PNAS RESEARCH ARTICLE | ENGINEERING

Closed-loop network of skin-interfaced wireless devices for quantifying vocal fatigue and providing user feedback

Hyoyoung Jeong[a,b,1], Jae-Young Yoo[a,1], Wei Ouyang[a,1], Aurora Lee Jean Xue Greane[c,d], Alexandra Jane Wiebe[c], Ivy Huang[a,e], Young Joong Lee[a,f,g], Jong Yoon Lee[a,h], Joohee Kim[a,i], Xinchen Ni[a], Suyeon Kim[d], Huong Le-Thien Huynh[d], Isabel Zhong[d], Yu Xuan Chin[d], Jianyu Gu[a], Aaron M. Johnson[j,k], Theresa Brancaccio[c], and John A. Rogers[a,d,e,f,l,m,n,2]

Contributed by John A. Rogers; received November 14, 2022; accepted January 12, 2023; reviewed by Huanyu Cheng and Martin Kaltenbrunner

Vocal fatigue is a measurable form of performance fatigue resulting from overuse of the voice and is characterized by negative vocal adaptation. Vocal dose refers to cumulative exposure of the vocal fold tissue to vibration. Professionals with high vocal demands, such as singers and teachers, are especially prone to vocal fatigue. Failure to adjust habits can lead to compensatory lapses in vocal technique and an increased risk of vocal fold injury. Quantifying and recording vocal dose to inform individuals about potential overuse is an important step toward mitigating vocal fatigue. Previous work establishes vocal dosimetry methods, that is, processes to quantify vocal fold vibration dose but with bulky, wired devices that are not amenable to continuous use during natural daily activities; these previously reported systems also provide limited mechanisms for real-time user feedback. This study introduces a soft, wireless, skin-conformal technology that gently mounts on the upper chest to capture vibratory responses associated with vocalization in a manner that is immune to ambient noises. Pairing with a separate, wirelessly linked device supports haptic feedback to the user based on quantitative thresholds in vocal usage. A machine learning-based approach enables precise vocal dosimetry from the recorded data, to support personalized, real-time quantitation and feedback. These systems have strong potential to guide healthy behaviors in vocal use.

closed-loop network | quantifying vocal fatigue | wearable electronics | haptic feedback | real-time machine learning

Significance

The absence of a quantitative monitoring method that can assess the health status of vocal folds increases the potential risk of incurring a vocal disorder. A closed-loop network system that combines a skin-interfaced wireless sensor technology and a haptic feedback module enables continuous monitoring of vocal fold activities related to vocal fatigue. Data analysis using real-time machine learning techniques separates and quantifies vocal dosimetry associated with speaking and singing, without confounding artifacts from ambient sounds, along with a breadth of information on cardiac and respiratory activities and overall physical exertion. This technology approach can help to guide healthy behaviors in vocal usage across a range of affected populations, from singers and teachers to coaches and telemarketers.

One in 13 adults experiences a voice problem each year in the United States, at an estimated cost of nearly 13 billion United States dollars and considerable negative effects on quality of life and mental well-being (1–3). A common complaint is vocal fatigue, a measurable form of performance fatigue resulting from high vocal demands and vocal dose (4). Vocal dose refers to cumulative exposure of the vocal fold tissue to vibration (5). Professionals such as singers and teachers who have high vocal demands are especially prone to heavy vocal loads and voice disorders (6). Singers depend on a high level of consistent vocal quality and sustainability for training and performing, while their voices are also ingrained in daily communication and social activities. Many are unaware of how much or how intensely they are using their voices, putting them at an elevated risk of vocal fatigue and injury. Because vocal performance majors in university and college programs, young professionals, and avocational singers are not always attuned to their daily vocal workload, external monitoring that provides accurate data on their vocal effort in quasi-real time can provide a critical tracking mechanism. The ability to view quantitative data and to receive alerts on vocal usage during varied vocal tasks forms the basis of establishing connections between behaviors and resulting levels of vocal fatigue, thus mitigating the risk of fatigue and vocal fold injury.

Students training to become professional singers in the field of classical music as well as contemporary popular music face great challenges in balancing the demands placed on their voices. Participants who rate themselves as highly talkative are more likely to experience mucosal lesions associated with vibratory trauma (7). Enthusiastic singers may fail to plan for days when rehearsals/performances are scheduled one after the other. Failure to modify habits on heavily scheduled days can lead to compensatory lapses in vocal technique and an increased risk of vocal fold injury, especially if complicated by factors such as upper-respiratory illness, allergies, dehydration, premenstrual syndrome, or reflux disease (8).

The absence of a feasible monitoring method that can assess the condition of vocal folds quantitatively and can continuously inform the user on the potential for overuse of the voice immediately increases the potential risk of incurring a vocal disorder (9). Hence, quantifying and monitoring ambulatory vocal dose to alert individuals in real time about potential overuse is an essential step toward mitigating vocal fatigue. Previous studies establish vocal dosimetry methods, that is, processes to calculate vocal fold vibration dose

Author contributions: H.J., J.-Y.Y., W.O., A.L.J.X.G., T.B., and J.A.R. designed research; H.J., J.-Y.Y., W.O., A.L.J.X.G., A.J.W., I.H., Y.J.L., J.Y.L., J.K., X.N., S.K., and T.B. performed research; H.J., J.-Y.Y., W.O., A.L.J.X.G., A.J.W., I.H., Y.J.L., X.N., H.L.-T.H., I.Z., Y.X.C., J.G., A.M.J., T.B., and J.A.R. analyzed data; and H.J., J.-Y.Y., W.O., A.L.J.X.G., A.J.W., T.B., and J.A.R. wrote the paper.

Reviewers: H.C., The Pennsylvania State University; and M.K., Johannes Kepler University, Linz.

The authors declare no competing interest.

[1]H.J., J.-Y.Y., and W.O. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: jrogers@northwestern.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2219394120/-/DCSupplemental.

Published February 21, 2023.

(10). Such methods in ambulatory voice monitoring and vocal dosimetry studies typically utilize devices that are bulky and require wired interfaces that are poorly suited to continuous monitoring during daily activities. Also, existing methods for postsignal processing often result in delays in notification that limit the user's ability to appropriately modify behaviors (11, 12). A small, lightweight, wireless monitor capable of quantifying vocal dose in any setting and of informing the user of their vocal load status in real time would enable singers and others to adjust behaviors and mitigate risk.

Previous efforts to quantify vocal effort validate the use of an external monitor to measure vocal fold tissue exposure to vibration over time. One study concluded that singers who wore an ambulatory voice monitor had a heightened awareness of voice use during intensive performance situations and were successful in maintaining their vocal health (13). Toles et al. reported the use of ambulatory voice monitors to obtain data surrounding the vocal use of forty-two singers with known phonotrauma (14). Earlier, Gaskill, Cowgill, and Many exploited related monitors to gather 5 d of data from students in their first 2 y of study in vocal performance, music education, and music theater (15). These and other devices require external hardware that is both cumbersome and susceptible to entanglement, to disruptions in data collection, and to artifacts from motion and ambient sounds. In general, these systems are also unable to provide rapid, continuous feedback to the users on cumulative and instantaneous vocal load.

This paper introduces an autonomous closed-loop device and data analytics approach that combines a soft, skin-interfaced wireless mechanoacoustic (MA) sensor and a separate haptic feedback actuator, paired with a real-time machine learning algorithm that operates through a graphical user interface on a smartphone. When mounted at a comfortable location on the upper chest, below the suprasternal notch (SN) (16), the measured signals correspond to broadband accelerations at the surface of the skin, from quasistatic to 3.3 kHz, associated with processes that range from body motions to cardiopulmonary activity (heart rate and respiratory rate) to high-frequency vibrations due to speaking and singing. A simple magnetic coupling scheme allows repeated application and removal of the devices from a fixed region of the skin without irritation over a timeframe of many days. Field data collected from professional singers (N = 16) with different vocal ranges serve as the basis for training and optimizing a computationally efficient machine learning algorithm that can operate in real time on a smartphone to classify singing and speaking events with >90% accuracy, and to quantify instantaneous and cumulative vocal dose. The results appear on a graphical user interface, with additional feedback to the user through a separate wireless device that includes vibrohaptic actuators.

Results

System Design and Closed-Loop Operation. Fig. 1*A* schematically illustrates the platform and its use in monitoring vocal load in a singer. A machine learning algorithm operating on a smartphone distinguishes between singing and speaking events from time series data collected by an MA device on the upper chest, and then quantifies vocal dose, all in real time. A graphical user interface presents the results to the user and a wirelessly paired vibrohaptic device provides user alerts when the cumulative vocal dose exceeds a predetermined personalized threshold (PTh_{VD}). In addition to metrics of vocal activity (such as intensity, pitch, and cumulative dose), the data also include information on cardiac (heart rate) and pulmonary (respiratory rate) activities (*SI Appendix,* Fig. S1). Representative data presented as a spectrogram recorded from a professional singer (soprano) appear in Fig. 1*B*. Haptic feedback can be activated upon detection not only of vocal dose above a threshold but also upon alternative triggering events associated with intensity, pitch, or other parameters (Fig. 2*A*). Fig. 2*B* shows the key components of the MA device including soft encapsulating layers and skin adhesives. The electrical components consist of a Bluetooth Low Energy system on a chip (SoC), flash memory, power-managing units, wireless charging interfaces, a lithium-polymer battery, and an inertial measurement unit that provides accelerometry data along the three axes. A silicone elastomer (Silbione 4420) forms the encapsulating structure. The bottom layer includes a collection of seven small magnets (neodymium magnets; 3 mm diameter, 0.5 mm thickness) that attach to a skin adhesive with a matching set of magnets built into a microfabric layer (Fig. 2*C* and *SI Appendix,* Fig. S2).

Magnetic Mounting Strategy. The most accurate measurements follow from calibrated signals recorded from a specific mounting location. The magnetic coupling scheme introduced here minimizes skin irritation that might otherwise result from repeated application and removal of a device from this single location using a conventional adhesive. Specifically, magnets in the adhesive pair with those integrated into the encapsulating structure of the MA sensor to automatically align and mechanically couple the device to the body. In this way, the user can remove and apply the device to precisely the same location, without disturbing the adhesion at the surface of the skin. Fig. 2*D* demonstrates use of the device and magnetic adhesive on the upper chest. Spectrogram analysis of data from trials confirms that this scheme yields results with similar information content to those collected with a traditional skin adhesive approach (*SI Appendix,* Fig. S3). Descriptions in the *Methods* section provide details.

Haptic Feedback. Information can be presented to the user in the form of graphical display and/or haptic feedback, via use of a smartphone and/or a vibrohaptic device, respectively. As shown in Fig. 2*E*, the latter consists of a wrist-mounted system that includes a Bluetooth 5.1 low-energy SoC for wireless communication (CC2640, Texas Instruments), a microcontroller (ATMega328P, Microchip Technology) to control each haptic actuator through pulse width modulation, and SoCs for wireless near-field communication charging. Four brush-type eccentric rotating actuators in a linear arrangement with a spacing (42 mm) slightly above the two-point discrimination threshold at the wrist (40 mm) provide vibrohaptic feedback with a force controlled over 45 levels by pulse width modulation and with a response time of less than 50 ms. The encapsulated device is shown in the upper image in Fig. 2*F*. Each actuator and the main flexible PCB body interface through serpentine interconnections to enhance the flexibility and stretchability of the system, thereby enhancing wearability at the wrist. The haptic feedback occurs in four different patterns depending on the vocal and cardiopulmonary status. The LED indicators provide a visual indication of these patterns, as shown in the images in Fig. 2*F*. Snap buttons secure the device onto the wrist (Fig. 2*G*).

Convolutional Neural Network Model for Vocal Dose Calculation. Fig. 3*A* illustrates the signal processing flow for vocal usage analysis. The algorithm segments the z-axis acceleration data into nonoverlapping windows with widths of 1 s and labels a window as a voiced frame if the first nonzero-lag peak in the normalized autocorrelation exceeds a threshold of 0.6 (12). A Convolutional neural network (CNN) model evaluates all voice frames and classifies them as either singing or speaking. A median

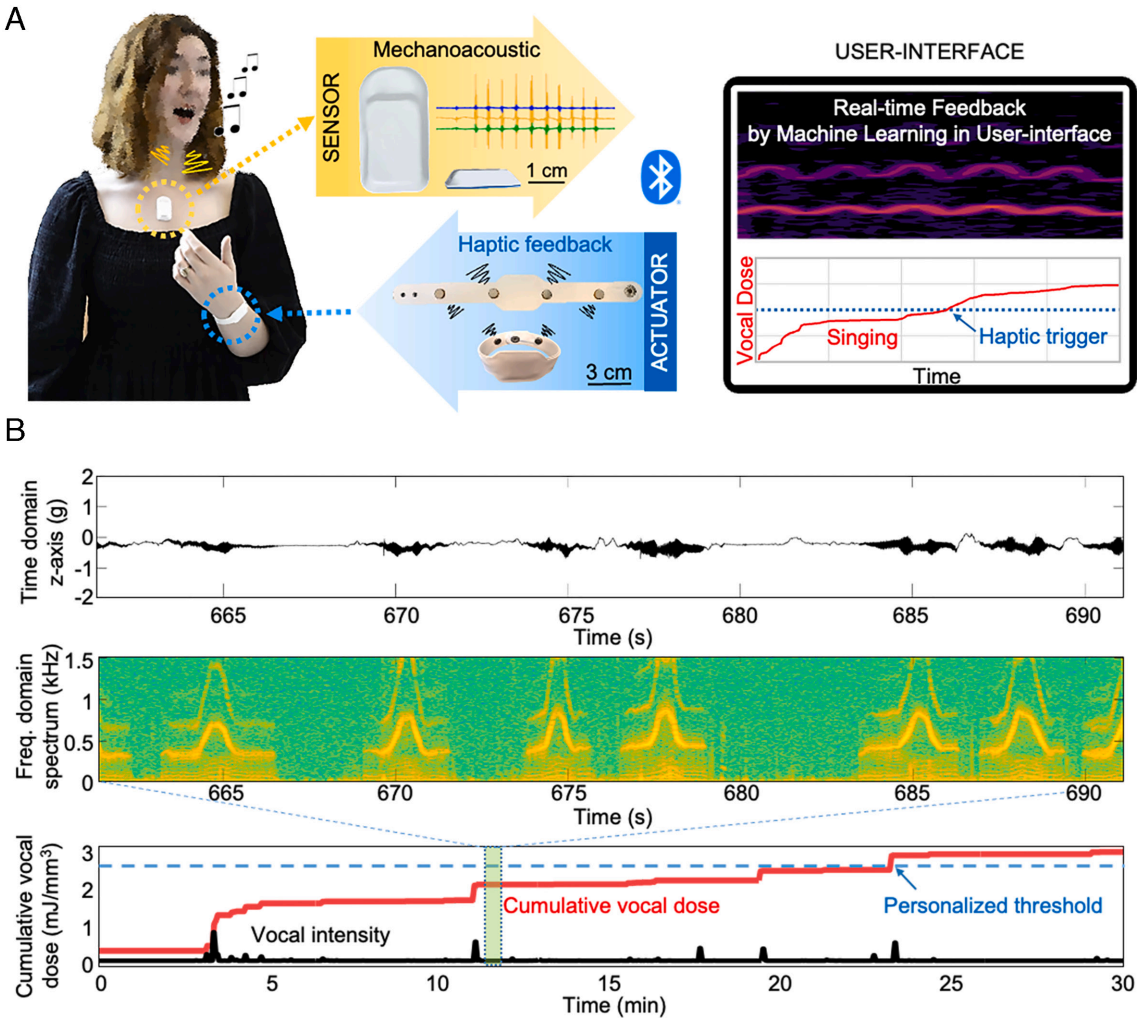

Fig. 1. Wireless, soft, skin-interfaced platform designed for vocal fatigue monitoring with haptic feedback. (*A*) Simplified illustration of the platform operation. A MA sensor mounts on the upper chest, below the SN, for real-time data acquisition of signals related to vocal, cardiac, respiratory, and overall physical activities. A wrist band provides haptic feedback. The user interface supports signal analysis by machine learning to monitor vocal fatigue and to generate graphical and haptic forms of feedback to the user. (*B*) Representative data from continuous monitoring: MA signal, spectrogram, cumulative vocal energy dose, vocal intensity.

filter smooths noise in the classification results. Across sliding windows with widths of 0.1 s, the algorithm computes the energy dose (10), fundamental frequency (f_0, reciprocal of the time lag of the first nonzero-lag peak), the difference between the first and second harmonic magnitudes (H1-H2) (12), and cepstral peak prominence (CPP) (17) of each 0.1-s window. Visualization of the analysis results involves plotting the mean and/or sum of singing/speaking time and energy dose in bins with durations of 5 min and plotting the mean of (H1-H2) and CPP in 1-min bins for singing and speaking, respectively.

Fig. 3 *B* and *C* show representative time-series data and spectrograms of singing and speaking, respectively. Signals associated with singing exhibit better periodicity/resonance and more regular harmonic features than those with speaking (18, 19). Singing is typically more continuous in the time domain and covers a wider range of frequencies in the frequency domain than speaking. These differences motivate the use of the spectrograms for distinguishing singing and speaking. CNN, a widely used classification model for two-dimensional image-like features, is a good candidate for classifying singing and speaking in this context. The CNN takes as input the spectrogram (shape: 168 × 100) determined using a short-time Fourier transform and a Hanning window with a width of 0.1 s moving in time steps of 0.01 s. The CNN starts with three stages of convolutions with a kernel size of 3 × 3, Rectified Linear Unit (ReLU) activation and max pooling, followed by two layers of fully connected neural networks with ReLU activation and one dropout layer ($P = 0.5$). The final output of the CNN model has two neurons with Softmax activation, which correspond to the probabilities of the event as singing or speaking (Fig. 3*D*).

Training of the CNN model uses data collected from 15 professional classical singers (2 male basses, 4 male baritones, 2 male tenors, 4 female sopranos, 3 female mezzo-sopranos), each generating approximately 2,500 1-s windows of singing and 2,500 1-s windows of speaking. A common method to validate the generalization performance of a machine learning model relies on a leave-one-out strategy, where one leaves a subject out of the training set (14 subjects for training) and then tests the trained model on this subject. Iterations apply this approach to each of the 15 subjects. Each training set consists of a random collection of 80% of the labeled events from the 14 subjects, thereby leaving the remaining 20% for validation. The training uses an Adam optimization algorithm. Fig. 3*E* shows the averaged confusion matrix of 15 leave-one-out testing cycles. The model achieves accuracies of 0.96 ± 0.02 for singing and 0.95 ± 0.03 for speaking. Fig. 3*F* shows the overall classification accuracies for each subject using a model trained on the other 14 subjects. The results are above 0.90 for all the 15 subjects. Fig. 3*G* presents the receiver operating characteristic (ROC) curves for each subject. The high area under the curve

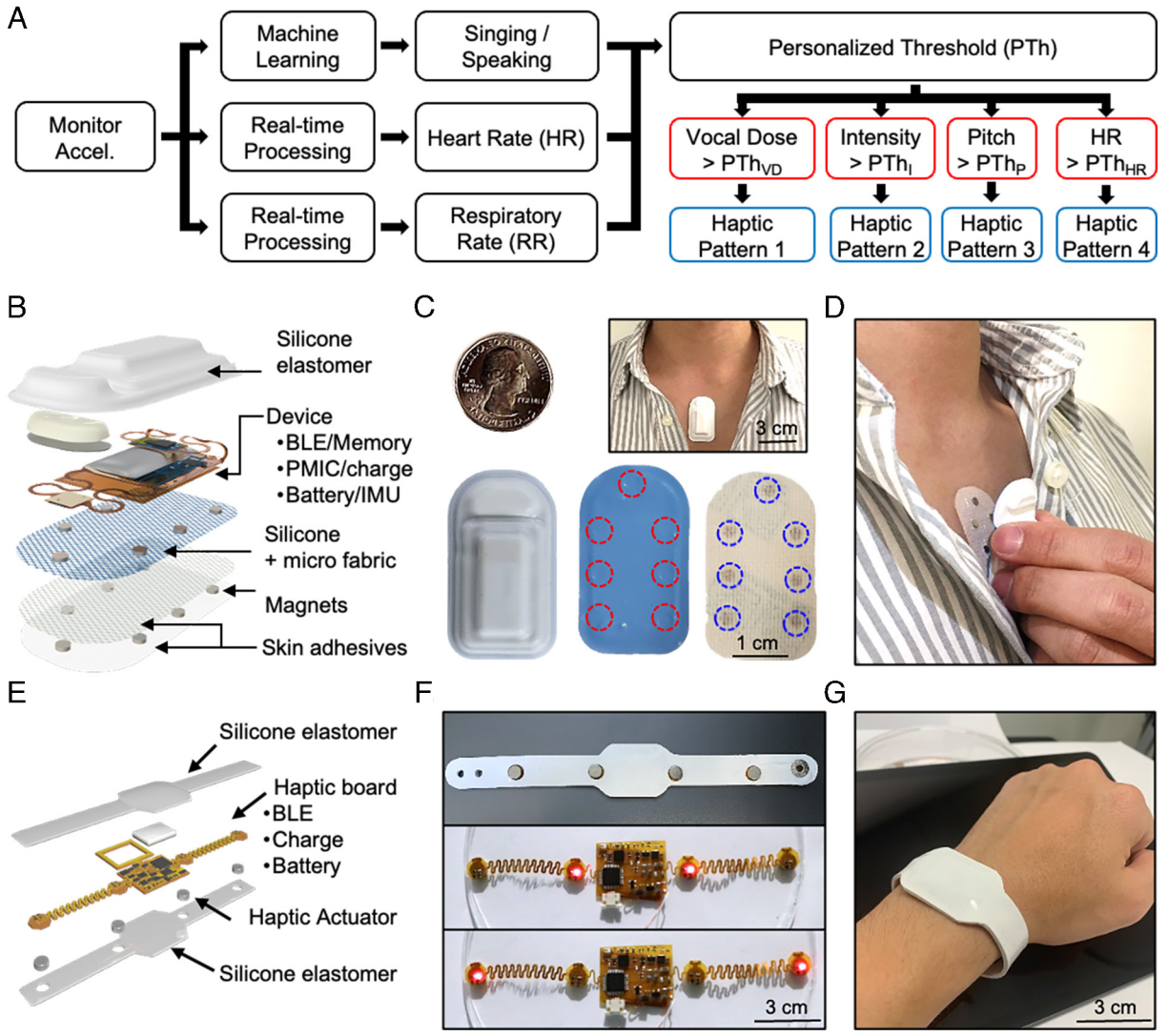

Fig. 2. Functional flowcharts, images, and haptic feedback mechanisms for real-time vocal fatigue detection. (*A*) Block diagram for signal analysis and haptic feedback for energy dose, vocal intensity, vocal pitch, and heart rate. (*B*) Exploded-view illustration of the MA device. (*C*) *Top*/*Bottom* views of an MA device, skin adhesives that support magnetic coupling, and a picture of a device mounted on a subject. (*D*) Mechanism for mechanical coupling of the device to the body via a collection of small magnets embedded in the bottom encapsulation structure of the MA device and the skin adhesive. (*E*) Exploded-view illustration of the haptic device. (*F*) *Bottom* views of a haptic device with indicator light-emitting diodes and actuators. (*G*) Picture of a haptic device on the wrist.

(AUC) >0.97 for all subjects indicates that the model achieves a good balance between sensitivity and specificity. This CNN-based machine learning algorithm is well suited for analyzing data collected during a typical day, where the results guide singers to manage vocal fatigue by providing detailed personalized daily vocal usage and health analysis reports.

Real-Time Machine Learning for Vocal Fatigue Assessment. Along with these daily reports determined using CNN-based machine learning, capabilities for real-time vocal fatigue monitoring can be valuable. The approach described here uses the k-nearest neighbors (KNN)-based image classification machine learning processing scheme to reduce the computation load. The result enables real-time classification of vocal usage and vocal energy dose, along with vital signs, using processing capabilities available on mobile devices. The process involves dividing each 1-s window of data into a voiceless frame and a voiced frame using an autocorrelation process. Additional computations yield spectrogram images of the voiced frames, for display in real-time. These images can be classified into singing and speaking with an accuracy of 91% through the KNN-based image machine learning model of the Core machine learning tool. Calculating the energy dose for each 1-s window of data enables real-time monitoring of vocal fatigue. Fig. 4*A* shows the graphical user interface for spectrograms and cumulative energy dose for singing and speaking. The respiratory rate and heart rate follow from application of a low-pass filter with 0.9 Hz cutoff frequency and a band-pass filter between 10 Hz and 40 Hz followed by application of peak finding algorithms, respectively.

Whenever the user feels discomfort in the vocal folds during daily monitoring, the cumulative energy dose and energy dose intensity data at that time are separately recorded by pressing the button on the real-time monitoring graphical user interface app. These values serve as personalized thresholds.

The vibrohaptic device provides real-time feedback based on vocal monitoring results and these personalized thresholds. As shown in Fig. 4 *B* and *C*, the cumulative energy dose and vocal intensity of singing and speaking activities while monitoring singers during a 30-min rehearsal period appear in real time. Upon exceeding corresponding thresholds, the vibrohaptic device produces patterns of feedback as shown in Fig. 4*D*. Specifically, when the cumulative energy dose exceeds its threshold, the actuators

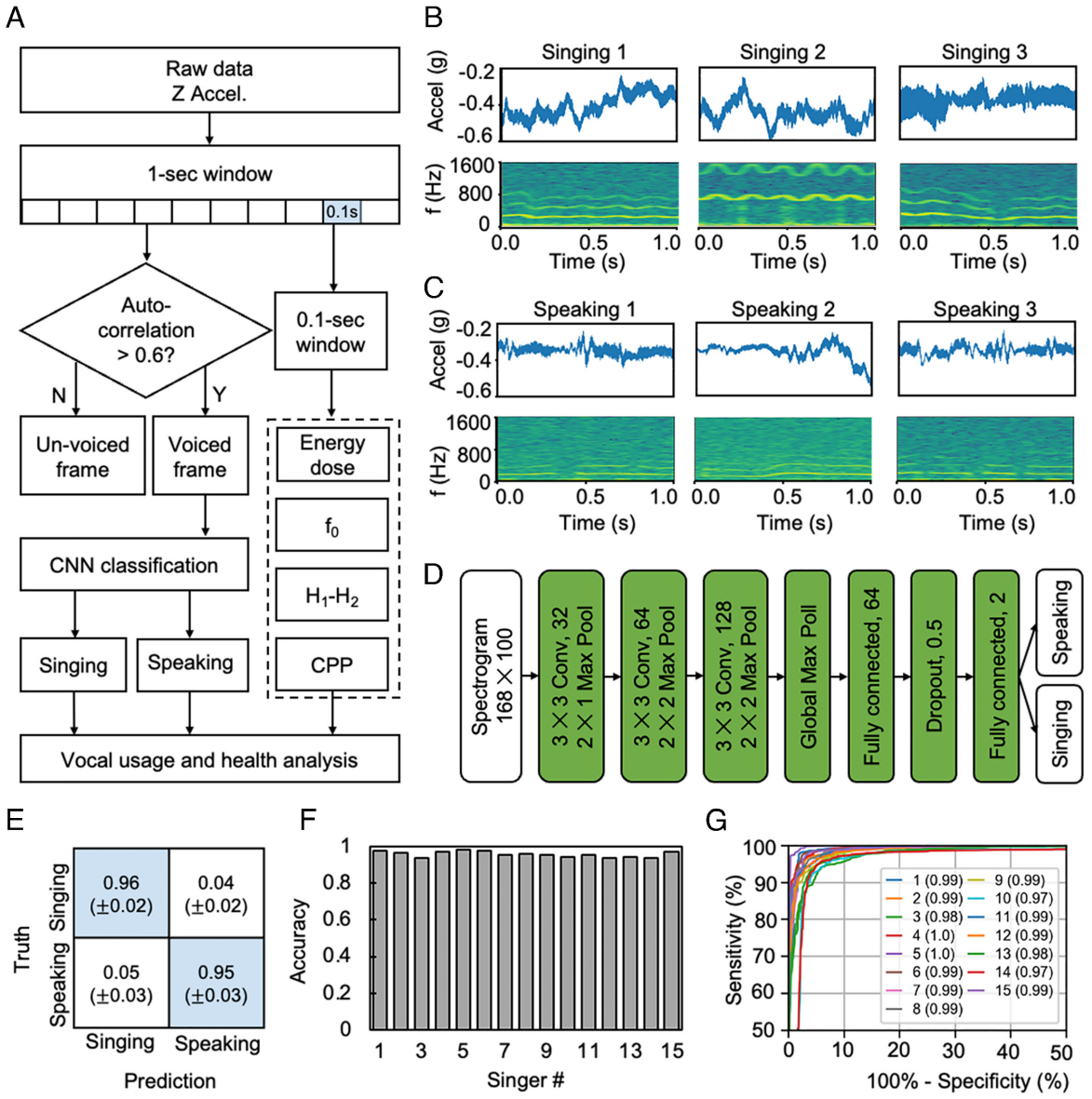

Fig. 3. Algorithms for singing and speaking analysis and representative results. (*A*) Flowchart of the algorithm for singing/speaking classification and computation of vocal measures (energy dose, fundamental frequency, H1-H2, CPP). (*B*) Representative raw data and spectrograms associated with singing. (*C*) Representative raw data and spectrograms associated with speaking. (*D*) Architecture of the convolutional neural network for singing/speaking classification with the spectrogram as the input. (*E*) Confusion matrix of singing/speaking classification by the CNN model. (*F*) Overall classification accuracy using a leave-one-out strategy iterated through all the 15 singers. (*G*) ROC curves of the classification performance on all 15 singers and the corresponding AUC values.

above and below the wrist sequentially vibrate at intervals of 500 ms, as shown in the left vibration detection graph and optical image in Fig. 4*D*. In parallel, the system tracks heart rate and respiratory rate as shown in Fig. 4 *E* and *F*. This information has the potential to enhance breath control training for for singers who must engage in acting, dancing, or other forms of physical exertion during a performance.

Field Study. These studies explore four different data collection protocols. The first focuses on training a machine learning algorithm to distinguish singing from speaking. The second demonstrates the functionality of the real-time interface with multiple sensors in a rehearsal setting. The third aims to confirm that the calculated vocal dose corresponded with users' perceived effort. The fourth seeks to align sensor data with self-reported tasks. Data for distinguishing between singing and speaking follow from sixteen singers wearing MA sensors on the upper chest while singing six different vocal exercises across their personal full vocal range. These exercises include hums, glides, legato scales, arpeggios, staccato scales, and monotones with varied dynamics, along with singing a song of their choice for 4 min and reading from a book for 10 min. These data serve as the basis for training the machine learning algorithm (*SI Appendix, Supporting text* and Figs. S9 and S11).

Devices used in a choir rehearsal setting demonstrate the capacity to capture data from an individual singer without influence from vocalization by other singers. The studies involve four singers (one soprano, one alto, one tenor, and one bass) during a rehearsal, with devices paired with smartphones or tablets for real-time dosimetry calculations, as shown in Fig. 5 *A* and *C* and Movie S1. In addition to vocal energy dose, the measurements also capture quantitative trends of heart rates (HR) and respiratory rates (RR). Collectively, these factors are relevant as indirect assessments of lung pressure and of overall body fatigue level. Processing the raw

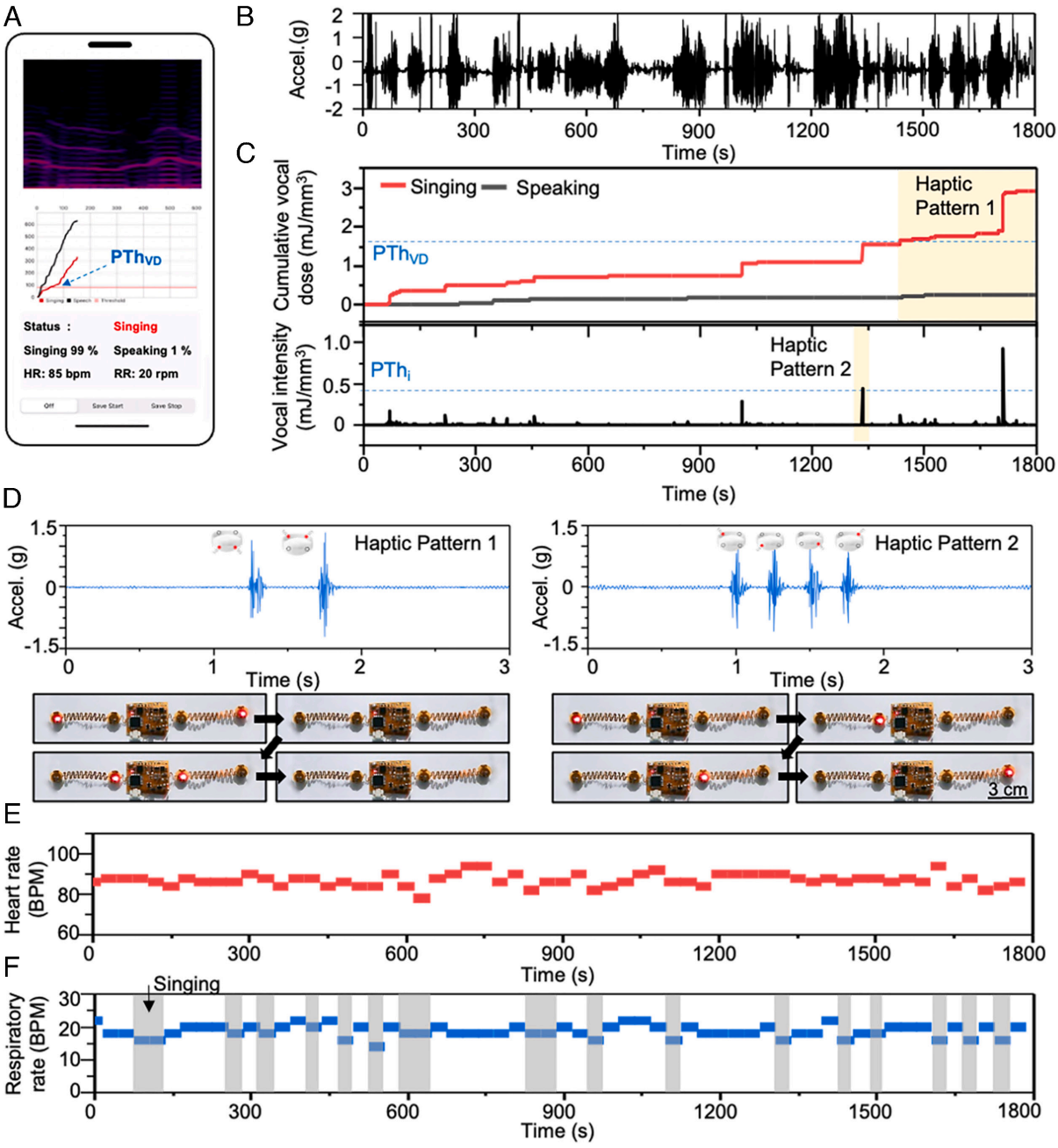

Fig. 4. Real-time vocal fatigue detection and haptic feedback. (*A*) The graphical user interface displays spectrograms, energy dose levels for singing and speaking, and respiratory rate and heart rate, all in real time. (*B*) MA data collected during a 30-min singing rehearsal. (*C*) Cumulative energy dose (upper graph) and energy dose intensity (under graph) associated with singing and speaking during a 30-min rehearsal. The blue dashed line of the upper graph is the personalized threshold of cumulative energy dose (PTh_{VD}). The blue dashed line of the lower graph is the personalized threshold of energy dose intensity (PTh_i). The red shaded region identifies the moment of haptic feedback. (*D*) Accelerations generated by haptic vibration pattern 1 and pattern 2 activated when the cumulative energy dose and energy dose intensity exceed the corresponding personalized thresholds, respectively. The photographs show the haptic vibration of the device via red LED indicators. (*E*) Calculated respiratory rate determined from raw data after band-pass filtering. (*F*) Calculated heart rate determined from raw data after band-pass filtering.

data yields cumulative vocal energy dose and time for each singer, including classifications of singing and speaking. Fig. 5 *D* and *G* summarizes the cumulative singing/speaking energy dose values and times, along with the HR and RR for each singer. Unlike microphone recordings, these data are not affected by ambient sounds from other singers or sources.

Long-Term Monitoring of Vocal Use. These technologies are effective for long-term vocal monitoring because they can be used almost anytime and anywhere, with minimal burden. Fig. 6 *A*–*C* shows results of analysis of a representative day of recordings from a singer, aligned with a manual log of vocal activity. The algorithm accurately discerns singing (e.g., vocalizing and choir rehearsal) and speaking (e.g., phone call, chatting) events. Listening to the MA data after conversion into audio forms further confirms the accuracy and reveals that alternations between singing and speaking occur from unlogged, intermittent speaking/singing during singing/speaking. This process also identifies laughing events, which are not yet incorporated into the model. Fig. 6 *B* and *C* shows the singing/speaking times color coded by mean energy dose and cumulative energy dose for singing and speaking, respectively. Fig. 6 *D* and *F* summarizes the results of vocal usage

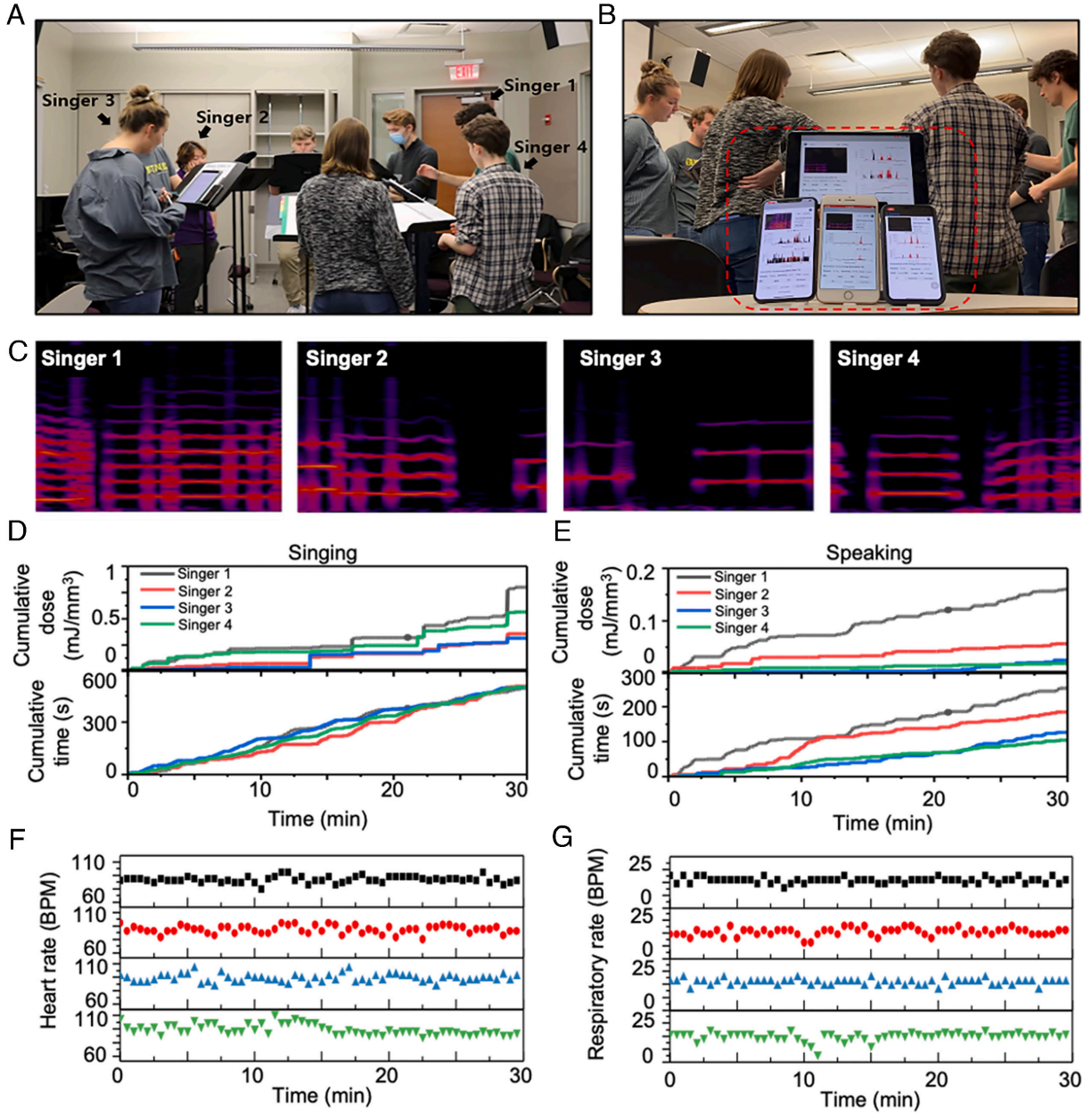

Fig. 5. Data collected from four singers in an acapella group simultaneously measure vocal energy dose, heart rate, and respiratory rate. (*A*) Real-time simultaneous data acquisition from four singers wearing MA devices. (*B*) Pictures of the user interface for devices for each of the singers. (*C*) Representative spectrograms from each singer: base (singer 1), countertenor (singers 2 and 3), soprano (Singer 4). Simultaneous measurements of (*D*) cumulative singing energy dose, cumulative singing time, (*E*) cumulative speaking energy dose, cumulative speaking time, (*F*) heart rate, and (*G*) respiratory rate for each singer.

monitoring for 5 d from four singers. Fig. 6*F* indicates that, compared to singers 1 and 2, singers 3 and 4 exhibit a relatively high dose associated with speaking, consistent with the lecturing activities of these two singers. These examples illustrate the value in continuously tracking the cumulative energy dose values and times for both singing and speaking, separately.

Discussion

The closed-loop, body-worn network reported here presents a class of skin-interfaced wireless technology that allows quantitative and continuous monitoring of vocal fold use with capabilities in both graphical and haptic feedback. The system provides information that can help guide healthy behaviors in singing and speaking. The sensing mechanism relies on vibratory responses digitally recorded at a given location on the upper chest with a soft, wireless device, to enable accurate voice dosimetry in a comfortable manner without confounding artifacts from ambient sounds. The data, when analyzed with machine learning approaches, yield detailed information on instantaneous and cumulative vocal load, along with a breadth of information on cardiac and respiratory activities, body orientation, and overall physical exertion. Operation is possible over extended periods from a single location on the skin enabled by a magnetic coupling scheme, with features that address practical use considerations, including those associated with rehearsal settings and workplace environments. Demonstrations in tracking of a vocal usage and cardiac activity of singers for 5 d illustrate the potential for persistent vocal management. While the current study demonstrates the feasibility of real-time monitoring and feedback notification for vocal fatigue status associated

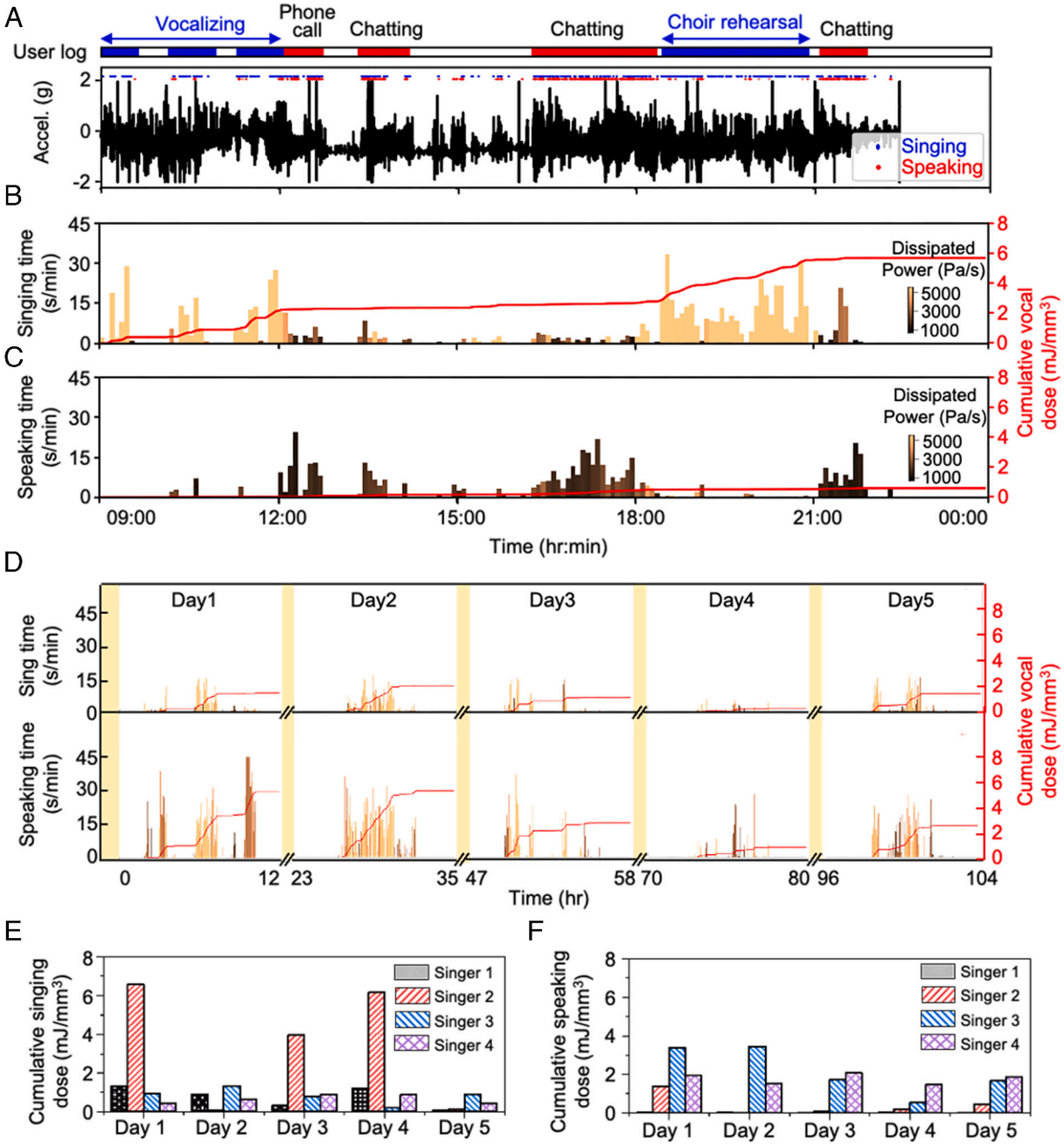

Fig. 6. Tracking of long-term vocal use in the form of daily cumulative singing/speaking energy dose and time. (*A*) Representative 1-d recording of a singer aligned with the singer's log and singing/speaking classification by the CNN model. (*B*) Singing time per minute and corresponding cumulative energy dose throughout the day. (*C*) Speaking time per minute and corresponding cumulative energy dose throughout the day. (*D*) Long-term singing and speaking measurements over 5 d. Variation of vocal energy dose from the singer while practicing for a rehearsal over a period of 5 d. The first set was measured from 11 a.m. to 11 p.m. on the first day. The second set was measured from 10 a.m. to 10 p.m. on the second day. The third set was measured from 10 a.m. to 9 p.m. on the third day. The fourth set was measured from 9 a.m. to 7 p.m. on the seventh day, and the fifth set was measured from 11 a.m. to 7 p.m. on the eighth day. The red line shows the cumulative vocal energy dose. (*E*) Daily cumulative singing energy dose monitoring and (*F*) cumulative speaking energy dose monitoring for 5 d of four singers.

with singers, these technology platforms can apply also those in other professions that rely heavily on the voice, from teachers and coaches, to telemarketers and salespeople.

Materials and Methods

Forming Magnetic Coupling Structures in the Device Encapsulation. A laser-cutting process defined a template on a glass slide. Spin coating and curing a layer of silicone, placing a thin layer of fabric on top, and spin coating and curing an overcoat of silicone yielded a bottom encapsulation structure with a collection of relief features in the circular geometries of the magnets. Peeling this structure from the glass slide, placing magnets in these features, and then executing the other steps in the encapsulation process according to procedures described elsewhere, completed the device. *SI Appendix*, Fig. S4 depicts glass substrate with the template, silicone, and magnets.

Comparing Adhesives and Magnetic Coupling Schemes. Tests of peel force and skin comfort provided a basis for comparing multiple adhesives shown in *SI Appendix*, Figs. S5 and S6. The adhesive found to be most comfortable out of those tested was a combination of a breathable fabric-based kinesiology tape (Food and Drug Administration approved and medical grade) for the skin-side interface, coupled with a strong adhesive (soft silicone tape) for the device-side interface (#2 in *SI Appendix*, Fig. S5), with magnets embedded in geometries to

match those in the bottom side of the device encapsulation. Tests of devices with a magnetic adhesive design and a medical silicone tape enabled comparisons of the quality of the corresponding data captured with participants sitting, standing, and walking for thirty seconds each. *SI Appendix,* Fig. S3 shows data from a sitting segment of trials with each adhesive. The results indicate comparable information content. *SI Appendix,* Fig. S8 shows 3 min of speaking and 3 min of singing data. A CNN-based machine learning algorithm extracted the fundamental frequencies as well as the total singing times and cumulative dose levels. As shown in *SI Appendix,* Fig. S8, the algorithm successfully distinguished between singing and speaking using data collected in all cases.

Calibrating the Mechano-Acoustic to Acoustic Power. Calibration of MA data to acoustic power enabled quantified measurements of vocal load. Data collected for this purpose involved an MA device mounted on the upper chest and a phone with a decibel meter app (Decibel X) placed 0.5 m away from the mouth. Synchronized recordings used the MA device and the decibel meter during repetition of a specific word 10 times at a variety of volumes: soft whispering, moderate whispering, loud whispering, soft speaking, moderately soft speaking, typical speaking, moderately loud speaking, loud speaking, moderate shouting, loud shouting, and extremely loud shouting. Representative data are shown in *SI Appendix,* Fig. S12. From left to right, the data segments correspond to the volumes listed above. The vocal dose algorithm used raw data collected from these calibration tests.

Convolutional Neural Network. The CNN starts with three stages in the following order: 32-channel 3 × 3 convolution, 2 × 1 Max pooling, 64-channel 3 × 3 convolution, 2 × 2 Max pooling, 128-channel 3 × 3 convolution, 2 × 2 Max pooling, global max pooling. The model subsequently consists of a fully connected neural network with 64 neurons at the input and a dropout layer ($P = 0.5$). At the final output are two neurons representing the probabilities of the two classes. All layers use the ReLU activation. The CNN uses an Adam optimizer for training. The training process follows a leave-one-out strategy, where one leaves a subject out of the training set (14 remaining subjects for training) and then tests the trained model on this subject. Each training set applies a fivefold crossvalidation procedure. This approach iterates through each of the 15 subjects. Comparisons of singing/speaking classification by the CNN model using the data collected at five different locations on the chest appear in *SI Appendix,* Figs. S13 and S14. Although the signal amplitude depends on location, the CNN model can distinguish singing from speaking with >90% accuracy in all cases, based on spectrogram-based classification (*SI Appendix,* Fig. S14) as shown in *SI Appendix,* Fig. S13 (blue dots: singing, red dots: talking).

Data Analytics. All analyses used Python 3.0 with SciPy and TensorFlow packages. External call of PRAAT (cite: https://www.fon.hum.uva.nl/praat/) from the Python script fulfills the computation of CPP.

Distinguishing Singing from Speech. Sixteen singers, ranging in age from 20 to 61 y, wore an MA device adhered to the sternum, just below the sternal notch. The participants sang six different vocal exercises through their full vocal range (hums, glides, legato scales, arpeggios, staccato scales, and monotones with varied dynamics) followed by a song of their choice for 4 min. The participants then read from the first chapter of the book Grit: The Power of Passion and Perseverance for 10 min (20). These samples served as the training set for the development of machine learning algorithms, capable of distinguishing singing from speaking with 91% accuracy.

Validation of Interference by Ambient Noise. Two types of control measurements examined the interference from ambient noise. A qualitative test involved an MA device on the upper chest of a subject singing with different ranges of pitch in an ambient with loud music (Beethoven Symphony No. 9, 62 to 79 dBA range). *SI Appendix,* Figs. S15 and S17 demonstrate that the influence of ambient music is almost negligible. The data collected during (0 to 12 s) with the same ambient music showed clear features of cardiac activity and respiratory cycles. During singing (14 to 42 s), clear fundamental frequency and harmonic signals can be observed as shown in *SI Appendix,* Fig. S15. Converting the data to audio files confirmed the expected nature of the data (Audio S1). Additional experiments in *SI Appendix,* Figs. S16 and S17 allowed for quantitative comparisons. Specifically, the signal-to-noise ratios (SNRs) were 34.6 dB or more (51.4 dB, 34.6 dB, and 35.1 dB at SN, SN–1″, and SN–2″, respectively) for data captured at three different locations including the SN, 1 inch below SN, and 2 inches below SN (*SI Appendix,* Fig. S17). The SNR was calculated based on the ratio of the speaking signal (moderate speaking level) and the noise signal (ambient music). Ambient music at an amplitude of 60 dBA appears in the MA data as signals with amplitudes less than 1/1,000th of that associated with whispering.

Known Tasks. To confirm the accuracy of the algorithm for determining vocal dose and effort, four singers (soprano, alto, tenor, and baritone) completed a set of between 10 and 11 tasks while recording with the MA device, each lasting for 1 min with 1 min of rest in between. A handy video recorder (Q8, Zoom) for all participants simultaneously recorded acoustic audio samples. The tasks included normal speaking, speaking over 60 dB of ambient noise, whispered speaking, moderately loud singing (low to mid range), moderately loud singing (mid to high range), very loud singing (low to mid range), very loud singing (mid to high range), singing without vibrato (low to mid range), staccato arpeggios throughout the vocal range, and strained speaking. Each participant selected one excerpt to use for the low- to mid-range tasks and another piece to use for the mid- to high-range tasks. For the speaking exercises, participants read from "Practicing Vocal Music Efficiently and Effectively: Applying 'Deliberate Practice to a New Piece of Music'" by Ruth Rainero (21). The strained speaking task was completed only by the alto participant.

Protocols for Field Study. The studies were approved by the Institutional Review Boards of Northwestern University (STU00207900). During data collection, the device was mounted on the SN of singers. All singers consented to procedures and provided written consent form for images.

Data, Materials, and Software Availability. Data; code data have been deposited in [University of Idaho Resource Computing and Data Services Data Repository] (https://doi.org/10.11578/1908656; TBD).

ACKNOWLEDGMENTS. This work was supported by the Querrey-Simpson Institute for Bioelectronics at Northwestern University.

Author affiliations: [a]Querrey Simpson Institute for Bioelectronics, Northwestern University, Evanston, IL 60208; [b]Department of Electrical and Computer Engineering, University of California, Davis, CA 95616; [c]Bienen School of Music, Northwestern University, Evanston, IL 60208; [d]Department of Biomedical Engineering, Northwestern University, Evanston, IL 60208; [e]Department of Materials Science Engineering, Northwestern University, Evanston, IL 60208; [f]Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208; [g]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02142; [h]Sibel Health, Niles, IL 60714; [i]Center for Bionics, Biomedical Research Institute, Korea Institute of Science and Technology, Seoul 02792, South Korea; [j]Department of Otolaryngology-Head and Neck Surgery, Grossman School of Medicine, New York University, New York, NY 10016; [k]Department of Rehabilitation Medicine, Grossman School of Medicine, New York University, New York, NY 10016; [l]Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208; [m]Department of Chemistry, Northwestern University, Evanston, IL 60208; and [n]Department of Neurological Surgery, Northwestern University, Evanston, IL 60208

1. N. Roy, R. M. Merrill, S. D. Gray, E. M. Smith, Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *Laryngoscope* **115**, 1988–1995 (2005).
2. S. M. Cohen, Self-reported impact of dysphonia in a primary care population: An epidemiological study. *Laryngoscope* **120**, 2022–2032 (2010).
3. N. Bhattacharyya, The prevalence of Dysphagia among adults in the United States. *Otolaryngol. Neck Surg.* **151**, 765–769 (2014).
4. E. J. Hunter *et al.*, Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. *J. Speech Lang. Hear. Res.* **63**, 509–532 (2020).
5. I. R. Titze, J. G. Švec, P. S. Popolo, Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. *J. Speech Lang. Hear. Res.* **46**, 919–932 (2003).
6. M. Behlau, F. Zambon, A. C. Guerrieri, N. Roy, Epidemiology of voice disorders in teachers and nonteachers in Brazil: Prevalence and adverse effects. *J. Voice* **26**, e9–e18 (2012).
7. R. W. Bastian, J. P. Thomas, Do talkativeness and vocal loudness correlate with laryngeal pathology? A study of the vocal overdoer/underdoer continuum. *J. Voice* **30**, 557–562 (2016).
8. P. W. Flint *et al.*, Eds., *Cummings Otolaryngology: Head and Neck Surgery* (Elsevier, ed. 7, 2021).
9. J. P. Assad, M. d. C. Magalhães, J. N. Santos, A. C. C. Gama, Dose vocal: Uma revisão integrativa da literatura. *Rev. CEFAC* **19**, 429–438 (2017).
10. I. R. Titze, E. J. Hunter, Comparison of vocal vibration-dose measures for potential-damage risk criteria. *J. Speech Lang. Hear. Res.* **58**, 1425–1439 (2015).
11. D. D. Mehta *et al.*, Using ambulatory voice monitoring to investigate common voice disorders: Research update. *Front. Bioeng. Biotechnol.* **3**, 155 (2015).
12. J. H. Van Stan *et al.*, Differences in weeklong ambulatory vocal behavior between female patients with phonotraumatic lesions and matched controls. *J. Speech Lang. Hear. Res.* **63**, 372–384 (2020).

13. M. J. Schloneger, Graduate student voice use and vocal efficiency in an opera rehearsal week: A case study. *J. Voice* **25**, e265–e273 (2011).
14. L. E. Toles *et al.*, Differences between female singers with phonotrauma and vocally healthy matched controls in singing and speaking voice use during 1 week of ambulatory monitoring. *Am. J. Speech Lang. Pathol.* **30**, 199–209 (2021).
15. C. S. Gaskill, J. G. Cowgill, S. Many, Comparing the vocal dose of university students from vocal performance, music education, and music theater. *J. Sing.* **70**, 11–19 (2013).
16. H. Jeong *et al.*, Differential cardiopulmonary monitoring system for artifact-canceled physiological tracking of athletes, workers, and COVID-19 patients. *Sci. Adv.* **7**, eabg3092 (2021).
17. R. Fraile, J. I. Godino-Llorente, Cepstral peak prominence: A comprehensive analysis. *Biomed. Signal Process. Control* **14**, 42–54 (2014).
18. A. J. Ortiz *et al.*, Automatic speech and singing classification in ambulatory recordings for normal and disordered voices. *J. Acoust. Soc. Am.* **146**, EL22–EL27 (2019).
19. W.-H. Tsai, C.-H. Ma, "Automatic speech and singing discrimination for audio data indexing" in *Big Data Applications and Use Cases*, P. C. K. Hung, Ed. (International Series on Computer Entertainment and Media Technology, Springer International Publishing, 2016), **1**, pp. 33–47.
20. A. Duckworth, *Grit: The Power of Passion and Perseverance* (Scribner/Simon & Schuster, 2016).
21. R. Rainero, Practicing vocal music efficiently and effectively: Applying "deliberate practice" to a new piece of music. *J. Sing.* **69**, 203–214 (2012).