Electronic Supplementary Information

Breath odor-based individual authentication by an artificial olfactory sensor system and machine learning

Chaiyanut Jirayupat,[1,2] Kazuki Nagashima,[1,3]* Takuro Hosomi,[1,3] Tsunaki Takahashi,[1,3] Benjarong Samransuksamer,[1] Yosuke Hanai,[4] Atsuo Nakao,[4] Masaya Nakatani,[4] Jiangyang Liu,[1] Guozhu Zhang,[1] Wataru Tanaka,[1] Masaki Kanai,[5] Takao Yasui,[3,6] Yoshinobu Baba,[6] and Takeshi Yanagida[1,2,5]*

[1] *Department of Applied Chemistry, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

[2] *Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, 6-1 Kasuga-Koen, Kasuga, Fukuoka, 816-8580, Japan*

[3] *PRESTO, Japan Science and Technology Agency, 4-1-8, Honcho, Kawaguchi-shi, Saitama 332-0012 Japan*

[4] *Panasonic Corporation, Industry Company, Sensing Solutions Development Center, Kadoma 1006, Kadoma, Osaka 571-8506, Japan*

[5] *Institute for Materials Chemistry and Engineering, Kyushu University, 6-1 Kasuga-Koen, Kasuga, Fukuoka 816-8580, Japan*

[6] *Department of Biomolecular Engineering, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan*

Experimental Details

Breath Odor Sample Preparation. The breath odor samples used in this study were collected from 6 healthy people (3 males and 3 females) with different nationalities (Thai, Chinese, Japanese) by using a 10 L gas sampling bag (Smart Bag PA CEK-10, GL Science Inc.) and from 20 healthy people (16 males and 4 females) with different nationalities (Thai, Chinese, Japanese) by using a 1 L gas sampling bag (Smart Bag PA AAK-1, GL Science Inc.). The five breath odor samples per person were collected from the tested subject at different timing (in the morning/afternoon) or in different day (at least three different days per each person). To exclude the influence of exogenous compounds originating from the diets and the tested environments, the breath samples were collected in the same room from the tested persons fasted for 6 h. The gas sampling bags filled with breath odor were stored for 24 h prior to the analyses, in order to stabilize the humidity condition inside the bag. For the breath component analysis, the gas sampling bag containing breath odor was connected to an adsorbent-filled sample tube (Packed Liner with Tenax GR, mesh 80/100 #2414-1021, GL Science Inc.) and 500 mL of breath odor was transferred to the sample tube with pumping at the rate of 50 mL/min. The sample tubes were sealed and stored in refrigerator at 4 ºC until conducting the gas chromatography–mass spectrometry (GC-MS) measurements.

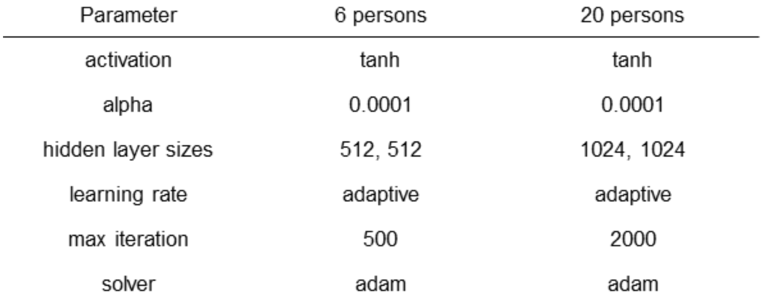
Breath Component Analysis by GC-MS. Component analysis of the collected breath odor samples were conducted by GC-MS (Shimadzu, GCMS-QP2020) equipped with inlet temperature control unit (OPTIC4). For the GC-MS measurements, the collected chemical compounds in the sample tube were desorbed by rapidly increasing the injection port temperature to 300 ºC with split-less mode. The oven temperature was kept at 40 ºC for 5 min, then increased to 280 ºC at the rate of 5 ºC/min, and kept at 280 ºC for 5 min. The capillary column of InertCap 5MS/NP (60 m length, 0.25 mm inner diameter, 1 μm thickness, GL Science Inc.) was used to separate the desorbed compounds prior to MS analysis. The column flow rate and the purge flow rate of helium gas (99.9999% pure) were set to be 1 mL/min and 5 mL/min, respectively. Both the ion source temperature and the interface temperature of mass-spectrometer were fixed at 200 ºC during the measurements. The characterized mass to charge ratio (*m/z*) in the range of 35-300. The obtained data was analyzed by GCMS Solution ver. 4.45 SP1. The 2D MS maps and the 2D feature score maps were obtained by using *NPFimg*, *i.e.* the data analysis program we developed recently.S1) For GC-MS analysis, 20 breath odor samples per person were collected from 3 persons at different timing or in different day.

Fabrication of Artificial Olfactory Sensor System. 16 types of GC stationary phase material (GCM)-carbon black (CB) nanocomposite were prepared and used for sensing materials. The details of fabrication procedure and its usage can be seen elsewhere.S2-S5) GCM-CB nanocomposites were made by mixing 10 mg carbon black (45μm Graphitized carbon black, Sigma) and 10 mg GC stationary phase materials (tetrahydroxyethylenediamine (THEED), GL Sciences/ *N*,*N*-Bis(2-cyanoethyl)formamide (BCEF), Tokyo Chemical Industry/LAC-3-R-728 (12% diethylene glycol succinate (DEGS), GL Sciences/DEGS, Supelco/ poly(ethylene succinate) (PES), Sigma/ UCON 75-H-90000, polyalkylene glycol (PAG) containing 75 wt% oxyethylene and 25 % oxypropylene groups, Shinwa Chemical Industries/1,2,3-Tris(2-cyanoethoxy)propane (TCEP), Supelco/SP-2330, poly (80% biscyanopropyl/20% cyanopropylphenyl siloxane), Supelco/SP-2340, poly (biscyanopropyl siloxane), Supelco/diglycerol, Tokyo Chemical Industry/Reoplex 400, GL Sciences/poly[di(ethylglycol)adipate] (PDEGA), Sigma/PEG4000, poly(ethyele glycol) 4000, Sigma/PEG20K, poly(ethyele glycol) 20000, United States Pharmacopeia (USP) Reference/PEG20M, poly(ethyele glycol) 20M, Shinwa Chemical Industries/free fatty acid phase (FFAP), Supelco) in 10 mL *N*,*N*-dimethylformamide (DMF, Wako). To uniformly disperse GCM and CB in solvent, the sonication was applied for 60 min at 38 kHz without addition of any dispersant. The as-prepared nanocomposite inks were deposited on an electrode-patterned Si substrate (*n*-type, with 100 nm-thick SiO_2 surface layer) to fabricate the 16-channel chemiresistive sensor array. Prior to the deposition, the comb-shaped Pt electrodes with Ti adhesive layer were first patterned on a 7 × 7 mm^2 sized substrate by photolithography and radio frequency (RF) sputtering. Gap distance and thickness of the electrodes were 40 μm and 400 nm, respectively. A SU-8 photoresist was then coated with 45 μm thickness on the electrode-patterned substrate by spin-coating and circular holes were made by photolithography. Each GCM-CB nanocomposite ink with 40 nL amount (40 shots at the rate of 1 nL/shot) was dropped at the circular holes by means of an ink-jet printing (custom-made, SIJ Technology Inc.). After depositing the GCM-CB nanocomposite inks, the device was annealed on a hotplate at 50 ºC for 60 min and subsequently in a vacuum oven (100 Pa) at 50 ºC for 60 min. The device was stored in the vacuum sealed bag until conducting the breath odor sensing measurements. The structural details of the fabricated sensor device were confirmed by an optical microscopy (OLYMPUS DP21).

Breath Odor Sensing Measurement. The breath odor sensing data were collected by a homemade sensing module, which consists of a gas flow chamber, solenoid valves, an air pump and a sensor operation/data collection system. Although a tube with solid phase adsorbent is often used for analyzing gaseous analytes, in this study, we conducted the sensing measurement by directly connecting the gas sampling bag with sensing system. This is because some molecules can evaporate during the sampling due to the weak interaction with adsorbent[S6)] and it may affect the breath print. For the measurements, the flow of breath odor into the chamber was controlled by the pump at a rate of 100 mL/min. The sensor response was collected as a variation of the voltage across the sensor by sequentially switching the flows of breath odor and N_2 carrier gas every 3 s with the solenoid valves. The sensor response was defined by the following equation: $\Delta V/V_0=(V-V_0)/V_0$, where V_0 and V are the output voltages under the flows of N_2 carrier gas with relative humidity value of 44% and breath odor, respectively. V and V_0 were calculated by using the average values of the sensor signal, e.g., $V=(\sum_{i=1}^{n} V_i)/n$, where n is the number of steps in time domain available within sensing period of 3 sec. All sensing measurements were performed at room temperature in air. Because some sensor materials were degraded within 30 days in air with RH 67% condition S7), in this study, we completed all sensing measurement before the sensor materials were degraded (also, the sensor device was stored in vacuum before and after the measurements to prevent the oxidation induced degradation).

Data Analysis with Machine Learning. Prior to the data analysis, the baseline correction was performed for the obtained sensing curves to make the baseline flat. The sensor responses $\Delta V/V_0$ were collected from 16-channel sensor array and used as dataset for machine learning. Sensing was cycled to each breath odor sample for increasing the number of training data for machine learning. Totally, ca. 1500 datasets of sensor response were obtained for the authentication of 6 persons and ca. 6000 datasets were obtained for the authentication of 20 persons. For machine learning, neural network algorithm was employed to build classifiers. The models were optimized by the hyper-parameters and ran with the parameters in Table S1. A 9-fold cross-validation was used to confirm the reproducibility of classifier. The reliability of classifier was characterized by the average area under curve (AUC) of receiver operating characteristic (ROC) curve. The prediction accuracy and the coefficient of variation in prediction accuracy were computed to evaluate the performance of breath odor sensing based individual authentication. The feature score for each sensor was evaluated using the permutation feature importance algorithm available in the library of scikit-learn (i.e. ‘permutation_importance’). The number of times to permute a feature was 30.

Table S1. Setting parameters for neural network algorithm.



Parameter	6 persons	20 persons
activation	tanh	tanh
alpha	0.0001	0.0001
hidden layer sizes	512, 512	1024, 1024
learning rate	adaptive	adaptive
max iteration	500	2000
solver	adam	adam

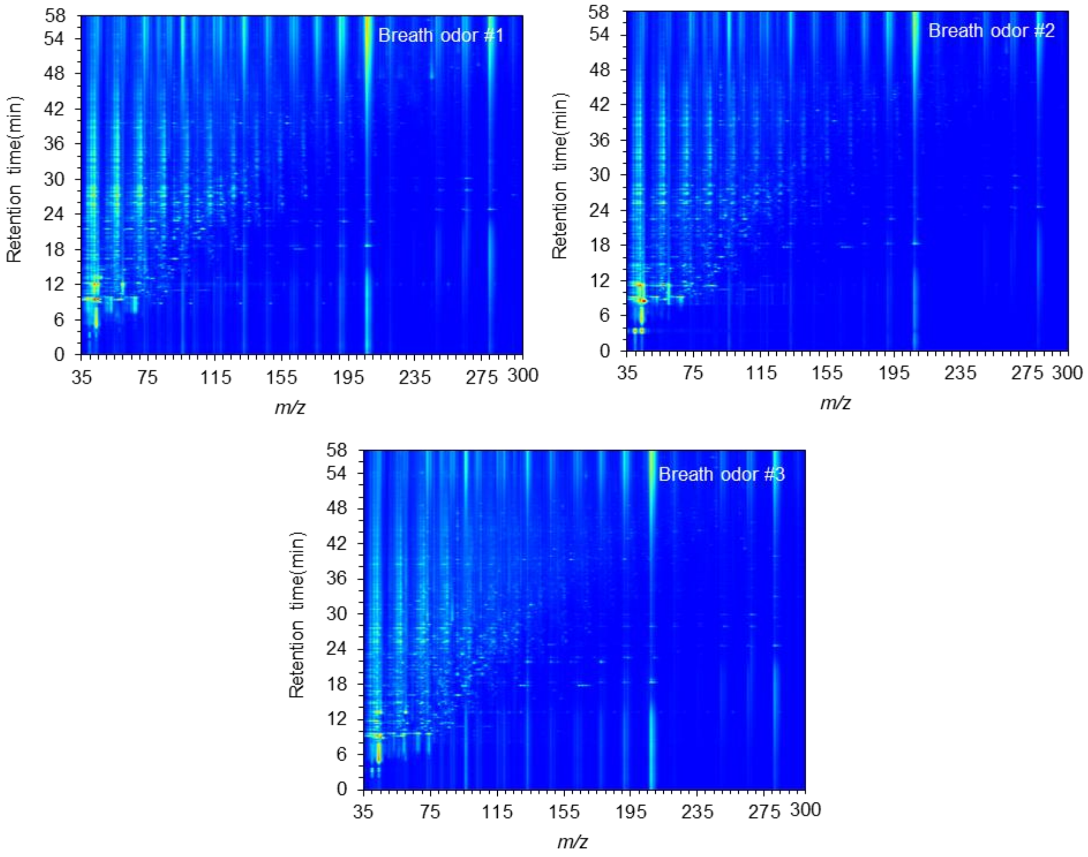
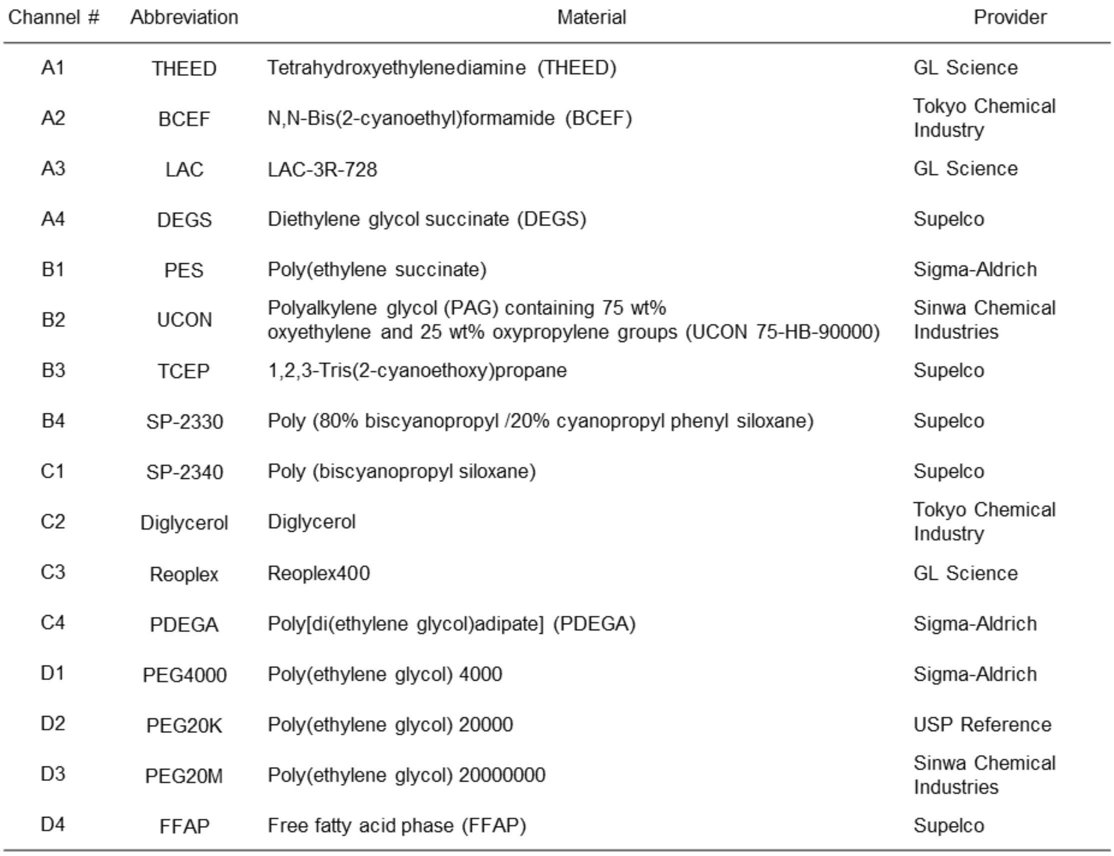


Figure S1. Full range 2D MS maps for the tested 3 persons (3 males), created by *NPFimg*.

Table S2. List of specific marker compounds for individual authentication

marker candidate	molecular weight	structure	molecular group	specificity
decanal	156.2	$C_{10}H_{20}O$	aldehye	1 person
octanal	128.2	$C_8H_{16}O$		1 person
pentanal	86.1	$C_5H_{10}O$		1 person
undecanal	170.3	$C_{11}H_{22}O$		3 persons
benzophenone	182.2	$C_{13}H_{10}O$	ketone	1 person
2-ethylhexyl octanoate	256.4	$C_{16}H_{32}O_2$	ester	1 person
viny benzoate	148.2	$C_9H_8O_2$	ester	1 person
methyl 3-ethylhexanoate	158.2	$C_9H_{18}O_2$	ester	1 person
methyl-2-butylhexanoate	186.3	$C_{11}H_{22}O_2$		1 person
2-phenyltridecane	260.5	$C_{19}H_{32}$	hydrocarbon	1 person
4-methyltridecane	198.4	$C_{14}H_{30}$		1 person
3,5 dimethyloctane	142.3	$C_{10}H_{22}$	hydrocarbon	1 person
5-ethyl-2methyloctane	156.3	$C_{11}H_{24}$		1 person
5-methyloctadecane	268.5	$C_{19}H_{40}$		1 person
6-methyloctadecane	268.5	$C_{19}H_{40}$		1 person
2,4-dimethyl-4-pentenoate	142.2	$C_8H_{14}O_2$	ester	3 persons
pentyl pentanoate	172.3	$C_{10}H_{20}O_2$		1 person
Isotridecanol-	200.4	$C_{13}H_{28}O$	alcohol and phenol	1 person
3-methylnonane	142.3	$C_{10}H_{22}$	hydrocarbon	1 person
4-methylpentadecane	226.4	$C_{16}H_{34}$		1 person
4-methyltetradecane	212.4	$C_{15}H_{32}$		1 person
diethyl ether	74.1	$C_4H_{10}O$	other	1 person
2-butanone	72.1	C_4H_8O	ketone	2 persons
nonadecane	268.5	$C_{19}H_{40}$	hydrocarbon	2 persons
octane	114.2	C_8H_{18}		2 persons
pentadecane	212.4	$C_{15}H_{32}$		2 persons
tetradecane	198.4	$C_{14}H_{30}$		2 persons
undecane	156.3	$C_{11}H_{24}$		2 persons

Table S3. The list of sensing materials used for the 16-channel sensor array.



Channel #	Abbreviation	Material	Provider
A1	THEED	Tetrahydroxyethylenediamine (THEED)	GL Science
A2	BCEF	N,N-Bis(2-cyanoethyl)formamide (BCEF)	Tokyo Chemical Industry
A3	LAC	LAC-3R-728	GL Science
A4	DEGS	Diethylene glycol succinate (DEGS)	Supelco
B1	PES	Poly(ethylene succinate)	Sigma-Aldrich
B2	UCON	Polyalkylene glycol (PAG) containing 75 wt% oxyethylene and 25 wt% oxypropylene groups (UCON 75-HB-90000)	Sinwa Chemical Industries
B3	TCEP	1,2,3-Tris(2-cyanoethoxy)propane	Supelco
B4	SP-2330	Poly (80% biscyanopropyl /20% cyanopropyl phenyl siloxane)	Supelco
C1	SP-2340	Poly (biscyanopropyl siloxane)	Supelco
C2	Diglycerol	Diglycerol	Tokyo Chemical Industry
C3	Reoplex	Reoplex400	GL Science
C4	PDEGA	Poly[di(ethylene glycol)adipate] (PDEGA)	Sigma-Aldrich
D1	PEG4000	Poly(ethylene glycol) 4000	Sigma-Aldrich
D2	PEG20K	Poly(ethylene glycol) 20000	USP Reference
D3	PEG20M	Poly(ethylene glycol) 20000000	Sinwa Chemical Industries
D4	FFAP	Free fatty acid phase (FFAP)	Supelco

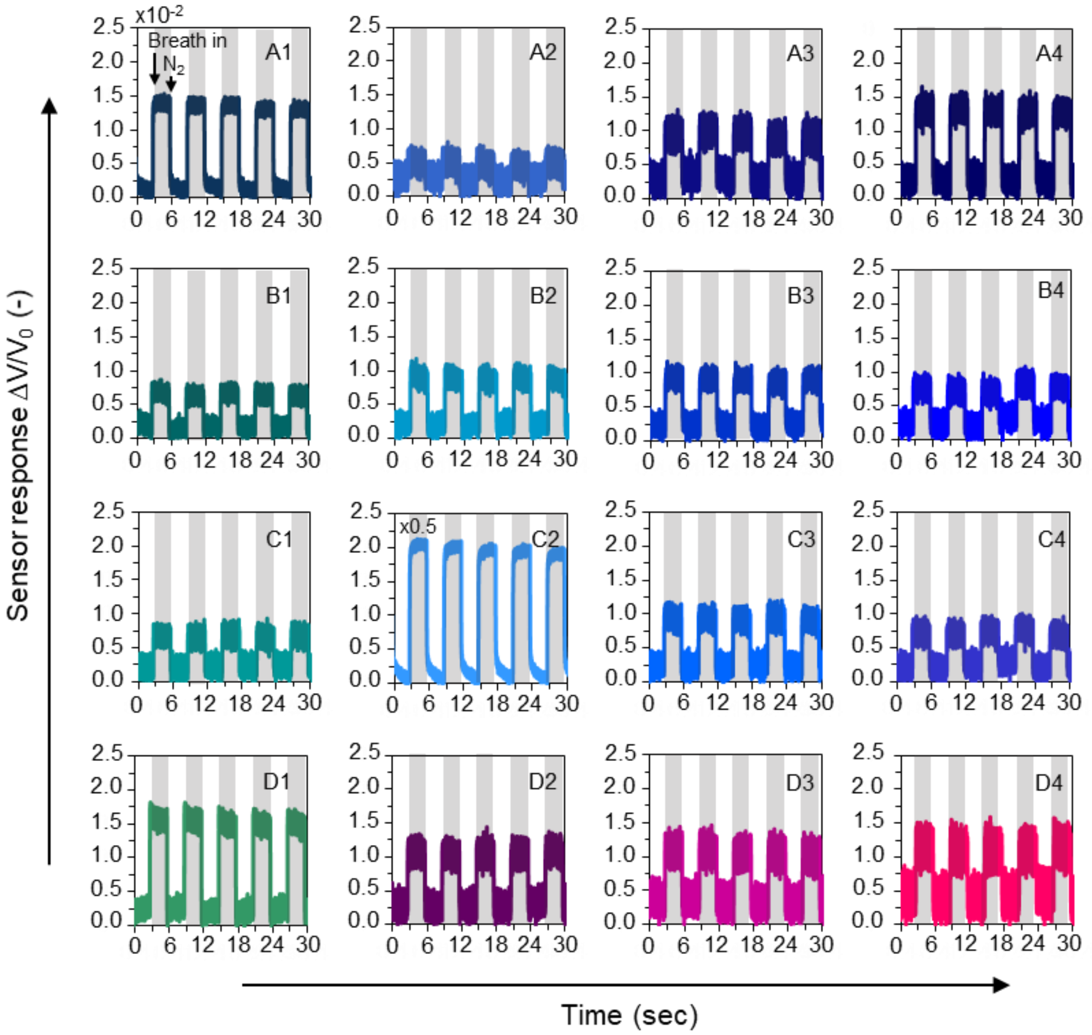


Figure S2. Sensing curves of 16-channel sensor array for the breath odor sensing of subject-V$^{\#2}$ after the baseline corrections.

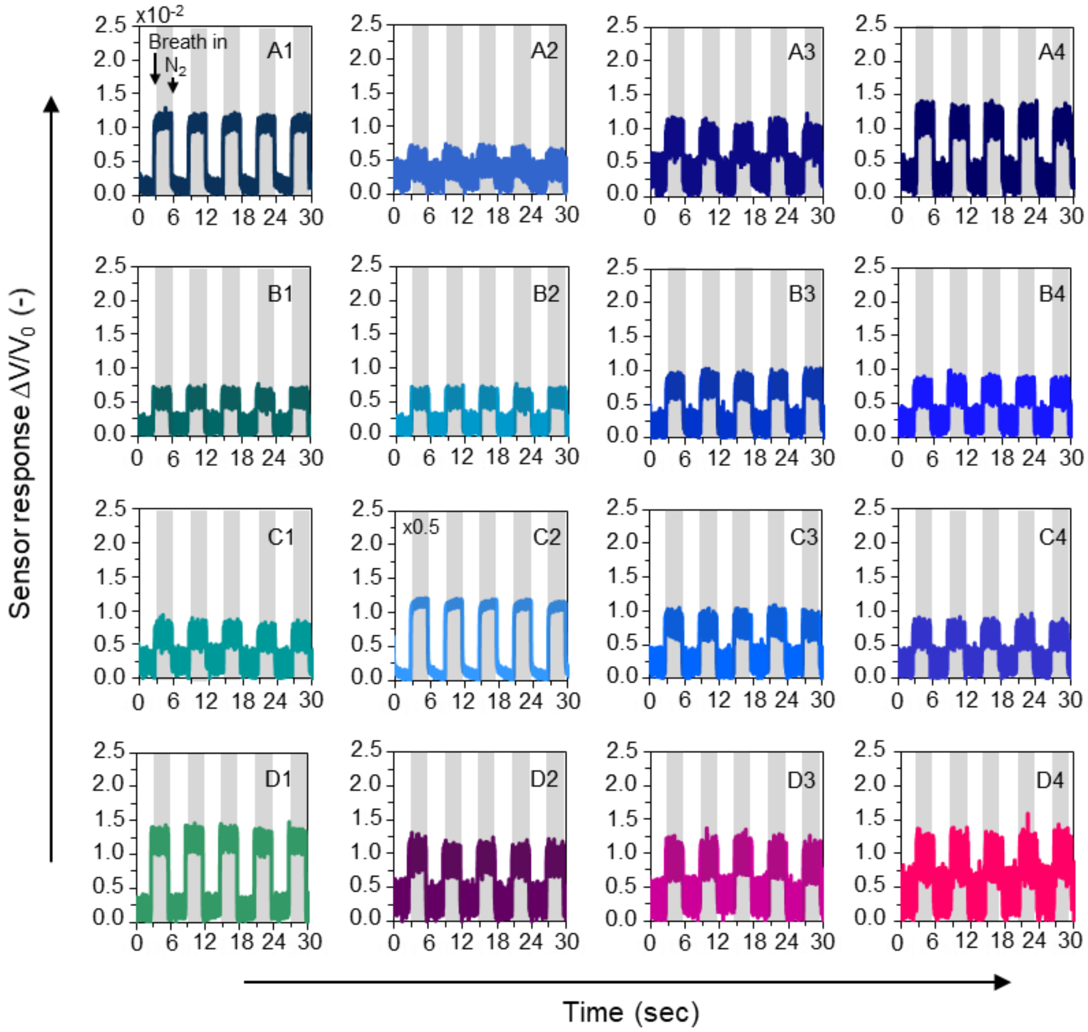


Figure S3. Sensing curves of 16-channel sensor array for the breath odor sensing of subject-V[#3] after the baseline corrections.

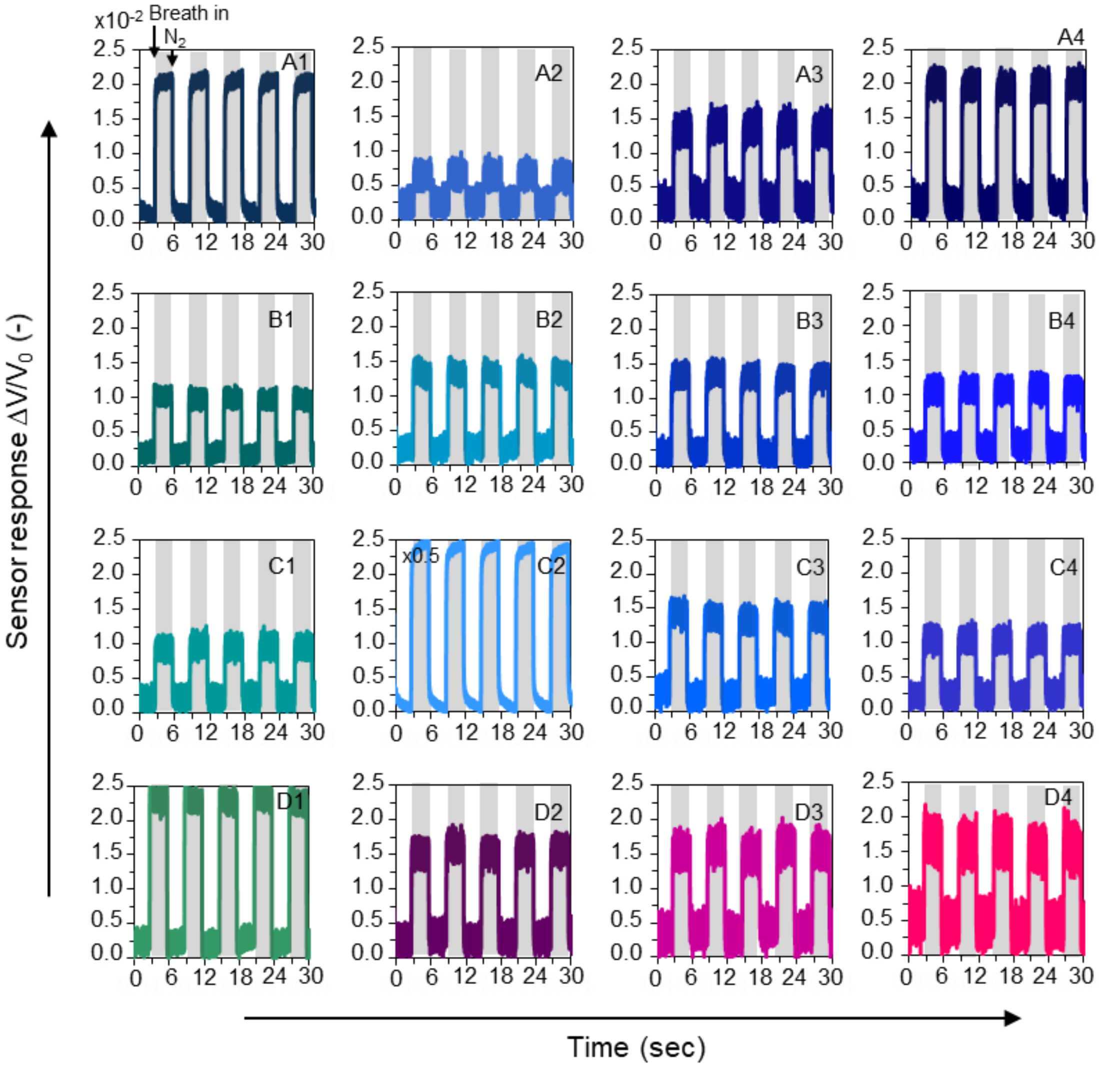


Figure S4. Sensing curves of 16-channel sensor array for the breath odor sensing of subject-V[#4] after the baseline corrections.

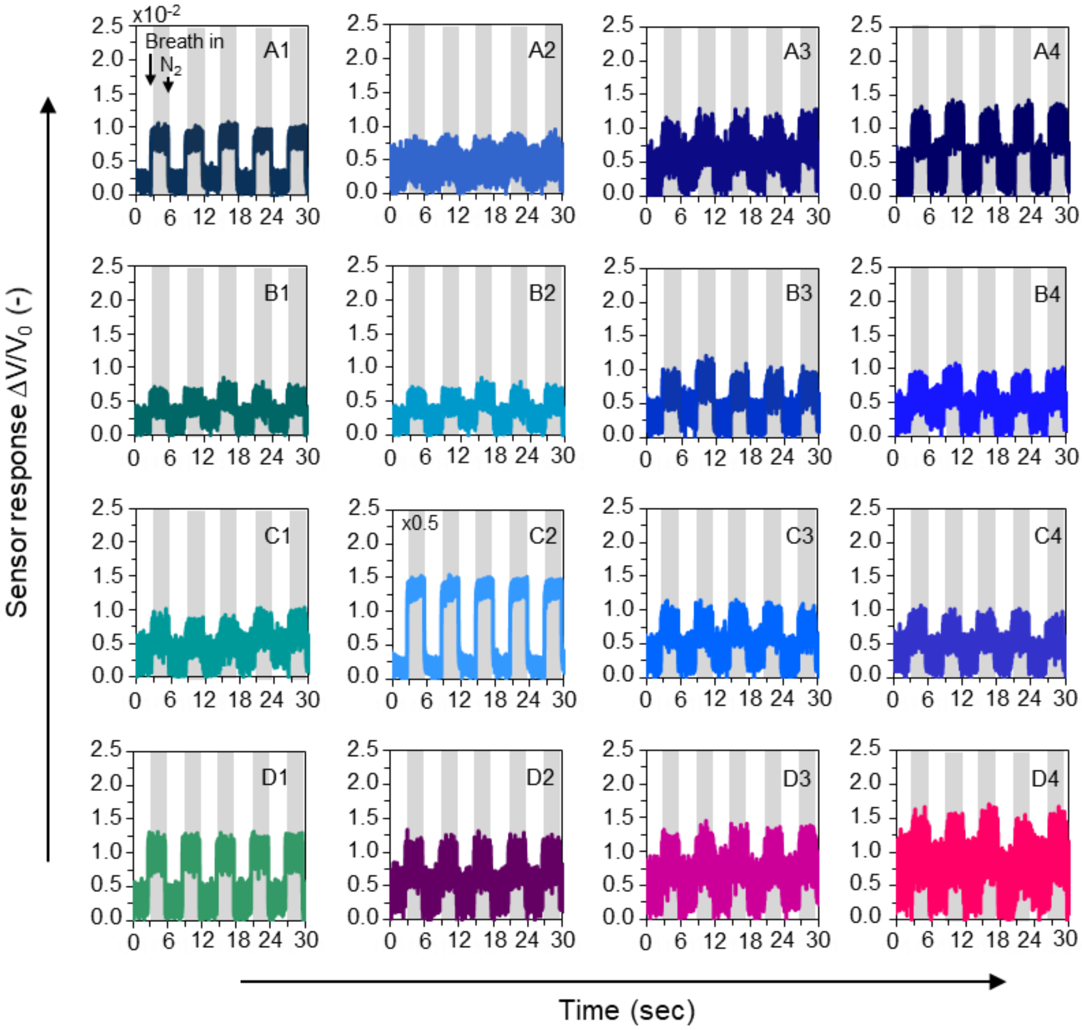


Figure S5. Sensing curves of 16-channel sensor array for the breath odor sensing of subject-V$^{\#5}$ after the baseline corrections.

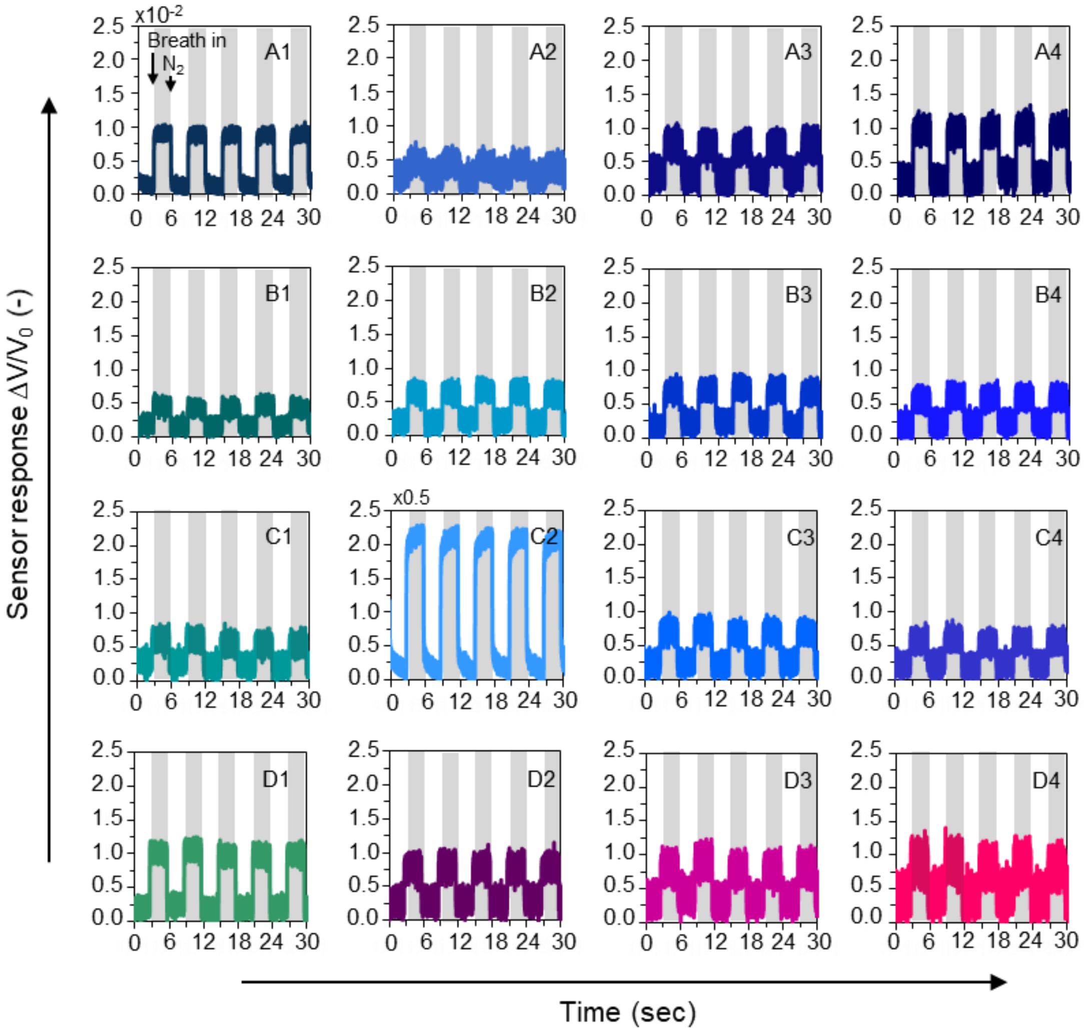
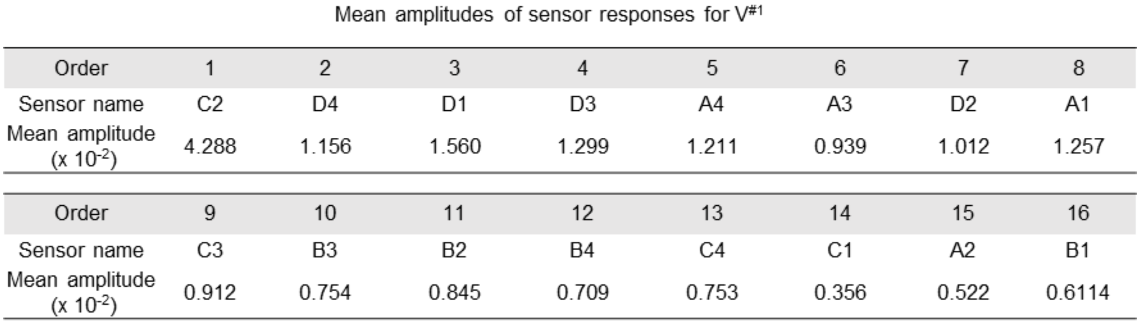


Figure S6. Sensing curves of 16-channel sensor array for the breath odor sensing of subject-V[#6] after the baseline corrections.

Table S4. Mean amplitudes of sensor responses of used sensors.



Mean amplitudes of sensor responses for $V^{\#1}$

Order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Sensor name	C2	D4	D1	D3	A4	A3	D2	A1	C3	B3	B2	B4	C4	C1	A2	B1
Mean amplitude ($\times 10^{-2}$)	4.288	1.156	1.560	1.299	1.211	0.939	1.012	1.257	0.912	0.754	0.845	0.709	0.753	0.356	0.522	0.6114

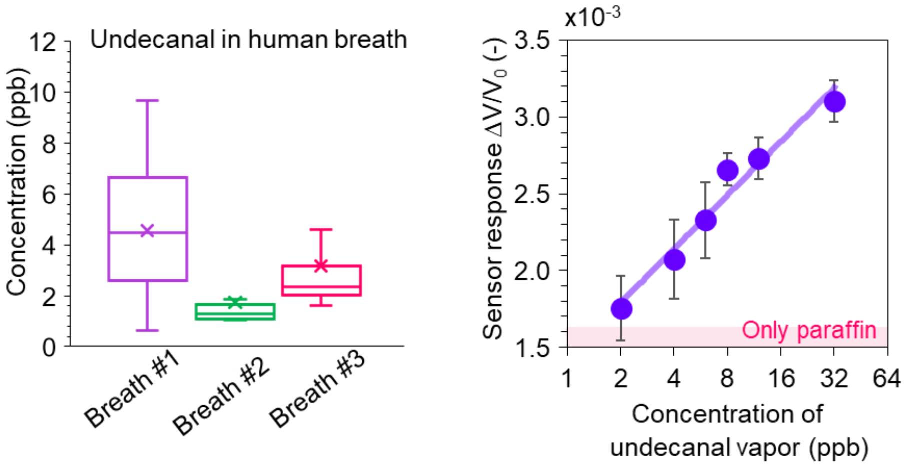
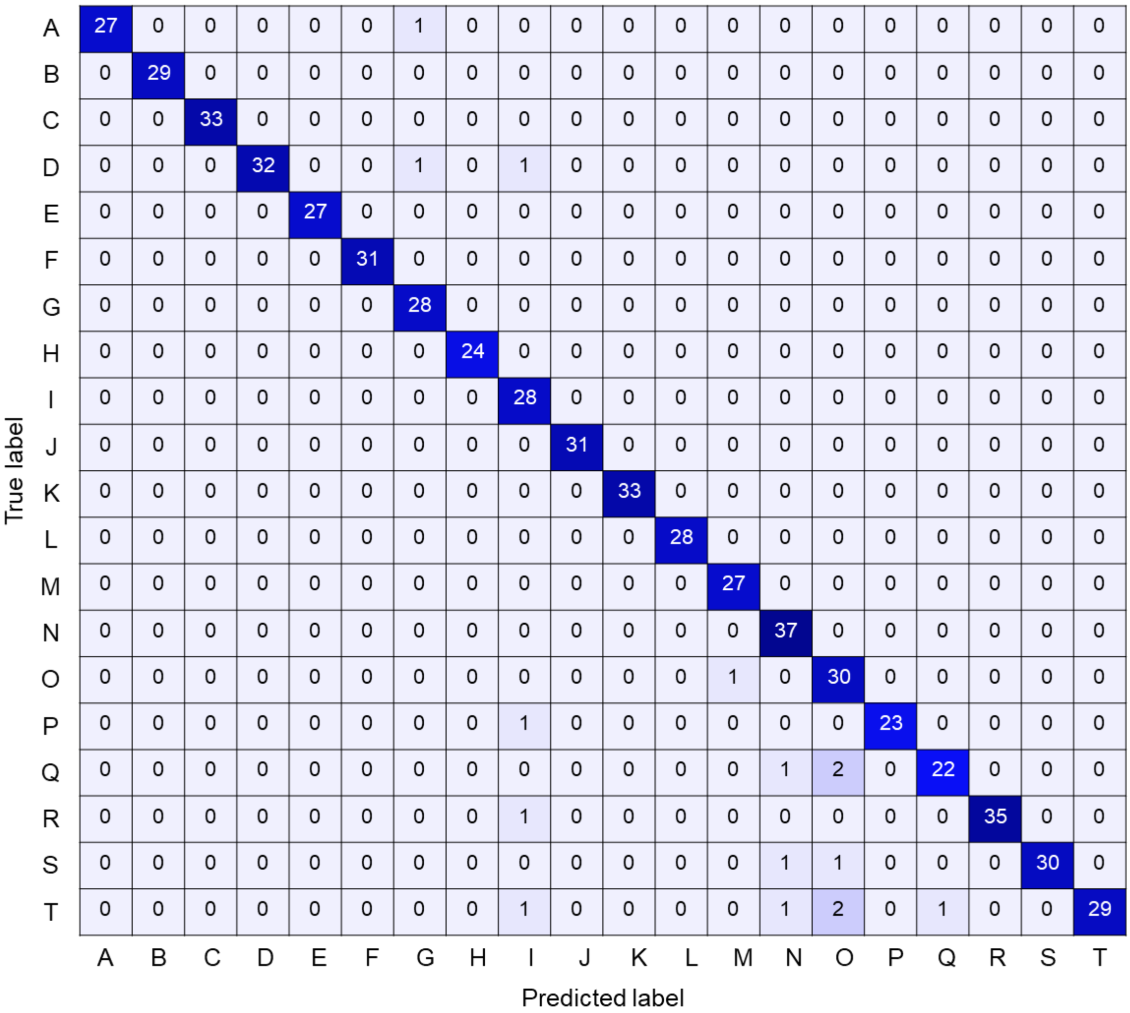


Figure S7. Undecanal concentration in breath odor samples (left). Sensor response to undecanal vapor (right).

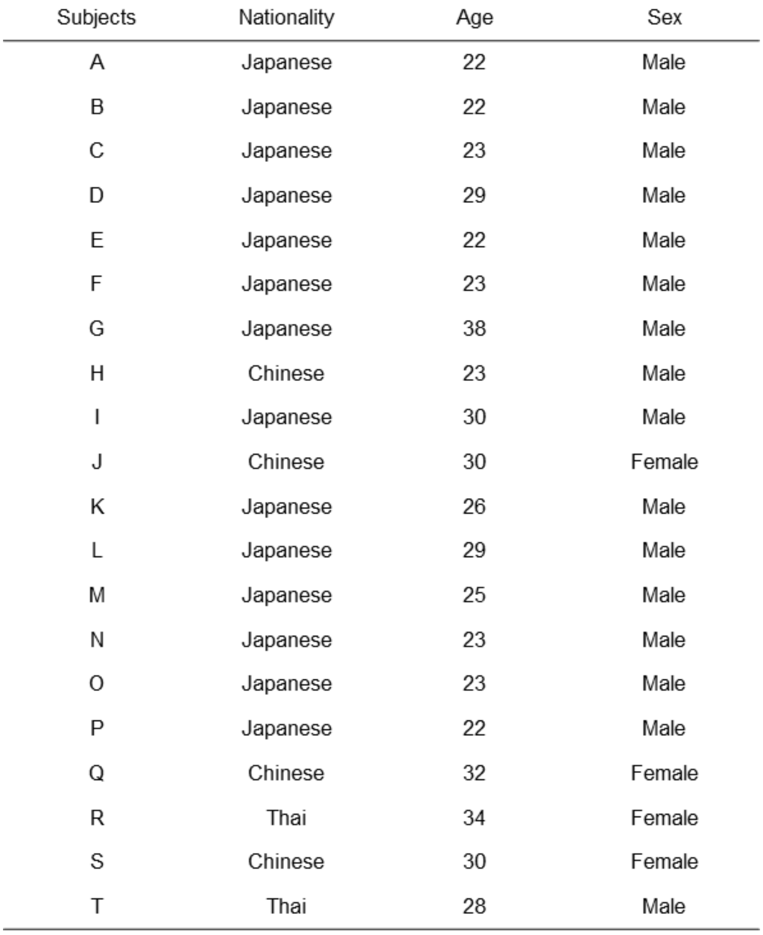
In order to confirm the feasibility of electrical detection of low concentration marker compound in breath, we analyzed the concentration of identified marker compound in breath odors and examined sensing of marker vapor at that concentration level. According to the GC-MS measurement, the concentration of undecanal in breath was in the range of 1-10 ppb. For the sensing measurement, we controlled the concentration of undecanal vapor by diluting its liquid concentrate with liquid paraffin, of which its vapor pressure is quite low. The vapor concentration of undecanal was estimated by GC-MS measurement. The sensing results exhibited that our sensor is capable of electrically detecting undecanal at least down to 2 ppb, covering the most of concentration range in breath odor.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	27	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	32	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	37	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	1	0	30	0	0	0	0	0
P	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	23	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	22	0	0	0
R	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	35	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	30	0
T	0	0	0	0	0	0	0	0	1	0	0	0	0	1	2	0	1	0	0	29

Figure S8. Confusion matrix for the breath odor sensing based individual authentication for 20 persons.

Table S5. The details of tested subjects for individual authentication for 20 persons.



Subjects	Nationality	Age	Sex
A	Japanese	22	Male
B	Japanese	22	Male
C	Japanese	23	Male
D	Japanese	29	Male
E	Japanese	22	Male
F	Japanese	23	Male
G	Japanese	38	Male
H	Chinese	23	Male
I	Japanese	30	Male
J	Chinese	30	Female
K	Japanese	26	Male
L	Japanese	29	Male
M	Japanese	25	Male
N	Japanese	23	Male
O	Japanese	23	Male
P	Japanese	22	Male
Q	Chinese	32	Female
R	Thai	34	Female
S	Chinese	30	Female
T	Thai	28	Male

References

S1) C. Jirayupat, K. Nagashima, T. Hosomi, T. Takahashi, W. Tanaka, B. Samransuksamer, G. Zhang, J. Liu, M. Kanai and T. Yanagida, *Anal. Chem.* 2021, **93**, 14708-14715.

S2) B. Wyszynski, R. Yatabe, A. Nakao, M. Nakatani, A. Oki, H. Oka and K. Toko, *Sensors* 2017, **17**, 1606.

S3) A. Shunori, R. Yatabe, B. Wyszynski, Y. Hanai, A. Nakao, M. Nakatani, A. Oki, H. Oka, T. Washio and K. Toko, 2*019 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN)*, 2019, pp. 1-3.

S4) R. Yatabe, A. Shunori, B. Wyszynski, Y. Hanai, A. Nakao, M. Nakatani, A. Oki, H. Oka, T. Washio and K. Toko, *IEEE Sensors Journal*, 2021, **21**, 2077-2083.

S5) Y. Liu, S. Furuno, S. Akagawa, R. Yatabe, T. Onodera, N. Fujiwara, H. Takeda, S. Uchida and K. Toko, *Chemosensors* 2021, **9**, 261.

S6) A. Marcillo, V. Jakimovska, A. Widding, and C. Birkemeyer, *J. Chromatogr. A* 2017, **1514**, 16-28.

S7) W. Li, K. Nagashima, T. Hosomi, C. Wang, Y. Hanai, A. Nakao, A. Shunori, J. Liu, G. Zhang, T. Takahashi, W. Tanaka, M. Kanai and T. Yanagida, *ACS Sens.* 2021, **7**, 151-158.