International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2017)
T. Huang (Editor)

Poster

Can Face Swapping Technology Facilitate Mental Imagery Training?

Haruka Matsumura[1], Hironori Watanabe[2], Tai Chih Chen[2], Takafumi Taketomi[1], Yasuhide Yoshitake[2], Alexandor Plopski[1], Christian Sandor[1] and Hirokazu Kato[1]

[1]Nara Institute of Science and Technology, Japan
[2]National Institute of Fitness and Sports in KANOYA, Japan

Abstract

In this research, we conducted a preliminary study to investigate the effectiveness of face swapping technology for mental imagery training. To confirm its effectiveness, we used transcranial magnetic stimulation for measuring motor evoked potential (MEP) as brain excitability during mental imagery training. In the experiment, we used three motions: wrist dorsiflexion as an easy-to-perform motion, and pen spinning and baoding balls rotating motions as difficult motions. In each target motion, we compared MEPs when watching own motion video, another person's motion video, and another person's motion video with the face swapped with own face. The results showed that there was a difference between MEPs in difficult motion video observations. Watching another person's motion video with face swapping showed higher MEP than simply watching another person's video.

CCS Concepts

•***Computing methodologies*** → ***Mixed / augmented reality;*** *Image processing;*

1. Introduction

Mental imagery training is a motor learning method that utilizes imagination in order to learn a particular motion. For example, in sports training, athletes usually observe an expert's motion to improve their performance during training. In the neurophysiology research field, it is known that brain activity can be observed from a person watching a video of an athlete performing a motion, even if the said person is not performing the actual motion [RC04]. For this reason, mental imagery training has been known to facilitate motor performance. In addition, it has been reported that observing one's own video produces stronger brain activity than observing another person's motion video [KBRF10]. From these physiological backgrounds, we make the hypothesis that "it becomes easy to imagine one's own motion by observing another person's motion video but with the face swapped with one's own face, and as such, face swapping facilitates athlete's motor learning." On the other hand, in the computer vision research field, face swapping technologies have been proposed to generate a natural face swapping result [BKD*08, DSJ*11]. Recently, this technology has been widely used in the mixed reality application domain such as SNOW [SNO].

In this study, we investigate an application of face swapping in a mental imagery learning scenario. In order to confirm the effectiveness of face swapping, we conducted a preliminary experiment. In the experiment, we measured motor evoked potential (MEP) as a brain activity in the motor cortex by using transcranial magnetic stimulation (TMS) method while watching videos. Differences of MEPs are measured between own video, another person's video, and another person's video with face swapping. In addition, we also investigated a method for quantitatively evaluating the quality of face swapping results. To the best of our knowledge, no research has been done yet on evaluating the quality of face swapping results quantitatively.

2. Face Swapping Algorithm

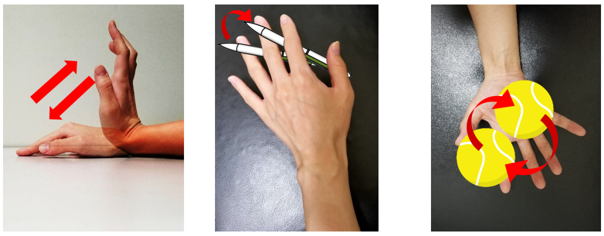
The purpose of the preliminary study is to confirm the effectiveness of face swapping for mental imagery training. Therefore, we use a simple configuration; a person sits in front of a camera, and then he/she performs a finger or arm motion. For this reason, in the experiment, the face swapped video is generated using a simple algorithm. The face swapping algorithm consists of an initialization process and a sequential process.

Initialization Process: In the initialization process, face part location differences $\boldsymbol{d}_i$ between a source image (trainee's face image) I_{src} and a target video (expert's face image) I_{dst} are calculated as follows.

$$\boldsymbol{d}_i = \left(\boldsymbol{A}_{fit}\boldsymbol{p}_i + \boldsymbol{b}_{fit}\right) - \boldsymbol{q}_i \quad (1)$$

where $\boldsymbol{A}_{fit}$ and $\boldsymbol{b}_{fit}$ are affine transformation parameters to align faces in I_{src} and I_{dst}. $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$ are facial landmark locations in I_{src} and I_{dst}, respectively.

DOI: 10.2312/egve.20171367



(a) *wrist dorsiflexion* **(b)** *pen spinning* **(c)** *baoding balls rotating*

Figure 1: *Observed motions.*

Sequential Process: In this process, facial landmark detection [KS14], facial landmark location compensation, and image warping and composition are repeatedly executed. In the facial landmark compensation process, detected landmark locations are compensated using $\boldsymbol{d}_i$ as follows.

$$\bar{\boldsymbol{q}}_i' = \boldsymbol{s}_j \boldsymbol{d}_i + \bar{\boldsymbol{q}}_i \tag{2}$$

$\boldsymbol{s}_j$ is a scaling factor, which is calculated as follows.

$$\boldsymbol{s}_j = \begin{bmatrix} w_j'/w_j & 0 \\ 0 & h_j'/h_j \end{bmatrix} \tag{3}$$

where w_j and h_j are bounding box width and height of a face j, and superscript represents the target video. After this compensation process, the source image I_{src} is warped into the target video I_{dst}. In order to warp the source image, first, a face area is divided into small regions using Delaunay triangulation for face landmarks. I_{src} is then warped into I_{dst} using affine transformation calculated from the corresponded regions. Finally, warped regions are composited using poisson image blending [PGB03].

3. Experiment

We conducted a user study to confirm the effectiveness of face swapping for mental imagery training. In the experiment, we measured MEPs using TMS while imagining a target motion by observing a target motion video. We used three motions as shown in Fig. 1. Wrist dorsiflexion motion can be performed by all participants. Therefore, we measured MEPs for own video, another person's video and face swapped video. On the other hand, pen spinning and baoding ball rotating motions cannot be performed by all participants. Therefore, we used another person's video and face swapped video only. The participant's own video, another person's video, and face swapped video were randomly shown to participants. It should be noted that we used mirrored videos to reduce the effect of mental rotation. The number of participants for each measurement are 6 for wrist dorsiflexion, 3 for pen spinning, and 4 for baoding balls rotating. MEP was obtained as a peak-to-peak value from averaged waveform of 15 trials. It should be noted that we also measured a normal MEP in relaxed state. The normal MEP is used as a control data, and then obtained MEPs are normalized by the normal MEP. Participants tried to imagine each target motion without actual motion by themselves.

Fig. 2 shows results of average MEPs in each motion. The results confirmed that average MEPs of face swapped video were higher than that of another person's motion video. Specifically,

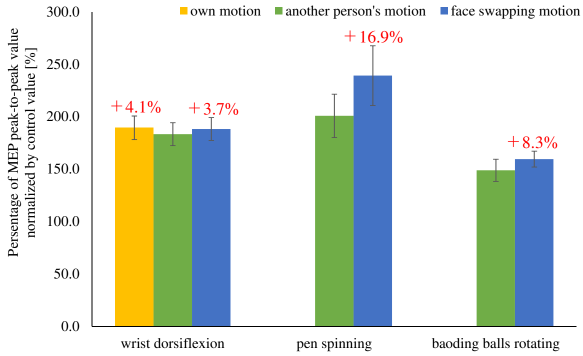


Figure 2: *Average MEPs for each video observation.*

there were large differences between another person's video observation and face swapped video observation in pen spinning and baoding balls rotating cases. From these results, it can be considered that the face is important to get good imagination in difficult motions. In addition, according to the interview, we could confirm that the relationship between MEP values and user's sensation. Thus, we believe that TMS can be used to evaluate face swapping algorithms quantitatively.

4. Conclusion

In this research, we conducted a preliminary user study to confirm the effectiveness of face swapping in mental imagery training. We measured MEP using TMS while watching videos from recorded motions. From the results of the experiment, we consider that face swapping is effective to get higher MEP. However, some participants feel unnatural sensation in the face swapped video, and their MEPs did not increase from another person's video. In the future, we are planning to confirm the effectiveness of face swapping with more participants. In addition, we are planning to develop a face swapping algorithm that can generate self-recognizable image.

Acknowledgements

This work was supported by the MIC/SCOPE #162107006.

References

[BKD*08] Bitouk D., Kumar N., Dhillon S., Belhumeur P., Nayar S. K.: Face swapping: automatically replacing faces in photographs. *ACM Transactions on Graphics 27*, 3 (2008), 39. 1

[DSJ*11] Dale K., Sunkavalli K., Johnson M. K., Vlasic D., Matusik W., Pfister H.: Video face replacement. *ACM Transactions on Graphics 30*, 6 (2011), 130. 1

[KBRF10] Keyes H., Brady N., Reilly R. B., Foxe J. J.: My face or yours? event-related potential correlates of self-face processing. *Brain and cognition 72*, 2 (2010), 244–254. 1

[KS14] Kazemi V., Sullivan J.: One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 1867–1874. 2

[PGB03] Pérez P., Gangnet M., Blake A.: Poisson image editing. *ACM Transactions on Graphics 22*, 3 (2003), 313–318. 2

[RC04] Rizzolatti G., Craighero L.: The mirror-neuron system. *Annu. Rev. Neurosci. 27* (2004), 169–192. 1

[SNO] SNOW. http://snow.me/. last checked: 2017-09-19. 1