Sensor Glove Approach for Japanese Fingerspelling Recognition System Using Convolutional Neural Networks

Tomohiko Tsuchiya[1], Akihisa Shitara[2], Fumio Yoneyama[1], Nobuko Kato[1], Yuhki Shirarishi[1]
[1]Tsukuba University of Technology, Japan, [2]University of Tsukuba, Japan
E-mail: a193102@a.tsukuba-tech.ac.jp

The Thirteenth International Conference on Advances in Computer-Human Interactions ACHI 2020
November 21, 2020 to November 25, 2020 – Valencia Spain

Resume of the presenter

Tomohiko Tsuchiya

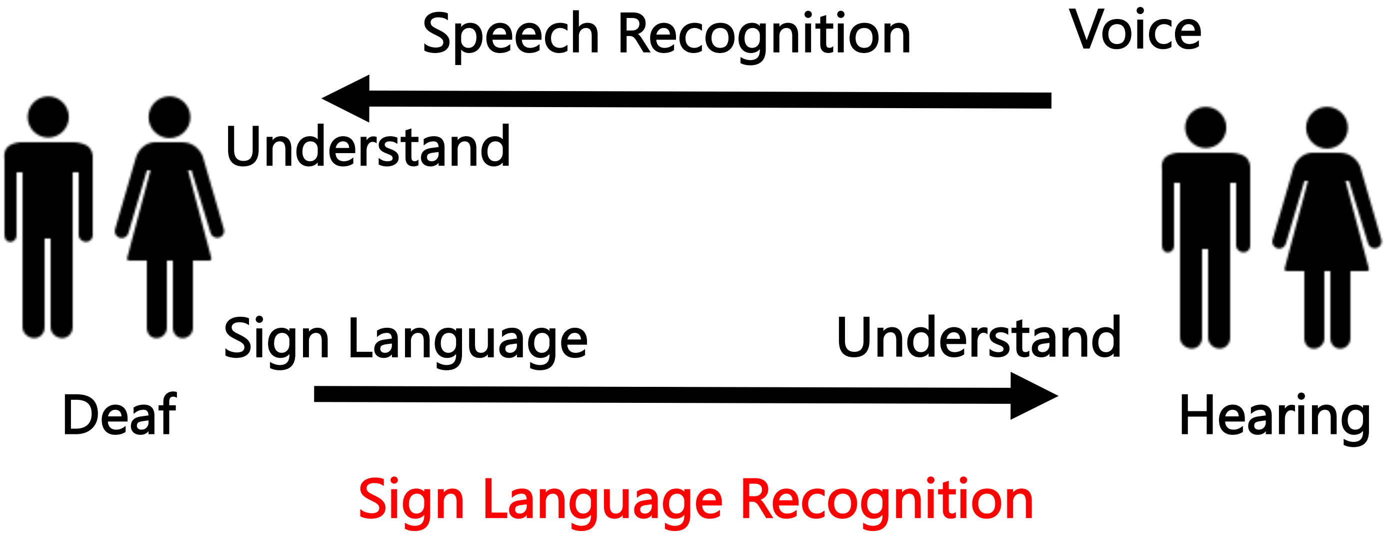
I am Deaf

2019.4~
Course of Information Science,
Division of Industrial Technology,
Graduate School of Technology and Science,
Tsukuba University of Technology, Japan

Contents

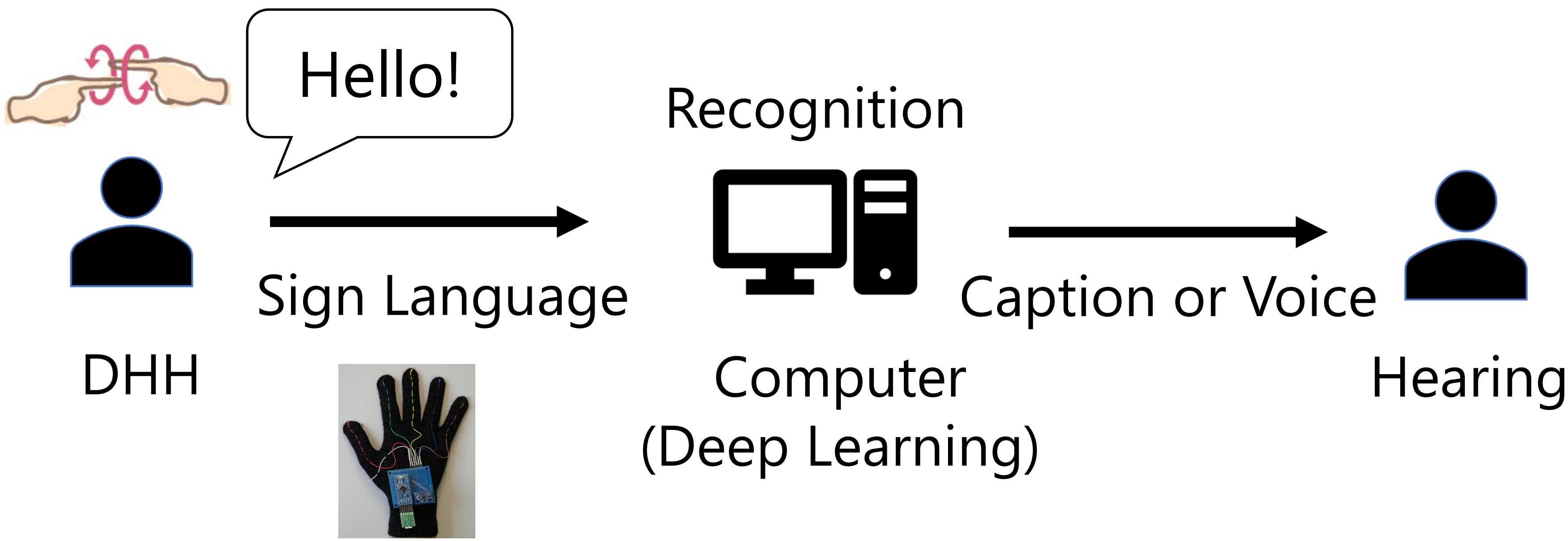
Contents

Background



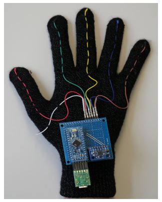
Speech Recognition
Voice
Understand
Sign Language
Understand
Deaf
Hearing
Sign Language Recognition

Realizing the Sensor Glove Approach for Japanese Fingerspelling Recognition System



Contents

Related work

Recognition System	features	points
Image recognition	Hand → Camera → PC	• Occlusion factors • Environmental factors
Sensor glove recognition	Hand → Sensor glove  → PC	• Sensors can be attached directly to the hands • Easy to perform

Related work

Number of fingerspelling in each country

Language	Dynamic	Static	Sum
American	2	24	26
French	3	23	26
Japanese	35	41	76

types of fingerspelling recognition method in each country

type	Recognition method	Fingerspelling target	Recogniti on rate
5DT Data Glove 5 Ultra＋accelerometer*1	Neural Network	Static fingerspelling in ASL	94.07%
IMU-based glove*2	Machine Learning	Static fingerspelling in French	92%

*1 M. E. Cabrera, J. M. Bogado, L. Fermin, R. Acuna, and D. Ralev,"Glove-based gesture recognition system," in Adaptive MobileRobotics. World Scientific, 2012, pp. 747–753.
*2C. K. Mummadi, F. P. P. Leo, K. D. Verma, S. Kasireddy, P. M.Scholl, and K. Van Laerhoven, "Real-time embedded recognitionof sign language alphabet fingerspelling in an imu-based glove," in Proceedings of the 4th International Workshop on Sensor-BasedActivity Recognition and Interaction, ser. iWOAR '17.New York,NY, USA: Association for Computing Machinery, 2017, pp. 1–6

Contents

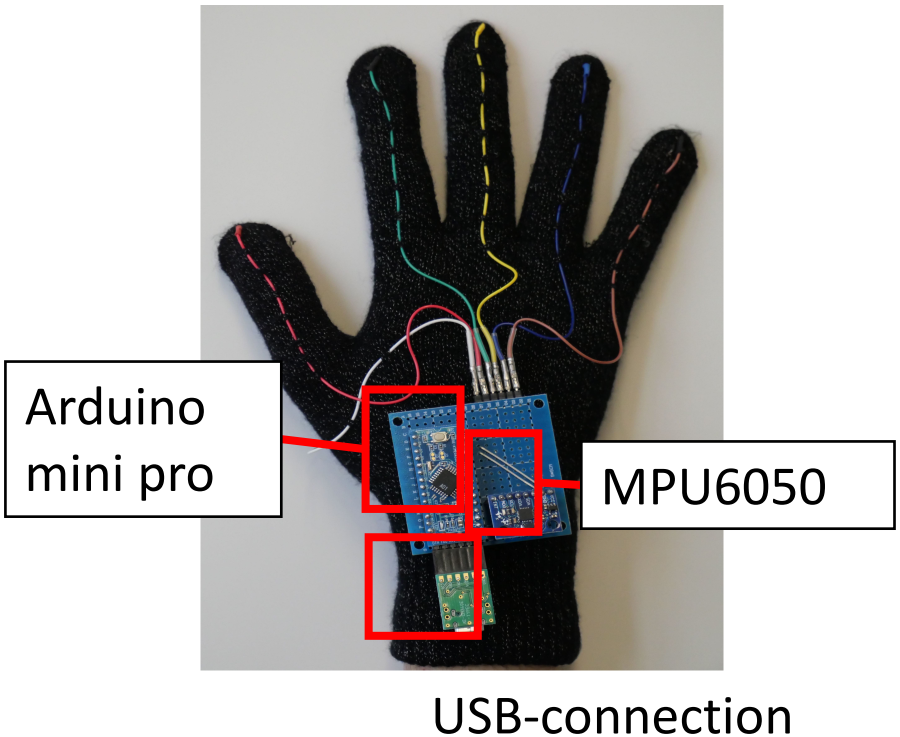
Objective

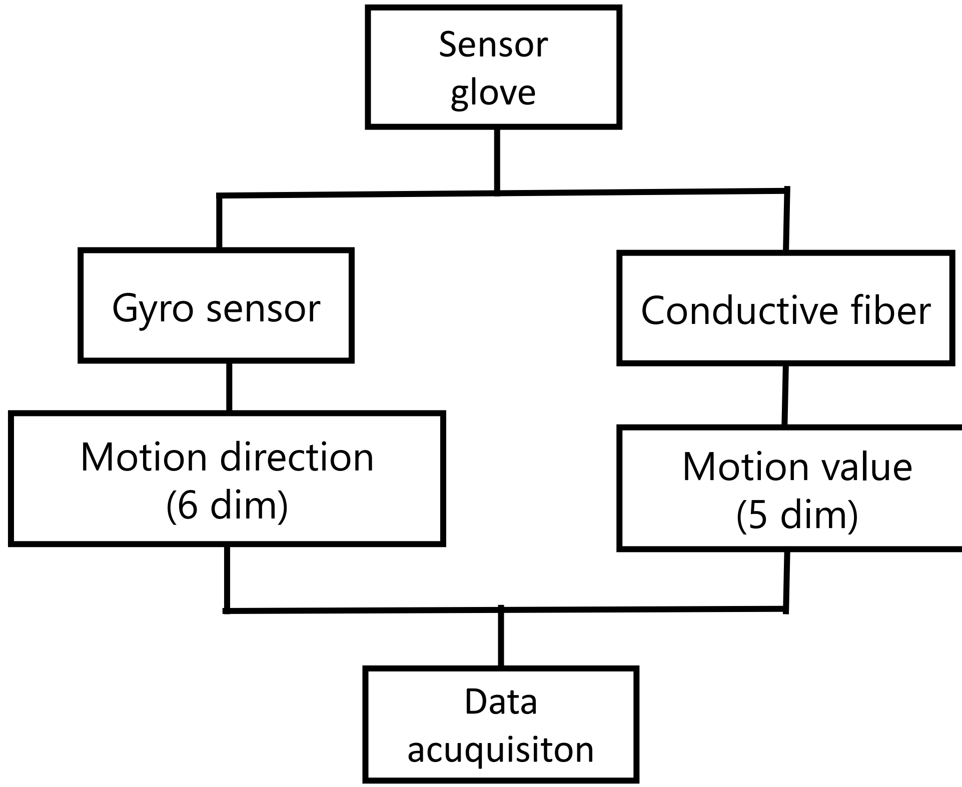
JFS recognition by adopting conductive fiber weaving technology, which can reduce the weight and cost of sensor gloves and simplify hand movement.

Evaluated our developed system by classifying 76 JFS characters, including dynamic (non-static) fingerspelling characters

Contents

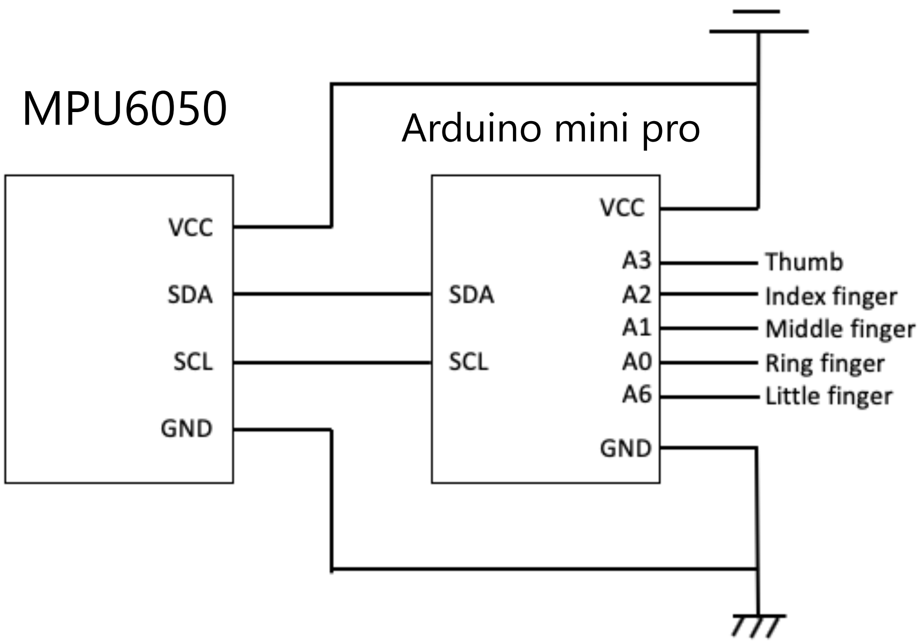
Method-Sensor glove

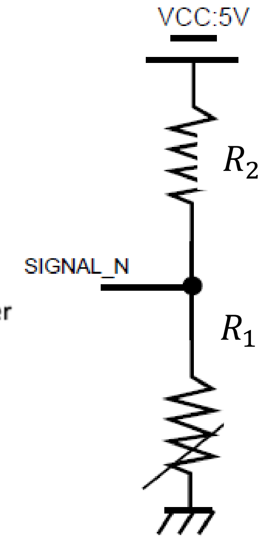




R. Takada, J. Kadomoto, and B. Shizuki, "A sensing technique for dataglove using conductive fiber," in Extended Abstracts of the 2019 CHIConference on Human Factors in Computing Systems, ser. CHI EA '19.New York, NY, USA: Association for Computing Machinery, 2019,pp. 1–4. [Online].

Method-Sensor glove





conductive sensor glove (5 dimensions)

$V_{in} = \frac{R_1}{R_1+R_2} * V_{out}$ (1)

R_1 : variable resistance by the conductive fibers

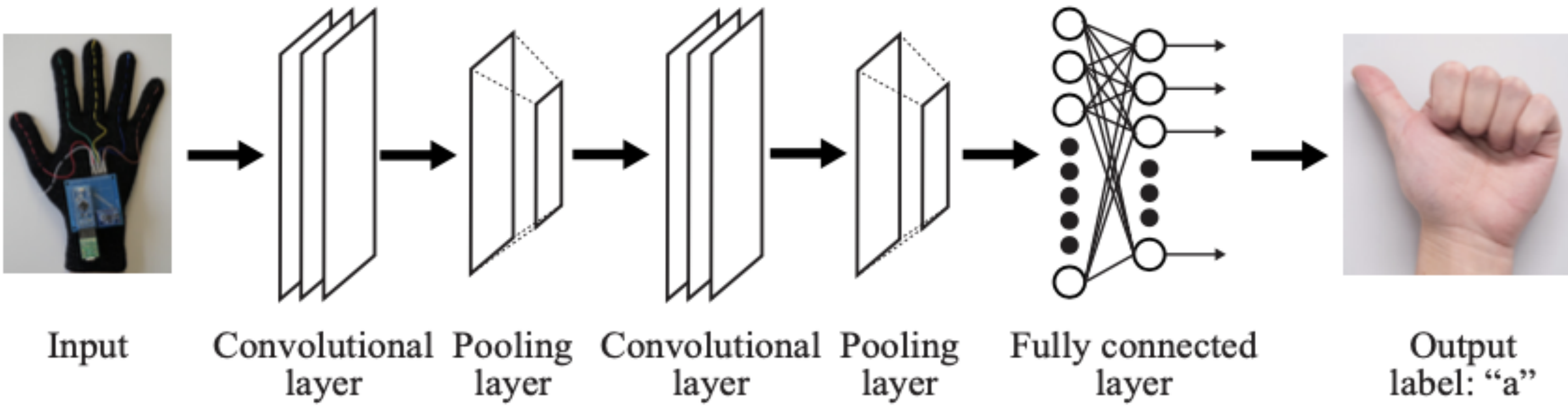
R_2 : fixed resistors ($10k\Omega$)

Motion amounts : $D_{out} = \frac{1}{N_{max}-N_{min}} * (D_{in} - N_{max})$

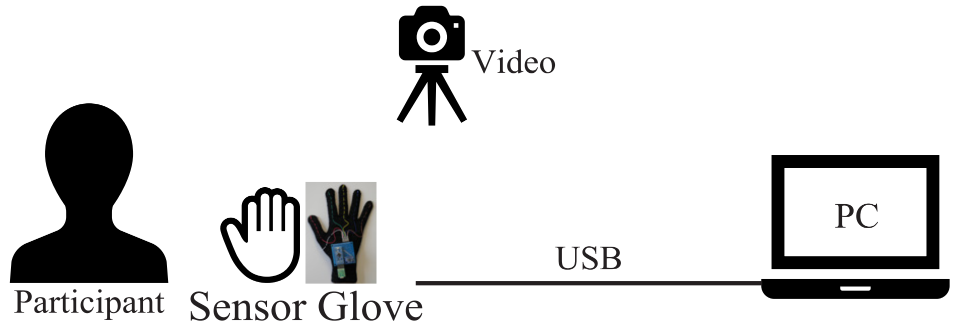
Motion direction (Accelerometer 3 dimensions) : $D_{out\_acc} = \frac{D_{in\_acc}}{2.0}$

Motion direction (Gyro 3 dimensions) : $D_{out\_gyro} = \frac{D_{in\_gyro}}{250.14}$

Convolutional Neural Network (CNN)



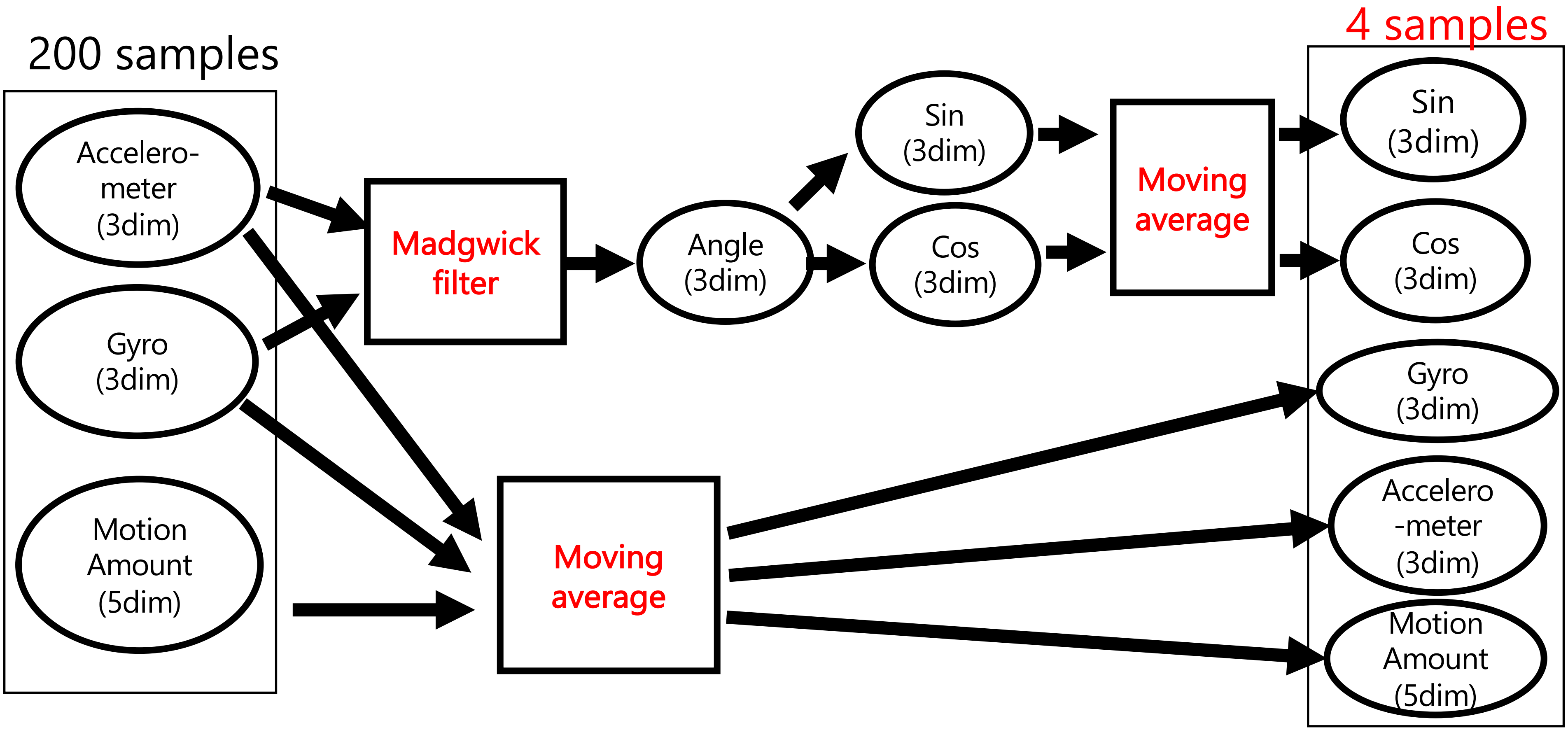
Experiment-Data collection



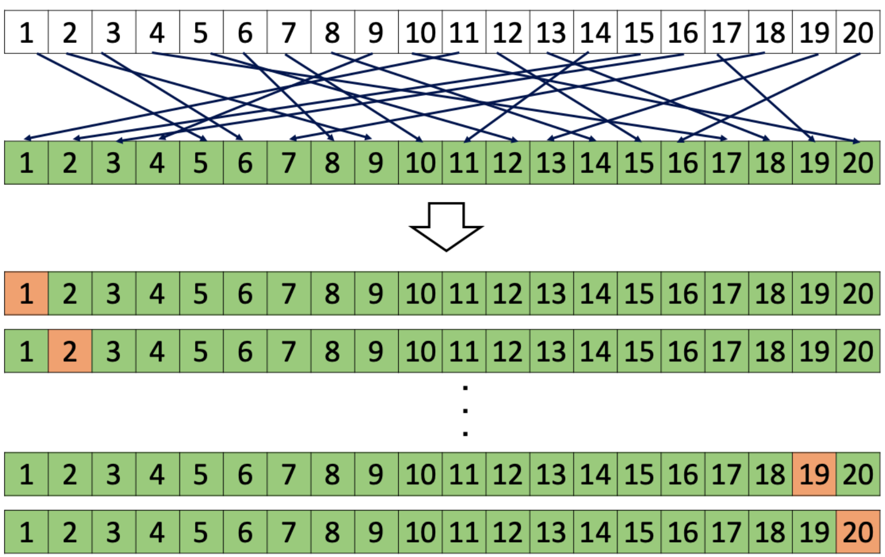
Participant	20 peoples	
Recognition subjects	76 characters Including dullness, semi-voiced sound, diphthong, and a long vowel	
Acquiring data	Motion amounts (5dimensions) Motion direction （Accelerometer：3dimensions・Gyro：3dimensions）	
Method	Motion	1 character
	Acquiring in second	1 second
	time	5 times

Contents

Method-madgwick filter and data reduction

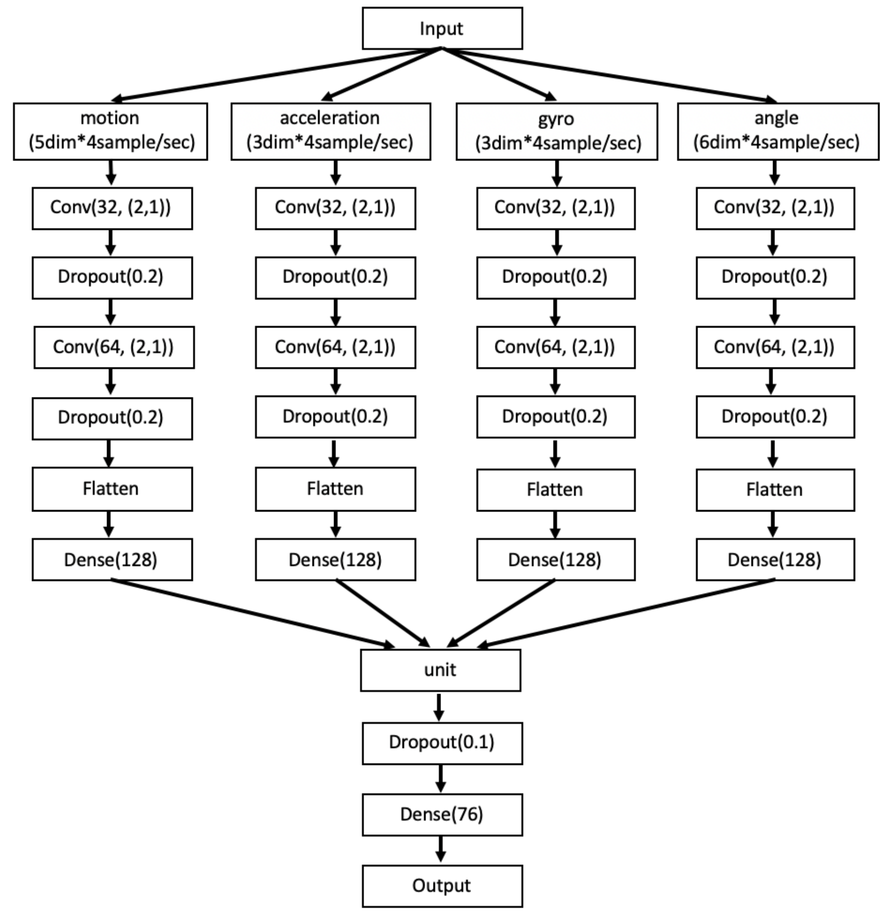


Method Recognition



Review of 20-fold cross validation

Architecture of convolutional neural network

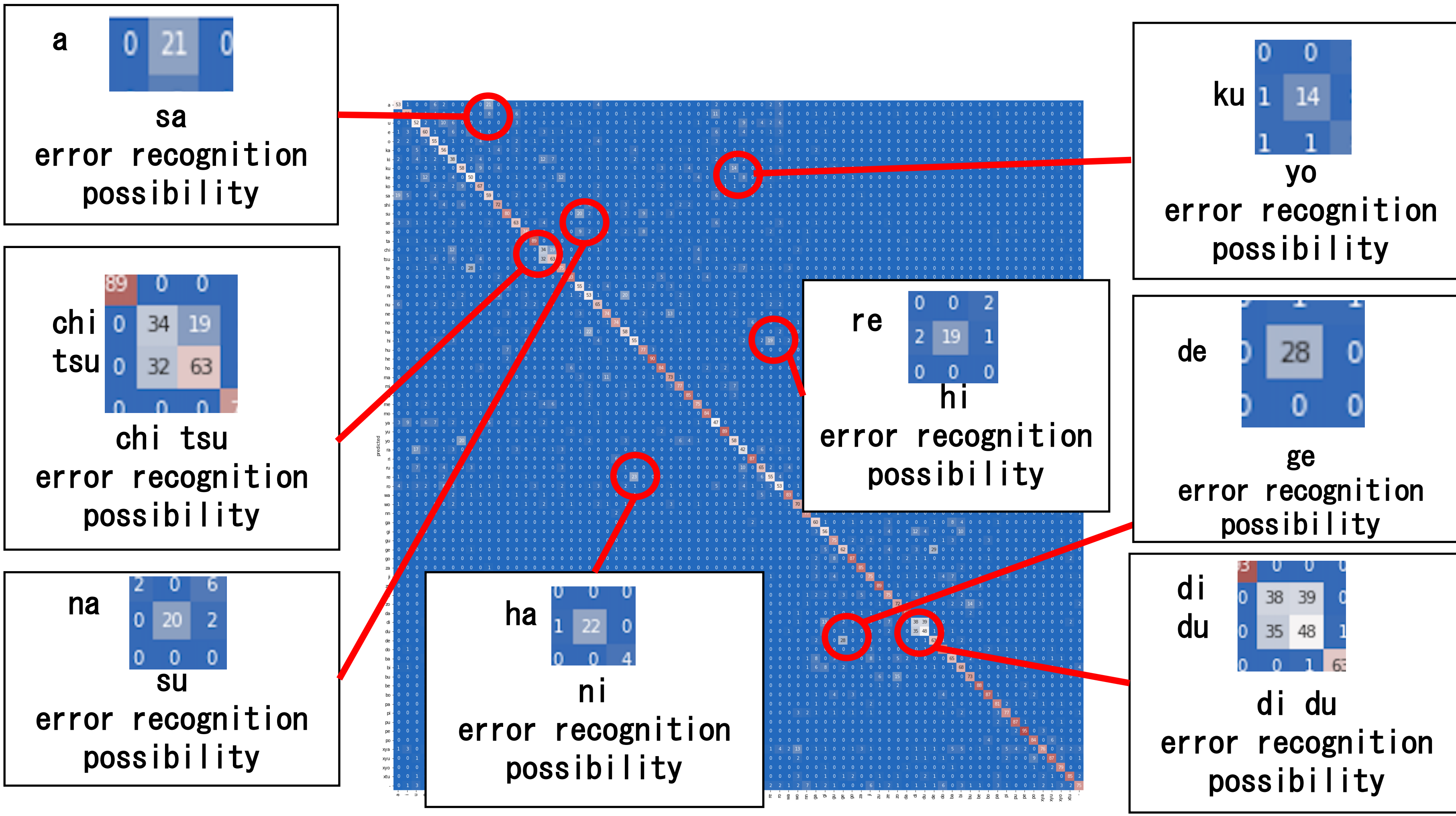


Contents

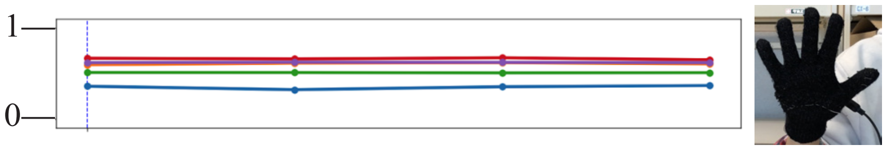
Experiment-Recognition Result

Number	Accuracy rate of training data	Accuracy rate of test data	Note
1	93.6%	65.0%	Min recognition rate
2	94.1%	75.5%	Max recognition rate
3	94.8%	68.7%	
4	93.1%	69.7%	
5	94.2%	66.3%	
6	93.9%	73.2%	
7	92.9%	67.9%	
8	93.5%	71.1%	
9	93.0%	67.4%	
10	94.6%	70.5%	

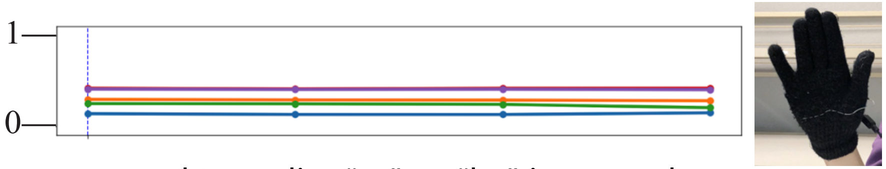
Number	Accuracy rate of training data	Accuracy rate of test data	Note
11	93.4%	71.6%	
12	93.0%	66.1%	
13	94.6%	68.9%	
14	94.3%	70.3%	
15	93.0%	69.7%	
16	93.4%	68.4%	
17	92.9%	71.3%	
18	93.1%	71.1%	
19	94.5%	74.2%	
20	94.5%	72.4%	
Average	93.7%	70.0%	


a
sa
error recognition possibility
chi
tsu
chi tsu
error recognition possibility
na
su
error recognition possibility
ha
ni
error recognition possibility
ku
yo
error recognition possibility
re
hi
error recognition possibility
de
ge
error recognition possibility
di
du
di du
error recognition possibility

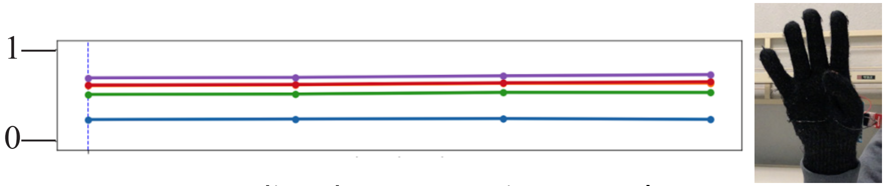
Teacher	a	sa	ku	yo	ke	te	ki	chi	chi
Prediction	sa	a	yo	ku	te	ke	chi	ki	tsu
Rate (%)	21.0	19.0	14.0	20.0	12.0	28.0	12.0	12.0	34.0
Teacher	tsu	ni	ha	ne	ma	hi	re	wo	xya
Prediction	chi	ha	ni	ma	ne	re	hi	xya	wo
Rate (%)	32.0	20.0	22.0	13.0	11.0	19.0	23.0	11.0	13.0
Teacher	gi	di	ge	de	di	du	zo	bu	
Prediction	di	gi	de	ge	du	di	bu	zo	
Rate (%)	12.0	13.0	29.0	20.0	39.0	35.0	14.0	15.0	



(a) predict “te” as “te” correctly



(b) predict “te” as “ke” incorrectly



(c) predict “ke” as “te” incorrectly

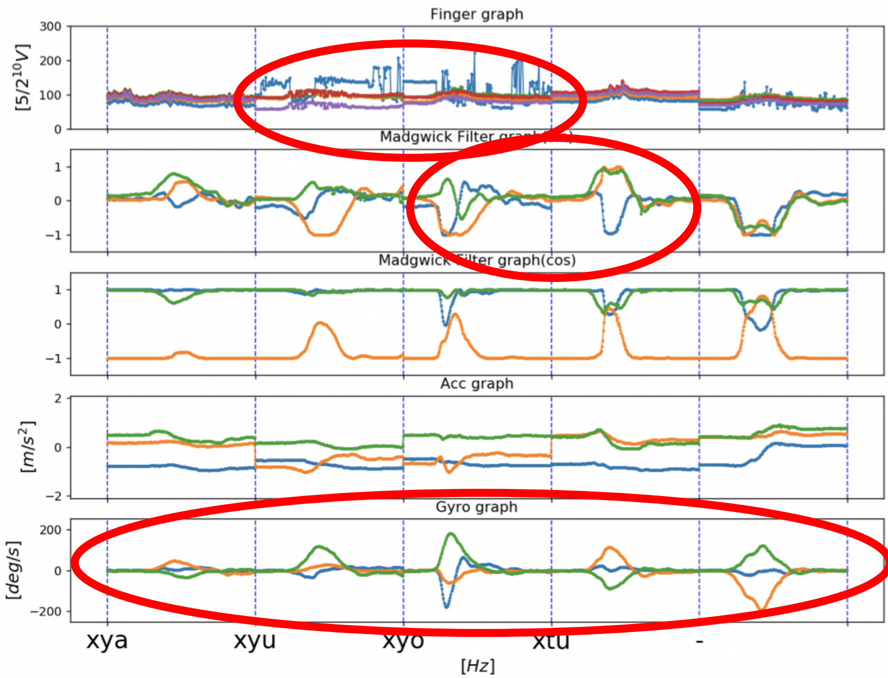
(d) predict “ke” as “ke” correctly

It was confirmed that close contact between the fingers caused these errors.

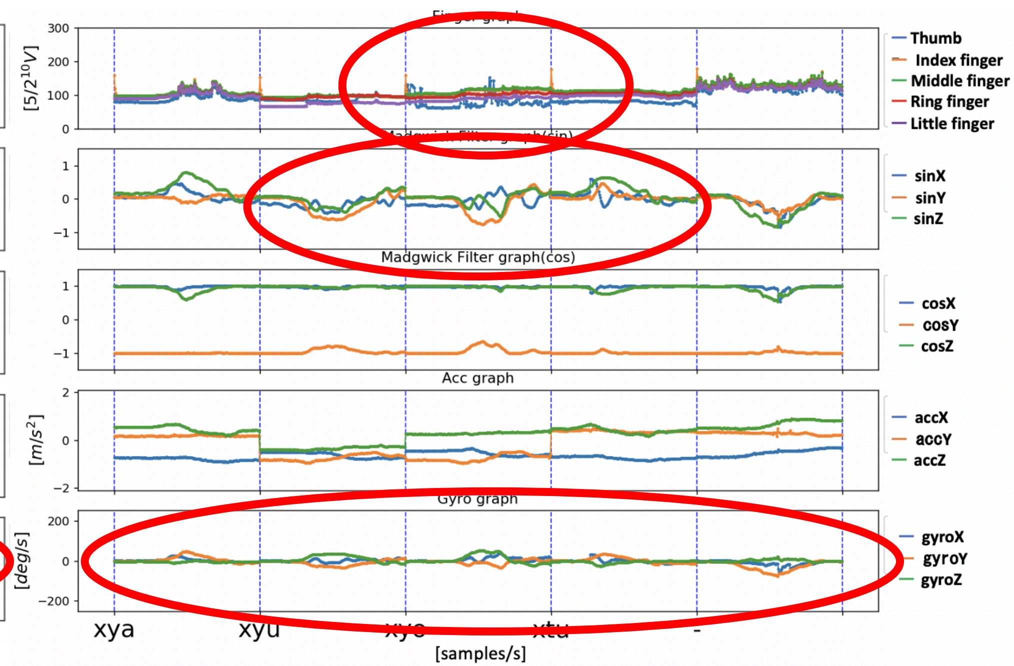
Notably, the thumb sometimes contacted the forefinger.

Additionally, depending on the participant, the hand may be widely opened or the fingers may be in close contact.

One person



Another person



This figure clearly highlights the individual differences in fingerspelling between participants, particularly in the strength of finger bending (including noisy signals), timing of hand move movement, and shape of the fingers.

Contents

Conclusions

- To realize smooth communication between the DHH and hearing people, adopted a lightweight sensor glove using CNN.

⇩

- Five dimensions of motion magnitude data, three dimensions of acceleration data, three dimensions of gyro data, and six dimensions of angle for inputs
- We calculated moving averages to reduce the frequency to 4 samples/s.
- A 20-fold cross validation evaluation experiment was conducted.

⇩

- The average recognition rate was approximately 70.0%
- The maximum recognition rate was approximately 75.5%
- The firm attachment of conductive fibers was a significant cause of misrecognition.

Future works

- Constructing improved sensor gloves and investigate methods to handle various problems.
- Planning additional experiments for data collection under more controlled conditions.
- Conducting continuous fingerspelling recognition experiments.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Number #19K11411 and the Promotional Projects for Advanced Education and Research in NTUT.
We would like to thank Editage (www.editage.com) for English language editing.