International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2019)
Y. Kakehi and A. Hiyama (Editors)

FaceDrive: Facial Expression Driven Operation to Control Virtual Supernumerary Robotic Arms

Masaaki Fukuoka[†1] and Adrien Verhulst[‡2] and Fumihiko Nakamura[§1] and Ryo Takizawa[¶1] and Katsutoshi Masai[‖1] and Maki Sugimoto[**1]

[1]Keio University, Japan
[2]The University of Tokyo, Japan

Abstract

Supernumerary Robotic Limbs (SRLs) can make physical activities easier, but require cooperation with the operator. To improve cooperation between the SRLs and the operator, the SRLs can try to predict the operator's intentions. A way to predict the operator's intentions is to use his/her Facial Expressions (FEs). Here we investigate the mapping between FEs and Supernumerary Robotic Arms (SRAs) commands (e.g. grab, release). To measure FEs, we used a optical sensor-based approach (here inside a HMD). The sensors data are fed to a SVM able to predict FEs. The SRAs can then carry out commands by predicting the operator's FEs (and arguably, the operator's intention). We ran a data collection study (N=10) to know which FEs assign to which robotic arm commands in a Virtual reality Environment (VE). We researched the mapping patterns by (1) performing an object reaching - grasping - releasing task using "any" FEs; (2) analyzing sensors data and a self-reported FE questionnaire to find the most common FEs used for a given command; (3) classifying the FEs in FEs groups. We then ran another study (N=14) to find the most effective combination of FEs groups / SRAs commands by recording task completion time. As a result, we found that the optimum combinations are: (i) Eyes + Mouth for grabbing / releasing; and (ii) Mouth for extending / contracting the arms (i.e. a along the forward axis).

CCS Concepts

• ***Computer systems organization*** → *External interfaces for robotics; Real-time operating systems;* • ***Software and its engineering*** → *Virtual worlds training simulations;*

1. Introduction

Supernumerary Robotic Limbs (SRLs) (or in our case, Supernumerary Robotic Arms, SRAs) can support people in physical activities [All18]. The SRAs are especially useful if the user wants to: (i) reduce physical workload; and (ii) use more than 2 arms (c.f. Fig. 1 and Fig. 2). They have been adopted in healthcare [Gmb18], and in several other sectors of activities (e.g. construction [PA16], fabrication).

Although SRAs can also be operated remotely, SRAs are usually operated by the user "wearing" them (whom we will refer to as the **operator**) [LAM18], with: (i) a direct mapping of the operator's own arms to the SRAs (e.g. robot arms strapped to the operator); and / or (ii) joysticks / buttons. The cooperation between the operator and the SRAs makes physical tasks much easier. For example, in the case of ceiling tiles, the operator positions the SRAs to support a ceiling tile, and then can focus on fixing it to the ceiling [BA14].

However, cooperation with SRAs is challenging, as it requires the operator to balance his/her arms movements and the SRAs movements. Indeed the SRAs are "moving with" the operator's arms, making it difficult to keep the SRAs in place while the operator moves his/her own arms. Moreover, the SRAs cannot guess the intends behind the operator movements. Recent works have been addressing the issue by either stopping the mapping (e.g. releasing a joystick on top of the robotic arms), by remapping the robotics arms to other body parts [ABB*16, DVV*19], by recording / playing commands, and / or using AI and task-recognition [BA14, VDV*19].

In this work, we investigated the use of Facial Expressions (FEs) based commands to control virtual SRAs (here *only* 2 arms). Indeed we argue that FEs are: (i) an usable control method to give continuous or discrete commands (e.g. to move the robotic arms along an axis, to record / play a command, etc.); and (ii) able to convey "emotion-based" commands (e.g. to have a kill-switch if the oper-

† e-mail: mskifukuoka@imlab.ics.keio.ac.jp
‡ e-mail: adrienverhulst@star.rcast.u-tokyo.ac.jp
§ e-mail: f.nakamura@imlab.ics.keio.ac.jp
¶ e-mail: r-takizawa@imlab.ics.keio.ac.jp
‖ e-mail: masai@imlab.ics.keio.ac.jp
** e-mail: sugimoto@imlab.ics.keio.ac.jp

DOI: 10.2312/egve.20191275

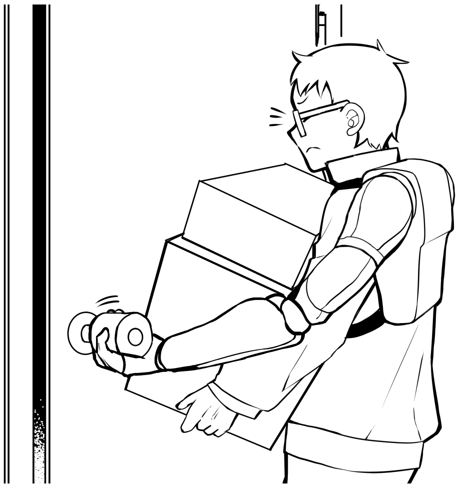

Figure 1: *The operator has his 2 hands busy with large boxes and uses a FE command to make a SRA grab the door handle.*

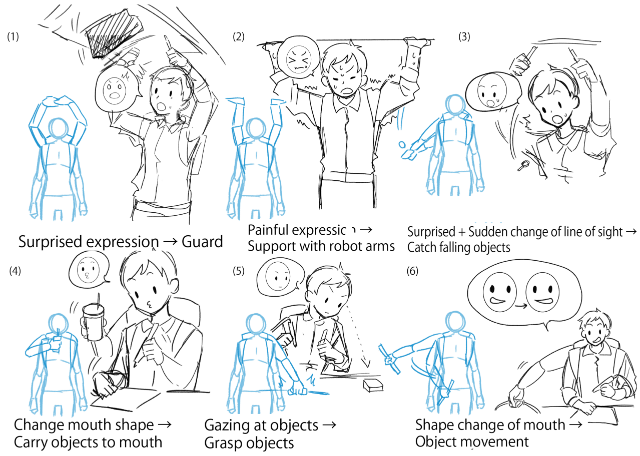

Figure 2: *From left to right: (1) Upon a surprised FE, the SRAs rise up to guard the operator against falling objects; (2) Upon a painful FE, the SRAs support the operator; (3) Using the gaze direction and a surprised FE, the SRAs try to catch the object in view; (4) Following a FE, the SRAs do a discrete action, here bringing a soda to the wearer; (5) Using the gaze direction, the SRAs grasp the object in view; (6) Following a FE, the SRAs do a continuous action, here moving alongside an axis.*

ator looks frighten, to move slowly if the operator looks tired, etc.). The Fig. 2 presents several applications of FEs and SRAs cooperation.

We used a VE with virtual SRAs instead of using real SRAs. Indeed VR is a cost-effective, safer and faster approach than developing with real SRAs. Since the HMD is covering the face (and a RGB camera cannot be used to get FEs), here facial movements are measured with 16 reflective optical sensors built-in the HMD, and classified in FEs by using a SVM (c.f. Sec. 3).

After this work, we plan to use reflective optical sensors built-in glasses [MSO*15] with the same sensing method than the HMD in order to adapt real SRAs environment.

The novelties of this work are:

- We used FEs commands to control SRAs, as a novel attempt to connect robots actions to human motions and emotions;
- We mapped a set of FEs to a set of SRAs actions and researched the optimum combinations of FEs commands / SRAs actions for a reaching / grabbing / releasing task (c.f. Sec. 4).

2. Related Work

2.1. Motion Based Control Method of Robot Arm

Among the numerous methods to control a robot arm, some rely on recognizing known human motions (e.g a hand gesture) and using them as a control method by associating them to a robot actions. This type of control method allows low dof commands (e.g. a set of 5 hand gestures) to control a high dof robot, and are used to facilitate Human-Machine Interaction (HMI).

Rogalla et al. [REZ*02] developed a control method using verbal and / or hand gestures based commands for a one arm robot. The gestures (recognized with a RGB camera and a contour-curvature approach) were used as a set of commands, such as: confirmation, stop, triggering scripts, etc. Similarly, Jindai et al. [JSYW06] proposed a control method using verbal and hand gestures (also recognized with a RGB camera and a contour-curvature approach) based commands for a "handing over" robot. Here the commands adapted the robot handing velocity and position. Lately, Tsarouchi et al. [TAM*16] proposed a method for simplifying one arm robot programming using human motions. They associated a set of body / hand gestures (recognized respectively thanks to the Kinect / Leap Motion API) to a set of programming instructions, and tested it in a case study where they programmed a one arm robot to do a picking up task.

While useful, these control methods have the issue to be somewhat difficult and time consuming to setup, as it is necessary to "develop" the set of human motions ("develop", as in write them in the software). Niwa et al. [NIAM12] developed a simple yet effective approach to this issue, called the "Tumori control":

1. The operator looks at the robot doing a recorded action (e.g. goes to the right);
2. At the time of a robot motion, the operator does an action (e.g. joystick tilts to the right);
3. The operator's action is now associated to the robot's action (e.g. here if the joystick tilts to the right, then the robot goes to the right).

2.2. "Alternative" Mapping of Supernumerary Robotic Arms

As mentioned previously, we are using FEs to control SRAs. One of the excepted output of our work is to investigate further user behaviour when controlling robots with body parts. Previous studies presented specifics mapping of SRAs to research the impact of body control over user behaviour / cognition / performance. We present some of them below.

Saraiji et al. [SSM*18], used teleoperated SRAs (= no mapping between the operator arms and the SRAs). The teleoperator had video feedback thanks to a 3 dof camera mounted on the SRAs and was able to teleoperate the SRAs. Sasaki et al. [SSF*17] controlled SRAs with the operator's legs (= "not a natural mapping" between the operator legs and the SRAs). Feuchtner et al. [FM17] used the well known long arm illusion [KNSVS12] to manipulate remote objects (= "not a 1:1 mapping" between the operator arms and the SRAs) in AR. A virtual long arm reached real far away objects (e.g. a rotating panel about 2 meters away), which were remotely activated when the SRAs "touched" them (e.g. the panel rotates as the virtual hand "touches" it).

2.3. User Interface Using Facial Expressions

FEs recognition techniques have been widely studied. It is common to recognized them thanks to an RGB camera [NNVW15] and IA classification (such as SVM). For precise measurement (in healthcare especially), it is also possible to use electromyography sensors [ABB*04]). Once reicgnozed, FEs can be used either to give some insight into the user's physiological / emotional state, or as input for HMI. We will focus on the later.

Ciftci et al. [CZY17] recognized mouth gestures (7 at most) when wearing a HMD (using a camera to record the mouth, a 3D edge map to detect the mouth area and a SVM to classify it as a gesture) and used them for interacting with a small VR application (move a player with a stretch mouth, eat an object with a smile, etc.). Matthies et al [MSU17] detected FEs (5 at most: smile, look away, turn head, open mouth, mouth in "shh", using facial muscles with an electrode-mounted earphone), and used them to control a smartphone application.

On the other hand, Masai et al. [MSO*15] used reflective photo-sensors embedded in an HMD (i.e. the "AffectiveWear"), and classified the sensors data into FEs (8 at most, c.f. Sec. 3). Reflective photo-sensors are smaller, have lower latency and lower power consumption than the RGB cameras used to recognize FEs. Transon et al. [TVN*17] evaluated FEs interactions inside a HMD (using an "AffectiveWear" approach), and showed that, for short time sessions (< 1 min), there was no significant differences in self-reported workload and usability when using FEs compared to pressing a button, even when used intensively.

3. Implementation

To operate SARs with FEs, we developed a control method using the operator's FEs to command virtual SARs (c.f. Fig. 3. We present the implementation of our system in this section.

3.1. Reflective Photo Sensor Embedded in the HMD

We used the "AffectiveWear" approach to capture FEs when wearing a HMD. Here, a set of 16 reflective photo-sensors is fixed inside an Oculus CV1 (c.f. Fig. 4). The distance value d of every sensors is calculated (c.f. Eq 1) and fed to a set of SVM classifiers (each of them trained to correspond to a FE, such as raising eyes, moving left cheek, etc.). The highest confident classifier predicts the corresponding FE [SNO*17].

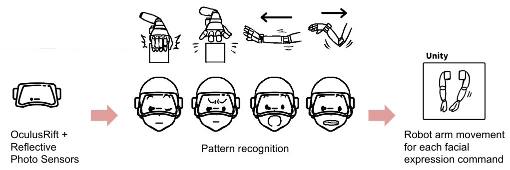

Figure 3: *System execution. A FE commands the virtual robot arm.*

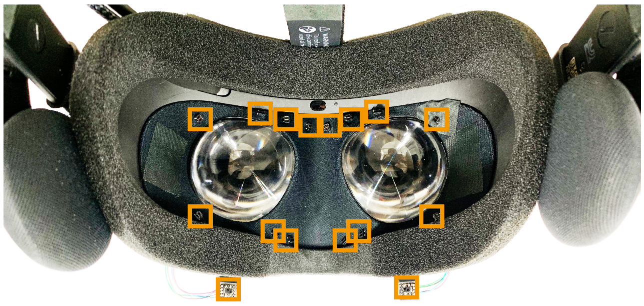

Figure 4: *The Oculus CV1 with the placement of every photo-sensor. Please note the photo-sensors at the bottom of the HMD, able to detect the skin surface around the user's mouth (and therefore better detect FEs done with his/her mouth).*

In more detail:

The distance value d can be approximated by the following equation (c.f. Eq 1):

$$d \approx (\frac{\alpha}{v})^{-\gamma} \tag{1}$$

The reflection value r of a sensor depends of the system characteristics (i.e. the constants α and γ, which are estimated by least square regression) and the distance sensor-object surface d (here sensor-skin surface). r can be approximated by the following equation (c.f. Eq 2):

$$r \approx \frac{\alpha}{d^{\gamma}} \tag{2}$$

To identify a FE from d we use a SVM with a one-vs-rest approach [Bis06], where making a decision means applying all classifiers to an unseen sample x and predicting the label k for which the corresponding classifier reports the highest confidence score (c.f. Eq 3):

$$y(X) = \max_{k} y_k(X) \tag{3}$$

We also calculated the probability of the identification results belonging to each class. Platt [Pla99] has proposed fitting a logistic sigmoid to the outputs of a previously trained SVM. Specifically, the conditional probability is assumed to be of the form (c.f. Eq 4) [Bis06]. The values for the parameters A and B are defined by minimizing the cross-entropy error function defined by a training set consisting of pairs of values $y(x_n)$ and t_n.

$$p(t=1|X) = \sigma(Ay(X)+B) \tag{4}$$

Also, $y(X)$ is the discriminant function represented by the Eq 5 ($\phi(X)$ is the feature space conversion function [Bis06]).

$$y(X) = W^T\phi(X)+b \tag{5}$$

3.2. "Tumori Control" Method

The task we developed (c.f. Sec. 4) uses 4 SRAs actions: (i) extending / contracting a SRA to reach an object (done by changing the elbow angle); and (ii) opening / closing the SRA hand to grab an object. To associate these actions to FEs, we used the "Tumori control" method (cf. Sec.2.1):

1. We playback SRA actions in the VE (here "reaching" actions and "grab" actions);
2. When the SRA does an action, the operator does a FE to associate the action to a FE, c.f. Fig. 5 (e.g. when the SRA extends to reach an object, the operator smiles, thus associating the "reaching action" to his/her smile).

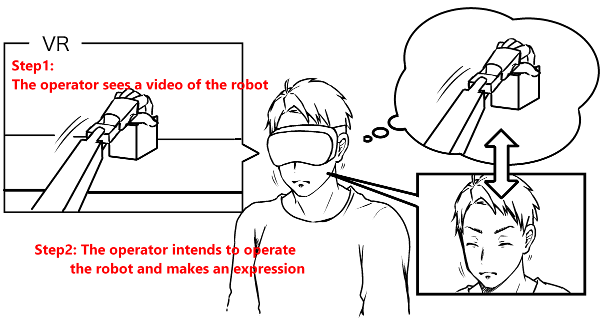

Figure 5: *An example of "Tumori control", the operator sees the SRA grab an object in the VE, and at the same moment does the FE he wants to associate to the "grab action"*

The "Tumori control" method work especially well between 1-1 associations (here 1 robot arm operation is linked to 1 FE). While this is an issue for advanced controls, here we are targeting simple controls (moreover FEs on their own must be kept simple to be done effectively). In our system, if there is a change of FE when a motion is executed, then the motion is automatically associated to the FE at the time of the expression change).

3.3. VE for linking robot arms and FEs

The VE was built in Unity 2019 and displayed in a Oculus Rift CV1. There are 2 SRAs localized on each side of the avatar's waist (c.f. Fig. 6). Since the operator can only control 1 SRA at a time, the right / left SRA is controllable when the operator is looking at respectively his/her right / left side (when grabbing an object, the right / left switching is automatically turned off to easily move the object). The "controlled" SRA follows the head direction thanks to Inverse Kinematics (IK). The IK, implemented with Unity's FinalIK plugin, targets a transform positioned about 1.5 meters in front of the HMD.

The VE receives the operator's current FE every 15ms from our FEs recognition system (sent by UDP). In addition, the FE commands (e.g. opening / closing) and the elapsed time since the beginning of the task are displayed against a wall in front of the operator.

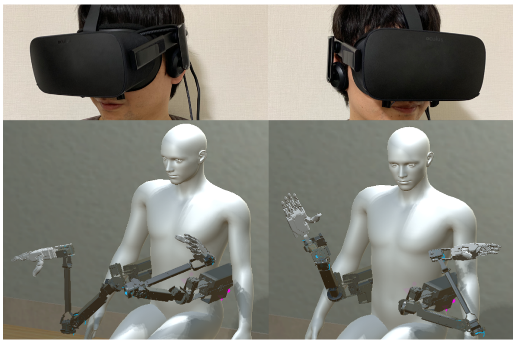

Figure 6: *The operator's avatar and the SRAs follow the operator's head direction 8only 1 SRA follow the head / can be commanded at a time).*

4. Study 1 - Assigning Facial Expressions to Supernumerary Robotic Arms Actions

We conducted a data collection study to know which FEs the participants chose to command a SRA to extend / contract its arm and open / close its hand.

4.1. Experiment Design

10 participants (invited one-by-one) joined the study: 7 males (age: M = 32.6, SD = 13.1) and 3 females (age: M = 41.3, SD = 22.5). Each of them received an explanation of the system and the task, and then put on the HMD. Throughout the study, they were helped by the experimenter and explicitly told to stop if there was any sign of discomfort (such as dizziness). Once in the VE (c.f. Sec. 3.3), after a quick calibration, participants looked at a SRA actions (i.e. extending / contracting arm and opening / closing hand) and were asked for each actions to associate the action to a FE of their choice by making said FE, cf. Sec 3.2). We collected 2,000 sensor data samples of every FEs (to find out the FEs used and also to prepare a "global" classifier).

After associating every SRA actions to a FE, the participants performed a cube reaching - grabbing - moving task (i.e. they had to

Table 1: *FEs associated to SRA actions for each participant*

	Close hand	Open hand	Extend arm	Contract arm
1	Eyes+Mouth	Eyes+Mouth	Mouth	Mouth
2	Mouth	Eyes	Cheek	Cheek
3	Eyes+Mouth	Eyes+Mouth	Mouth	Mouth
4	Eyes+Mouth+Cheek	Eyes+Mouth+Cheek	Eyes+Mouth+Cheek	Eyes+Mouth+Cheek
5	Eyes+Mouth+Cheek	Eyes+Mouth+Cheek	Eyes+Mouth+Cheeks	Eyes+Mouth+Cheek
6	Eyes+Eyebrow	Eyes+Eyebrow	Mouth	Mouth
7	Eyes+Eyebrow	Eyes+Eyebrow	Mouth	Mouth
8	Eyes+Eyebrow	Eyes+Eyebrow	Eyes	Mouth
9	Eyes	Mouth	Eyes	Mouth
10	Cheek+Mouth	Cheek+Mouth	Eyes	Eyes

Table 2: *Command set used by 2+ participants*

	Close hand	Open hand	Extend arm	Contract arm
Group 1	Eyes+Eyebrow	Eyes+Eyebrow	Mouth	Mouth
Group 2	Eyes+Mouth+Cheek	Eyes+Mouth+Cheek	Eyes+Mouth+Cheek	Eyes+Mouth+Cheek
Group 3	Eyes+ Mouth	Eyes+ Mouth	Mouth	Mouth

reach for a sphere positioned at a point A by extending a SRA arm, then grab the sphere by closing the SRA hand, move the sphere inside a cube positioned at a point B and release it by opening the SRA hand, c.f. Fig 7). Then, they removed the HMD and answered the "FE questionnaire" regarding the FEs they chose (c.f. Fig 8), before exiting the study room.

4.2. Results

Here, we show the results of the data collection to find out which FEs and FEs combination the participants chose to command a SRA.

4.2.1. Frequent Pattern of FEs combination

The Tab. 1 shows the FEs used to command a SRA. If FEs were used simultaneously, all were listed. Also, we do not separate left and right FEs.

4.2.2. Pattern analysis of FEs combination

Following the results of the Tab. 1 and our own observations, we combined the FEs used by 2+ participants (c.f. Tab. 2) and grouped them into 3 FEs groups.

5. Study 2 - Operational Efficiency for Each FEs Group

Based on the result of the study 1, we constructed 3 groups of FEs to command SRA (c.f. Tab. 2). In this study, we compared the 3 groups in terms of task completion time, time that each FE group was used. And we did the System Usability Scale(SUS) [GBKP13] just one time in order to investigate the whole systems usability.

5.1. Experiment Design

14 participants (invited one-by-one) joined the study: 12 males (age: M = 28.83, SD = 6.75) and 2 females (age: M = 25.5, SD = 0.7). Just like in the previous study, participants received an explanation on the system and the task, put on the HMD and were helped by the experimenter. Once in the VE, after a quick calibration, the participants did the following:

- The participants did the 4 FEs of a group of FEs (c.f. Tab. 2), as well as a neutral FE. The participants repeated each FE until at least 2.000 samples were recorded;
- The experimenter asked the participant to make each of those 4 FEs at least 10 times, to verify the accuracy of the classifiers (we obtained 89.7% of recognition in average);
- The participants then did the reaching - grabbing - release task (c.f. Sec. 4) 3 times, 2 times as a training, and 1 last time where his/her performance was recorded for further analysis (cf. Fig. 9).

The participants repeated this process for every group of FEs, here 4 times (to avoid order effect, the order of the group of FEs changed between participants). Then, they removed the HMD and answered a SUS questionnaire [GBKP13] regarding the use of FEs to command a SRA, before exiting the study room.

5.2. Result

Here, we show the performance and SUS result to know which group of FEs is the most efficient / usable.

5.2.1. Analysis of Task Completion Time for Each Combination of FEs

We calculated the time of task completion for each group of FEs (Fig. 9). Note that we removed 1 data set because the accuracy of its FEs recognition decreased during the study (the HMD moved) and commanding the SRAs became too demanding.

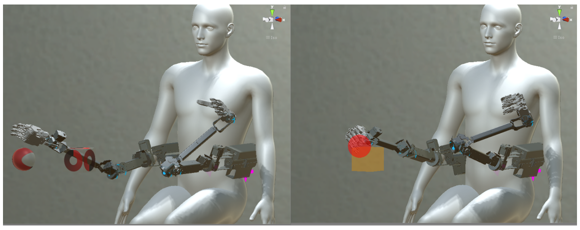

Figure 7: *The reaching / grabbing / releasing task. Left: the participant is going to grab a sphere. Right: the participant is going to release the sphere inside a cube.*

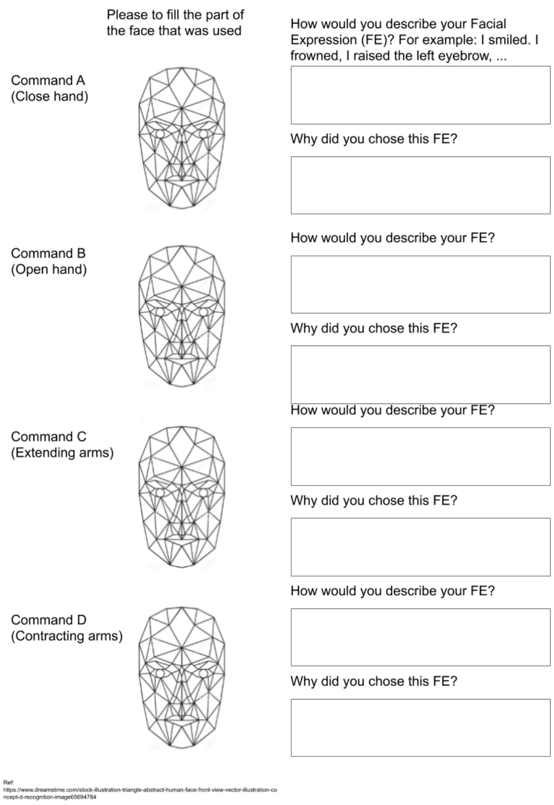

Figure 8: *The FE questionnaire focus on the facial regions [Sig16] used for each commands (translated in English from ANONYMOUS LANGUAGE for this submission) as well as the reasons the participant chose them. We combined the results with the sensors data samples collected.*

We found a significant difference in task completion time between the group of FEs with ANOVA: F (2, 36) = 3.259, **p <.046**. Also, As a result of multiple comparison by Fishers Least Significant Difference (LSD) test, it was found that Group 1 has significance longer average task completion time than Group 2 and Group 3 (as seen in the Fig. 9).

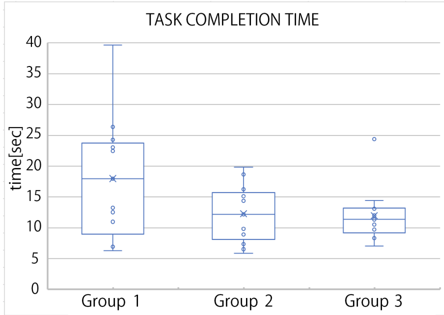

Figure 9: *Box plot of task completion time of the study 2.*

5.2.2. Usability Evaluation

We asked each participant to fill a SUS questionnaire relatif to the use of FEs to command a robot arm, SUS score: $M = 69.28; SD = 18.92$. We show the response histogram of SUS contents in Fig. 10.

6. Discussion

In many case the same facial region was used to command 2 similar actions (e.g. the mouth was use for both extending and contracting the arm). Indeed participants associated similar actions with similar FEs. The other way around, participants did not associate non similar actions with similar FEs (except some, like participant 9).

We observed 2 limits to the use of FEs inside a HMD: (1) *a material one* - The HMD made it difficult to move the upper half of the face, therefore participants preferred to use the lower half (i.e. the mouth); and (2) *a physical one* - Fine tuned eyebrows-based FEs are difficult to do [Rin84] (or even impossible for some people), therefore participants moved their eyebrows simultaneously with their eyes. Also, we observed that after participants issued a command by opening their mouth, they often forgot to close it. Since the mouth was kept open. it increased the FEs false recognition rate. Moreover, some participants had difficulty to open their mouth wide, it made it difficult to classify the FE and impacted their task completion time.

Regarding FEs, when moving the lower half of the face (e.g. mouth), the left side moves following commands from the right side of the brain, and the right side moves following commands from the left side of the brain. But when moving the eyebrows, the brain commands are transmitted to both the left and right eyelid. Thus, even if people try to move only one eyelid, the other one moves in conjunction. Raising the corner lowers the lower eyelid and narrows the closing range of the upper eyelid. So it becomes easier to wink. This is considered to be the reason some people used the eyes and other parts at the same time.

In this study, we investigated a correspondence between basic SRA actions and FEs using VE in order to survey SRA manupulation method. In the next stage, verification in a real environment will be performed using a glasses-type device (c.f. Sec. 1). The potential gap of adapting real SRAs are as follow:

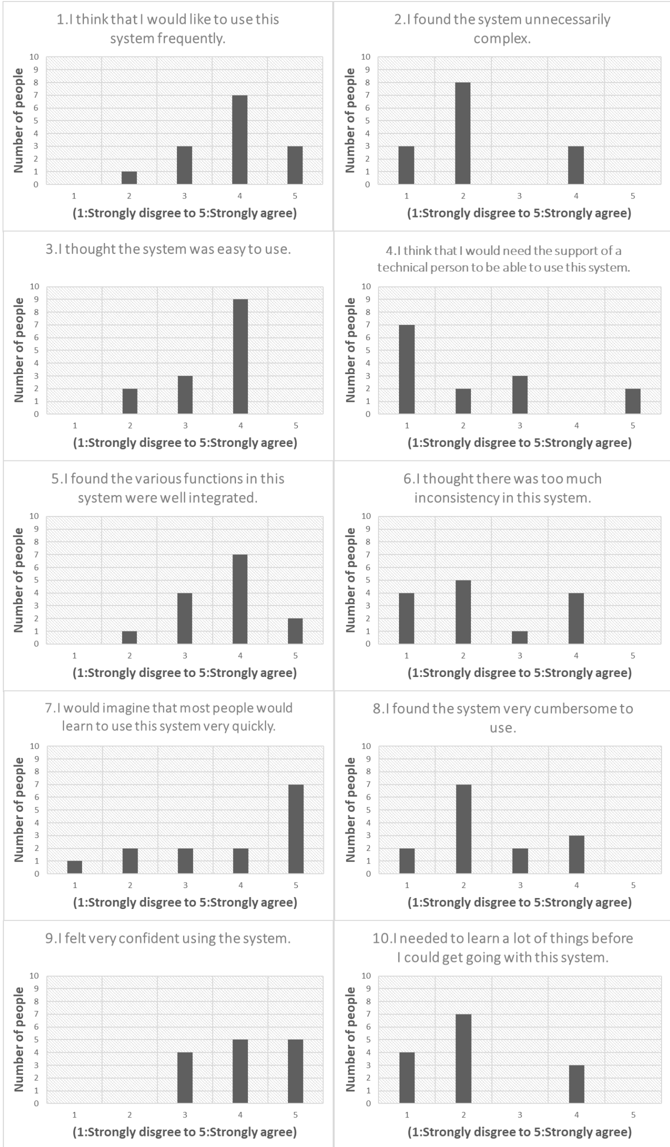

Figure 10: *SUS result*

- Deterioration of FEs detection accuracy due to device misalignment.
- The influence of the social situation (situations where the user should not laugh or should not frown under the influence of interpersonal relationships) could changes FEs associated with SRAs action.

There is far less variation in the task completion time in the Group 3 (Eyes + Mouth) than in the Group 1 (Eyes + Eyebrow), likely because the Group 3 have FEs geometrically close to each other (i.e. the mouth is shared between the actions in the Group 3), making it more natural to use. This suggest that it is more efficient to use FEs close to each other (in term of facial geometry)

In terms of system usability, the evaluation varies widely among individuals (in average, it is slightly above average). Again, it is likely because of the HMD made it difficult to move the upper half of the face, some FEs also, are more tiring to do than others.

7. Conclusion

In this study, we proposed a SRA control method using FEs. The purpose of the studies were to find out how FEs can be associated to SRAs actions. We realized those studies in a VE, with virtual SRAs and an AffectiveWear setup (i.e. here an HMD with 16 reflective photo-sensors able to classify FEs with a SVM).

The FEs were then associated to SRAs actions, such as opening / closing the hand. We examined people preferences in associating a given FE to a given SRA action in a reaching - grabbing - releasing task (c.f. 4.2. We then grouped the FEs into relevant group (based on people's preferences) recorded their performance (i.e. completion time) in a reaching / grabbing / releasing task.

We observed that there is lower variance in task completion time when using the eyes and the mouth simultaneously rather than using the eyes and the mouth separately. We also observed that the operator often forgets to close his/her mouth, leading to FE false recognition. Based on these findings, when applying FEs manipulation to a SRA in a real environment, we will consider associated similar FEs to similar SRA actions, as well as "linking" eyes-based FEs to mouth-based FEs. Furthermore, while here we mainly dealt with conscious change of FEs, but in future work, we will investigate the possibility of using SRAs under unconscious FE changes, such as surprise.

8. Acknowledgments

This work was partially funded by INAMI JIZAI Body Project, ERATO, JST (Grant No. JPMJER1701).

References

[ABB*04] Ang L. B. P., Belen E. F., Bernardo R. A., Boongaling E. R., Briones G. H., Coronel J. B.: Facial expression recognition through pattern analysis of facial muscle movements utilizing electromyogram sensors. In *2004 IEEE Region 10 Conference TENCON 2004.* (Nov 2004), vol. C, pp. 600–603 Vol. 3. doi:10.1109/TENCON.2004.1414843. 3

[ABB*16] Abdi E., Burdet E., Bouri M., Himidan S., Bleuler H.: In a demanding task, three-handed manipulation is preferred to two-handed manipulation. *Scientific Reports 6* (2016), 1–11. URL: http://dx.doi.org/10.1038/srep21758, doi:10.1038/srep21758. 1

[All18] Allison D.: Delta air lines may outfit some employees with 'wearable robotics', Mar 2018. URL: https://www.cnbc.com/2018/03/23/delta-air-lines-may-outfit-some-employees-with-wearable-robotics.html. 1

[BA14] Bonilla B. L., Asada H. H.: A robot on the shoulder: Coordinated human-wearable robot control using Coloured Petri Nets and Partial Least Squares predictions. In *2014 IEEE International Conference on Robotics and Automation (ICRA)* (may 2014), IEEE, pp. 119–125. URL: http://ieeexplore.ieee.org/document/6906598/, doi:10.1109/ICRA.2014.6906598. 1

[Bis06] BISHOP C. M.: *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. 3, 4

[CZY17] CIFTCI U. A., ZHANG X., YIN L.: Partially occluded facial action recognition and interaction in virtual reality applications. In *2017 IEEE International Conference on Multimedia and Expo (ICME)* (July 2017), pp. 715–720. doi:10.1109/ICME.2017.8019545. 3

[DVV*19] DROGEMULLER A., VERHULST A., VOLMER B., THOMAS B., INAMI M., SUGIMOTO M.: Remapping a third arm in virtual reality. In *Osaka, IEEE VR 2019: the 26th IEEE Conference on Virtual Reality and 3D User Interfaces* (02 2019). 1

[FM17] FEUCHTNER T., MÜLLER J.: Extending the body for interaction with reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2017), CHI '17, ACM, pp. 5145–5157. URL: http://doi.acm.org/10.1145/3025453.3025689, doi:10.1145/3025453.3025689. 3

[GBKP13] GRIER R. A., BANGOR A., KORTUM P., PERES S. C.: The system usability scale: Beyond standard usability testing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 57*, 1 (2013), 187–191. URL: https://doi.org/10.1177/1541931213571042, arXiv:https://doi.org/10.1177/1541931213571042, doi:10.1177/1541931213571042. 5

[Gmb18] GMBH S. S. D. H.: Helping people walk with exoskeletons | the growth of wearable robots in healthcare, Jul 2018. URL: https://www.dr-hempel-network.com/digital-health-technolgy/wearable-robots-in-healthcare/. 1

[JSYW06] JINDAI M., SHIBATA S., YAMAMOTO T., WATANABE T.: A study on robot-human system with consideration of individual preferences. *JSME International Journal Series C Mechanical Systems, Machine Elements and Manufacturing 49*, 4 (2006), 1033–1039. doi:10.1299/jsmec.49.1033. 2

[KNSVS12] KILTENI K., NORMAND J.-M., SANCHEZ-VIVES M. V., SLATER M.: Extending body space in immersive virtual reality: A very long arm illusion. *PLOS ONE 7*, 7 (07 2012), 1–15. URL: https://doi.org/10.1371/journal.pone.0040867, doi:10.1371/journal.pone.0040867. 3

[LAM18] LEIGH S.-W., AGRAWAL H., MAES P.: Robotic symbionts: Interweaving human and machine actions. *IEEE Pervasive Computing 17* (04 2018), 34–43. doi:10.1109/MPRV.2018.022511241. 1

[MSO*15] MASAI K., SUGIURA Y., OGATA M., SUZUKI K., NAKAMURA F., SHIMAMURA S., KUNZE K., INAMI M., SUGIMOTO M.: Affectivewear: Toward recognizing facial expression. In *ACM SIGGRAPH 2015 Posters* (New York, NY, USA, 2015), SIGGRAPH '15, ACM, pp. 16:1–16:1. URL: http://doi.acm.org/10.1145/2787626.2792632, doi:10.1145/2787626.2792632. 2, 3

[MSU17] MATTHIES D. J. C., STRECKER B. A., URBAN B.: Earfieldsensing: A novel in-ear electric field sensing to enrich wearable gesture input through facial expressions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2017), CHI '17, ACM, pp. 1911–1922. URL: http://doi.acm.org/10.1145/3025453.3025692, doi:10.1145/3025453.3025692. 3

[NIAM12] NIWA M., IIZUKA H., ANDO H., MAEDA T.: Tumori control : Manipulate robot with detection and transmission of an archetype of behavioral intention. *Transactions of the Virtual Reality Society of Japan 17*, 1 (2012), 3–10. doi:10.18974/tvrsj.17.1_3. 2

[NNVW15] NG H.-W., NGUYEN V. D., VONIKAKIS V., WINKLER S.: Deep learning for emotion recognition on small datasets using transfer learning. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (New York, NY, USA, 2015), ICMI '15, ACM, pp. 443–449. URL: http://doi.acm.org/10.1145/2818346.2830593, doi:10.1145/2818346.2830593. 3

[PA16] PARIETTI F., ASADA H.: Supernumerary Robotic Limbs for Human Body Support. *IEEE Transactions on Robotics 32*, 2 (2016), 301–311. doi:10.1109/TRO.2016.2520486. 1

[Pla99] PLATT J. C.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *ADVANCES IN LARGE MARGIN CLASSIFIERS* (1999), MIT Press, pp. 61–74. 4

[REZ*02] ROGALLA O., EHRENMANN M., ZÃŰLLNER R., BECHER R., DILLMANN R.: Using gesture and speech control for commanding a robot assistant. In *Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication* (Sep. 2002), pp. 454–459. doi:10.1109/ROMAN.2002.1045664. 2

[Rin84] RINN W.: The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological bulletin 95* (02 1984), 52–77. doi:10.1037//0033-2909.95.1.52. 6

[Sig16] SIGOVA M.: Triangle abstract human face, front view. vector illustration stock vector - illustration of identification, biometric: 65694784, Jan 2016. URL: https://www.dreamstime.com/kluvdia_latest-illustrations-vectors-clipart_pg1. 6

[SNO*17] SUZUKI K., NAKAMURA F., OTSUKA J., MASAI K., ITO Y., SUGIURA Y., SUGIMORO M.: Affectivehmd: Facial expression recognition and mapping to virtual avatar using embedded photo sensors. *Transactions of the Virtual Reality Society of Japan 22*, 3 (2017), 379–389. doi:10.18974/tvrsj.22.3_379. 3

[SSF*17] SASAKI T., SARAIJI M. Y., FERNANDO C. L., MINAMIZAWA K., INAMI M.: Metalimbs: Multiple arms interaction metamorphism. In *ACM SIGGRAPH 2017 Emerging Technologies* (New York, NY, USA, 2017), SIGGRAPH '17, ACM, pp. 16:1–16:2. URL: http://doi.acm.org/10.1145/3084822.3084837, doi:10.1145/3084822.3084837. 3

[SSM*18] SARAIJI M. Y., SASAKI T., MATSUMURA R., MINAMIZAWA K., INAMI M.: Fusion: Full body surrogacy for collaborative communication. In *ACM SIGGRAPH 2018 Emerging Technologies* (New York, NY, USA, 2018), SIGGRAPH '18, ACM, pp. 7:1–7:2. URL: http://doi.acm.org/10.1145/3214907.3214912, doi:10.1145/3214907.3214912. 3

[TAM*16] TSAROUCHI P., ATHANASATOS A., MAKRIS S., CHATZIGEORGIOU X., CHRYSSOLOURIS G.: High level robot programming using body and hand gestures. *Procedia CIRP 55* (2016), 1 – 5. 5th CIRP Global Web Conference - Research and Innovation for Future Production (CIRPe 2016). URL: http://www.sciencedirect.com/science/article/pii/S2212827116309544, doi:https://doi.org/10.1016/j.procir.2016.09.020. 2

[TVN*17] TRANSON A., VERHULST A., NORMAND J.-M., MOREAU G., SUGIMOTO M.: Evaluation of facial expressions as an interaction mechanism and their impact on affect, workload and usability in an ar game. In *2017 23rd International Conference on Virtual System & Multimedia (VSMM)* (2017), IEEE, pp. 1–8. 3

[VDV*19] VOLMER B., DROGEMULLER A., VERHULST A., VOLMER B., THOMAS B., INAMI M., SUGIMOTO M.: Towards robot arm training in virtual reality using partial least squares regression. In *Osaka, IEEE VR 2019: the 26th IEEE Conference on Virtual Reality and 3D User Interfaces* (03 2019). 1