Supporting information

Decoding Silent Speech Commands from Articulatory Movements Through Soft Magnetic Skin and Machine Learning

*Penghao Dong, Yizong Li, Si Chen, Justin T. Grafstein, Irfaan Khan, Shanshan Yao**

Penghao Dong, Yizong Li, Si Chen, Justin T. Grafstein, Shanshan Yao

Department of Mechanical Engineering, Stony Brook University, Stony Brook, New York 11794, USA

E-mail (Shanshan Yao): shanshan.yao@stonybrook.edu

Irfaan Khan

Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, New York 11794, USA

Method for eliminating signal interferences:

The geomagnetic field and human motion artifacts are sources of interferences that affect magnetic signals. The Kabsch algorithm is a method utilized for computing the optimal rotation matrix that minimizes the root mean squared deviation between two corresponding sets of points. [1] In this particular study, the rotation matrix between the working magnetometer coordinates (X, Y, and Z in Fig. S1a) and the reference magnetometer coordinates (x, y, and z in Fig. S1b) was computed. Initially, the two magnetometers were affixed to the temporal bones, while the magnetic skin was not attached. Data collection was first performed without the magnetic skin to calculate the rotation matrix necessary for rotating the data from the reference magnetometer coordinate into the working magnetometer coordinate. The acquired data from the working and reference magnetometers (without magnetic skin) can be represented as matrix $\boldsymbol{P}$ and matrix $\boldsymbol{Q}$, respectively, as shown in Equations S1 and S2.

$$\boldsymbol{P} = \begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n \end{bmatrix} \tag{S1}$$

$$\boldsymbol{Q} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{bmatrix} \tag{S2}$$

where the first, second, thirds columns of $\boldsymbol{P}/\boldsymbol{Q}$ are the data points for the X/x, Y/y, and Z/z directions, respectively.

The first step is the translation. All the data points of $\boldsymbol{P}$ were subtracted from the average of the whole column. The same procedure was also applied to $\boldsymbol{Q}$. The second step involves the computation of the covariance matrix $\boldsymbol{H}$ as described by Equation S3.

$$\boldsymbol{H} = \boldsymbol{Q}^T \boldsymbol{P} \tag{S3}$$

The third step is using singular value decomposition (SVD) to calculate the optimal rotation matrix. In this step, the SVD of $\boldsymbol{H}$ was first conducted. The equation for SVD is shown in Equation S4.

$$\boldsymbol{H} = \boldsymbol{U\Sigma V}^T \tag{S4}$$

where $\boldsymbol{\Sigma}$ denotes the diagonal matrix of which the singular value is equal to $\boldsymbol{H}$. $\boldsymbol{U}$ and $\boldsymbol{V}$ are the left and right singular vectors for the corresponding singular values. The SVD can be performed using the numpy.linalg.svd function in the Python numpy package.

Then, the optimal rotation matrix can be calculated by Equation S5.

$$\boldsymbol{R} = \boldsymbol{VU}^T \tag{S5}$$

Finally, the noise induced by interferences can be removed using Equation S6.

$$\boldsymbol{P}_{\boldsymbol{denoise}} = \boldsymbol{P} - \boldsymbol{RQ} \tag{S6}$$

Fig. S18 depicts the denoising process utilizing the Kabsch algorithm. Initially, signals were collected during the subject's motion, as shown in Fig. S18a. The root sum square analysis (Fig. S18b) demonstrates that noise levels for both the working and the reference magnetometers are highly similar, indicating that appropriate rotations can effectively eliminate the noise. Fig. S18c presents the original signals captured by the working magnetometer and signals obtained from the reference magnetometer after applying the rotation. Notably, these two sets of signal patterns exhibit high similarity. Subsequently, by subtracting magnetic signals captured by the reference magnetometer from signals captured by the working one, the influence of noises on the magnetic signals detected by the working magnetometer can be significantly reduced, as depicted in Fig. S18d.

Table S1. Comparison of various silent speech systems based on contact-based approaches.

Method	Algorithm	Language	Soft/Rigid sensing interface	Sensor amount	Classificati on task	Recogniti on Accuracy	Number of Classes	Natural speech	Interface location
Magnetic signal [This work]	LDA	English	Soft	1	Phoneme/ Word/Phra ses/Extend ed word list	93.2%/ 93.5% 96.7% 85.7%	9/8/6/54	Yes	SRT
Pressure[2]	Morse code related	English	Soft	1	Alphabet	95%	26	No	Throat
Strain[3]	SVM	English	Soft	5	Alphabet	98.63%	11	No	Hand
Strain[4]	RNN	English	Soft	4	Word	85.2%	5	Yes	Face
Strain[5]	RF	English	Soft	1	Word	86.0%	11	Yes	Face
Strain[6]	DTW	English Chinese	Soft	5	Word	80%	8	Yes	Face
Strain[7]	DNN	English	Soft	4	Word	87.5%	100	Yes	Fac
Strain[8]	CNN	English	Soft	8	Word	84.4%	21	Yes	Face
EMG[9]	LDA SVM	English	Soft	8	Word	94.8%	11	Yes	Face Neck
EMG[10]	WPT	English	Soft	4	Word	92.6%	110	Yes	Face
EMG[11]	WPT	English	Soft	3	Word	89.04%	6	Yes	Face, Neck
Strain[12]	RNN	English Chinese	Soft	2	Word	94.5%	20	Yes	Face
EPG[13]	HMM	English	Soft	124	Word	97.0%	107	Yes	Oral palate
Proximity signal [14]	CNN, LSTM, CTC	English	Rigid	2	Word	90.0%	32	Yes	Ear canal
Angular velocity, Acceleration [15, 16]	Particle filter	English	Rigid	2	Phoneme/ Word	95.1%/ 91.0%	9	Yes	TMJ
EMG[17]	CNN	English	Rigid	7	Word	92.0%	42	Yes	Face, Neck
EMG[18]	KALDI	English	Rigid	8	Word	89.7%	\	Yes	Face, Neck
Magnetic signal[19, 20]	RNN	English	Rigid	4	Word	92.0%	\	Yes	Lips, Tongue
Magnetic signal, Proximity signal [21]	HMM	English	Rigid	3	Word	90.5%	11	Yes	Tongue, Ear canal
EEG[22]	k-NN	English	Rigid	128	Phoneme	68.8%	2	Yes	Head
EEG MEG[23]	Least squares	English	Rigid	164	Word	90.0%	7	Yes	Head

Note: The blue part in the table is for rigid sensing interfaces, while the black part is for soft interfaces. Articles in these two parts are ranked based on the unobtrusiveness from high to low, respectively. The word list and extended word list contain word pairs with similar pronunciations. The word classification accuracy of this work indicated in the table is the result for the drone control words. The Extended list is the word list containing 54 words. Skin between ramus and temporal (SRT); Temporomandibular joint (TMJ); Electromyography (EMG); Electropalatography (EPG); Linear discriminant analysis (LDA); Support vector machine (SVM); Dynamic time regularization (DTW); Wavelet packet tree (WPT); Convolutional neural networks (CNN); Deep neural network (DNN); Recurrent neural network (RNN); Gaussian mixture models (GMM); Long short-term memory networks (LSTM); Connectionist temporal classification

(CTC); Frequency shift detection model (FSDM); Hidden Markov model (HMM); Bidirectional long-short term memory networks (BiLSTM); Random forests (RF). k-nearest neighbors algorithm (k-NN)

Table S2. Selected words with similar pronunciations under each viseme group including 'bilabial', 'alveolar', 'velar', 'labiodental', 'palato-alveolar', 'dental', 'retracting-spread', 'spread', 'neutral', 'protruding-rounded', 'rounded', and 'closed'. Detailed viseme pictures can be seen in references.[9, 24]

Bilabial		Alveolar		Velar	
Pay	**B**ay	**T**ea	**S**ea	**K**ay	**G**ay
Labiodental		**Palato-alveolar**		**Dental**	
Fan	**V**an	**Ch**oke	**J**oke	**Th**in	**Th**en
Retracting-spread		**Spread**		**Neutral**	
Way		Sh**ee**p	Sh**i**p	B**i**te	B**u**t
Protruding-rounded		**Rounded**		**Closed**	
B**ir**d		B**oo**k	B**oa**t	#	

Table S3. Silent speech recognition accuracies for five subjects.

Subject No.	Accuracies for nine phonemes	Accuracies for a list of words containing similar words
1	92.7%	85.6%
2	92.4%	88.1%
3	94.4%	89.7%
4	89.6%	84.5%
5	96.7%	88.4%
Average	93.2%	87.3%
Standard Deviation	2.62%	2.14%

Table S4. Silent speech recognition accuracy of the first subject using models trained by various data sets.

Training data set	Testing data set	Recognition accuracy
Subject 2	Subject 1	3.7%
Subjects 2, 3	Subject 1	8.2%
Subjects 2, 3, 4	Subject 1	6.2%
Subjects 2, 3, 4, 5	Subject 1	9.7%

Table S5. Features for the classification.

Features	Formulas
	Time domain
Mean ($\bar{x}$)	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Standard deviation (σ)	$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
Max	Max value
Min	Min value
25th percentile (P_{25})	$P_{25} = value\ of\ \frac{25}{100}(n+1)th\ item$
50th percentile (P_{50})	$P_{50} = value\ of\ \frac{50}{100}(n+1)th\ item$
75th percentile (P_{75})	$P_{75} = value\ of\ \frac{75}{100}(n+1)th\ item$
Skew ($\tilde{\mu}_3$)	$\tilde{\mu}_3 = \frac{\sum_i^n (x_i - \bar{x})^3}{(n-1)\sigma^3}$
Kurtosis ($Kurt$)	$Kurt = \frac{\mu_4}{\sigma^4}$
Root mean square (RMS)	$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2}$
Covariance of x and y directions ($cov_{x,y}$)	$cov_{x,y} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$
Covariance of y and z directions ($cov_{y,z}$)	$cov_{y,z} = \frac{\sum_i^n (y_i - \bar{y})(z_i - \bar{z})}{n-1}$
Covariance of x and z directions ($cov_{x,z}$)	$cov_{x,z} = \frac{\sum_i^n (x_i - \bar{x})(z_i - \bar{z})}{n-1}$
	Frequency domain
The real part of the first eight coefficients of the discrete Fourier transform (X_k)	$X_k = \sum_{i=1}^{n-1} x_i e^{-\frac{j2\pi}{n}ki}, where\ k = 0, \cdots, 7$

Note: n represents the number of data points. x_i represents the value of each data point. In the covariance calculation, x_i, y_i, and z_i represent values of data points in x, y, and z directions respectively. j represents the imaginary unit $\sqrt{-1}$. k represents the frequency index.

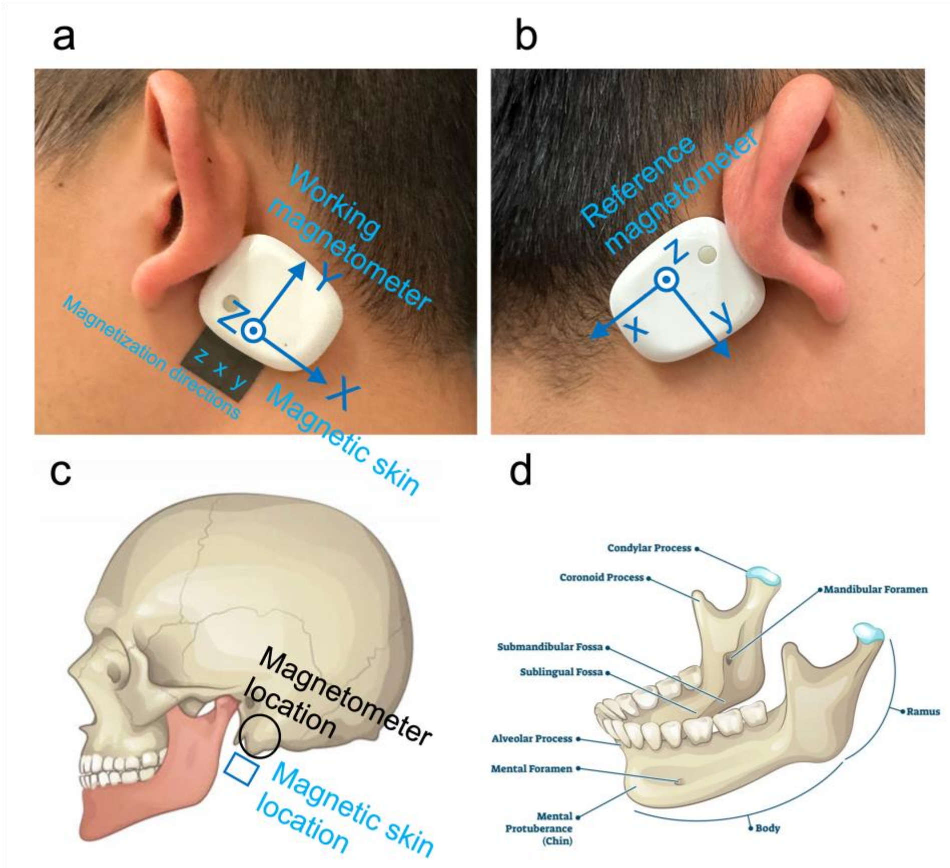


Fig. S1. Placement of the silent speech interface. a) Position of the working magnetometer and the magnetic skin. b) Position of the reference magnetometer. c) Structure of the human head skeleton and the position of the magnetometer and the magnetic skin. d) Structure of the mandible. Fig. S1c and Fig. S1d were reproduced from the web page[25] with permission.

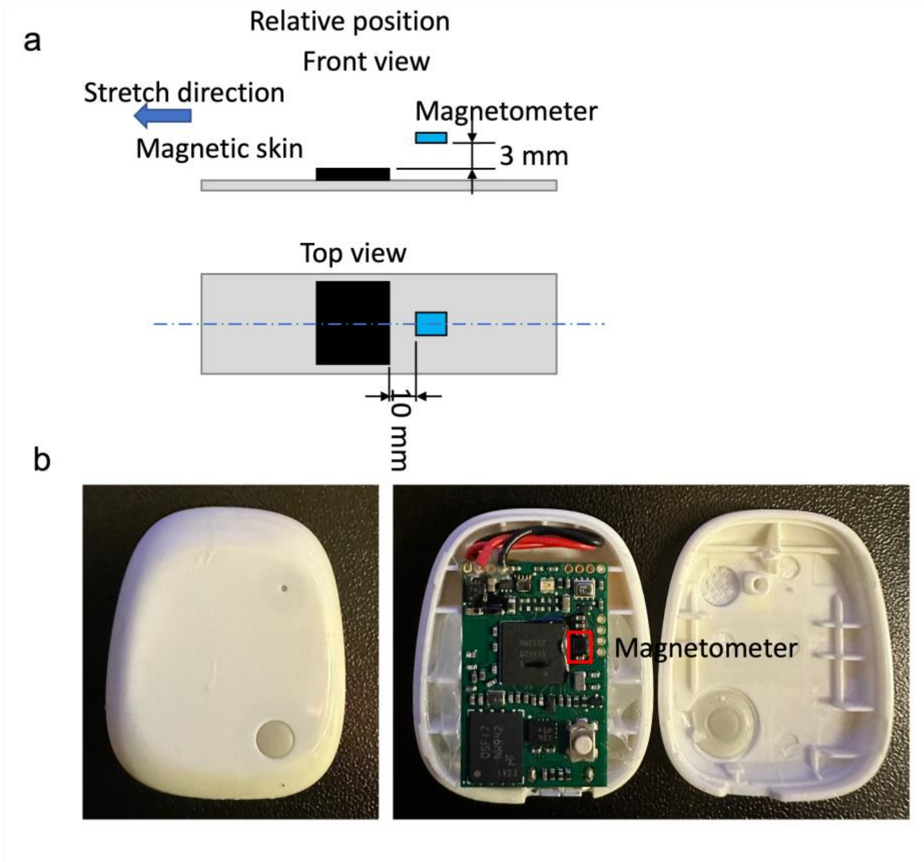


Fig. S2. Setup for measuring magnetic flux density changes under strain. a) Illustration of the relative position of each part corresponding to Fig. 3l. Not true to scale. b) Photographs of the complete wireless data acquisition system and the magnetometer inside.

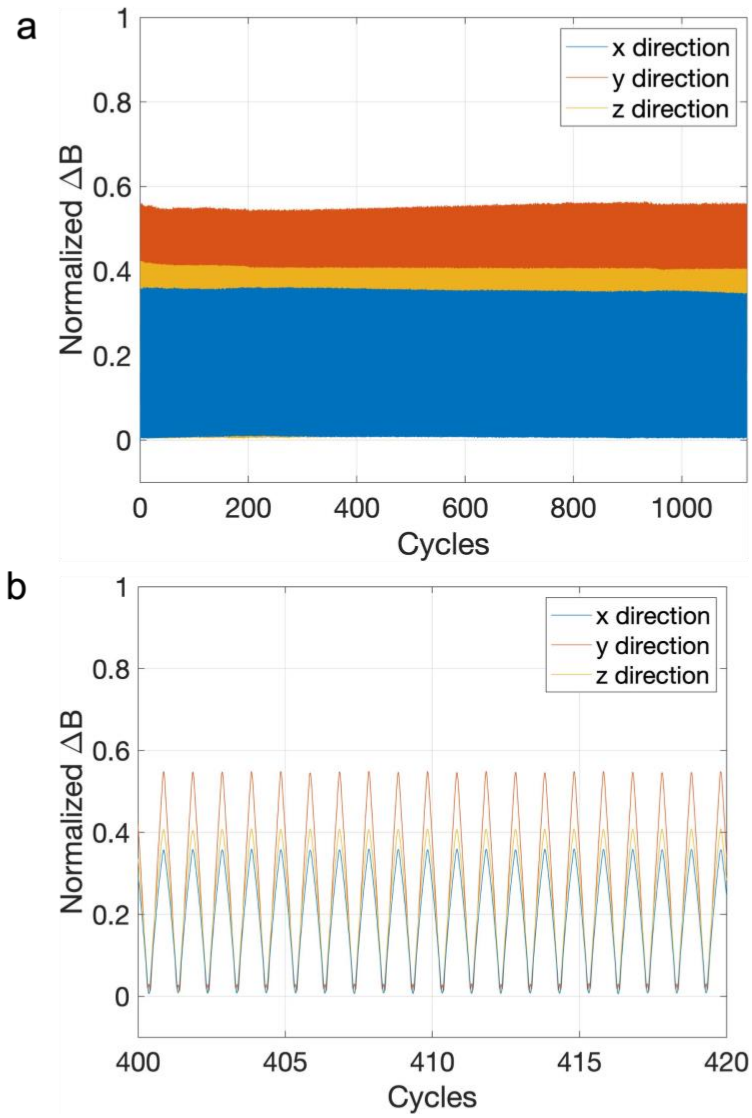


Fig. S3. Repeatability and reliability of the magnetic skin. a) Magnetic flux density changes along three directions during repeated stretching/releasing cycles. B) Results for the 400^{th} – 420^{th} stretching/releasing cycles as examples.

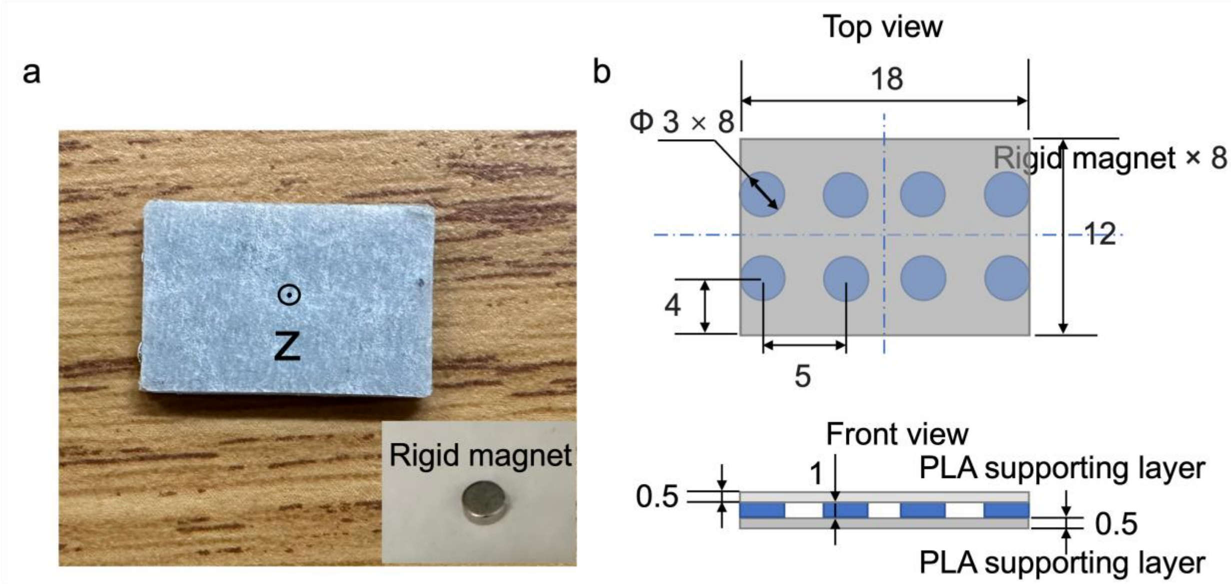
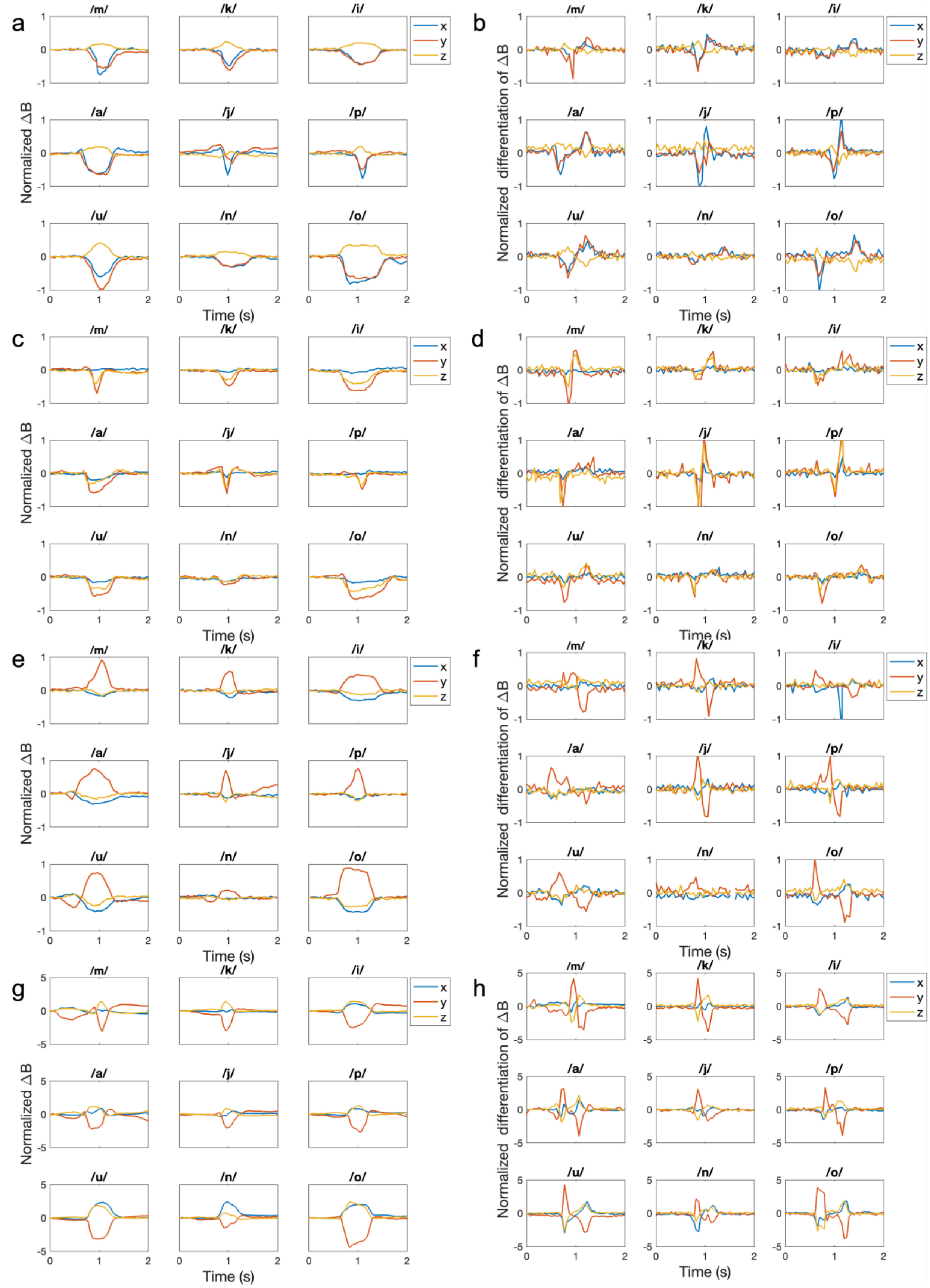


Fig. S4. Photographs of the rigid magnet. a) Photograph of eight rigid magnets (shown in the insert) attached to the PLA support. b) Dimension and inner structure.



a
b
c
d
e
f
g
h
/m/
/k/
/i/
/a/
/j/
/p/
/u/
/n/
/o/
x
y
z
Normalized ΔB
Normalized differentiation of ΔB
Time (s)

Fig. S5. Time series signals and signals after differentiation for subject 1. Time series signals of the nine phonemes from a) XXX sample, c) YYY sample, e) ZZZ sample, and g) rigid magnet. The signals are normalized by dividing all values by 30 μT. Time-series signals of the nine phonemes after differentiation from b) XXX sample, d) YYY sample, f) ZZZ sample, and h) rigid magnet. The signals after differentiation are normalized by dividing all values by 150 μT/s.

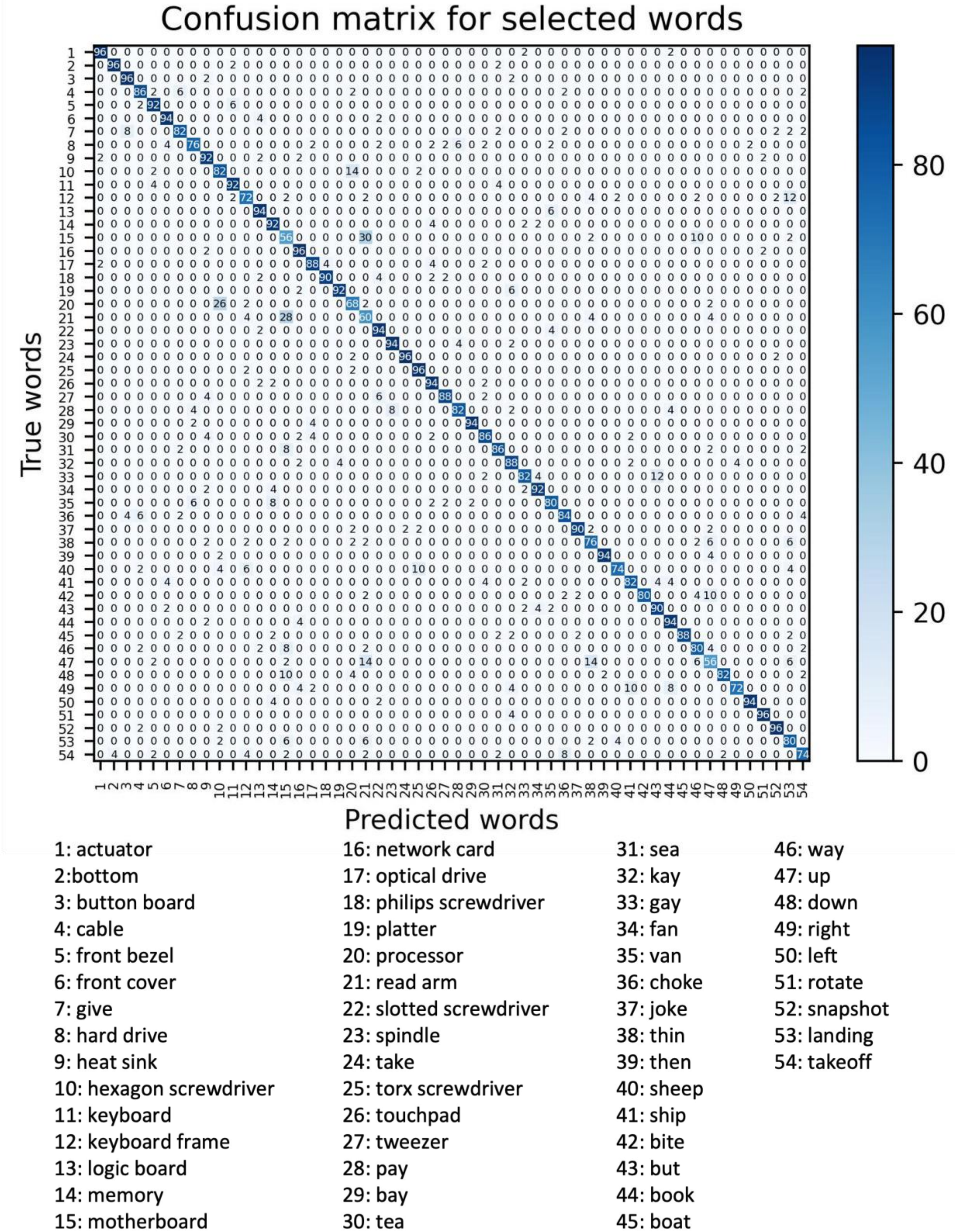


Fig. S6. Confusion matrix of the extended word list for the first subject.

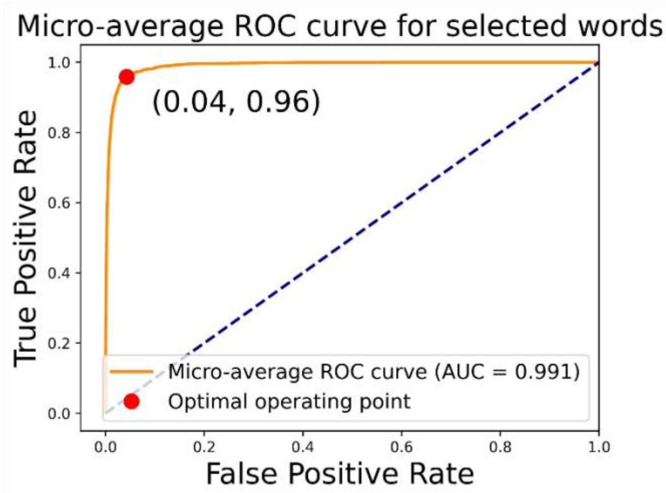


Fig. S7. Micro-average ROC curve of the extended word list for the first subject

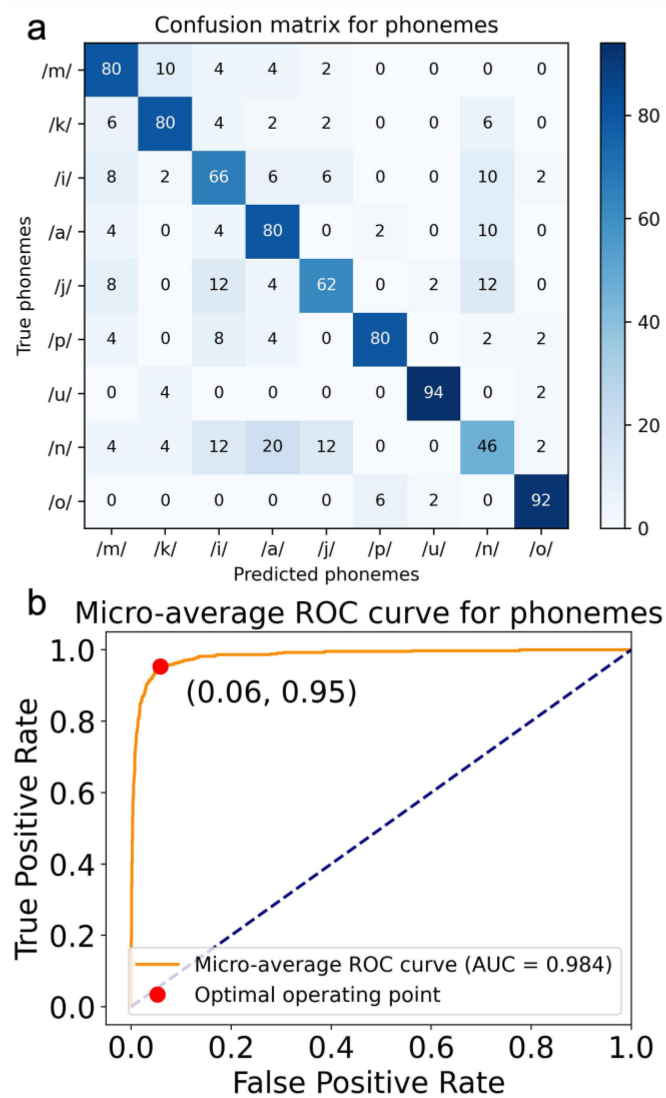


Fig. S8. Classification results for the first subject using the rigid magnet. Confusion matrix (a) and micro-average ROC curve (b) for nine phonemes.

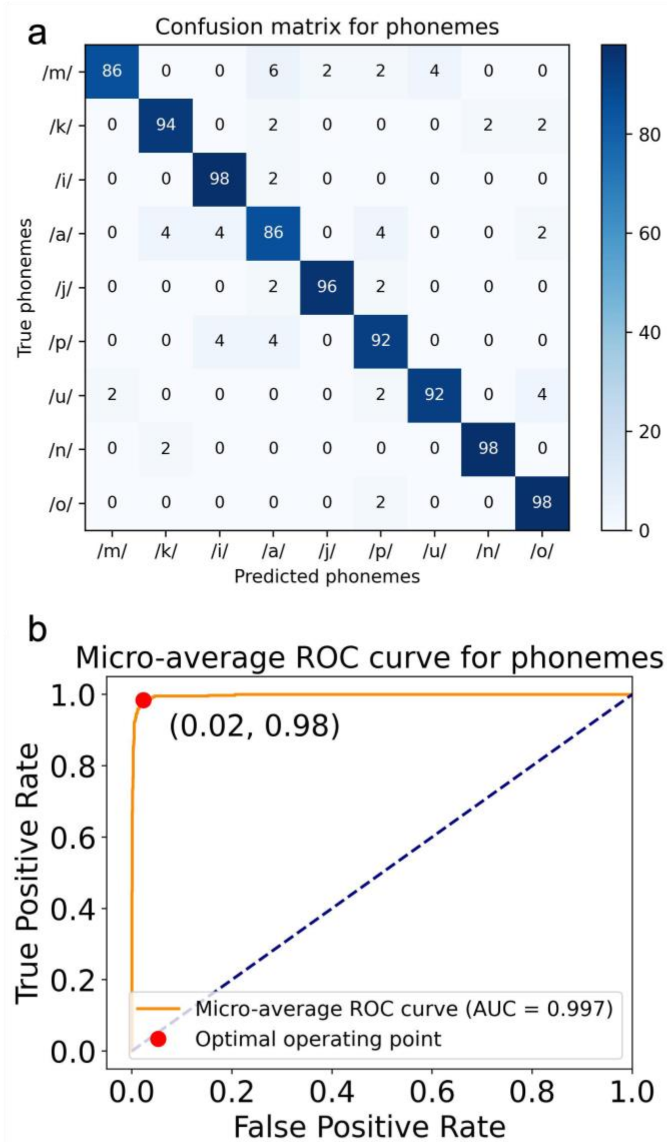


Fig. S9. Classification results for the first subject under noisy environments. Confusion matrix (a) and micro-average ROC curve (b) for nine phonemes.

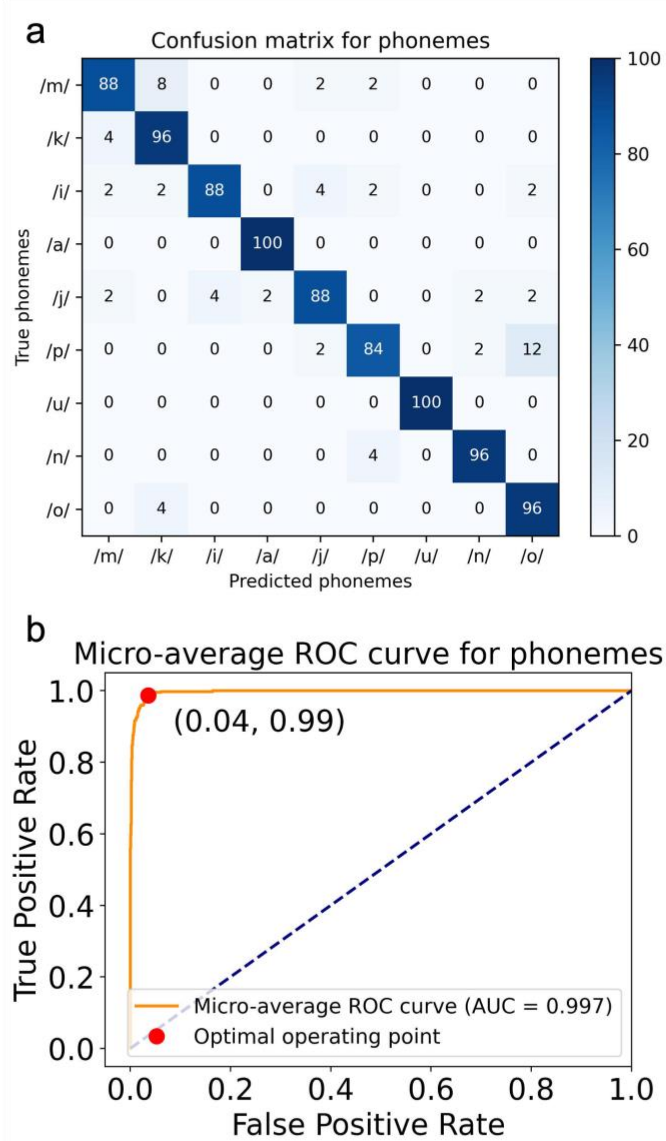


Fig. S10. Classification results for the first subject under dark environments. Confusion matrix (a) and micro-average ROC curve (b) for nine phonemes.

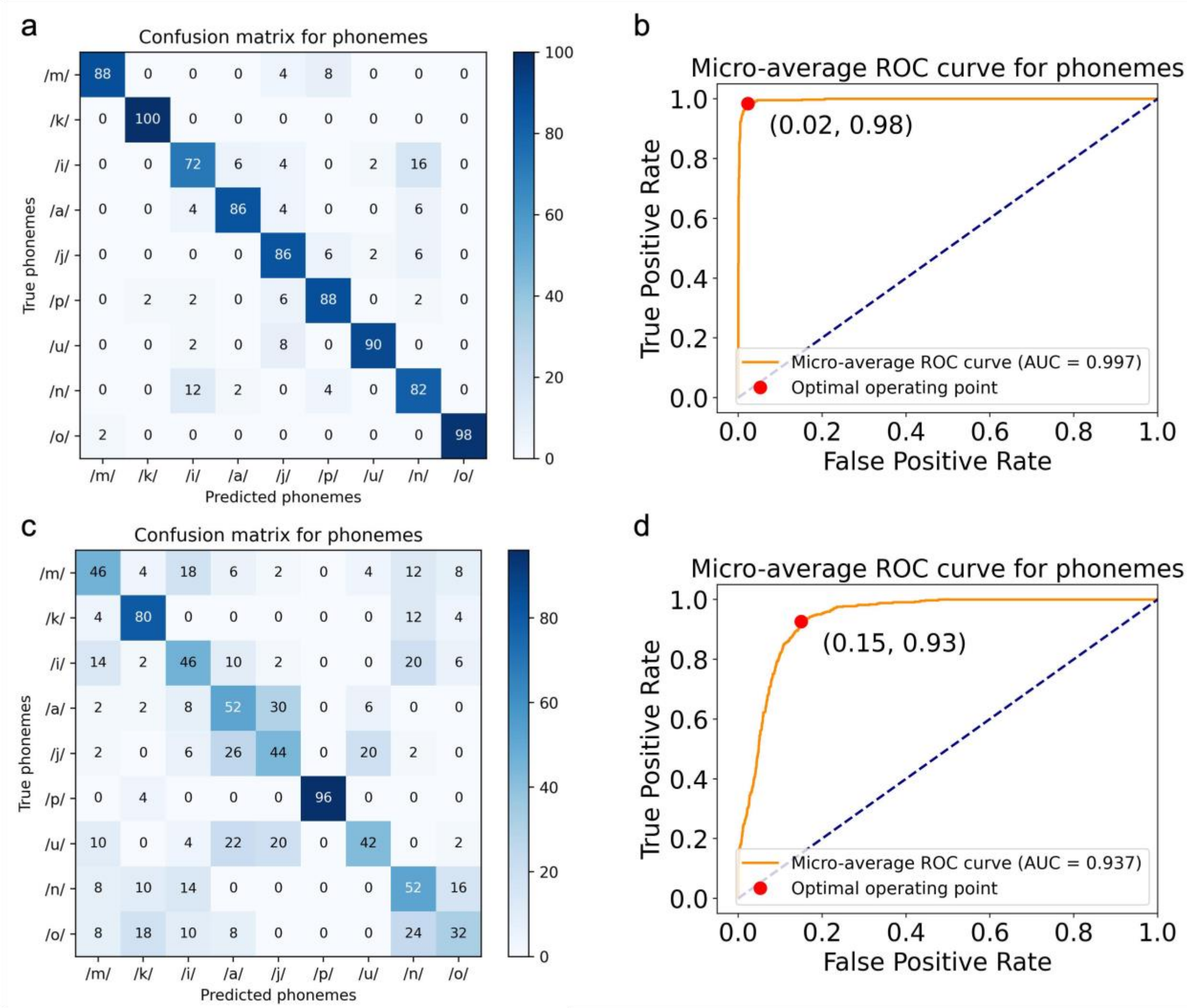


Fig. S11. Classification results of signals before and after calibration for the first subject during walking. Confusion matrix (a) and micro-average ROC curve (b) for signals corresponding to nine phonemes after calibration. Confusion matrix (c) and micro-average ROC curve (d) for signals corresponding to nine phonemes before calibration.

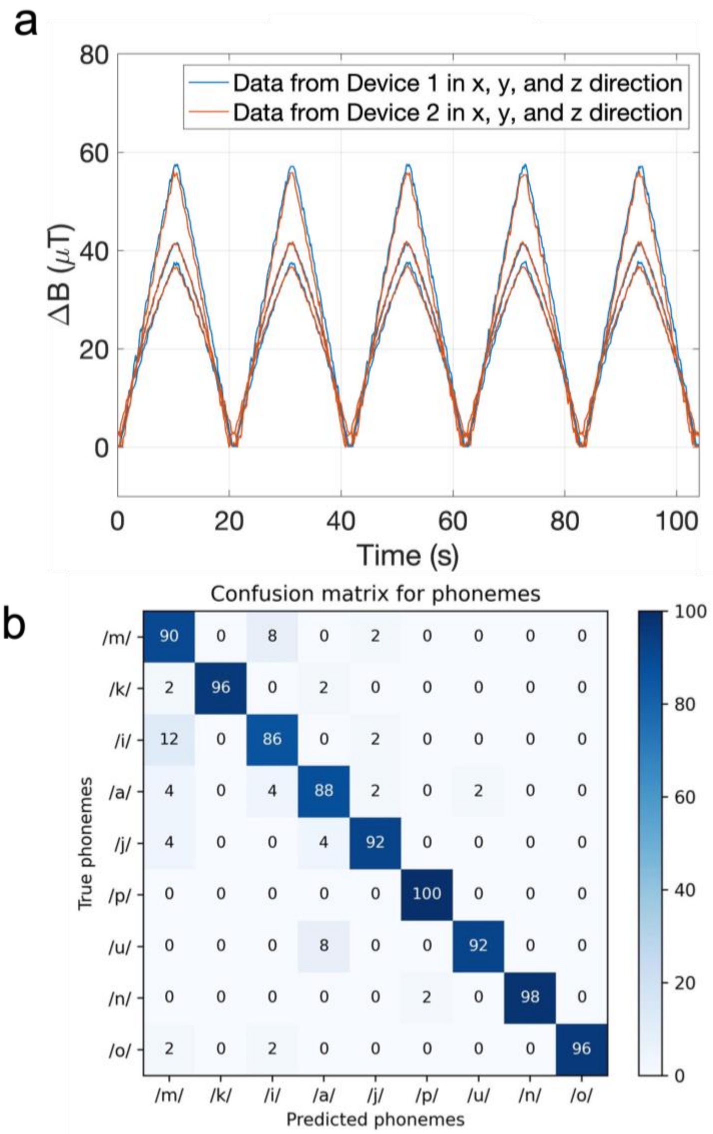


Fig. S12. Device variations and the influence on the recognition accuracy. a) Magnetic flux density changes during stretching/releasing cycles obtained using two different devices. b) Confusion matrix for the first subject when adding new data training set acquired from different devices.

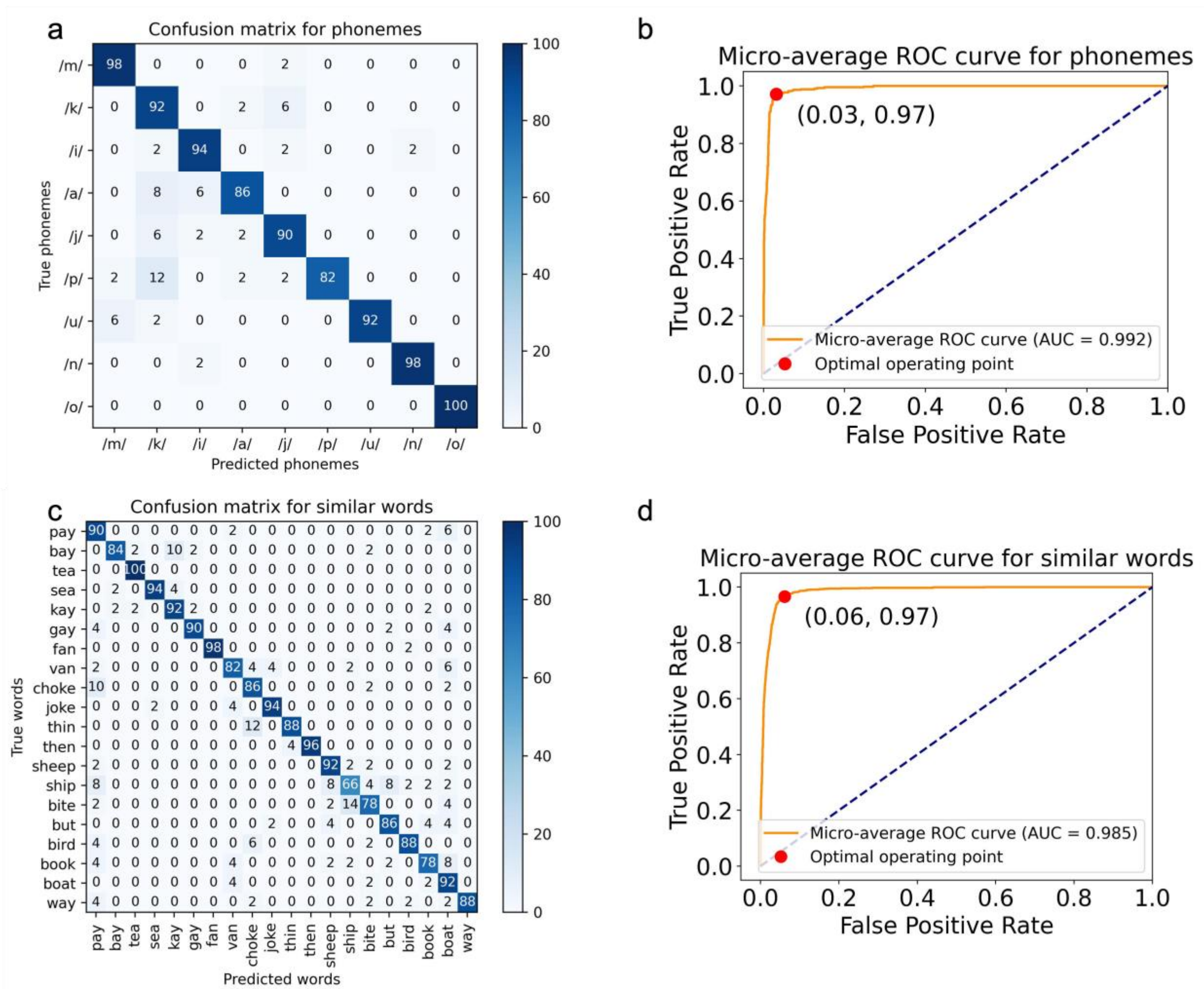


Fig. S13. Classification results for the second subject. Confusion matrix (a) and micro-average ROC curve (b) for nine phonemes. Confusion matrix (c) and micro-average ROC curve (d) for a list of words containing word pairs with similar pronunciations (from the same viseme group).

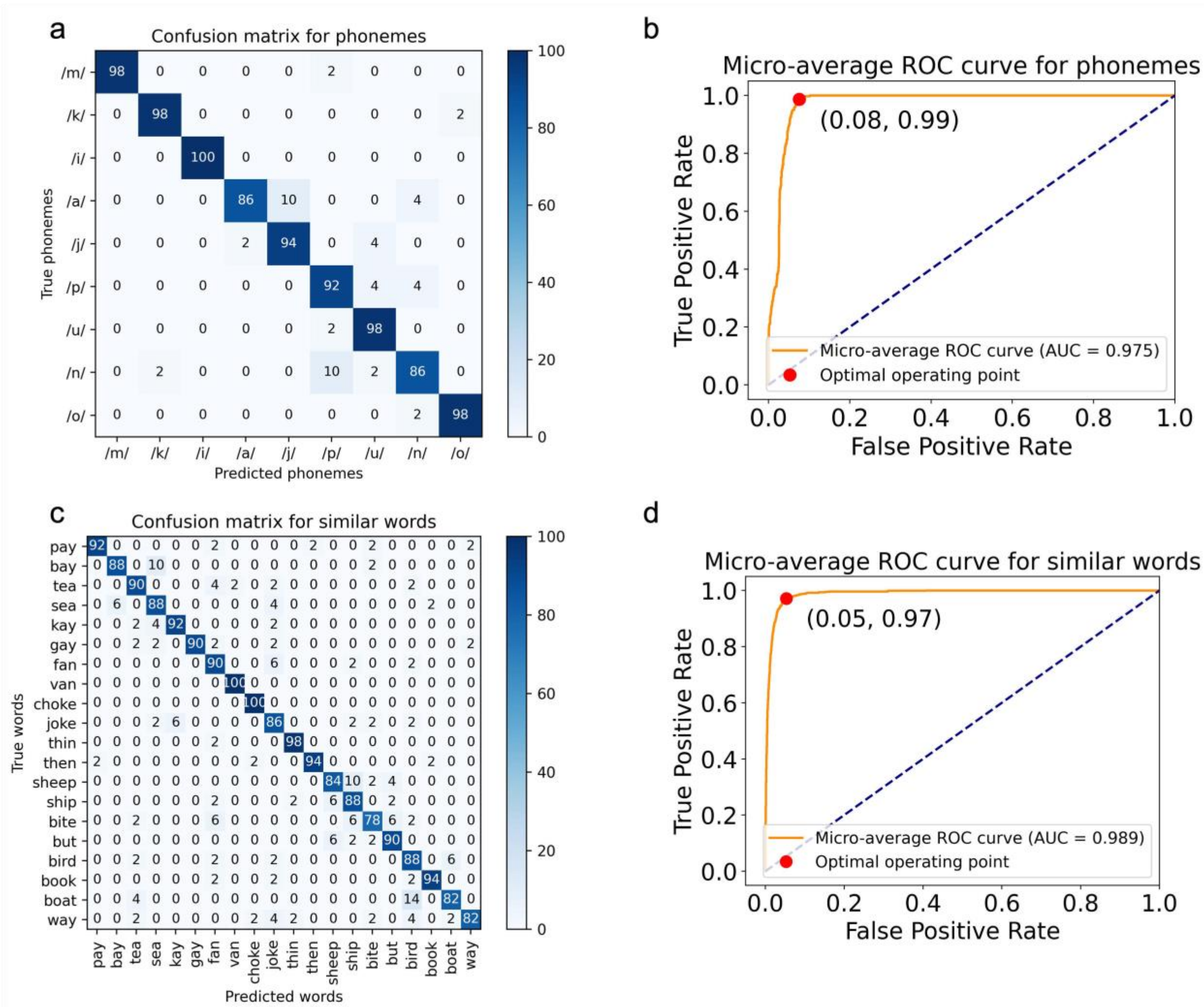


Fig. S14. Classification results for the third subject. Confusion matrix (a) and micro-average ROC curve (b) for nine phonemes. Confusion matrix (c) and micro-average ROC curve (d) for a list of words containing word pairs with similar pronunciations (from the same viseme group).

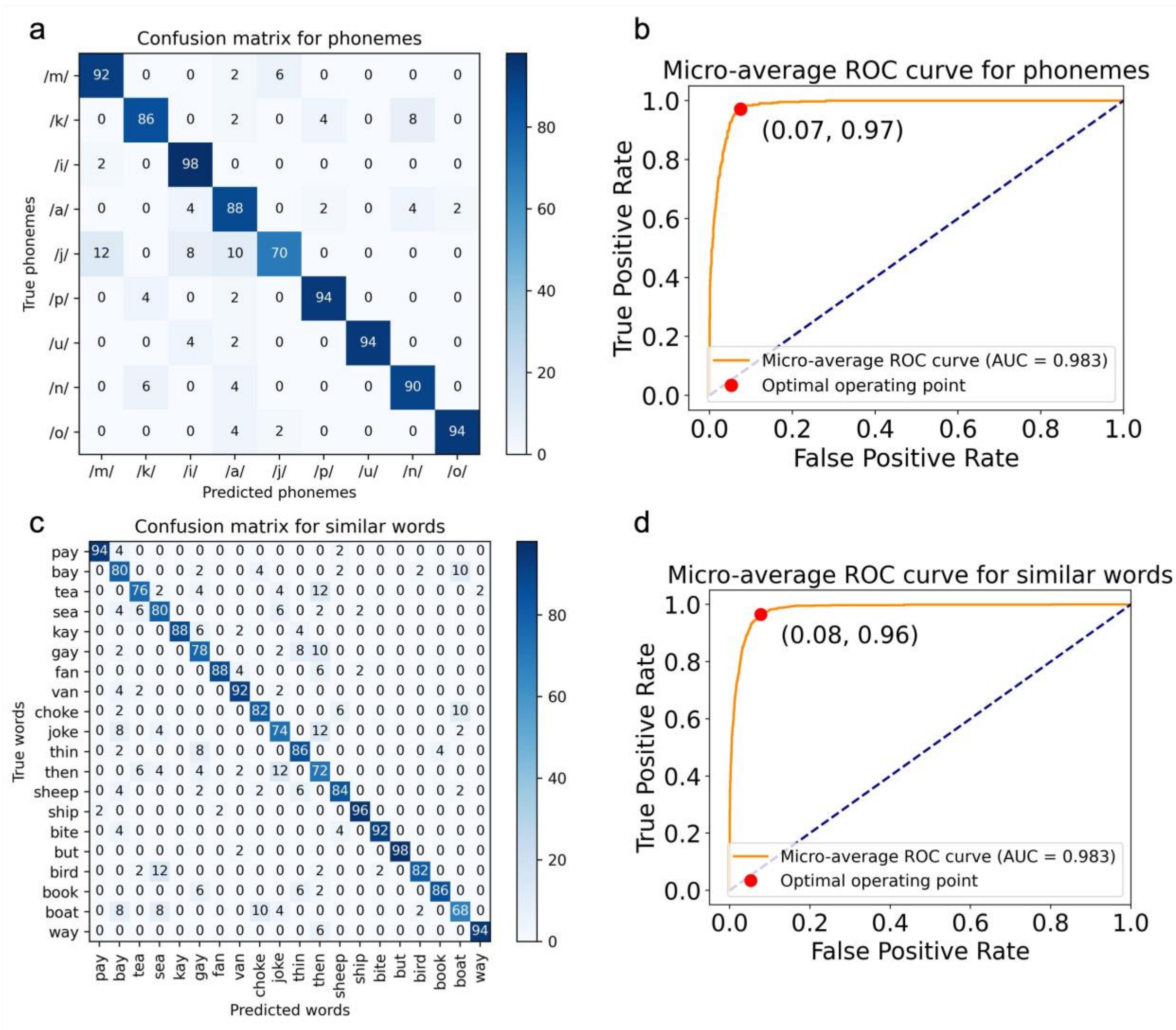


Fig. S15. Classification results for the fourth subject. Confusion matrix (a) and micro-average ROC curve (b) for nine phonemes. Confusion matrix (c) and micro-average ROC curve (d) for a list of words containing word pairs with similar pronunciations (from the same viseme group).

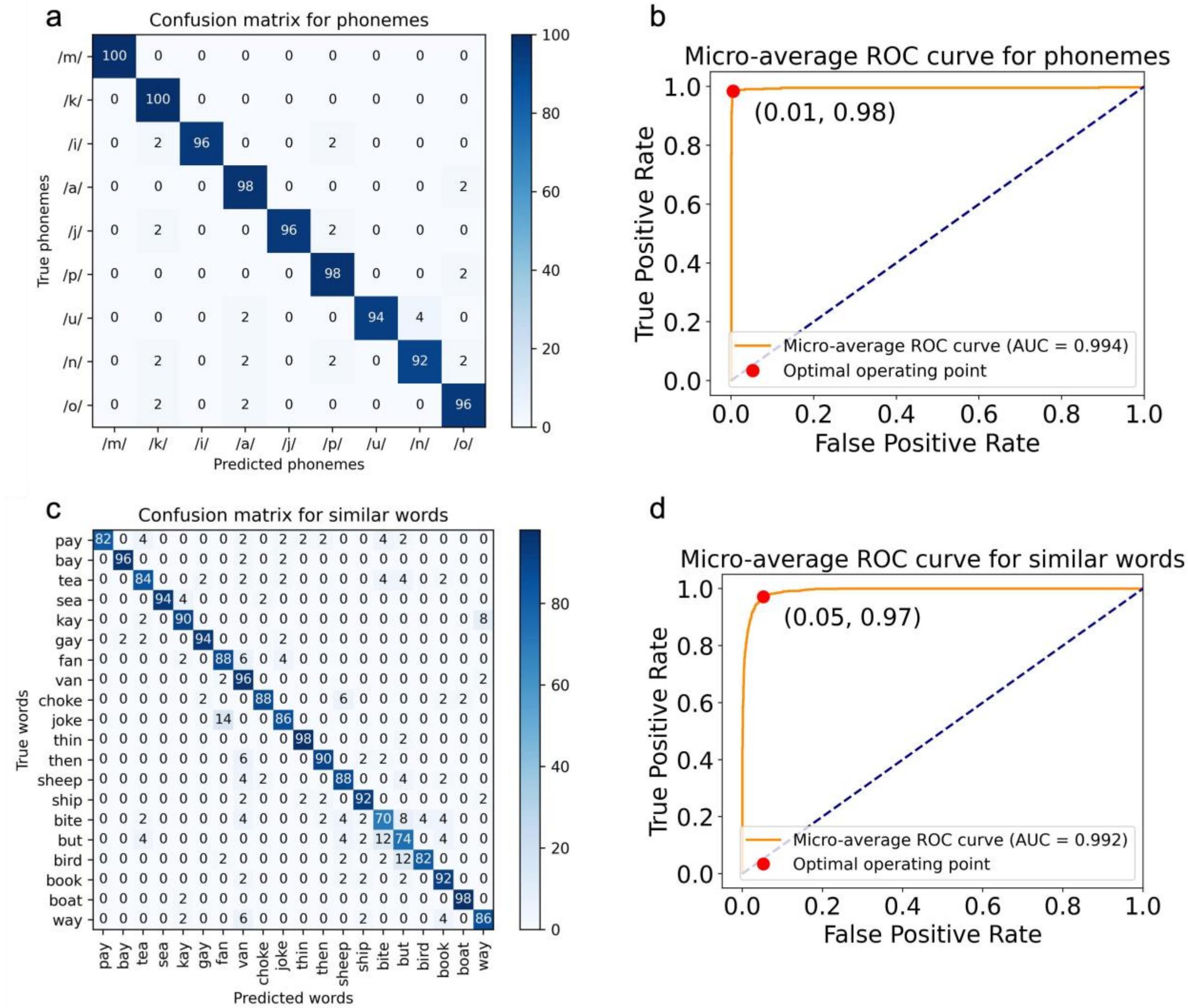


Fig. S16. Classification results for the fifth subject. Confusion matrix (a) and micro-average ROC curve (b) for nine phonemes. Confusion matrix (c) and micro-average ROC curve (d) for a list of words containing word pairs with similar pronunciations (from the same viseme group).

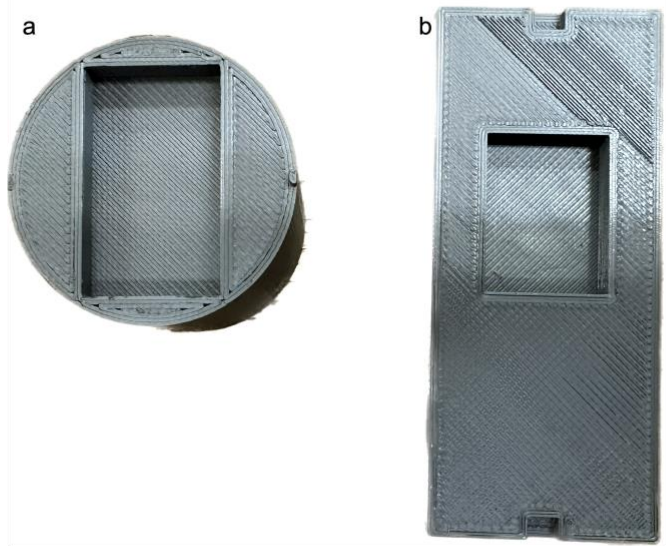


Fig. S17. 3D-printed molds for magnetization. a) Mold for the magnetization in the z-direction. b) Mold for the magnetization in the x and y directions.

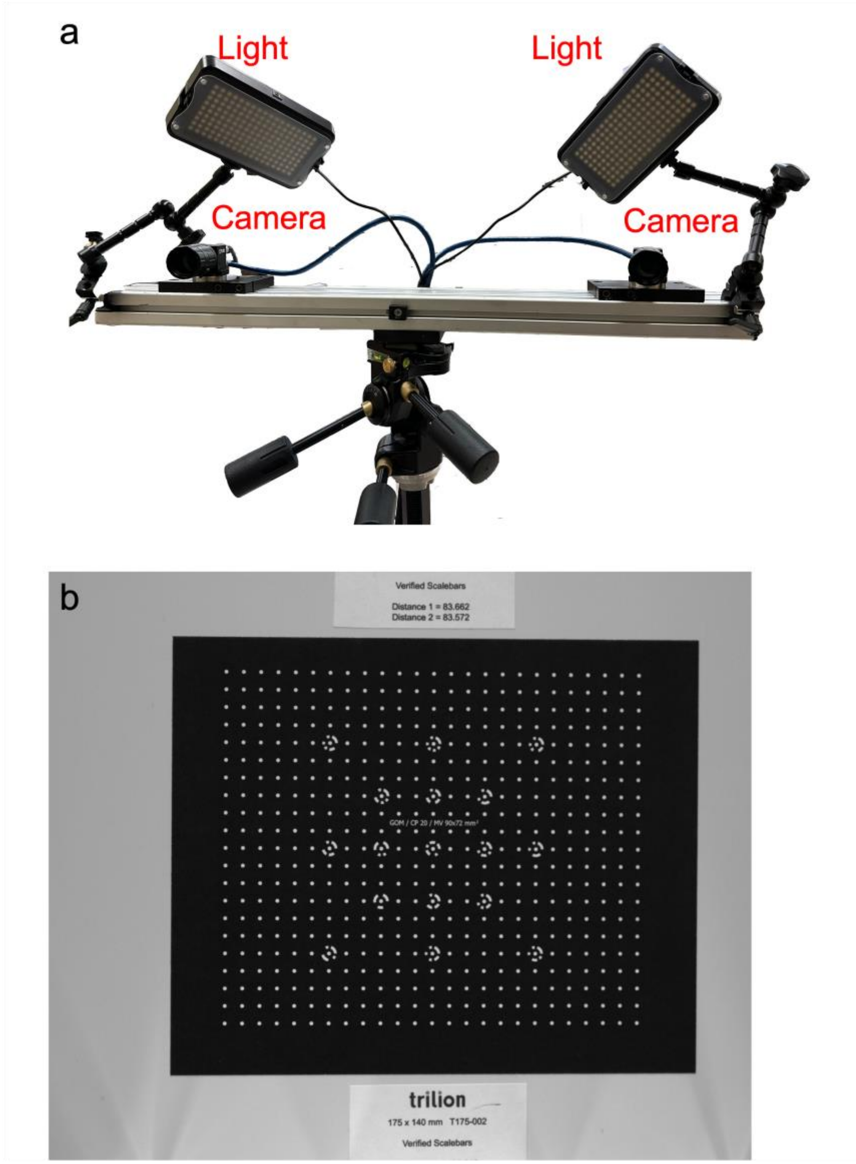


Fig. S18. DIC system. a) Camera setup. b) Calibration pad.

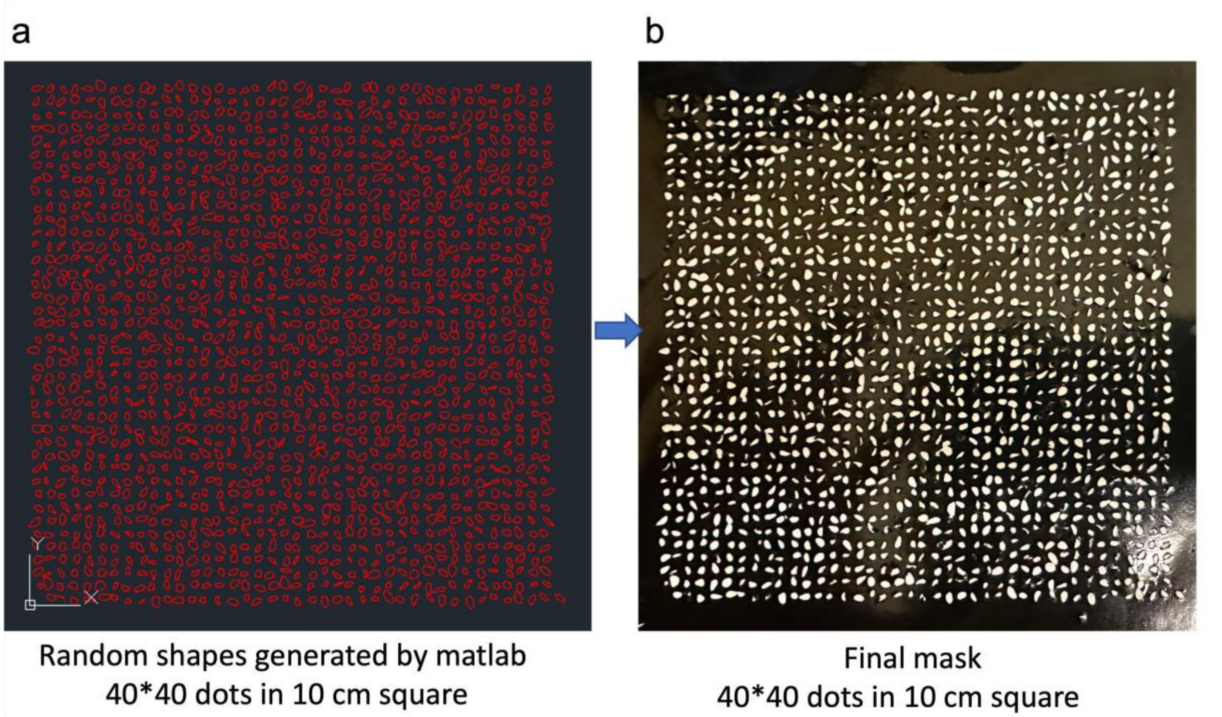


Fig. S19. Mask with randomly shaped patterns for the DIC measurement. a) Patterns in DXF format generated by MATLAB. b) Photograph of the final mask.

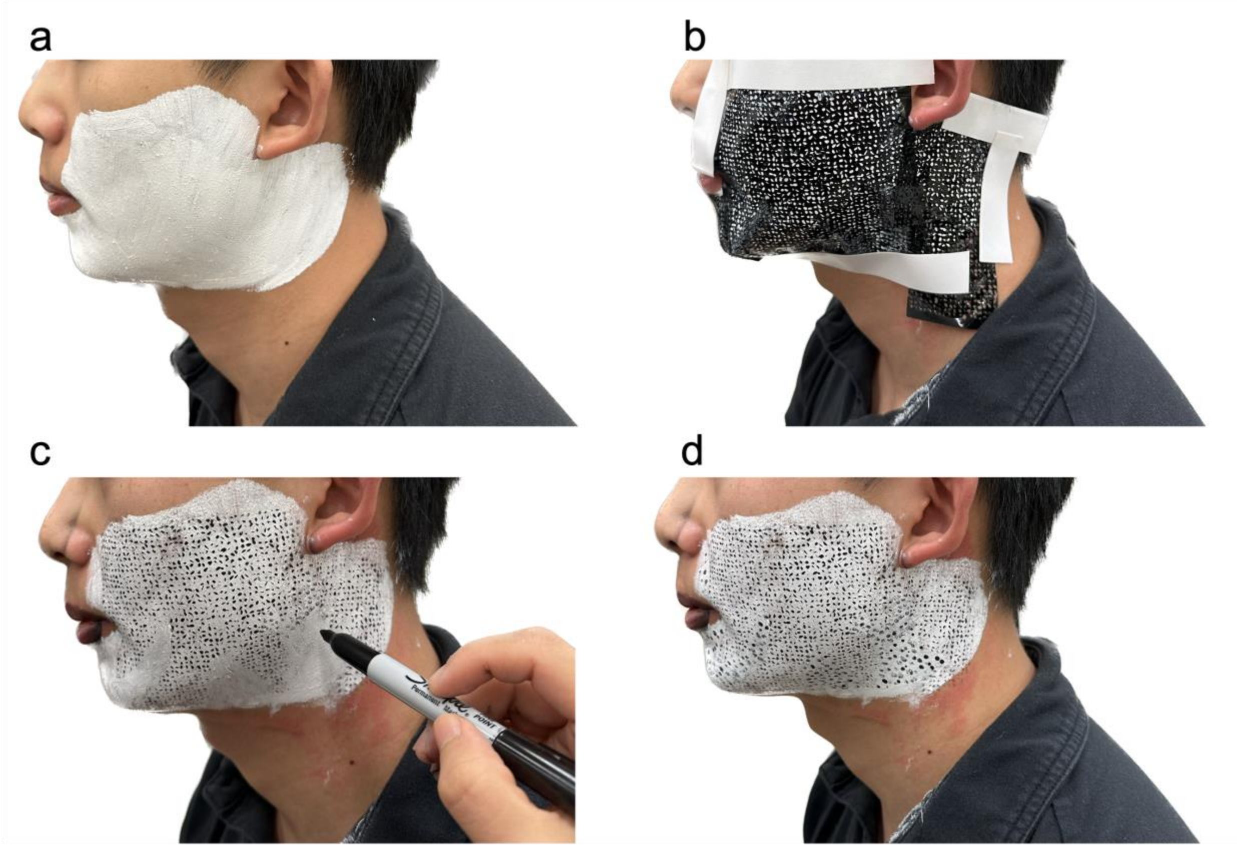


Fig. S20. Process of painting randomly shaped dots on the facial skin. a) Painting clown white on the facial skin. b) Attaching the mask to the clown white and spray-coating black paint. c) Using a marker pen to add dots and fill in the remaining blank areas. d) Final appearance.

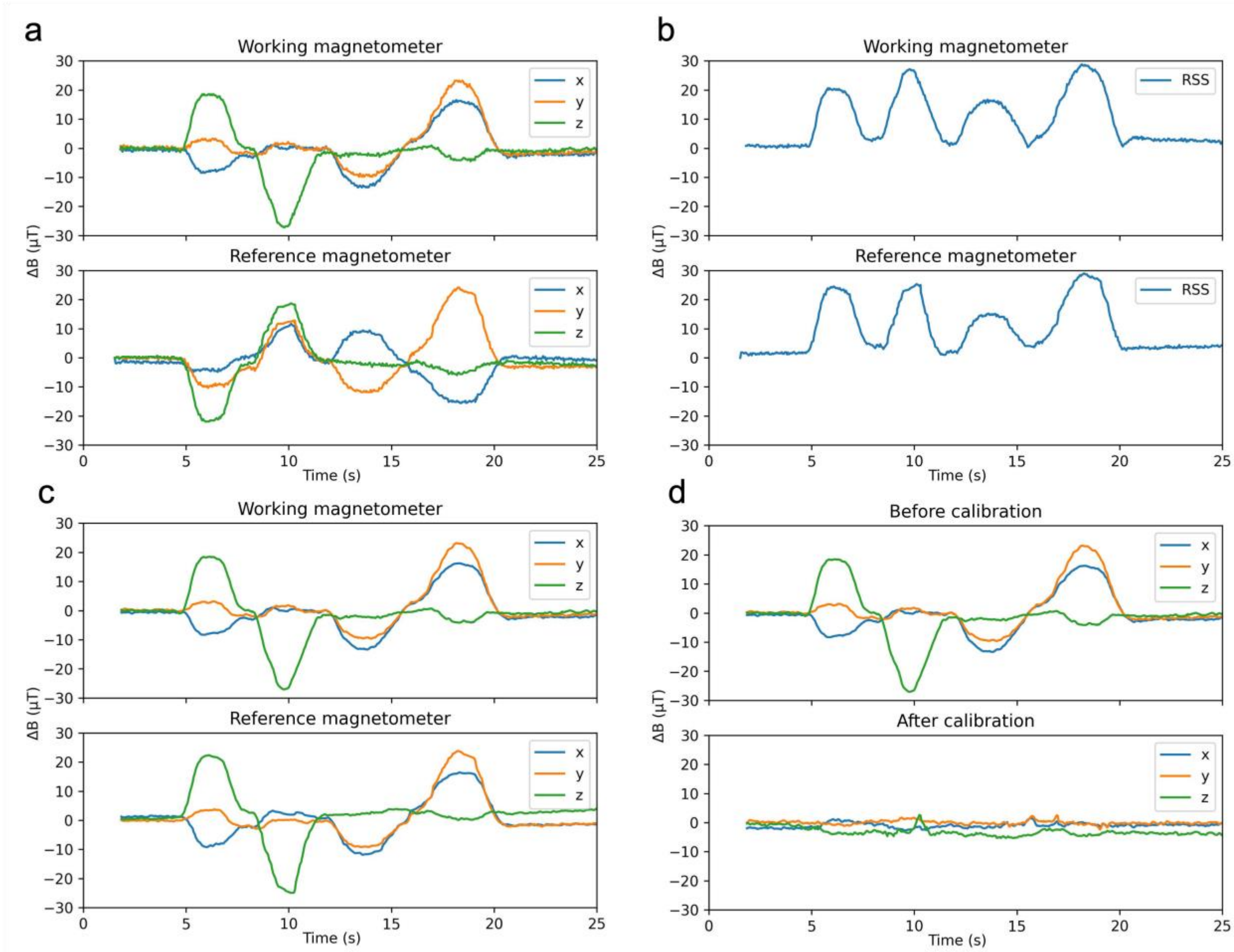


Fig. S21. Process of finding the rotation matrix between the working magnetometer coordinate and the reference magnetometer coordinate. a) Acquired signals during the subject's motion. b) Root sum square of the signal. c) Signals after a coordinate transformation. d) Signals before and after calibration.

References

1. Wikipedia, Kabsch algorithm, https://en.wikipedia.org/wiki/Kabsch_algorithm, (accessed May, 2023).
2. S. Xu, J. X. Yu, H. Guo, S. Tian, Y. Long, J. Yang and L. Zhang, *Nat. Commun.*, 2023, **14**, 219.
3. Z. Zhou, K. Chen, X. Li, S. Zhang, Y. Wu, Y. Zhou, K. Meng, C. Sun, Q. He, W. Fan, E. Fan, Z. Lin, X. Tan, W. Deng, J. Yang and J. Chen, *Nat. Electron.*, 2020, **3**, 571-578.
4. H. Yoo, E. Kim, J. W. Chung, H. Cho, S. Jeong, H. Kim, D. Jang, H. Kim, J. Yoon, G. H. Lee, H. Kang, J. Y. Kim, Y. Yun, S. Yoon and Y. Hong, *ACS Appl. Mater. Interfaces*, 2022, **14**, 54157-54169.
5. J. Wang, C. Pan, H. Jin, V. Singh, Y. Jain, J. I. Hong, C. Majidi and S. Kumar, *Proc. ACM interact. mob. wearable ubiquitous technol.*, 2019, **3**, 1-24.
6. L. Cheng, D. Q. Ruan, Y. W. He, J. Y. Yang, W. Qian, L. W. Zhu, P. D. Zhu, H. P. Wu and A. P. Liu, *J. Mater. Chem. C*, 2023, **1**, 1.
7. T. Kim, Y. Shin, K. Kang, K. Kim, G. Kim, Y. Byeon, H. Kim, Y. Gao, J. R. Lee, G. Son, T. Kim, Y. Jun, J. Kim, J. Lee, S. Um, Y. Kwon, B. G. Son, M. Cho, M. Sang, J. Shin, K. Kim, J. Suh, H. Choi, S. Hong, H. Cheng, H. G. Kang, D. Hwang and K. J. Yu, *Nat. Commun.*, 2022, **13**, 5815.
8. Y. Kunimi, M. Ogata, H. Hiraki, M. Itagaki, S. Kanazawa and M. Mochimaru, Proceedings of Augmented Humans 2022, Kashiwa, Japan, 2022.
9. P. Dong, Y. Song, S. Yu, Z. Zhang, S. K. Mallipattu, P. M. Djuric and S. Yao, *Small*, 2023, **19**, e2205058.
10. Y. H. Wang, T. Y. Tang, Y. Xu, Y. Z. Bai, L. Yin, G. Li, H. M. Zhang, H. C. Liu and Y. A. Huang, *Npj Flex. Electron.*, 2021, **5**, 1-9.
11. H. Liu, W. Dong, Y. Li, F. Li, J. Geng, M. Zhu, T. Chen, H. Zhang, L. Sun and C. Lee, *Microsyst. Nanoeng.*, 2020, **6**, 16.
12. Y. Lu, H. Tian, J. Cheng, F. Zhu, B. Liu, S. Wei, L. Ji and Z. L. Wang, *Nat. Commun.*, 2022, **13**, 1401.
13. N. Kimura, T. Gemicioglu, J. Womack, R. Li, Y. Zhao, A. Bedri, Z. Su, A. Olwal, J. Rekimoto and T. Starner, Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, USA, 2022.

14. Y. Jin, Y. Gao, X. Xu, S. Choi, J. Li, F. Liu, Z. Li and Z. Jin, *Proc. ACM interact. mob. wearable ubiquitous technol.*, 2022, **6**, 1-28.
15. T. Srivastava, P. Khanna, S. Pan, P. Nguyen and S. Jain, *Proc. ACM interact. mob. wearable ubiquitous technol.*, 2022, **6**, 1-26.
16. P. Khanna, T. Srivastava, S. Pan, S. Jain and P. Nguyen, Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications, Virtual, United Kingdom, 2021.
17. A. Kapur, S. Kapur and P. Maes, Proceeding of 23rd International Conference on Intelligent User Interfaces, Tokyo, Japan, 2018.
18. G. S. Meltzner, J. T. Heaton, Y. Deng, G. De Luca, S. H. Roy and J. C. Kline, *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2017, **25**, 2386-2398.
19. J. A. Gonzalez, L. A. Cheah, A. M. Gomez, P. D. Green, J. M. Gilbert, S. R. Ell, R. K. Moore and E. Holdsworth, *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2017, **25**, 2362-2374.
20. R. Hofe, S. R. Ell, M. J. Fagan, J. M. Gilbert, P. D. Green, R. K. Moore and S. I. Rybchenko, *Speech Commun.*, 2013, **55**, 22-32.
21. H. Sahni, A. Bedri, G. Reyes, P. Thukral, Z. Guo, T. Starner and M. Ghovanloo, Proceedings of the 2014 ACM International Symposium on Wearable Computers, Seattle, USA, 2014.
22. K. Brigham and B. V. Kumar, Proceedings of the 2010 4th International Conference on Bioinformatics and Biomedical Engineering, Chengdu, China, 2010.
23. P. Suppes, Z. L. Lu and B. Han, *PNAS*, 1997, **94**, 14965-14969.
24. D. Jachimski, A. Czyzewski and T. Ciszewski, *Multimed. Tools. Appl.*, 2017, **77**, 16495-16532.
25. V. Ayaga, The Mandible (Lower Jaw), https://www.newmouth.com/oral-health/mandible/, (accessed May, 2023).