A Tongue Training System for Children with Down Syndrome

Masato Miyauchi **Takashi Kimura** **Takuya Nojima**
Graduate School of Information Systems, University of Electro-Communications
1-5-1 Chofugaoka, Chofu
Tokyo, 182-8585 Japan
{m_miyauchi, tkimura}@vogue.is.uec.ac.jp, tnojima@computer.org

ABSTRACT
Children with Down syndrome have a variety of symptoms including speech and swallowing disorders. To improve these symptoms, tongue training is thought to be beneficial. However, inducing children with Down syndrome to do such training is not easy because tongue training can be an unpleasant experience for children. In addition, with no supporting technology for such training, teachers and families around such children must make efforts to induce them to undergo the training. In this research, we develop an interactive tongue training system especially for children with Down syndrome using SITA (Simple Interface for Tongue motion Acquisition) system. In this paper, we describe in detail our preliminary evaluations of SITA, and present the results of user tests.

Author Keywords
Tongue training; myofunctional therapy; Down syndrome; depth camera; tongue-computer interface

ACM Classification Keywords
H5.2 [Information interfaces and presentation]: User Interfaces. - Input devices and strategies.

General Terms
Human Factors

INTRODUCTION
Many children with Down syndrome have several symptoms such as speech and swallowing disorders. To improve these symptoms, myofunctional therapy (MFT) [1,6], which is a collection of mouth and tongue training, is considered to be beneficial. However, such training is too simple and monotonous to maintain the attention of children.

One of the MFT methods for training the tongue muscle, called the "lip tracer" method, involves protruding the tongue and then moving it along the contour of the upper lip, as if the tongue were tracing the lip. Another training method is to open the mouth with the tongue at the roof of the mouth. In addition to these methods, MFT includes a variety of training sets, such as "swallowing training" (i.e., swallowing water while opening the mouth and biting a straw), "orbicularis oris muscle and buccinator muscle training" (i.e., opening the mouth horizontally, closing it, and then inflating the cheeks). It would seem that these MFT training sets are so simple that children with Down syndrome would be able to easily train their tongues using such methods. However, this simplicity means that the training does not have any mentally stimulating or interesting aspects from the perspective of children, such as visual or acoustic effects. During the training, children simply move their tongues. This results in little motivation for children to engage in the training. In addition, to motivate children with Down syndrome to do such training, the support of their teachers or families is necessary; they must constantly check to ensure that the children are able to do the training properly.

To increase children's motivation to perform the training, and to simultaneously reduce the amount of work required to monitor and supporting the training, an interactive system to detect tongue motion is needed. For such an interactive system, tongue motion should be detected in a non-contact way. In particular, because the target users of the system are children with Down syndrome, no devices should be worn or put around the mouth (1) to avoid accidental ingestion, (2) to maintain hygiene, (3) and to reduce any unnecessary effort. We developed the SITA (Simple Interface for Tongue motion Acquisition) system to satisfy these conditions [5]. In addition, we also developed a tongue training game to better attract the attention and interest of such children.

RELATED WORK
To measure tongue motion, a number of methods have been suggested. Huo et al. proposed a tongue-computer interface using a magnet on the tongue and magnet sensors outside of the mouth [3]. The magnet is placed on the tip of the tongue, and its magnetic field changes according to the motion of the tongue, which is measured by magnet sensors around the mouth. Saponas et al. have proposed a tongue-sensing dental retainer used as a computer interface for patients with paralyzing injuries [8]. The system is able to recognize simple tongue gestures by using proximity sensors embedded on a dental retainer. Salem et al. have proposed a tongue pointing system using a pressure sensitive isometric joystick [7]. Furthermore, Slyper et al. have proposed three types of tongue input devices: a sip/puff switch, a bite sensor, and a tongue joystick [10]. However, while these research

UIST'13, October 08 - 11 2013, St Andrews, United Kingdom

http://dx.doi.org/10.1145/2501988.2502055

systems are capable of detecting various types of tongue motion, they require users to put certain devices inside and around the mouth, which runs the risk of accidental ingestion. In addition, maintaining these devices hygienically can be a burden for people around children with Down syndrome. Therefore these systems are not very suitable for use by such children. Hueber et al. have proposed a lip and tongue motion recognition system based on ultrasound and optical images [2]; originally designed as a segmental vocoder, the system is able to recognize tongue motion without installing any special devices on a user's body. However, the system requires an ultrasound sensor that is not commonly accessible to the public. Li et al. also proposed a non-contact method using an RGB camera [4]. Their method can detect six states of the tongue and mouth, but it is not able to measure continuous tongue motion, which is important in tongue training.

PRELIMINARY SYSTEM

Tongue motion recognition method

Figure 1 shows the hardware setup for our proposed SITA system. A Kinect (Microsoft) is placed 70cm in front of the user and captures the user's facial RGB image and depth data.

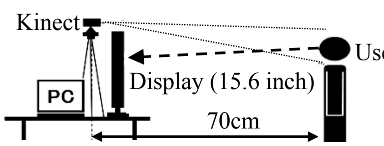

Figure 1: Hardware setup for preliminary system.

Eye recognition followed by facial recognition is applied to the captured image to search the eye area using the OpenCV Haar-like feature detection library. Then, the nose and mouth areas are determined from the structural characteristics of the face. As shown in Figure 2(a), a rectangular mouth area is placed approximately 20 pixels below the highest point of the nose area, which is placed at the center of the eye area.

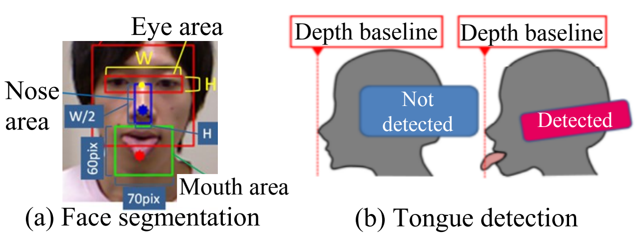

Figure 2: Basic face segmentation and tongue detection.

Inside the mouth area, the SITA system continuously seeks the highest point. If this highest point inside the mouth area is higher than that of the nose area, then SITA assumes that the tongue has been thrust out of the mouth (Figure 2(b)).

Pointing accuracy evaluation

Experimental setup

An overview of the experimental setup is shown in the left side of Figure **3**. The setup consists of the Kinect, a chin support, and an acrylic plate, though the chin support and the acrylic plate are not necessary during actual use. The chin support is used to fix the participant's head. Then the acrylic plate is placed in front of the participant in a position that can barely be reached by the participant's protruded tongue. Then the Kinect is fixed at 70 cm away from the acrylic plate. On the plate, a matrix of 4 by 7 dots is engraved at 1cm intervals(Figure **3** right, lower, "Normal view"). The participants are asked to indicate one of the dots on the plate with their tongue by using the haptic cues of their tongue. Then, we measure the position of the tip of the tongue based on our proposed method. The acrylic plate is transparent under visible light as well as infrared light, allowing the core part of the tongue motion acquisition algorithm to work properly. Before conducting the experiment, we captured the dot positions with the Kinect's RGB camera to be used as a reference. Because these dots are small, we lighted up the acrylic plate from the side to illuminate them. Then the camera is able to capture a clear image of the dots (Figure **3** right, upper, "Illuminated dots").

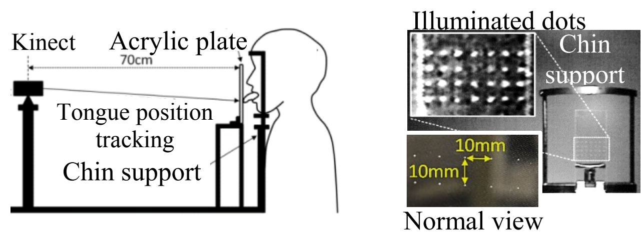

Figure 3: Left: Special setup for evaluation experiments; Right: Chin support and acrylic plate.

Accuracy evaluation procedure and results

Five males in their 20s participated in this experiment. After each participant's head was fixed onto the chin support and the Kinect was placed in an appropriate position, the participants pointed out certain dots on the plate with their tongue and maintained the motion for 5 seconds. Each point was selected randomly. No dots were selected that a particular participant would not be able to indicate with his tongue. All points were measured 3 times for each participant. The results are shown in Figure 4.

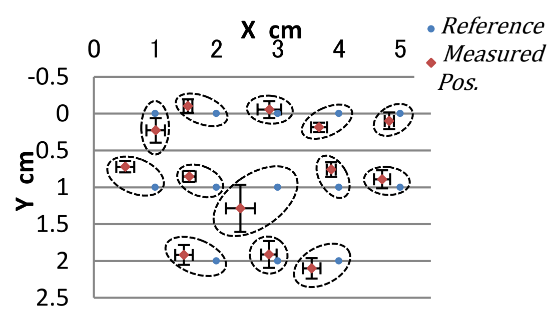

Figure 4: Accuracy of measured positions.

In the figure, it may appear that several points of data are missing. This is because some of the matrix points were too far away from the participants' tongues to comfortably indicate them, and thus these points were eliminated from the results. Nonetheless, from the obtained results, the proposed algorithm has the ability to measure the position of the tip of the tongue with an accuracy of within 0.5cm. Among the various above mentioned training exercises in the MFT, the "lip tracer" is one that requires accurate motion. It is an exercise that traces the upper lip by the tongue. A tongue motion measuring system should therefore be able to measure such motion. Considering the typical thickness of the lip (about 1cm), the accuracy of our proposed system (about 0.5cm) should be sufficient for this purpose.

Preliminary user test

Following the accuracy evaluation, a simple shooter game controlled by tongue motion was developed (Figure 5). The center green circle indicates a player (participant). The circle automatically shoots bullets at regular intervals in a specific direction controlled by the player's tongue. The game's objective is to shoot the surrounding target circles. Once all of the targets on the screen are shot and destroyed, then smaller targets appear.

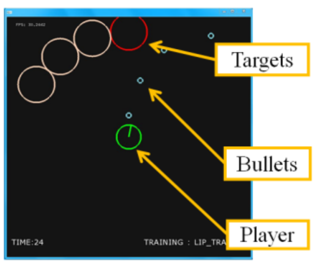

Figure 5: Screenshot of preliminary game application.

For the user test, 7 students of different types of disorders (four males and three females, including two Down syndrome) attending the Tokyo Metropolitan Chofu Special Support School were allowed to play this application freely while we and their teachers observed them. From the observations, we identified several points to be improved, as follows. (1) A larger display was preferred; small displays and small images tended to cause the children to physically approach the display, which resulted in sensor detection problems due to the sensor range. (2) Robustness with respect to the orientation of a user's face needed to be improved. In addition, the gap between the positions of the Kinect camera and the display easily caused errors in face recognition to occur. (3) The shooting game application did not seem sufficient to attract the participants' attention. Based on these observations, therefore, we developed a second version of our game system.

SECONDARY DEVELOPED SYSTEM

In response to the observations made during the first user test, we developed a second version and conducted the test again.

System components

In this second, revised version of our system, we used a more compact RGB and depth camera (Xtion PRO LIVE, ASUS). In Figure 6 left, the camera was placed in the center of the 21.5 inch wide LCD display, using a transparent mounting device, to recognize the user's face and tongue effectively.

Figure 6: Left: Hardware layout of secondary system; Right: Screenshot of the revised second application.

Improvements in tongue motion recognition method

We applied the following improvements to our tongue motion recognition method. We changed the face recognition module from the OpenCV library to Facetracker [9], to recognize the nose and mouth areas directly. At the same time, face orientation data provided by the library enabled the proposed method to adjust the face rotation (Figure 7, left). A face orientation notifier was also implemented (Figure 7, right). This notifier informs the player (participant) whether his/her face orientation is within acceptable limits. If the face position exceeds any acceptable limit, then the game is paused and a notification image is shown to notify the player to correct their face orientation.

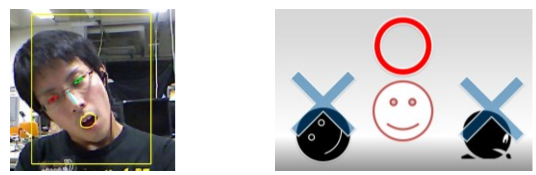

Figure 7: Left: robust recognition for rotation. Right: face orientation notifier (red circle: OK, blue cross: FAIL).

Application and user test

The objective of this revised, second application is for the user to catch fish through tongue movement (Figure 6, right). The pink circle in the center of the figure is a tongue-controlled cursor used to catch the fish. A fish can be caught if the distance between the cursor and the fish becomes less than 50 pixels. When all of the fish have been caught, a sound effect is generated and the user proceeds to the next stage. A second user test was conducted at the same special support school as the first test. The subjects were 4 children (two 5th grade (10-11 years old) girls and a boy and a girl of the 1st grade (6-7 years old)) with Down syndrome. The experiment was carried out in 3 separate days. All of the children could freely move their tongue forward, right and left, as well as up and down. The experiment duration was 3 to 5 minutes. During the experiment, the children's behavior was observed and notes were taken. The results are as follows.

Student 1 (Girl with moderate Down syndrome, 5th grade): She seemed to be very interested in playing the game. Every time she caught a fish, she smiled and gave out a scream of delight. However, sometimes the facial recognition failed because she physically approached the display. Both of her teachers agreed that she displayed substantial tongue movement and enjoyed this application. ***Student 2*** (Girl with mild Down syndrome, 1st grade): At the beginning of the trial, she was hesitating to play the game. Then we did a demonstration to show her how to play that. After the demonstration, she played the game happily and caught all the fish. Her teacher commented that she was enjoying the game and has a good understanding of the application. ***Student 3*** (Boy with moderate Down syndrome, 1st grade): He did not play the game at all. He kept his head hanged down during the experiment. His teacher commented that he is too shy to play the game in unfamiliar situations. However, he sometimes watched the display and got excited. His teacher also noted that he is sure interested in the game. ***Student 4*** (Girl with severe Down syndrome, 5th grade): Although she appeared interested in the game, she pushed her tongue forward only a few times during her 3 minute session. Her teacher noted that keeping her tongue protruded is difficult for her. In addition, she appeared distracted by the people around her.

This secondary experiment shows that the proposed system is completely safe and secure from both physic and hygienic aspects. In addition, Children with Down syndrome are able to play the game if they could understand the rules of the game. However, all the children had some sort of difficulty in understanding the rule and how to play that. Then more appropriate instructions are necessary for easier understanding. Actually, Student 1 and 2, children with moderate or mild Down syndrome, are able to understand and play the game but they need an additional short instruction from their teachers. Student 3 and 4 had more difficulty in understanding the game. In the case of Student 3, it was not because of his symptoms but because of his shyness. In his case, several trials should help him to understand the game. Adding to that, carefully designed instruction should be necessary especially for the child with severe Down syndrome, Student 4. The system encounters difficulty in measuring tongue motion when a player move his/her head and body aggressively. Thus the design of the game should be improved that could reduce unwanted player's motion, along with an improved detection algorithm.

Teacher's comments

After the conclusion of the tests, the teacher observers provided the following feedback. Several stages of the game with different levels of difficulty should be created. Each child has a different degree of ability to control their tongue. Thus, the availability of different difficulty levels which match each student's ability would help motivate them to play the game. Background images should be simplified because children can be easily distracted. To allow them to better concentrate on playing the game, simpler background images should therefore be used. Distinct screen images as well as a cursor icon that suits game context is preferable for easier comprehension of the game situation. Finally, more impressive visual and auditory effects should be added to better attract the children's attention.

Teachers also commented that this revised, second version of the game was more attractive than the first version. It was also mentioned that some of the students could use this application for their daily training exercises.

CONCLUSION AND FUTURE WORK

In this paper, the basic mechanisms of our newly developed non-contact tongue motion acquisition (SITA) system were described. Our accuracy evaluations showed that our SITA system has sufficient accuracy to be used as a game interface. Then, we developed a tongue training system for children with Down syndrome using the proposed SITA system. Through user tests, we could confirm the applicability of the SITA system as useful for tongue training. In our future work, we plan to develop more attractive and easier applications for children with Down syndrome to use in their independent play. To this end, we plan to conduct long-term user tests to evaluate the usability and training effects of our system.

ACKNOWLEDGEMENTS

We thank all of the teachers, staff, and students at Tokyo Metropolitan Chofu Special Support School, and the students' parents who supported our user tests.

REFERENCES

1. Gallerano, G., Ruoppolo, G. and Silvestri, A. Myofunctional and speech rehabilitation after orthodontic-surgical treatment of dento-maxillofacial dysgnathia. *Progress in Orthodontics 13,*1(2012), 57-68.
2. Hueber, T., Benaroya, E., Chollet, G., Denby, B., Dreyfus, G. and Stone, M. Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips. *Speech Communication 52*, 4 (2010), 288-300.
3. Huo, X., Wang, J. and Ghovanloo, M. A Magneto-Inductive Sensor Based Wireless Tongue-Computer Interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 16*, 5 (2008), 497-504.
4. Liu, L., Niu, S., Ren, J. and Zhang, J. Tongible: a non-contact tongue-based interaction technique. In *Proc. ASSETS 2012*, ACM Press, (2012), 233-234.
5. Miyauchi, M., Kimura, T. and Nojima, T. Development of a non-contact tongue-motion acquisition system. *Proc. UIST Adjunct 2012*, ACM Press (2012), 75-76.
6. Profitt, W.R. and Mason, R.M. Myofunctional therapy for tongue-thrusting: background and recommendations. *The Journal of the American Dental Association 90*, 2 (1975), 403-411.
7. Salem, C. and Zhai, S. An isometric tongue pointing device. In *Proc. CHI 1997*, ACM Press (1997), 538-539.
8. Saponas, T.S., Kelly, D., Parviz, B.A. and Tan D.S. Optically Sensing Tongue Gestures for Computer Input. In *Proc. UIST 2009*, ACM Press (2009), 177-180.
9. Saragih, J., Lucey, S. and Cohn, J. Deformable Model Fitting by Regularized Landmark Mean-Shift, *International Journal of Computer Vision 91*, 2 (2011), 200-215.
10. Slyper, R., Lehman, J., Forlizzi, J. and Hodgins, J. A Tongue Input Device for Creating Conversations, In *Proc. UIST 2011*, ACM Press (2011), 117-126.