Classification of mouth movement for Meal Support Robot

Hiroshi Narazaki^a, Kanya Tanaka^b, Shenglin Mu^c, Shota Nakashima^{a,*}

^aYamaguchi University, 2-16-1, Tokiwadai, Ube, Yamaguchi, 755-8611, Japan

^bMeiji University, 1-1-1, Higashimita, Tama, Kawasaki, Kanagawa, 214-8571, Japan

^cEhime University, 3, Bunkyo-cho, Matsuyama, Ehime, 790-8577, Japan

*Corresponding Author: s-naka@yamaguchi-u.ac.jp

Abstract

The population of people who need assistance in daily life is increasing year by year in Japan. The number of upper limb disabled people, who have various difficulties in daily lives even cannot have meals by themselves, is increasing to 690,000. They have difficulty eating on their own. The number of caregivers is short for the daily care work. The situation in welfare field is that the work burden on the caregivers is increasing continuously.

To relief the burden of caregivers, the meal support robots which are able to give assistance to users automatically, are developed. In this research, a mouth movement recognition method is proposed to improve the automation of meal support robots. This system use the mouth movement of a person who needs assistance and In the proposed method, the information of the user's mouth movement will be transferred to a server PC. The algorithms in the server are implemented to recognize the chewing and speech movement. In addition, mouth movement shows chewing movement and speech movement. In addition, chewing movement and speech movement call themselves chewing movement, speech afterward.

Keywords: meal support robots, mouth movement, check the situation of user

1. Introduction

Currently, the number of people with disabilities who need assistance is increasing year by year⁽¹⁾. In addition, the number of upper limb disabled persons is about 690,000 people, and among them, there are about 570,000 people who have disabilities with disabilities class 2 or higher with severe disability⁽²⁾. Assistance for upper limb disabled people is indispensable in their daily lives, especially, since it is difficult to have meals by themselves, meal support

activities are regarded. Furthermore, for physically handicapped people with upper limb disorder, ingesting themselves on their own is a matter of quality of life (QOL)⁽³⁾, which is an important factor. On the other hand, caregivers are short. For this reason, the situation in welfare field is that the burden on the caregivers are increasing continuously. To relief the burden on the caregivers, various meal support robots are currently being developed. There are some existent meal support robots such as, iARM⁽⁴⁾, My spoon⁽⁵⁾ or HANDY1⁽⁶⁾ and so on.

By using these meal support robots, even those who have disabilities in upper limbs can do their own meals, it is thought that QOL can be improved. It also helps to reduce the burden per assistant. However, in most meal support situation, the caregiver cannot really get away unless the user's situation can be obtained in case some chewing or drinking trouble happens. Therefore, it is necessary to construct a system to obtain the user's situation. To obtain the usage information, the chewing movement information of user is used. Conventionally, as a method of detecting chewing, there are a method using a bone conduction microphone, and a method of detecting chewing using an acceleration sensor or a gyro sensor⁽⁷⁾. However, in order to detect chewing in contact with people, it is troublesome to wear around eyes when eating. Also it is difficult to distinguish between talking and chewing. In order to solve this problem, a method to detect mastication and utterance in non-contact using Kinect was developed⁽⁸⁾. In this method, information on mastication and utterance for each user is required before use, and there is a problem that it is necessary to perform calibration. Also, as the start command of the detection method, since we use the action of "hand" to eat when we eat, we can not use it with those obstacles in the upper limbs. Also, with the development of various diet support robots, even those with obstacles in upper limbs can now do their own meals. In this research, we use the chewing

Fig. 1. Cartesian coordinate meal support robot.

movement of the care recipient to grasp usage situation. For grasping the usage situation, use the chewing movement of the care recipient. For chewing detection, we use a Microsoft Kinect sensor, and a server is constructed so that a caretaker can obtain the use situation of a user with a mobile terminal or the like. The accuracy of the chewing detection was verified by experiments.

2. Meal Support Robot

3.1 Orthogonal Coordinating Type Meal Support Robot

Figure 1 shows the orthogonal coordinating type meal support robot. The operation departments of this robot consist of 4 section a pusher, a spoon, a plate and a shutter. The plate is designed as a platform for the chosen dish to be move by the shutter. As for the shutter, from the user viewpoint, it is designed to move downwards and placed behind the chosen dish. The pusher is a tool to push the food on the plate towards the spoon. The spoon is a tool to carry the food into the user's mouth. As each axis move independently of their translator movement, the vibration causing the food to fall can be reduced. In addition, LEDs are attached under the plate helps the user recognized the chosen food visually.

2.2 Interface

For the input device of this robot, there is various interface can be personalized depending on the user state of assistance required. In this study, to measure the movement quantity of the user's head turn, a non-contact interface device; Kinect is used. The non-contact, user's head turn movement quantity interface is explained as below. This interface is an input device that follows the face image obtained by the camera installed in Kinect and detects the rotation amount of the face. By tilting the face towards the plate on which the food user want to eat is on, the LED at the bottom of the plate lights up. They can select by turning your

face up with the LED on. The input device used in this research uses the head rotation amount interface.

2.3 Support Procedure

Explain the flow to support by meal support robot.

First, when the switch is turned on, the plate, pusher, shutter, spoon moves to the initial position. Also, the 1st to 5th plate selection LEDs arranged at the bottom of the plate light up according to the movement of the head. When the LED is lit, the plate is selected and the dish is moved so that the selected plate comes to the front of the pusher. The pusher stretches and lowers the shutter behind the food. Also, once the pusher elongates, extrude the first piece of food and place it on a spoon. When the food is put on the spoon, it is carried to user's mouth by the spoon. Also, as soon as the spoon finishes elongation, the dish returns to its initial position. After that, dishes are selected again. In addition, since the robot memorizes how many times each dish was selected, the position where the shutter descends when food is pushed out is changed. You can select the back row after feeding the front row of food for each dish.

3. Chewing Detection

3.1 Identification of Chewing And Speech

In order to obtain the use situation of the care recipient use the chewing movement. The muscles to use when user speech or user chewing are defferent. Muscles used for chewing are called mastication muscles and exist around the cheeks of the face. Since this muscle was used to move the lower jaw and crush the food, we considered to detect the chewing movement using the characteristic points of the lower jaw. However, there is a possibility that a requiring assistant speaks other than chewing as a situation where the lower jaw moves using a meal support robot. Therefore, it is necessary to distinguish between speech and chewing. When speaking, the muscles called mouth orbicular muscles move around. Therefore, there are also studies that perform speech simulation using the movement of the orbicularis muscle⁽⁹⁾. The positions of the masticatory muscle and the orbicularis muscle are shown in Fig. 2. To distinguish chewing and speech, use the characteristic points of the lower jaw moving by the masticatory muscles and the characteristic points around the mouth moved by the orbicularis muscle.

3.2 Feature Point Extraction Chewing and Speech

To extract feature points of mastication and utterance, we use the Kinect sensor Hereinafter, Kinect sensor is called Kinect. It is equipped with a distance sensor, a microphone, an image sensor, etc. Also, Kinect is equipped with "Developer" provided from "Kinect SDK v1.5" Obtained by using the library (FaceTrackLib.lib) in "Face Tracking SDK" of "Toolkit". Face tracking detects and tracks the face. The state of face recognition by Kinect is shown in Fig. 2.

3.3 Distinguish speech and chewing

Since the position of the nose in the face does not change, it is based on the midpoint under the nose are shown in Fig. 3. The place to use and the midpoint under the nose are shown in Fig. 3. If the coordinates of the midpoint of the nose (no_x, no_y) and the feature point of the mandible are (ch_{nx}, ch_{ny}), the sum W_{ch} of distance from the reference to the feature points of the lower jaw can be obtained by the following calculation formula.

Fig. 2. Facial feature points by Face Tracking.

Fig. 3. Distance between nasal and lower jaw.

Fig. 4. Characteristic points up and down, right and left in mouth.

Fig. 5. Three-dimensional graph of chewing and speech.

Characteristic points around the mouth are shown in Fig. 4. Four feature points are used from the center of the lip, up, down, left, and right. W_{pitch} is the distance between the upper and lower feature points, and W_{yaw} is the distance between the left and right feature points. Using these three pieces of information, a three-dimensional graph in speech and chewing is shown in Fig. 5.

$$W_{ch} = \sqrt{(ch_{1x} - no_x)^2 + (ch_{1y} - no_y)^2} + \sqrt{(ch_{2x} - no_x)^2 + (ch_{2y} - no_y)^2} + \dots + \sqrt{(ch_{nx} - no_x)^2 + (ch_{ny} - no_y)^2} \quad (1)$$

3.4 Use of Support Vector Machine

In this research, two classes of speaking and chewing can not be discriminated linearly because the discrimination plane is not known by the above method. In such a case, the soft margin SVM corresponds to the distribution of classes. The procedure will be described. First, face tracking using kinect is performed, feature points are extracted as shown in Fig. 2. We calculate W_{ch} , W_{yaw} , W_{pitch} from each feature point extracted in Fig. 2, and obtain 100 data each of them and analyze it. Based on this data, identification is made on the identification surface created using SVM, judgment on correctness is made, and the identification rate is calculated.

3.5 SVM Training

Training data is necessary to distinguish chewing and speech by SVM. There is a problem of flowing to a large class if there is a great bias in the training data of mastication and speech. This problem is called an unbalanced data

problem. Unify the number of mastication and speech data N so that the problem of unbalanced data does not occur. Another problem is that the skeleton is different depending on the individual, and the movement of mastication and speech is not constant. Therefore, we acquire mastication and speech of multiple people and prepare training data.

3.6 The state-of-use Recognition System

It is necessary to construct a system that can confirm the information of the requiring assistant who is using the meal support robot.

Information that a caregiver can acquire includes mastication and speech of a care recipient, date and time of use of the robot, number of times of selection, elapsed time since the start of meal, and the like. In addition, it is necessary to notify the danger according to the case where there is a change in the situation of the care recipient. In this study, as a pattern to report danger as a pattern, while the meal support robot is in operation, it is dangerous not to detect a requiring assistant or to select food, but do not move the mouth at all. We created a server to construct a system that helpers can easily obtain these information. Fig. 6 shows the created usage situation grasping system.

4. Experimental Results

4.1 Experimental Environment and Method

The experimental environment is shown in Fig. 7. As shown in Fig. 7, the distance from the Kinect to the user was fixed at 65 cm. Man sat on the front of the Kinect and set it with the eyes at the same height. The meal support robot developed in this research was able to put 3×5 mass food material on one meal, so we performed mastication and utterance 15 times each. As training data of mastication and utterance used in this experiment, three dimensional data of

Date	Situation	Time	Count
2018-1-28 20:50:36	Chewing	34	1
2018-1-28 20:50:58	Speech	57	2
2018-1-28 20:51:17	Danger	75	3
2018-1-28 20:51:33	Speech	91	4

Fig. 6. Usage situation grasping system.

mastication and utterance of three men in their twenties (W_{ch} , W_{pivh} , W_{yaw}) learned 1600 data. At this time, training data was prepared using the linear kernel with parameter C of SVM = 0.1. This experiment used bottle gum to take chewing data.

The experimental environment is shown Fig. 7 After the meal support robot supports food to the mouth of the user, the learning data is acquired for 10 seconds and the learning data is compared with the training data, whereby the probability of mastication and utterance is outputted. Therefore, in this experiment, when the probability of output chewing is 80% or more, it is chewing, and in the case of 79% or less, the utterance is taken. The reason for this is to recognize a person who is chewing as a speech and call attention to a caregiver not think that it will be a problem because it can obtain the usage situation by double as confirming the care recipient visually. However, when judging that the utterance is mastication and giving notice, the caregiver may not check the situation of the user, and there was a fear that correspondence would be delayed when there was a big trouble. Therefore, criteria for judging chewing compared to utterance were strict.

4.2 Experimental Result

The recognition rate of chewing and speech is shown in Table 1 and the probability of being chewing output by SVM is shown in Table 2. Three subjects A, B, C contain speech and chewing data in the training data.

Five of the mastication and speaching data not included in the training data were misrecognized in chewing. In the speaching, it was found that one person who contained mastication and speaching data in the training data and one person who did not contain mastication and speech misrecognized once each. Although it is understood from the experimental result that there was a misrecognition, as a whole it was able to recognize correctly at a rate of over 90% in both chewing and speech.

Fig. 7. Experiment Environment.

Table 1. Recognition rate of mastication and utterance.

Subject	Chewing	Speech
A	15/15	15/15
B	15/15	15/15
C	15/15	14/15
D	14/15	15/15
E	14/15	15/15
F	15/15	15/15
G	12/15	15/15
H	15/15	15/15
I	13/15	15/15
J	15/15	15/15
K	12/15	15/15
L	15/15	14/15
Recognition rate	94 %	99 %

Table 2. Probability of mastication output by SVM.

C:Chewing [%] S:Speech [%]

A		B		C		D		E		F		G		H		I		J		K		L	
C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S	C	S
95	32	94	59	98	54	89	27	89	2	81	40	83	6	90	49	97	47	81	45	81	65	81	50
93	34	100	38	100	50	89	1	82	35	96	12	82	7	100	70	90	44	83	38	82	26	98	50
88	33	100	58	100	75	86	2	83	0	95	10	81	16	100	57	86	45	81	16	77	1	89	37
94	36	100	65	100	81	85	3	77	1	80	11	71	12	97	50	71	52	93	22	71	47	96	54
92	16	98	69	100	39	87	0	83	1	82	11	86	7	96	62	80	36	91	25	80	20	97	69
95	40	98	53	99	63	91	20	82	3	94	10	76	12	100	63	81	44	94	20	74	22	99	51
97	16	100	57	100	51	94	6	80	9	98	10	87	16	100	41	86	33	89	19	80	22	99	68
98	50	99	65	100	76	87	58	81	20	94	10	76	17	99	49	83	44	99	65	81	18	97	63
94	48	84	35	100	65	81	1	95	0	92	25	82	11	97	22	83	14	84	23	82	26	96	65
96	33	95	40	100	71	86	21	93	3	91	19	84	36	100	40	82	55	94	33	81	1	100	52
94	29	100	60	100	77	56	5	91	0	99	17	85	22	99	74	88	28	92	4	80	2	98	54
94	48	84	36	99	74	84	15	89	9	93	17	85	19	98	62	81	78	92	26	80	15	96	68
94	33	97	69	100	46	88	12	92	2	92	10	82	19	99	31	65	37	88	40	83	7	94	82
96	42	100	53	100	73	89	5	95	1	93	14	85	28	95	31	81	54	90	29	84	15	94	72

In this experiment, even when there were many people who did not have chewing and speaking data, we found that it was able to recognize it at a high rate. When mastication was performed from Table 2, it was found that the probability of chewing as a whole is lower than those who do not have misrecognition, if the number of misrecognition is large. This is chewing while opening her mouth, I think that the cause is determined to be close to the speaking. In addition, there are individual differences in the skeleton, chewing and speech movement, so it is necessary to increase the number of people who refer to various patterns of mastication and speech and learn.

As future prospects, the distance to the face is constant at 65 cm and can not be correctly recognized unless it faces the front, so it is necessary to build a system with a high degree of freedom that is not bound by the orientation and distance of the face.

5. Conclusion

We constructed a system to obtain the user's situation when having meal. chewing and speaking were discriminated using SVM which is one type of machine learning. Experimental results showed misrecognition in mastication and speech, respectively, but there was recognition rate of 90% or more, and validity have been confirmed.

In this research, we built a server to grasp usage situation of meal support robot. In addition to information on mastication and utterance, as well as notification when a requiring caregiver is dangerous, it has become possible for a carer to grasp usage situation simply by holding a portable terminal or the like. Therefore, the user's situation can be confirmed by using the proposed method without caregiver's monitoring. The burden on caregivers are able to be relieved.

References

- (1) Ministry of Health, Labor and Welfare: 2014 welfare administration reported cases Overview , 2015, in Japanese
- (2) Ministry of Health, Labor and Welfare: Handicapped grade table, 2015, in Japanese
- (3) Yuriko Doi : General concept - the concept of QOL and the importance of QOL research , J.Natl.Inst.Public Health, Vol.53, No.3, pp. 176-177, 2004, in Japanese
- (4) J.Sijs, F.Liefhebber and G.Romer: "Combined Position & Force Control for a robotic Manipulator", IEEE 10th International Conference on Rehabilitation Robotics 2007, pp. 106-111, 2007,
- (5) Zyunya Ishiya:Meal Support Robot My Spoon, The Robotics Society of Japan Journal, pp. 44-47, 2003,in Japanese
- (6) Masayuki Yamamoto, Yoshiro Sakai, Takakazu Ishimatsu, Fumio Kasagami : Assist robot hand with bilateral human-interface, Vol.50, FNo2, p.107, 2000,in Japanese
- (7) Shuji Uno, Ryo Ariizumi, Shigeo Kaneda, Hirohide Haga : Advising the Number of Mastication by Using Bone-Conduction Microphone, The 24th Annual Conference of the Japanese Society for Artificial Intelligence, pp.1-4, 2010,in Japanese
- (8) Kokei bin, Tomo Fujii: Non-contact masticatory sensing eating environment Creating chewing log, Image laboratory, Nippon Kogyo Shuppan, pp.9-16, 2013,in Japanese
- (9) Yoshinobu Ide, Chewing of the encyclopedia, Asakura Shoten Co., Ltd., pp.77-78, 2007