Evaluation of Facial Expressions as an Interaction Mechanism and their Impact on Affect, Workload and Usability in an AR game

Alan Transon*, Adrien Verhulst†, Jean-Marie Normand†‡,
Guillaume Moreau†‡, Maki Sugimoto*
*Interactive Media Lab, Keio University, Japan.
alan.transon@gmail.com, maki.sugimoto@keio.jp
†AAU UMR CNRS 1563, École Centrale de Nantes, France.
firstname.lastname@ec-nantes.fr
‡Inria Hybrid, France.

Abstract—With the recent development of Head Mounted Display (HMD) for Virtual Reality (VR) allowing to track and recognize user's Facial Expression (FE)s in real-time, we investigated the impact that the use of FEs as an *action-trigger* input mechanism (e.g. a FE mapped to a single action) has on our emotional state; as well as their workload and usability compared to the use of a controller button. We developed an Augmented Reality (AR)-based memory card where the users select virtual cards using a wand and flip them using either a FE (smiling; frowning) or a Xbox controller button. The users were separated into three groups: (1) flipping the card with a smile (n = 10); (2) flipping the card with a frown (n = 8) and (3) flipping the cards with the Xbox controller button (n = 11). We did not see any significant differences between our groups in: (i) the positive and negative affect of the participants and (ii) the reported workload and usability, thus highlighting that the FEs could be used inside a HMD in the same way as a controller button.

Index Terms—Human computer interaction, Augmented Reality, Facial Expressions

I. INTRODUCTION

Recent developments allowed for tracking and recognizing user's FEs in real-time even when wearing a HMD (see Sec. II-C). Following those developments, we aim at investigating the influence of FEs when used as an *action-trigger* input mechanism (e.g. like a mouse button) in a HMD-based AR application. We specifically study (i) emotional influence on users and (ii) effort and usability of such a system.

a) Emotional influence: In 1872, Darwin [7] observed the universality of a set of facial expressions. Following this work, Ekman [9] proposed 6 universal *basic* FEs, each expressing a specific emotion: *Happiness*, *Sadness*, *Surprise*, *Fear*, *Anger* and *Disgust*. Later, Segerstrale and Molnar [26] demonstrated that the face was not only able to show affect, but also to modify the intensity of the affect. From those observations, the Facial Feedback Hypothesis (FFH) stated that facial movements can influence emotional experience, e.g. smiling enhance pleasant feelings (for a more extensive review the reader is referred to [1]).

Yet, to the best of our knowledge, the impact of FEs used as an *action-trigger* input on the participant's emotional state has not been studied. It is worth noting that several studies focused on how visual feedback of the user's own FEs could affect his/her emotional state. For example, Tsujita and Rekimoto [32] developed the "HappinessCounter", a device (e.g. a fridge) registering and providing a visual feedback of the user's FE (smiling and sad) over several days. The purpose was to make the user self-aware of his/her own mood, in order to force him/her to smile along the day (i.e. when using the appliance). The participants mentioned that "they tried to smile more and [...] thus felt that it improved their mood". Also, Yoshida et al. [38] deformed in a plausible way the user's mirrored FE in order to make him/her more "smiley" or "sad" than he/she was (the user was asked to keep a neutral FE). The experimental design was such that participants did not consciously notice the changes in their mirrored FEs; and the results showed that the "smiley" visual feedback could enhance positive affect and the "sad" visual feedback could enhance negative affect. Nevertheless, no studies did it using FEs as an *action-trigger* input.

b) Effort and usability: The use of FEs in an HMD-based application is not straightforward since the HMD covers part of the user's face thus making face capture through Computer Vision (CV) techniques difficult. Equally important, the HMD may constrain the user's FEs due to (i) the pressure it exerts on the user's face and (ii) the fact that it limits the user's range of movement (in particular the cheeks). Doing FEs with a HMD may therefore require more effort from the user, making the overall system less usable, justifying the investigation of the users' evaluations of effort (which will be referred to as workload in the remaining of the paper) and usability.

Finally, studying the influence of FEs on users' emotional state and their usability when used inside a HMD as an *action-trigger* input mechanism is not only relevant to the field of VR or AR, but could also be beneficial to facial therapy. Indeed, facial paralysis requires physical therapy, such as facial exercises (which show improvement in moderate facial paralysis [3]). Because of the relevance of VR and AR for physical rehabilitation (for a more extensive review the reader

is referred to [15]), facial rehabilitation could also benefit from our findings.

In the remainder of this paper, we present a literature review of the use of FEs as an interaction method in 2D and 3D applications as well as in facial therapy, then present our experiment and results before discussing them.

II. RELATED WORK

In this section, we draw a review of the work being carried out in the literature regarding Facial Activity (FA) and FE. Using FAs as an input mechanism usually relies on the detection of muscle(s) activation via ElectroMyoGram (EMG) and is mainly used to emulate a button click to allow physically challenged persons (e.g. paraplegics) to interact with a 2D User Interface (UI) (where FEs are usually detected by CV and act on several facial muscles). Even though this paper focuses on FEs, using FAs as an input mechanism still presents some similarities and we believe it is worth detailing here.

We first present in Sec. II-A work where facial expressions are used an input interface for computer systems. In Sec. I we concentrate on how FEs have been used for rehabilitation.

Moreover, let us mention that while emotions' body response is observable with Brain Computer Interface (BCI) and others physiological devices, in this paper we only focus on FEs and FAs.

A. Voluntary facial expression and affective control in HCI

This section is dedicated to detailing how FAs and FEs can be used as input interfaces for computer systems.

1) Facial Activities: Surakka, Illi and Isokoski [28] used Eye Gaze Tracking (EGT) and FA (smiling) as an interface to allow physically challenged persons (e.g. paraplegics) to point and click without using their hands. The authors noted that using this FA as a selector introduced a delay (without giving any actual measure) compared to a mouse button. Suraka et al. [29] studied the FAs "smile" and "frown" and advanced that "smiling" is a faster selector than "frowning".

Rantanen et al. [24] showed that the smiling FA did not affect the accuracy of the EGT, but, like most studies focusing on FAs, the Head Mounted Unit (HMU) they used to detect FAs was less invasive than a typical VR HMD. Tuisku et al. [33] compared different FAs (smiling, frowning and raising eyebrows) as a selector and showed that "smiling" is overall rated more positively than the others (especially in accuracy), the authors suggested that it is because "smiling" is easier to perform.

Vanhala and Surakka [34] used a 2D slider that moved depending on the FA's (smiling, frowning) intensity (low, moderate, high). Their findings show that moderate intensity of FAs was the most effective to detect changes in Heart Rate (HR) and Heart Rate Variability (HRV) while remaining easy to use. Those physiological changes then showed a linear trend with the subjective rating of emotional valence.

2) Facial Expressions: To the best of our knowledge, studies on FEs all use a very similar approach: (i) detecting facial features via CV; (ii) labeling and classifying (with a classifier: e.g. Support Vector Machine (SVM), Convolutional Neural Network (CNN), etc.) facial features in an expression set (or emotion set, see Sec. I); and (iii) retrieving the best rated FE.

Ilves et al. [13] classified applications using FEs as an input interface in three categories: (i) interactive mirror, where an avatar's FE follows the user's FE; (ii) action trigger control, where the application reacts to a given user's FE; and (iii) affective control, where the application constantly adapts itself to the user's FE (the most common adaption being to adapt a game's difficulty to the user's FE). Our own work falls in the second category.

a) Interactive mirror: Mourão and Magalhães [19] developed *NovaEmõtions*, a competitive multiplayer game where users are filmed trying to match a target FE (e.g. happiness), the one "closest" to the target emotion win the game. Their FEs set is from the Emotional Facial Action Coding System (EMFACS) and their findings show that some of those FEs are difficult to reproduce (such as contempt). Tan, Sapkota and Rosser [30] developed *BeFaced*, a "Candy-Crush"-like tablet game where the user must align three blocks with the same FE (by swiping them on the tablet), and upon alignment must do the aligned FE. The aim is to collect an "in the wild" FEs data set. With a similar objective Li et al. [17] developed a "Tower-defense"-like desktop game where the user must do several FEs. Their FEs set is from Ekman's 6 *basic* FEs.

b) Action trigger control: Ilves et al. [13] developed *Take Drunkard Home*, a desktop game where the user controls a character along a path and must avoid obstacles. The control is done either through: (i) Facial Tracking (FT) via CV (left and right direction) and 2 FEs (smiling and frowning to avoid moving and standing obstacles); or (ii) a joystick. Their findings shows that CV-based controls are more challenging and immersive than a joystick but less functional (the joystick is faster, easier and more accurate). Bayatpour et al. [2] used the intensity of the "happiness" and "sadness" FEs (either via CV alone or through CV and EMG) to control the movements (up and down) of a 3D flight-simulation game. Their findings show that users prefer the CV alone, since it is less invasive, but that it is also more difficult to control.

c) Affective control: Obaid, Han and Bilinghurst [20] developed "Feed the Fish", a game where the user control a fish with a keyboard and must meet or avoid others fishes. The user's FEs (happy, neutral or frustrated) are used to adapt the game's difficulty. Subjects reported that they "might feel stupid [...] to try to smile all the time" and needed "an indicator for the reaction from the facial tracking". In a similar way, Xiang, Yang and Zhang [37] used "frustrated", "excited", "relax" and "bored" FEs to adapt the difficulty of a Tetris-like game.

To sum up, there have been several studies using FEs as an input mechanism. Their findings show that most of the universal FEs are effective inputs (the best being smiling), but are less functional than common desktop interfaces (such as

a joystick). To the best of our knowledge, none of the studies on FEs have yet tackled VR or AR systems.

B. Voluntary Facial Expression and facial therapy

As mentioned in the introduction, FEs interactions are also relevant in facial therapy.

We previously mentioned some studies [28], [29], [33] targeting physically challenged persons and researching the use of FAs as a replacement for common input interfaces. FEs are also being used, for example with children with Autism Spectrum Disorders (ASD) to help them recognize, respond to and practice FEs:

- Cockburn et al. [6] developed *SmileMaze* a 2D game where a child has to guide a character through a maze with several obstacles that can only be removed if the child maintain a smile for a fixed duration of time.
- Deriso et al. [8] developed *Emotion Mirror* where children receive in real-time a visual feedback of their own FEs. They have to make all 6 Ekman's *basic* FEs, which are then captured and mapped in real-time to a character.
- Harrold et al. [12] developed *CopyMe*, a tablet game where children with ASD must match several FEs (again from Ekman's 6 *basic* FEs). Their pilot showed that the application is promising for recognizing FEs.

In a similar way, current facial rehabilitation techniques (targeting people suffering from facial nerve paralysis) focus either on mime therapy (series of massage and exercises where the patient does FEs) or EMG biofeedback therapy (see [22], [23]). For instance, Breedon et al. [4] recently developed *Face to Face*, a therapeutic game where patients try to match a FE displayed on a screen while their own FEs are tracked by a Kinect.

C. Facial Expressions and Head Mounted Display in VR

Among the recent development in FEs tracking inside a HMD, we would like to mention the followings:

- Li et al. [16] who proposed a system to capture FEs of users wearing an HMD to allow for "face-to-face" immersive communication in a Virtual Environment (VE). They combined surface strains on the HMD's foam liner (resulting from the user's FE) with RGB-D cameras mounted on the HMD to capture lips movements. Although their system requires a calibration, it results in realistic FEs.
- Thies et al. [31] that developed *FaceVR* to "synthesize" in real-time the user's FE while she/he is wearing a HMD. A RGB-D camera captures the user's face while an IR camera inside the HMD captures his/her gaze direction; the system then synthesizes a FE. Their approach allows for an accurate tracking of FEs while wearing an HMD.
- Olszewski et al. [21] who adopted a different approach with a similar objective using the FOVE[1] HMD with a built-in IR camera to track the gaze direction and a RGB camera to record mouth movements. Unlike [31], they

trained a CNN based on the gaze and mouth movements, successfully producing real-time realistic FEs.

Those recent works show that the tracking of FEs inside HMDs will be soon available for wider use thus highlighting their potential usability as an input mechanism for VR and AR.

III. Experiment

As a reminder, in this experiment we aim at studying, in an AR context, (i) whether interaction through FEs has an influence on the participant's affect and (ii) the workload and usability of FEs as an input mechanism. For that, we rely on a memory card game [2].

A. Apparatus and Participants

In the experiment, participants were sitting on a stool in front of a small table (50×40 cm) and had to play an AR memory card game.

Each participant was wearing a custom-made Video-See-Through HMD composed of an Oculus Rift (DK2) with a front-facing camera (Logitech HD Pro C920 with a resolution of 1920×1080 pixels) and an embedded AffectiveWear device [18], see Fig. 1. The image of the camera was displayed in the DK2 using Vuforia's[3] Unity3D plugin.

The participant was holding a wand (deeper than the lenses, see Fig. 3) which position was tracked by an 11-camera Optitrack system[4].

Fig. 1: Left: The AffectiveWear device embedded in the Oculus Rift DK2. Right: The front-facing camera attached to the DK2.

29 participants (21 male and 8 female) took part in the experiment in March 2017. They were all from a school of engineering and were aged from 21 to 30 years ($M = 22.62$, $SD = 1.74$). Of all the participants, 23 were East Asian (79.31%), the remaining being: 1 Central Asian, 1 South Asian, 1 North American, 1 West European, 1 East European and 1 North African. Moreover, 3 (9.38%) had no previous experience with AR, 4 (12.50%) used it *Once or twice*, 13 (40.63%) used it *3-10 times* and 12 (37.50%) used it *Frequently*.

---

[1]https://www.getfove.com/ (consulted 2017.03.15)

[2]https://en.wikipedia.org/wiki/Concentration_(game)

[3]https://www.vuforia.com

[4]http://optitrack.com/products/flex-3/

B. The AR Memory Game

The implementation of the game was straightforward. Selection and flipping animations combined with sound effects ensured visual and audio feedback to the participants. Upon selection, cards turn red and slightly increase in size. To ease the gameplay, the flipping-input is only expected by the system if two cards are selected (see Fig. 2).

During the game, 18 cards (9 pairs, 3.4×6.8 cm by card) with natural landscape images, organized on 3 rows of 6 columns, cover the tracking space (45.5×32.5 cm). To select a card the participant has to touch it with the custom-wand (a 30-cm long wooden hammer held by its head used to ease the selection process) tracked by the Optitrack system (see Fig. 3).
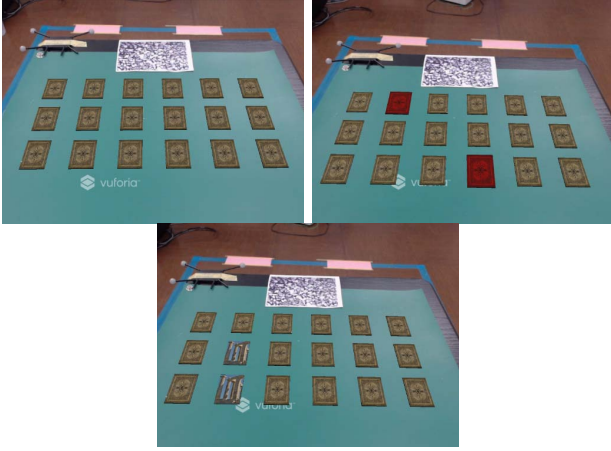

Fig. 2: The game with (Left) no card selected; (Center) two cards selected and (Right) a pair of flipped cards.
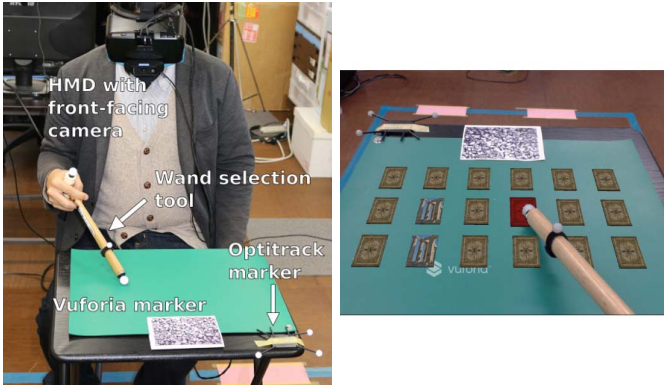

Fig. 3: Left: A participant playing the AR memory card game. Right: First person view of the selection with a wand. NB: the table is also tracked in order to use it as the reference of the 3D space's origin.

This game was implemented with Unity v5.5 and Vuforia v6.2. The camera stream is sent to Vuforia which tracks an AR marker (see Fig. 3) and positions the AR objects (i.e. the virtual memory cards) on Unity's camera plane. Since the virtual memory cards are located in the plane of the table, an occlusion problem happens whenever the subject's hand or the wand goes over the tracking space for selection. To solve this issue, a green mat is disposed on the table and is used as an occlusion mask (see Fig. 4).
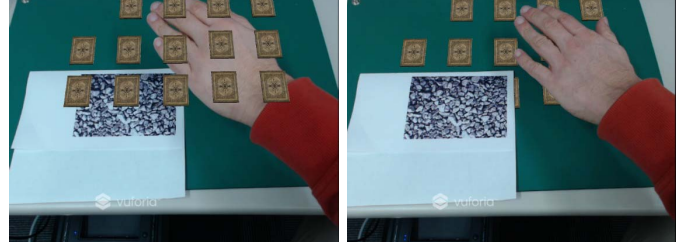

Fig. 4: Hand occlusion issue before (Left) and after (Right) using an occlusion mask. Note that the cards were bigger during the experiment.

C. The AffectiveWear

In our setup, we use the *AffectiveWear*, proposed by Masai et al. [18], embedded inside an Oculus DK2. The AffectiveWear includes 16 photo reflective sensors (SG-105[5]). The sensors measure the change of distance (provoked by FA) between the eye-wear and the skin surface; those changes are sent to a classifier which classifies them into a set of FEs. For more technical details the reader should refer to [18].

While FE tracking is not as accurate as the work previously mentioned (see Sec. II-C), we are not targeting FE reconstruction, but only FE detection. In this regard, the *AffectiveWear* is able to detect 7 expressions (the 6 "*basic*" expressions and a neutral expression) in real-time which is sufficient for our setup.

The AffectiveWear embedded in the Oculus DK2 is slightly deeper than the lenses (see Fig. 1), thus the use of FEs inside the HMD was mainly constrained by the HMD alone rather than the couple AffectiveWear and HMD.

In our setup, we only train our classifier to recognize smiling, frowning and neutral FEs.

D. Experimental Design

The experiment is designed as a between-groups study where participants were randomly assigned to one of the three groups (denoted G1, G2 or G3), each corresponding to a different input method to control the flipping of the cards:

G1 : input is performed with a large smile (smiling) ($n = 10$)

G2 : input is performed with an unhappy face (frowning) ($n = 8$)

G3 : input is performed via an Xbox controller (pushing the A button) ($n = 11$)

Thus, for every group, the card selection mechanism remains identical and is achieved by using the tracked wand. The difference comes into play when participants want to flip two cards to check if they are identical. At that point, some

[5]http://www.kodenshi.co.jp/english/products/photointerrupter_ref/ photointerrupter_ref.html

participants will use a specific FE or simply push a button on a physical controller.

Let us note that both types of flipping-input mechanism (i.e. FE and Xbox Controller) are an *arbitrary mapping* (i.e. flipping a card via a FE or with a button is an arbitrary mapping, unlike flipping a card with a hand motion, which is a more natural mapping). This is important since Skalski et al. [27] previously showed that natural mappings were more enjoyable and enhance spatial presence (the sense of being physically located in a virtual environment) than arbitrary mappings. Since both types of flipping-inputs are arbitrary, a change of affect is more likely to be due to the FFH than to the natural mapping of the flipping-input mechanism in our game.

In order to make sure the groups were homogeneous, we computed the chi-squared (for the non-metric variable: sex) as well as Analysis Of Variance (ANOVA)s (for the metric variables: age and AR familiarity). The 3 groups were homogeneous considering sex ($\chi^2 = 4.085$; $p-value = 0.1297$), age ($F(2,27) = 0.997$; $p = 0.326$) as well as AR familiarity ($F(2,27) = 0.049$; $p = 0.826$).

Every groups wore the HMD with the AffectiveWear device embedded and performed its calibration even if it was not used for G3. Moreover all groups had to use the wand to select the cards (therefore the G3 was holding the wand with one hand and using the controller with the other one).

During the experiment, the participant was supervised by an assistant that was here to help him/her throughout the whole process. There was nobody else in the experimentation room. The instructions as well as all the questionnaires were available in English and Japanese (the participant could choose his/her preferred version according to his/her most proficient language). All participants followed a similar procedure, that we sum up here (and detail in Sec. III-E):

1) Read the scenario
2) Fill the consent form and demographic questionnaire
3) Confirm the understanding of the scenario
4) Sit on the chair in front of the table and wear the HMD
5) Test the tracking and try the cards selection mechanism
6) Follow the AffectiveWear calibration procedure
7) Test the flipping-input mechanism (according to their group)
8) Play the game
9) Fill the post-experimental questionnaires

E. Task & Procedure

Upon entry in the experiment room, the participant was presented with the scenario reported in Table I (translated in Japanese by a Japanese native speaker) and had to fill a profile questionnaire. An assistant was present to answer any question the participant might have regarding the scenario. The assistant then helped the participant to wear the HMD and demonstrated how to hold the wand properly to ensure an optimal tracking.

The participant was asked to look around the tracking space (e.g. the green mat) in order to make sure that the AR tracking was steady, and to select three to four cards with the wand

(without using the flipping-input mechanism) in order to make sure they understood how to select cards.

Two participants, due to their small height, were unable to see the entire table in the HMD while seated and had to do the experiment standing (when seated their field of view was too close to the table which led to tracking issues). Since the rest of the experiment was relatively short (see Sec. IV) and the questionnaires only targeted the flipping-input mechanism (i.e. smiling; frowning; or pressing a button) we are confident this had no impact on our results.

The participants then followed the AffectiveWear calibration procedure. They were presented with pictures displayed in the HMD asking them to do and maintain three FEs: *Neutral* (no FA), *Smile* (a very visible smile) and *Frown* (a very visible frown); for five seconds with a three second break between each FEs (see Fig. 5).
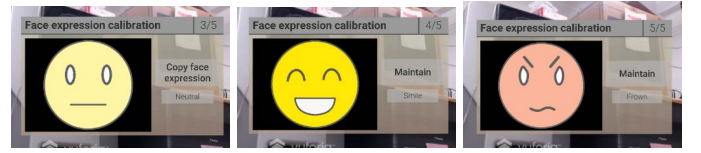

Fig. 5: Calibration instructions. From left to right: *Neutral*, *Smile* and *Frown*.

Then, they had to try the flipping-input mechanism to make sure that it was working properly and that they understood how to use it. Once again, they followed instructions displayed directly in the HMD. For this short training session, each group was presented with a different image.

Once the training was completed, the game started. The participant had to select two hidden cards among 18. Once two cards were selected, the participant had to confirm his/her choice by flipping them (the validation mechanism depending on the group he/she belonged to and being referred to as flipping-input mechanism). If the two highlighted cards matched, they both remained visible and could not be selected anymore. Otherwise, they returned to a hidden state. The game was over once all pairs had been discovered (i.e. all the cards are visible). After finishing the game, the participant had to answer a series of post-experimental questionnaires, see Sec. III-F.

F. Measurements and Questionnaires

Objective measurements include the time required to complete the game, the number of flipping-inputs performed and the AffectiveWear data (timestamps and detected facial expressions). Regarding subjective data, the participants answered the following questionnaires either in English or in Japanese (according to their most proficient language):

- The NASA-Task Load Index (TLX) questionnaire, used to evaluate the workload of the AR flipping-input mechanism along the following sub-scale: *Mental load*; *Physical load*; *Effort* and *Frustration*. We used the Japanese translation proposed by Haga and Mizukami [11].

TABLE I: Our experimental scenario, whenever participants had doubts they could ask precisions to the assistant that were here to help them throughout the experiment.

In the following experiment, you will use a setup that can detect your facial expressions. In this experiment, we want to test whether our setup could be used in an Augmented Reality (AR) context.
The experiment will proceed as follows: first a calibration of the device that detects facial expressions will be performed. Then, a short training session will introduce you with the setup and how to interact with the AR environment. The main experiment which consists of a "Memory Game" will then start.
In this game, you will be presented with 18 virtual cards in an AR environment. The goal is to find matching pairs. To select the virtual cards, use the pointer to touch them. You will know a card is selected once it changes color. When two (2) cards are selected you need to flip them to check if they match. To flip cards, use the input system that was introduced during the training session.

- The Positive and Negative Affect Schedule (PANAS) questionnaire (Watson, Clark and Tellegen [36]) used to evaluate the positive and negative affects of the participant after playing the game. We used the Japanese translation proposed by Watson and Clark [35].
- The private scale of the self-consciousness questionnaire (Scheier and Carver [25]), used to correlate the participant's affect and its self-consciousness. A Japanese native speaker did the translation from English to Japanese.
- The System Usability Scale (SUS) [5] questionnaire, used to evaluate the usability of the AR flipping-input. We used the Japanese translation proposed by Fukasawa and Mizukami [10].

Finally, the participants were required to write (in either English or Japanese) their opinion(s), remark(s) and suggestion(s) regarding the experiment if they had any.

IV. RESULTS

In this section we present results from our questionnaires. For each questionnaire we performed an ANOVA to point out differences amongst our groups.

The game sessions were relatively short ($M = 212.2$; $SD = 17.43$ seconds) with close to 30 inputs performed on average ($M = 29.44$; $SD = 6.27$).

The length of game sessions did not a follow a normal distribution (after a Shapiro-Wilk test, with $p = 0.036$), we therefore did a non-parametric Kruskal-Wallis test ($\chi^2 = 0.241$, $p - value = 0.8865$) which did not highlight any significant differences in our groups. An ANOVA revealed that there was no significant difference in the number of inputs between our groups ($F(2, 27) = 2.242$, $p - value = 0.146$).

A. Input Mechanism and its impact on participant's emotion

From the PANAS questionnaire we estimated the positive and negative participants' affects. In spite of our expectation, there was no significant difference among the three conditions (see Table II).

We then tested the correlation between the participants PANAS's Positive ($t = 1.065$, $df = 27$, $p - value = 0.2961$) and Negative ($t = -0.6159$, $df = 27$, $p - value = 0.5431$) affects and their private self-consciousness estimation with a Pearson's product-moment, but again there was no significant differences.

TABLE II: Mean, Standard Deviation (SD) and computed ANOVAs for the PANAS Positive and Negative scale

| Conditions | 0 (n=10) | | 1 (n=8) | | 2 (n=11) | | ANOVA | |
	Mean	SD	Mean	SD	Mean	SD	F(2,27)	p
PANAS Positive	35.6	113.15	34.25	55.071	32.81	92.763	0.47	0.499
PANAS Negative	17	39.11	17.25	23.071	17	18.8	0	0.998

TABLE III: Mean, Standard Deviation (SD) and computed ANOVAs for the NASA-TLX's mental demand, physical demand, effort and frustration subscales as well as the SUS

| Conditions | 0 (n=10) | | 1 (n=8) | | 2 (n=11) | | ANOVA | |
	Mean	SD	Mean	SD	Mean	SD	F(2,27)	p
Mental demand	3.2	8.4	1.875	0.6964	1.454	0.4727	4.923	**0.0351**
Physical demand	4.1	7.0	2	0.8571	2.4545	2.8727	3.434	0.0748
Effort	3.8	8.4	2	4.810	0.5714	2.636	1.325	0.26
Frustration	3.5	8.72	2.625	0.5535	3.0909	4.8909	0.162	0.691
SUS	78.5	129.44	77.18	157.92	72.954	134.772	4.719	**0.0388**

B. Input Mechanism and its workload and usability

From the NASA-TLX and SUS questionnaires we estimated the workload (on the tested NASA-TLX subscales) as well as the usability of the input mechanism. We explicitly mentioned to the participants that the questions were about the flipping-input mechanism only (i.e. smiling; frowning or pressing the controller button).

The ANOVAs revealed a significant difference among the three conditions for the NASA-TLX's mental demand as well as for the SUS's score (see Table III).

We conducted a Levene's test of equality of variance for those elements to verify the homogeneity of variance (NASA-TLX's mental demand: $F(2, 26) = 2.1794$, $p - value = 0.1333$; SUS's score: $F(2, 26) = 0.3477$, $p - value = 0.7095$). Since both elements were accepted, we performed Student's t-tests but none highlighted any significant difference. Therefore, there are no significant differences in our three conditions regarding the input mechanism's subjective workload (on the tested NASA-TLX subscales) and input's usability.

C. Participant's feedback

Most of the comments were originally written in Japanese and were translated into English. Overall, they appreciated it

("*It was fun to use*[6]"; "*The experiment was really enjoyable*[6]") and easy to understand ("*Very simple and convenient*[6]"; "*Really easy to understand*[6]"; "*The content was simple and easy to grasp*[6]"; "*It was easy to flip the cards*[6]"). Yet, some participants reported some difficulty to see the cards ("*The cards were too small*[6]") as well as some tracking issues ("*The cards disappeared too often*").

Some interesting points were raised regarding the FE input mechanism:

- Several participants were unable to wear their glasses with the HMD and had to remove them. Because of their eyesight and the relatively low resolution of the HMD, it raised an important issue: a false detection of the FE "frowning" due to the user narrowing his eyes to enhance his vision. Indeed one participant reported the following: "*Because of I have a bad eyesight, in order to see the card I had to narrow my eyes. Since narrowing eyes and the recognition of the "angry" expression are somehow similar, the cards were sometimes flipped*[6]".
- Some participants instinctively smiled and laughed as a natural reaction to stress. Even if the FE at that moment was not conscious, it was detected as an input. One participant reported the following: "*Since there is a tendency to laugh upon failure, even without willingly choosing to flip the cards, the cards were flipped*[6]". Note that in our setup it was not so much of a problem and happened rarely, since the input was only validated (e.g. the cards flipped over) if two cards were selected.
- One participant reported that while using FEs was easy, it may become painful to use them for too long: "*It is easy to use facial expressions, but because it may become painful to use them for a long time, I think it is suitable only for a short time*[6]".

V. DISCUSSION

We showed that there was no significant difference in positive and negative affects when using either a FE (i.e. smiling or frowning) or a controller button as an input mechanism in a HMD-based AR game. This is an interesting result regarding the FFH as "smiling" had no significant effect on the participants' affect even though they reportedly enjoyed the memory card game. This could be explained by the game being *too* enjoyable and the possible changes of affect from the FEs being too weak to be noticeable. Despite our expectations that this game would raise only a slight interest from the participants, the difficulty setting (9 pair of cards) may have nonetheless created an interesting challenge for them.

We also noted that there was no correlation between the positive or negative affects and private self-consciousness (see Sec. IV-A). This result also stands out considering the findings of [14] who showed that people with high private self-consciousness are more responsive to mood-inducing experiences. Again, this difference could be explained by the game being *too* enjoyable.

[6]Translated from Japanese

Our results also showed no significant difference between the tested input methods on the mental demand, physical demand, effort, frustration as well as usability (see Sec. IV-B). This seems to indicate that FEs inside a HMD when used as an input mechanism have the same workload impact (on the tested NASA-TLX subscales) than a controller button. We were not expecting this result since the HMD partially restrained the use of FEs and our system was calibrated so that only "strong" FEs would be detected. This result can be explained because we chose FEs easy to detect (i.e. smiling; frowning) and that require little effort while holding a controller in addition to the wand might have been cumbersome for some participants (even if mentioned only by one participant).

Finally, while our game was easily understandable, and relatively short, it was still long enough ($\sim$ 210s) to gather a sufficient number of inputs, especially compared to other applications using FEs as inputs: $\sim$ 120s for *Take Drunkard Home* [13]; 30s by image for *NovaEmötions* [19] and 30s by trial followed by a 30s break for [34]). We also noted that it was more stressful and enjoyable than excepted.

We report two issues with the use of FEs as an "Action trigger control" mechanism: (i) due to stress or enjoyment, participants can unconsciously perform FE that can be detected as inputs; and (ii) participants with bad eyesight may unconsciously adjust their neutral FE which can disrupt FEs detection. While we did not focus on proposing a game input mechanism using FEs, those issues are worth mentioning.

VI. CONCLUSION

We investigated the impact of FEs, used as an *action-trigger* input in a HMD-based AR game, on users' emotional state as well as their reported workload and usability.

We found that there is no significant difference in users' emotional state when using FEs (i.e. smiling; frowning) or a Xbox controller button. This is different from what we excepted from the FFH and may be explained by our AR game being *too* challenging. Moreover there was no difference in the reported workload and usability. Once again, this result is unexpected, but may be due to the relative *short* game session.

Yet, this study has a couple of drawbacks. First we had a limited number of participants ($n = 29$ distributed into three groups). Second, our AR game was probably too challenging (too many cards that may have been too small). The high challenge provoked unconscious facial activity (nervous laughter) which may have perturbed our results.

More importantly, we reported a couple of limitations of FEs used as an *action-trigger* input mechanism: the duration (or the frequency of inputs) may strain the users and involuntary facial movements may perturb the system. Furthermore, because FEs need to be visible (and in practice a bit "forced" by the users), it may impact their duration and in turn the system's usability.

We were not able however to see this impact in our study, since we only focused on short gaming sessions. We believe future work using FEs as an input in a HMD-based application should focus on slight and weak FEs, like Vanhala and Surakka [34] mentioned for FAs. Yet, for short durations

(an average of 210 seconds in our case) FEs require the same effort (i.e. workload) and have the same usability as a Xbox controller button. We would also like to propose a less challenging setup and test it for a longer period of time. Finally, we would also like to extend our findings to the facial rehabilitation field.

ACKNOWLEDGMENT

REFERENCES

[1] P. Andréasson. *Emotional Empathy, Facial Reactions, and Facial Feedback*. PhD dissertation, Uppsala University, Sweden, 2010.

[2] S. Bayatpour, U. Bernardet, S. Dipaola, A. Kitson, and B. E. Riecke. Exploring Facial Expressions for Human-Computer Interaction: Combining Visual Face Tracking and EMG Data to Control a Flight Simulation Game. In *Proceedings of ISEA 2015: the 21st international symposium on electronic art*, 2015.

[3] C. H. G. Beurskens and P. G. Heymans. Positive Effects of Mime Therapy on Sequelae of Facial Paralysis: Stiffness, Lip Mobility, and Social and Physical Aspects of Facial Disability. *Otology & Neurotology*, 24(4):677–681, jul 2003.

[4] P. Breedon, A. Russell, P. Logan, O. Newell, B. OBrien, J. Edmans, D. Baines, and P. Hall. First for Stroke : Using the Microsoft Kinect' as a stroke rehabilitation tool for facial weakness. *International Journal of Integrated Care*, 14(8):20004, nov 2014.

[5] J. Brooke. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[6] J. Cockburn, M. Bartlett, J. Tanaka, J. Movellan, M. Pierce, and R. Schultz. SmileMaze : A Tutoring System in Real-Time Facial Expression Perception and Production in Children with Autism Spectrum Disorder. *Computer*, pages 978–986, 2008.

[7] C. Darwin. *The expression of the emotions in man and animals.* American Psychological Association (APA), 1872.

[8] D. Deriso, J. Susskind, L. Krieger, and M. Bartlett. *Emotion Mirror: A Novel Intervention for Autism Based on Real-Time Expression Recognition*, pages 671–674. Springer Berlin Heidelberg, 2012.

[9] P. Ekman. Cross-cultural studies of facial expression. *Darwin and facial expression: A century of research in review*, pages 169–222, 1973.

[10] N. Fukasawa and N. Mizukami. Evaluation of guidance system for visually disabled. In *Proc. of the 14th World Congress on Intelligent Transport System (ITS)*, 2007.

[11] S. Haga and N. Mizumaki. Japanese version of NASA Task Load Index: Sensitivity of its workload score to difficulty of three different laboratory tasks. *The Japanese journal of ergonomics*, 32(2):71–79, 1996.

[12] N. Harrold, C. T. Tan, D. Rosser, and T. W. Leong. CopyMe: A Portable Real-Time Feedback Expression Recognition Game for Children. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*, pages 1195–1200, New York, New York, USA, 2014. ACM Press.

[13] M. Ilves, Y. Gizatdinova, V. Surakka, and E. Vankka. Head movement and facial expressions as game input. *Entertainment Computing*, 5(3):147–156, aug 2014.

[14] C. L. Kleinke, T. R. Peterson, and T. R. Rutledge. Effects of self-generated facial expressions on mood. *Journal of Personality and Social Psychology*, 74(1):272–279, 1998.

[15] K. E. Laver, S. George, S. Thomas, J. E. Deutsch, and M. Crotty. Virtual reality for stroke rehabilitation. In K. E. Laver, editor, *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd, feb 2015.

[16] H. Li, L. Trutoiu, K. Olszewski, L. Wei, T. Trutna, P.-L. Hsieh, A. Nicholls, and C. Ma. Facial performance sensing head-mounted display. *ACM Transactions on Graphics*, 34(4):47:1–47:9, jul 2015.

[17] W. Li, F. Abtahi, C. Tsangouri, and Z. Zhu. Towards an In-the-Wild Emotion Dataset Using a Game-Based Framework. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1526–1534. IEEE, jun 2016.

[18] K. Masai, Y. Sugiura, M. Ogata, K. Suzuki, F. Nakamura, S. Shimamura, K. Kunze, M. Inami, and M. Sugimoto. AffectiveWear. In *ACM SIGGRAPH 2015 Posters on - SIGGRAPH '15*, pages 1–1, New York, New York, USA, 2015. ACM Press.

[19] A. Mourão and J. Magalhães. Competitive affective gaming: winning with a smile. In *Proceedings of the 21st ACM international conference on Multimedia - MM '13*, pages 83–92, New York, New York, USA, 2013. ACM Press.

[20] M. Obaid, C. Han, and M. Billinghurst. "Feed the Fish": an affect-aware game. *Proceedings of the 5th Australasian Conference on Interactive Entertainment - IE '08*, pages 1–6, 2008.

[21] K. Olszewski, J. J. Lim, S. Saito, and H. Li. High-fidelity facial and speech animation for VR HMDs. *ACM Transactions on Graphics*, 35(6):1–14, nov 2016.

[22] L. Pereira, K. Obara, J. Dias, M. Menacho, E. Lavado, and J. Cardoso. Facial exercise therapy for facial palsy: systematic review and meta-analysis. *Clinical Rehabilitation*, 25(7):649–658, jul 2011.

[23] A. A. Pourmomeny and S. Asadi. Management of synkinesis and asymmetry in facial nerve palsy: a review article. *Iranian journal of otorhinolaryngology*, 26(77):251–6, oct 2014.

[24] V. Rantanen, J. Verho, J. Lekkala, O. Tuisku, V. Surakka, and T. Vanhala. The effect of clicking by smiling on the accuracy of head-mounted gaze tracking. In *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12*, page 345. ACM Press, 2012.

[25] M. F. Scheier and C. S. Carver. The Self-Consciousness Scale: A Revised Version for Use with General Populations1. *Journal of Applied Social Psychology*, 15(8):687–699, dec 1985.

[26] U. Segerstrale and P. Molnar, editors. *Nonverbal communication: where nature meets culture*. Lawrence Erlbaum Associates, 1997.

[27] P. Skalski, R. Tamborini, A. Shelton, M. Buncher, and P. Lindmark. Mapping the road to fun: Natural video game controllers, presence, and game enjoyment. *New Media & Society*, 13(2):224–242, mar 2011.

[28] V. Surakka, M. Illi, and P. Isokoski. Gazing and frowning as a new human–computer interaction technique. *ACM Transactions on Applied Perception*, 1(1):40–56, jul 2004.

[29] V. Surakka, P. Isokoski, M. Illi, and K. Salminen. Is it better to gaze and frown or gaze and smile when controlling user interfaces? In *Proceedings of the 11th International Conference on Human-Computer Interaction (HCI '05)*, 2005.

[30] C. T. Tan, H. Sapkota, and D. Rosser. BeFaced: a Casual Game to Crowdsource Facial Expressions in the Wild. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*, pages 491–494, New York, New York, USA, 2014. ACM Press.

[31] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner. FaceVR: Real-Time Facial Reenactment and Eye Gaze Control in Virtual Reality. *arXiv*, oct 2016.

[32] H. Tsujita and J. Rekimoto. Smiling makes us happier: enhancing positive mood and communication with smile-encouraging digital appliances. *Proceedings of the 13th international conference on Ubiquitous computing*, pages 1–10, 2011.

[33] O. Tuisku, V. Rantanen, O. Špakov, V. Surakka, and J. Lekkala. Pointing and Selecting with Facial Activity. *Interacting with Computers*, 28(1):1–12, jan 2016.

[34] T. Vanhala and V. Surakka. Facial activation control effect (face). In *Proceedings of the 2Nd International Conference on Affective Computing and Intelligent Interaction*, ACII '07, pages 278–289, Berlin, Heidelberg, 2007. Springer-Verlag.

[35] D. Watson and L. A. Clark. The Japanese Positive and Negative Affect Schedule: Factor-based scales for the assessment of mood, 1989.

[36] D. Watson, L. A. Clark, and A. Tellegen. Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54(6):1063–1070, 1988.

[37] N. Xiang, L. Yang, and M. Zhang. Dynamic Difficulty Adjustment by Facial Expression. In W. Du, editor, *Lecture Notes in Electrical Engineering*, volume 206 of *Lecture Notes in Electrical Engineering*, pages 761–768. Springer London, London, 2013.

[38] S. Yoshida, T. Tanikawa, S. Sakurai, M. Hirose, and T. Narumi. Manipulation of an emotional experience by real-time deformed facial feedback. *Proceedings of the 4th Augmented Human International Conference on - AH '13*, pages 35–42, 2013.