29] J. M. Vojtech, M. D. Chan, B. Shiwani, S. H. Roy, J. T. Heaton, G. S. Meltzner, P. Contessa, G. De Luca, R. Patel, J. C. Kline, *J. Speech Lang. Hear. Res.* **2021**, *64*, 2134.

[30] Living with ALS: Changes in Speech and Communication Solutions, https://arc.phhp.ufl.edu/wordpress/files/2017/06/Living-with-ALS-_-Speech.pdf (accessed: June 2022).

[31] K. Brigham, B. V. Kumar, in *Proc. 2010 4th Int. Conf. on Bioinformatics and Biomedical Engineering*, IEEE, Chengdu, China **2010**, p. 1, https://ieeexplore.ieee.org/abstract/document/5515807.

[32] P. Suppes, Z. L. Lu, B. Han, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 14965.

[33] T. Sun, F. Tasnim, R. T. McIntosh, N. Amiri, D. Solav, M. T. Anbarani, D. Sadat, L. Zhang, Y. Gu, M. A. Karami, C. Dagdeviren, *Nat. Biomed. Eng.* **2020**, *4*, 954.

[34] D. Gaddy, D. Klein, (Preprint) arXiv:2010.02960, v1, submitted: Oct **2020**, https://arxiv.org/abs/2010.02960.

[35] S. S. Yao, Y. Zhu, *JOM* **2016**, *68*, 1145.

[36] S. Yao, J. Yang, F. R. Poblete, X. Hu, Y. Zhu, *ACS Appl. Mater. Interfaces* **2019**, *11*, 31028.

[37] Q. Qin, J. Q. Li, S. S. Yao, C. Y. Liu, H. Huang, Y. Zhu, *IEEE Access* **2019**, *7*, 20789.

[38] Y. H. Wang, T. Y. Tang, Y. Xu, Y. Z. Bai, L. Yin, G. Li, H. M. Zhang, H. C. Liu, Y. A. Huang, *npj Flexible Electron.* **2021**, *5*, 20.

[39] H. Liu, W. Dong, Y. Li, F. Li, J. Geng, M. Zhu, T. Chen, H. Zhang, L. Sun, C. Lee, *Microsyst. Nanoeng.* **2020**, *6*, 16.

[40] Phoneme, https://en.wikipedia.org/wiki/Phoneme (accessed: May 2022).

[41] Viseme, https://en.wikipedia.org/wiki/Viseme (accessed: May 2022).

[42] S. Dupont, J. Luettin, *IEEE Trans. Multimedia* **2000**, *2*, 141.

[43] A. V. Nefian, L. H. Liang, X. B. Pi, X. X. Liu, K. Murphy, *Eurasip J. Appl. Signal Process.* **2002**, *2002*, 1274, https://link.springer.com/article/10.1155/S1110865702206083.

[44] T. Schultz, M. Wand, T. Hueber, D. J. Krusienski, C. Herff, J. S. Brumberg, *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 2257.

[45] M. Jiang, A.-M. Rahmani, T. Westerlund, P. Liljeberg, H. Tenhunen, in *Proc. 2015 IEEE Int. Conf. on Computer and Information Technology*, IEEE, Liverpool, UK **2015**, p. 981, https://ieeexplore.ieee.org/abstract/document/7363189.

[46] G. S. Meltzner, J. T. Heaton, Y. Deng, G. De Luca, S. H. Roy, J. C. Kline, *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 2386.

[47] Y. B. Deng, J. T. Heaton, G. S. Meltzner, *Interspeech* **2014**, *1*, 1164, https://www.isca-speech.org/archive_v0/archive_papers/interspeech_2014/i14_1164.pdf.

[48] G. S. Meltzner, J. Sroka, J. T. Heaton, L. D. Gilmore, G. Colby, S. Roy, N. Chen, C. J. De Luca, in *Proc. Interspeech 2008*, ISCA, Brisbane, Australia **2008**, p. 2667, https://www.bu.edu/nmrc/files/2011/01/c_010.pdf.

[49] L. Maier-Hein, F. Metze, T. Schultz, A. Waibel, in *Proc. 2005 IEEE Workshop on Automatic Speech Recognition and Understanding*, IEEE, Cancun, Mexico **2005**, p. 331, https://ieeexplore.ieee.org/abstract/document/1566521.

[50] S. Yao, A. Myers, A. Malhotra, F. Lin, A. Bozkurt, J. F. Muth, Y. Zhu, *Adv. Healthcare Mater.* **2017**, *6*, 1601159.

[51] S. Yao, Y. Zhu, *Adv. Mater.* **2015**, *27*, 1480.

[52] J. H. Kim, J. W. Park, *ACS Appl. Mater. Interfaces* **2015**, *7*, 18574.

[53] S. K. Ameri, R. Ho, H. Jang, L. Tao, Y. Wang, L. Wang, D. M. Schnyer, D. Akinwande, N. Lu, *ACS Nano* **2017**, *11*, 7634.

[54] C. J. Wang, M. Cai, Z. M. Hao, S. Nie, C. Y. Liu, H. G. Du, J. Wang, W. Q. Chen, J. Z. Song, *Adv. Intell. Syst.* **2021**, *3*, 2100031.

[55] J. Tian, X. Zhao, X. D. Gu, S. Chen, in *Proc. 2020 IEEE Int. Conf. on Robotics and Automation (ICRA)*, IEEE, Paris, France **2020**, p. 10067, https://ieeexplore.ieee.org/abstract/document/9197457.

[56] J. Tian, M. Li, Z. Han, Y. Chen, X. D. Gu, Q. J. Ge, S. Chen, *Comput. Methods Appl. Mech. Eng.* **2022**, *389*, 114394.

[57] S. Yao, P. Ren, R. Song, Y. Liu, Q. Huang, J. Dong, B. T. O'Connor, Y. Zhu, *Adv. Mater.* **2020**, *32*, 1902343.

[58] J. W. Jeong, W. H. Yeo, A. Akhtar, J. J. Norton, Y. J. Kwack, S. Li, S. Y. Jung, Y. Su, W. Lee, J. Xia, H. Cheng, Y. Huang, W. S. Choi, T. Bretl, J. A. Rogers, *Adv. Mater.* **2013**, *25*, 6839.

[59] S. D. Wang, M. Li, J. Wu, D. H. Kim, N. S. Lu, Y. W. Su, Z. Kang, Y. G. Huang, J. A. Rogers, *J. Appl. Mech.* **2012**, *79*, 031022.

[60] Y. Fang, Y. Li, Y. Li, M. Ding, J. Xie, B. Hu, *ACS Appl. Mater. Interfaces* **2020**, *12*, 23689.

[61] K. A. Ludwig, J. D. Uram, J. Yang, D. C. Martin, D. R. Kipke, *J. Neural Eng.* **2006**, *3*, 59.

[62] S. S. Yao, W. X. Zhou, R. Hinson, P. H. Dong, S. Wu, J. Ives, X. G. Hu, H. Huang, Y. Zhu, *Adv. Mater. Technol.* **2022**, *7*, 2101637.

[63] A. Searle, L. Kirkup, *Physiol. Meas.* **2000**, *21*, 271.

[64] J. Wang, C. Pan, H. Jin, V. Singh, Y. Jain, J. I. Hong, C. Majidi, S. Kumar, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 1, https://dl.acm.org/doi/abs/10.1145/3369812.

[65] Collins Dictionary, https://www.collinsdictionary.com/us/ (accessed: March 2022).

[66] Lips Don't Lie, https://wolfpaulus.com/lipsynchronization/ (accessed: February 2021).

[67] S. Lee, D. Yook, in *Proc. Pacific Rim Int. Conf. on Artificial Intelligence*, Springer, Tokyo, Japan **2002**, p. 563, https://link.springer.com/chapter/10.1007/3-540-45683-X_60.

[68] P. Yenigalla, A. Kumar, S. Tripathi, C. Singh, S. Kar, J. Vepa, in *Proc. 19th Annual Conf. of the Int. Speech Communication Association*, ISCA, Hyderabad, India **2018**, p. 3688, https://abhayk1201.github.io/files/paper2.pdf.

[69] U. Hunkeler, H. L. Truong, A. Stanford-Clark, in *Proc. 2008 3rd Int. Conf. on Communication System Software and Middleware and Workshops*, IEEE, Bangalore, India **2008**, p. 791, https://ieeexplore.ieee.org/abstract/document/4554519.

[70] L. Zhang, K. S. Kumar, H. He, C. J. Cai, X. He, H. Gao, S. Yue, C. Li, R. C. Seet, H. Ren, J. Ouyang, *Nat. Commun.* **2020**, *11*, 4683.

[71] ADInstrument, https://www.adinstruments.com/support/knowledge-base/can-data-be-streamed-labchart-windows-matlab-while-sampling (accessed: April 2022).

[72] Gpvigano, M2MqttUnity, https://github.com/gpvigano/M2MqttUnity (accessed: April 2022).