Eating chocolate, smelling perfume or watching video advertisement: Does it make any difference on emotional states measured at home using facial expressions?

Benjamin Mahieu, Michel Visalli, Pascal Schlich, Arnaud Thomas

Title

Eating chocolate, smelling perfume or watching video advertisement: does it make any difference on emotional states measured at home using facial expressions?

Authors

Benjamin Mahieu[a,b], Michel Visalli[a], Pascal Schlich[a], Arnaud Thomas[b]

[a]Centre des Sciences du Goût et de l'Alimentation, CNRS, INRA, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France.

[b]SensoStat, Dijon, France.

Corresponding author:

Benjamin Mahieu

benjamin.mahieu@inra.fr

Centre des Sciences du Goût et de l'Alimentation, CNRS, INRA, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France.

Abstract

The recording of facial expressions allows for implicit measurement of emotional states over time. The present study investigated whether these recordings can be acquired, using computer webcams, when testing products at home. Three types of product spaces (chocolates, perfumes and video advertisements) were evaluated at home by 44 subjects using a facial expression measurement protocol. Each product space was composed of three products. The first objective examined the feasibility of such a home-based protocol. The second objective investigated whether several products in the same product space could be characterized and discriminated using facial expression measurements. The third objective investigated potential differences in emotional responses between the different types of products. This study showed that a protocol of facial expression measurements at home was feasible and provided conclusive results. Perfumes and video advertisements could be discriminated between them, but chocolates were not discriminated. Perfumes and video advertisements elicited a temporal pattern of implicit emotions. These findings were obtained using a new method for performing a temporal analysis of facial expression measurements that accounts for individual baselines. The strength of the emotional response depended on the product type. Watching video advertisements elicited more emotions than smelling perfumes, which elicited more emotions than eating chocolates. These results showed that facial expression measurements are more adapted to certain product types.

Keywords

- Facial expression measurements
- Implicit emotions
- Temporal analysis
- Home Used Test

1. Introduction

The study of emotions in sensory and consumer research has grown rapidly in the past 10 years (Meiselman, 2015). Emotions help understanding consumers' choices and add information to liking and sensory methods (Gutjar et al., 2015). Methods to evaluate and measure emotions in sensory and consumer studies remain controversial and strongly depend on the aim of each study (Meiselman, 2015). However, emotion measurement methods can be classified in two categories: explicit (or declarative) and implicit. Examples of explicit methods include EsSense Profile™ (King & Meiselman, 2010) and PrEmo® (P. M. A. Desmet, Hekkert, & Jacobs, 2000), which measure emotion in a static manner immediately after product evaluation. Implicit methods include physiological measurements, such as heart rate, skin conductance and skin temperature (R. A. de Wijk, He, Mensink, Verhoeven, & de Graaf, 2014; René A. de Wijk, Kooijman, Verhoeven, Holthuysen, & de Graaf, 2012; He, Boesveldt, Delplanque, de Graaf, & de Wijk, 2017), and facial expression analysis, using an automated facial expressions analyser (Crist, Duncan, & Gallagher, 2016; Danner, Haindl, Joechl, & Duerrschmid, 2014; Leitch, Duncan, O'Keefe, Rudd, & Gallagher, 2015; Lewinski, Fransen, & Tan, 2014). This use of this last method video records subjects during their evaluation of each product. The video records are analysed frame-by-frame using an automated facial expression analyser at a given frequency. The recording duration and frequency used to analyse video records depend on the aim of the study and left to the discretion of the panel leader. Automated facial expression analysers are trained classifiers that take as entry an image containing faces and return emotional intensity scores for each face based on its expressions. Images in sensory and consumer research contain only one face that corresponds to a subject. Automated facial expression analysers are often based on Ekman's theory of emotions (Ekman, 1982) and use the Facial Action Coding System (FACS) (Ekman & Frisen, 1978; Frisesen & Ekman, 1983) to classify emotional responses

(Crist et al., 2016; Kuilenburg, Wiering, & Uyl, 2005). A baseline is highly recommended when using facial expression measurements to have a neutral emotion reference for the subjects (Crist et al., 2016). Two types of baselines can be distinguished. The first baseline is “baseline time without stimulus” (Crist et al., 2016), which consists of recording subjects without any stimulus and considering it as a neutral emotion state (Danner et al., 2014; Garcia-Burgos & Zamora, 2013; He et al., 2017; Rocha-Parra, García-Burgos, Munsch, Chirife, & Zamora, 2016). The second baseline is “baseline treatment stimulus” (Crist et al., 2016), which consists of recording subjects while evaluating an assumed emotion neutral stimulus of the same nature as the interest stimuli (e.g. water for liquid food products) (Crist et al., 2016; René A. de Wijk et al., 2012; Leitch et al., 2015).

Emotions exhibit a temporal dynamic component (Sander, Grandjean, & Scherer, 2005). However, this dynamic component was primarily only evaluated using Temporal Dominance of Emotions (TDE) (Jager et al., 2014), which is an extension to emotions of the Temporal Dominance of Sensations (TDS) (Pineau et al., 2009) method that measures emotions in a dynamic declarative manner during product evaluation (Jager et al., 2014; Peltier, Visalli, & Thomas, 2019). Only a few studies considered the dynamic information of facial expression measurements (Zhi, Wan, Zhang, & Li, 2018), and these studies only considered a short duration (< 10 seconds) of dynamic information of facial expression measurements (Crist et al., 2016; R. A. de Wijk et al., 2014; Leitch et al., 2015).

Previous evaluations of emotions using facial expression measurements focused on food products or advertisements (Leitch et al., 2015; Lewinski et al., 2014). To our knowledge, there are no studies evaluating emotions elicited by cosmetics using facial expression measurements. The eventual differences in emotions elicited by different product types have not been investigated experimentally using facial expression measurements.

The present study investigated emotions elicited at home by three different types of products: food products (chocolates), cosmetics (perfumes) and advertisements (video advertisements). The hypothesis was that video advertisements would elicit a greater emotional response than perfumes, which would elicit a greater emotional response than chocolates. The choice of these types of products enabled results comparisons to previous explicit dynamic emotion characterization of chocolates and video advertisements that were performed using the TDE protocol (Jager et al., 2014; Peltier et al., 2019). Emotions were measured implicitly via facial expression analyses using an automated web service. The first objective of this study assessed the feasibility of a facial expression measurement protocol using home-based video recording. The second objective investigated whether this protocol discriminated between products of the same type. The third objective examined whether the products elicited a temporal emotional response. The final objective compared the emotional responses elicited by the three different types of products and reached conclusions on potential differences.

2. Material and methods

2.1 Participants

To create a situation as close as possible to everyday consumption situations, the study was performed at home using 44 naïve subjects (15 men and 29 women), 18 to 65 years old. They were recruited from a population registered in the SensoStat's consumer database. This database has been declared to the relevant authority (Commission Nationale Informatique et Libertés - CNIL - n° d'autorisation 1655750). Subjects were selected based on their dark chocolate consumption frequency (at least once every two weeks), their agreement to be video recorded at home using their own webcam, their ability to read on a computer screen without glasses to avoid noise in facial expression measurements (Danner et al., 2014) and their possession of a computer equipped with a webcam. Subjects were economically rewarded for their participation in the study. Some data were not recorded or lost due to technical problems (described later). Twenty-two to 34 (mean of 29) evaluations were available for a given product.

2.2 Products

Three types of products were evaluated in this study: chocolates, perfumes and video advertisements for Nespresso® coffees. The subjects evaluated three products in each product category.

2.2.1 Chocolates

Three commercially available dark chocolates by Barry Callebaut® were selected:

- Choc 1 was a 73 % Ecuador cocoa chocolate that was generally perceived as bitterer than the other two chocolates.
- Choc 2 was a 67 % Madagascar cocoa chocolate that was generally perceived as sourer than the other two chocolates.
- Choc 3 was 54 % multiple origins cocoa chocolate that was generally perceived as sweeter than the other two chocolates.

Descriptions of the three chocolates were based on previous TDS studies (not reported here). These chocolates were selected because they were discriminated according to their sensory description in these previous studies.

2.2.2 Perfumes

Three commercially available perfumes by l’Occitane en Provence® were selected:

- Perf 1 was floral, fruity and feminine.
- Perf 2 was gourmet, aromatic and feminine.
- Perf 3 was fresh and mixed.

Descriptions of the three perfumes were based on l'Occitane en Provence®'s website. An expert in fine fragrances from l'Occitane en Provence ® Company also provided advice to the authors. This expert assured the authors that perfumes were from very different sensory spaces. Therefore, the perfumes used in the present study were different in their sensory properties.

2.2.3 Video advertisements

Three video advertisements for the same coffee brand, Nespresso®, were selected:

- Ad 1 (https://www.youtube.com/watch?v=DfyeXrdZZ1o)
- Ad 2 (https://www.youtube.com/watch?v=JwLSZ4y0GDY)
- Ad 3 (https://www.youtube.com/watch?v=ONVMIqZzw34)

These video advertisements are described in more detail in Peltier et al. (2019) using the same corresponding numbers.

2.3 Procedure

2.3.1 General procedure

Subjects participated in three home-based sessions on their own computer equipped with a webcam and running TimeSens software 2.0 (INRA, Dijon, France). To access the sessions, subjects clicked on a link sent via e-mail, which opened their browsers in the session's URL. Each session corresponded to one type of product and lasted approximately 10 minutes. The interval between two sessions was at least 24 hours. Within each session, products were presented following a Williams Latin Square experimental design. At the beginning of each session and before product evaluations, subjects were video recorded for 20 seconds without any stimuli. During this recording, subjects were instructed to face their webcam and to keep their head still. The aim of this recording was to approximate, as best as possible, a neutral state for each subject for use as a baseline. Subjects were video recorded during the evaluation of each product, after receiving instructions depending on the product type. Participants knew that they were video recorded. However, they did not know that their videos would be used to analyse their facial expressions and measure their felt emotions implicitly.

2.3.2 Chocolates

The subjects were instructed to click a “START” button when putting the chocolate in their mouth and watch the screen for 60 seconds. They had to put the same amount of chocolate in their mouth for all chocolates and rinse their mouth between chocolates during a 60-second forced break. No instructions were given about how to process the chocolate in their mouth. Subjects were free to process chocolates in their usual manner.

2.3.3 Perfumes

The subjects were instructed to click a “START” button before spraying the perfume on a paper stick then watch the screen for 60 seconds. During this time, they were free to smell the perfume as they wanted, but they were not allowed to spray the paper stick again. A 120-second forced break between each perfume evaluation enabled the subjects to “reset” their smell via smelling the crook of their elbow.

2.3.4 Video advertisements

The subjects were instructed to click a “PLAY” button to start the video advertisements, which lasted 48 to 54 seconds (ad duration + 3 seconds). Between two evaluations of video advertisement, a 60-second forced break that proposed a “Did you know that” question was displayed on the screen to avoid the subject’s boredom (Galmarini, Symoneaux, Visalli, Zamora, & Schlich, 2016). A “Did you know that” question was only added for video advertisements because subjects had nothing to do between two video advertisement evaluations, in contrast to the chocolate and perfume evaluations.

2.4 Facial expression measurements

All videos were analysed using an automated facial expressions analyser web service (Microsoft Cognitive Services Emotion API ©) at a 6 Hz frequency (6 images per second). For each frame analysed, the web service calculated an emotion intensity score (from 0 = absent to 1 = fully expressed) for each basic emotion according to Ekman (1982), namely happiness, surprise, fear, disgust, anger and sadness, plus contempt, as reported by Ekman and Frisesen (1986), and neutral. According to the web service documentation, the sum of emotion intensity scores adds up to one. In cases of no emotion expressed by a subject, an emotion intensity score of neutral equalled one, and the other emotion intensity scores equalled zero. The automated facial expression analyser Microsoft Cognitive Services Emotion API © is a trained classifier. This analyser uses a cloud-based emotion recognition algorithm and a deep learning neural network (Howard, Zhang, & Horvitz, 2017). It was chosen because of its ability to communicate with TimeSens© (the data acquisition software used in this study) and its high accuracy (Deshmukh & Jagtap, 2017). To our knowledge, this study is the first time that this web service was used in the field of sensory and consumer research, but it was used successfully in other fields (Kang, Wang, Wang, Angsuesser, & Fei, 2017; Mateus, Weber, & Rodriguez, 2016; Takáč, Mach, & Sinčák, 2016). Some technical problems (micro cuts in personal Wi-Fi connection involving an incomplete upload of the videos) were encountered and resulted in the loss of some videos (42% for chocolates, 27% for perfumes and 35% for advertisements).

2.5 Data analyses

Data analyses were conducted using R 3.5.1 (R Core Team, 2018). A non-temporal and a temporal approach were considered. The three types of products were analysed separately.

2.5.1 Baseline

The mean of each emotion score was computed for all subjects, per session, using the first 20 seconds of "neutral" recording, which resulted in one baseline score per emotion per subject per session. As expected, the baseline score of neutral was the highest at an average of 0.9326 across subjects and the three sessions. This means that the other seven emotional attributes shared the remaining 0.0674 (1-0.9326) to add up to 1. Each individual baseline score was subtracted to the corresponding raw score (i.e. same subject, same session, same emotion) returned by the facial expressions analyser software (every 1/6 second of evaluation). This procedure yielded emotion scores that ranged between -1 and 1, which were used in subsequent analyses.

2.5.2 Non-temporal approach

The use of mean emotion scores is a common method to analyse facial expression measurements in a non-temporal manner (Garcia-Burgos & Zamora, 2013; Leitch et al., 2015; Zhi et al., 2018). The maximum emotion scores (Danner et al., 2014; Zhi et al., 2018) were also considered in the non-temporal approach, but these results are not presented because they less discriminated the products. Other approaches, such as "frequency parameter" (number of times an emotion is the most expressed compared to others), "dynamic parameter" (weighted maximum emotion scores) (Zhi et al., 2018) or the use of values only above the 90th percentile (Lewinski et al., 2014) could have been considered. However, the mean emotion scores produced sufficient interesting results, especially for comparisons of emotional responses from the three types of products.

The mean emotion score was computed for each product and each emotion, over all evaluation times and subjects. The nullity of each of these means was tested using a bilateral t-test ($\alpha = 5$ %). The mean of emotion score was also computed for each evaluation (product × subject) and each emotion over all evaluation times. A two-way ANOVA (REML criterion) was performed on the mean emotion scores with product as a fixed factor and subject as a random factor. In case of significant differences ($p < 0.05$) in mean emotion scores between product, means were compared using Tukey's HSD post-hoc test ($\alpha = 5\%$).

2.5.3 Temporal approach

The mean emotion score was computed for each product and each emotion over the panel at each time (every 1/6 s). The computed mean nullity at each time and each emotion was tested using a bilateral t-test ($\alpha = 5$ %). If the mean was significantly different from zero, then the mean value was kept. If the mean was not significantly different from zero, the mean value was reduced to zero. An emotions temporal curve was drawn for each product to display the significant means over evaluation time.

3. Results

3.1 Non-temporal approach

Table 1 shows the mean values of the emotion scores. The F-values (Kenward-Roger method) correspond to the ANOVA performed for each type of products by emotion. Table 1 shows that Choc 1 (0.0099) and Choc 2 (0.0162) elicited a bit more sadness on average than baseline. Choc 2 (0.0099) elicited a bit more surprise. Although some mean emotion scores were significantly greater than zero for some chocolates these mean emotion scores stay very close to zero. None of the mean emotion scores of neutral was significantly different from zero for chocolates. These results show a quasi-absence of emotional response elicited by chocolates. Perf 1 (0.0388), Perf 2 (0.0458) and Perf 3 (0.0312) elicited more sadness on average than baseline Perf 3 (0.0001) elicited slightly more fear than baseline. Perf 1 (-0.0409) and Perf 3 (-0.0436) were less neutral than baseline on average. These results show that perfumes elicited an emotional response. This emotional response was primarily composed of sadness. Ad 1 (0.0165) and Ad 2 (0.0245) elicited more sadness than baseline on average, and Ad 3 (0.1447) elicited more happiness than baseline. Ad 1 (-0.0640) and Ad 3 (-0.1466) were less neutral than baseline on average. Ad 2 (-0.0001) and Ad 3 (-0.0001) elicited slightly less fear than baseline, and Ad 3 (-0.0039) elicited a little less surprise. These results show that video advertisements also elicited an emotional response. This response was primarily composed of happiness and sadness.

Table 1 shows that chocolates did not elicit different mean emotion scores ($p > 0.05$). Perfumes were different in their elicited mean score of happiness ($F = 3.84$, $p < 0.05$), and Perf 3 (0.0155) elicited more happiness than Perf 2 (0.0009). Video advertisements were different in their elicited mean scores of happiness ($F = 9.23$, $p < 0.001$), neutral ($F = 4.97$, $p < 0.05$) and sadness ($F = 8.16$, $p < 0.001$). The happiness elicited by Ad 3 (0.1447) was greater than Ad 1 (0.0397) and Ad 2 (0.0030). The sadness elicited by Ad 3 (0.0009) was less than Ad 1 (0.0165) and Ad 2 (0.0245). Ad 2 (-0.0384) was more neutral than Ad 3 (-0.1466). It is logical that Ad 3, with two apparent funny moments in its sequence (https://www.youtube.com/watch?v=ONVMIqZzw34), elicited more happiness and was less neutral than the other two video advertisements, which included only one apparent funny moment.

Table 1: ANOVA of mean emotion scores per emotion, multiple comparison of product means and t-test of means.

	F-choc	Num DF	Den DF	Choc 1	Choc 2	Choc 3	F-perf	Num DF	Den DF	Perf 1	Perf 2	Perf 3	F-ad	Num DF	Den DF	Ad 1	Ad 2	Ad 3
Anger	0.50	2	45,04	0.0004	0.0015	-0.0001	0.10	2	56,06	-0.0001	-0.0004	-0.0002	0.16	2	48,96	0.0004	0.0037	0.0026
Contempt	1.17	2	45,76	0.0059	0.0015	0.0030	0.27	2	55,22	-0.0045	-0.0059	-0.0031	0.64	2	83	0.0069	0.0105	0.0025
Disgust	0.63	2	43,05	0.0005	0.0005	0.0000	0.39	2	93	0.0004	0.0002	0.0004	0.14	2	56,43	0.0000	0.0000	0.0000
Fear	1.36	2	43,41	0.0008	0.0010	0.0009	0.12	2	55,68	0.0003	0.0003	**0.0001**	1.29	2	46,86	0.0000	**-0.0001**	**-0.0001**
Happiness	0.51	2	44	-0.0212	-0.0130	-0.0123	**3.84***	2	49,57	0.0047(ab)	0.0009(a)	0.0155(b)	**9.23*****	2	49,75	0.0397(a)	0.0030(a)	**0.1447(b)**
Neutral	0.89	2	46,11	-0.0128	-0.0175	-0.0341	0.15	2	60,7	**-0.0409**	-0.0384	**-0.0436**	**4.97***	2	49,92	**-0.0640(ab)**	-0.0384(b)	**-0.1466(a)**
Sadness	0.67	2	50,52	**0.0099**	**0.0162**	0.0179	0.26	2	60,73	**0.0388**	**0.0458**	**0.0312**	**8.16*****	2	50,75	**0.0165(b)**	**0.0245(b)**	0.0009(a)
Surprise	1.02	2	47,76	0.0165	**0.0099**	0.0245	1.40	2	63,1	0.0011	-0.0025	-0.0004	1.18	2	46,62	-0.0006	-0.0032	**-0.0039**

Significance Codes: '***' ($p < 0.001$) '**' ($p < 0.01$) '*' ($p < 0.05$) for ANOVA F-test (Kenward-Roger method). Means with different letters are significantly different (Tukey's HSD test, $\alpha = 5$ %). Means in bold are significantly different from zero (t-test, $\alpha = 5\%$).

3.2 Temporal approach

Fig. 1 presents the emotions temporal curve of each product. No emotional temporality was shown for chocolates because the curves were near close zero for the three chocolates. Each perfume elicited its own temporal pattern of emotion scores. Perf 1 elicited a low peak of sadness from beginning to approximately 20 seconds and another low peak at approximately 41 seconds. Perf 2 elicited one low peak of sadness from approximatively 20 seconds to 50 seconds. Perf 3 elicited two low peaks of sadness from 10 to 20 seconds and from 30 to 40 seconds. Each video advertisement elicited a different emotional temporality. Ad 1 elicited one high peak of happiness from approximately 47 seconds to the end. This peak seemed to correspond to the one apparent funny moment of Ad 1, which is the final joke. Ad 2 showed no emotional temporality because the curves were close to zero. Ad 3 elicited two large high peaks of happiness: the first one between approximately 12 and 28 seconds and the second from approximately 30 seconds to the end. These peaks seemed to correspond to the two apparent funny moments of Ad 3. The second large peak was higher than the first peak. This difference may be explained by a funnier final joke compared to the previous apparent funny moment in this video.

4. Discussion

4.1 At-home video recording feasibility

To our knowledge, this study is the first time that a protocol of data collection for facial expression analyses was performed at home. Subjects' recruitment was a difficult task. Recording them at their home using their own webcam worried many subjects, who consequently declined participation in the study. Some technical restrictions in the operating system and web browser used for data collection also added further complications. Only "Mozilla Firefox" and "Google Chrome" on a Windows© or Android©-based computer supported the video recording API when the experiment was performed. The relatively low number of subjects in this study was partially due to these two difficulties of recruitment. In addition, some instability in internet connections resulted in the loss of videos. The pre-tests worked well using a wired connection. The wireless issues were not anticipated and only discovered during the tests. An automatic save of the videos on the local drive of the subjects' devices has been implemented in the software for future applications to enable the recovery of videos by the subjects in case of network interruption. In conclusion, a protocol of facial expression analysis performed at home was difficult to establish. However, this study obtained coherent and interpretable signals using facial expression measurements at home, despite all of these problems and difficulties.

4.2 Data treatment of automated facial expressions analyser (frame rate, baseline and analyses)

The choice of a 6-Hz frequency for analysing recorded videos via an automated facial expression analyser was made for computational time reasons. It may seem rather small compared to 50 Hz (René A. de Wijk et al., 2012; Garcia-Burgos & Zamora, 2013) and 25 Hz (R. A. de Wijk et al., 2014), but this choice was made after testing different frequencies (6 Hz, 12 Hz and 24 Hz) on a sub-dataset of this study. No influence of the frequency on result interpretation was found so it was decided to analyse the full dataset using a 6-Hz frequency.

The “baseline time without stimulus” approach (Crist et al., 2016) was chosen because the same baseline approach for the three types of products was desired and using a “baseline treatment stimulus” (Crist et al., 2016) would have been too complicated for chocolates and video advertisements. Indeed, finding a solid comestible stimulus with the same texture as chocolate and without taste was nearly impossible. For the video advertisements, a video eliciting no emotion would have been very subjective. Additionally, product evaluation duration was too long to ask subjects for a baseline without stimulus that lasted the same duration as product evaluation. Indeed, a too long time without a stimulus could have elicited subject's boredom and led to a biased baseline. This possible bias would have made a direct comparison frame by frame between baseline and evaluations impossible. Finally, an average neutral state over a baseline video record of 20 seconds was chosen. This choice was a compromise between the subject's boredom and a sufficient duration of the baseline video. A duration of 20 seconds was assumed sufficient to establish a representative neutral state for each subject.

Representations of facial expression measurement using curves were proposed by Zhi et al. (2018) and R. A. de Wijk et al. (2014). Leitch et al. (2015) and Crist et al. (2016) proposed using time series to analyse facial expression measurements. However, to our knowledge, the present study is the first time that emotions temporal curves were represented considering individual baselines. In addition, this is the first time this was done with a duration of product evaluation greater than 10 seconds. In this representation, curve height is an important aspect to examine and its interpretation should strongly depend on it. It was shown that the use of a baseline produced a real interest using a facial measurement protocol because the interpretation of emotions temporal curve of Ad 1 and Ad 3 were consistent with the sequence of events present in these videos.

4.3 Product discrimination and pertinence of facial expression analyses

The three types of products elicited different emotional responses, which confirmed our study hypothesis. Indeed, considering a static approach, it is remarkable that the video advertisements were less neutral than perfumes, which were less neutral than chocolates. Further, there were more differences in baseline for video advertisements compared to perfumes and perfumes compared to chocolates. Additionally, facial expression measurements enabled the discrimination of video advertisements on three emotions, perfumes on one emotion and chocolates on no emotions. It is also remarkable that F-values increased across Table 1 (3.1) from chocolates to video advertisements, which supports the hypothesis of an emotional response gradient due to product type. This emotional response gradient was confirmed using the temporal approach because the peaks of video advertisements reached higher values than perfumes. The curves for chocolates were very close to zero, which shows a quasi-absent emotional response. However, it is important to keep in mind for the temporal approach that video advertisements were the only standardized stimuli. This factor can explain why the peaks of video advertisements reached higher values than perfumes and chocolates. Consensus was easier to reach for video advertisements. Anyway, the conclusions drawn using the non-temporal approach remain true because they were not affected by the standardization of evaluation time.

It is hard to know if the quasi-absence of an emotional response for chocolates was due to an effective absence of emotion elicited by chocolates or to the difficulty encountered by the automated facial expression analyser to analyse facial expressions due to face movements induced by chewing and swallowing (Crist et al., 2016; Danner et al., 2014). Even if the chocolates' temporal curves were hardly interpretable or directly comparable due to an absence of control in the way the chocolates were processed in mouth, the results from the non-temporal approach (not affected by non-standardization) remain true. This result is consistent with some previous studies (Leitch et al., 2015; Zhi et al., 2018) that showed the impossibility of discriminating liquid food products, which induce fewer facial movements, via comparisons of their mean emotion scores over evaluation time. Maximum emotion scores are not always conclusive on liquid food products (Zhi et al., 2018) even if some studies found them conclusive (Danner et al., 2014; R. A. de Wijk et al., 2014). With this in mind, it seems that the facial expression emotions elicited by food products are almost useless to discriminate several food products of the same type. Nevertheless, it is important to remember that the analyses presented here were performed at a panel level. Investigations of individual data were performed, and differences were found between subjects. This result means that emotional responses exist and differ at the individual level, but there were not enough subjects to segment the panel based on these differences.

The static approach for perfumes revealed that products were different in eliciting happiness, but the effect size of these differences is quite small. Therefore, these results should be interpreted with caution. The temporal approach showed that each perfume exhibited its own pattern. However, these patterns were based on low peaks. Low peaks may be due to a weak consensus of the panel, which may be explained by the fact that subjects were free to smell perfumes in their own way with no standardization of any kind. This lack of standardization makes the comparison between perfume temporal curves difficult. Standardizing the moments when the subjects smelled the perfumes may improve future similar protocols and facilitate consensus at the panel level. The patterns of perfumes were based on one emotion: sadness. Some of the videos recorded during the subjects' evaluation of perfumes (not all) revealed to the authors that the expressed emotion was not sadness, but rather disgust, because many of the subjects exhibited a rearward movement with "lip corner depressor" (Ekman & Frisen, 1978). "Lip corner depressor", and others action units, are found both in sadness and disgust expressions (Du, Tao, & Martinez, 2014; Frisesen & Ekman, 1983). Therefore, it is possible that the sadness emotion curves observed for perfumes are disgust emotion curves, or at least a mixture of sadness and disgust curves. This last point makes the temporal curves of perfumes more interpretable and may help explain the weak consensus on the temporal curves of perfumes.

The results of video advertisements are promising. Indeed, video advertisements were well discriminated using the static approach. Ad 3 elicited more happiness than Ad 1 and Ad 2 and was less neutral than Ad 2. Ad 1 was between the other two video advertisements in neutral scores. The elicited peaks of happiness occurred at funny moments of the video advertisements using the temporal approach. Therefore, the use of an automated facial expressions analyser to investigate emotions elicited from the video advertisements provided interpretable and coherent results. However, these hopeful results for video advertisements should be balanced because according to Reisenzein, Studtmann, and Horstmann (2013), there is very high coherence between amusement and smiling or “lip corner puller” (Ekman & Frisen, 1978), which only occurs in expressions of happiness (Frisesen & Ekman, 1983). Lewinski et al. (2014) also showed that happiness was an efficient emotion in discriminating some video advertisements made to elicit amusement using facial expression measurements. In addition, video advertisements were standardized stimuli, which made it easier to reach a consensus at the panel level. This standardization may explain why the emotions temporal curves of the video advertisements fit the video advertisements' sequence. Furthermore, watching videos is less occlusive for the automated facial expressions analyser than the other tasks asked in this study, as also suggested by Crist et al. (2016), which supports the good conclusive results obtained for video advertisements compared to chocolates and perfumes.

4.4 Explicit vs. implicit measurements

It is possible to compare directly the results of video advertisements to the results of TDE in Peltier et al. (2019) because the video advertisements used were the same in the two studies. The non-temporal approach results of this study were comparable to Peltier et al. (2019) using ANOVA of the durations of dominance of “No emotion” and “Joy” attributes. For temporal approach, the happiness temporal curves of Ad 1 and Ad 3 exhibited nearly identical development as the “Joy” attribute of TDE curves (Jager et al., 2014; Pineau et al., 2009). This direct comparison suggests that for standardized stimuli made to elicit emotion, both implicit and explicit methods produce almost identical information.

It is also interesting to compare the results for chocolates with these obtained using TDE in Jager et al. (2014) because both studies included dark chocolates, even if the chocolates were not exactly the same. With TDE, “classic” dark chocolates (70% cocoa & 85% cocoa) were discriminated using a non-temporal approach (Canonical Variate Analysis) and a temporal approach (TDE curves). In contrast, the present study used an implicit protocol and demonstrated that neither the non-temporal nor the temporal approach discriminated dark chocolates. This result suggests that explicit protocols (like TDE) can exacerbate the emotion elicited by chocolates. This exacerbation is a well-known “problem” of self-reported emotion, which are subject to social desirability (Arnold & Feldman, 1981) and “cognitive bias” (Poels & Dewitte, 2006). Future studies should be performed to compare temporal implicit and explicit protocols on the same set of chocolates, and eventually on perfumes, to examine the benefits and downsides of these two types of protocols.

4.5 Automated facial expressions analyser limitations

Notably, the use of an automated facial expression analyser implies some limitations. The major limitation is the selection of the set of emotions measured during product evaluations. To our knowledge, automated facial expressions analysers are often based on Ekman’s theory (Ekman, 1982), which seems limited in the case of consumer emotion research (P. M. Desmet & Schifferstein, 2008; Laros & Steenkamp, 2005; Richins, 1997). The main limitation is that Ekman’s theory includes more negative than positive emotions, which results in measuring an unbalanced set of emotions from a valence point of view. The other limitation is that Ekman’s theory includes emotions that are not often expressed, and it does not include emotions that are often expressed in consumer emotion research, especially for food products (P. M. Desmet & Schifferstein, 2008). Therefore, Ekman’s theory is not well adapted to food products, which provides an explanation for the quasi-absent emotional response using the implicit protocol. This last point is likely also true for perfumes because Ekman’s theory of emotions is only a sample of the possible felt emotions (Laros & Steenkamp, 2005; Scherer, Schorr, & Johnstone, 2001).

5. Conclusion

This new type of home-based protocol of implicit facial expression measurements is feasible. This study showed promising results despite some technical problems in data acquisition. Conclusive results were obtained under conditions as close as possible to natural use conditions. Video advertisements and perfumes were discriminated between them using a facial expression measurements protocol, and a temporal emotional response was found for these products. The chocolates were not discriminated between them and did not elicit a temporal emotional response. The hypothesis of the study was confirmed; video advertisements elicited a greater emotional response than perfumes, which elicited a greater emotional response than chocolates. A low frequency (6 Hz) was sufficient to obtain an interpretable signal of facial expression measurements. The new temporal approach introduced in this study provided very interesting and coherent results by taking into account subjects' individual baselines. This approach showed the usefulness of the baseline because it led to relevant results. This temporal approach must be investigated further because it provides an advantage in facial expression measurements analyses and adds information to non-temporal analyses, which are often performed. Facial expression measurements were difficult for food products because of mastication movements and the limited number of measured emotions. Further, the reality of emotions elicited by food products remains questionable. Facial expression measurements are more adapted to standardized stimuli that are made to elicit emotion, such as video advertisements. Drawing conclusions at the panel level is complicated using non-standardized stimuli. Facial expression measurements may be a useful tool for cosmetics with a standardized signal. However, future studies must be performed to confirm these conclusions.

Acknowledgment

The authors would like to thank Barry Callebaut® and l'Occitane en Provence® for providing their products for the study and Nespresso® for its interest in the study.

References

Arnold, H. J., & Feldman, D. C. (1981). Social Desirability Response Bias in Self-Report Choice Situations. *Academy of Management Journal, 24*(2), 377-385.

Crist, C. A., Duncan, S. E., & Gallagher, D. L. (2016). Protocol for Data Collection and Analysis Applied to Automated Facial Expression Analysis Technology and Temporal Analysis for Sensory Evaluation. *J Vis Exp*(114).

Danner, L., Haindl, S., Joechl, M., & Duerrschmid, K. (2014). Facial expressions and autonomous nervous system responses elicited by tasting different juices. *Food Res Int, 64*, 81-90.

de Wijk, R. A., He, W., Mensink, M. G., Verhoeven, R. H., & de Graaf, C. (2014). ANS responses and facial expressions differentiate between the taste of commercial breakfast drinks. *PLoS One, 9*(4), e93823.

de Wijk, R. A., Kooijman, V., Verhoeven, R. H. G., Holthuysen, N. T. E., & de Graaf, C. (2012). Autonomic nervous system responses on and facial expressions to the sight, smell, and taste of liked and disliked foods. *Food Quality and Preference, 26*(2), 196-203.

Deshmukh, R., & Jagtap, V. (2017). A Survey: Software API and Database for Emotion Recognition.

Desmet, P. M., & Schifferstein, H. N. (2008). Sources of positive and negative emotions in food experience. *Appetite, 50*(2-3), 290-301.

Desmet, P. M. A., Hekkert, P., & Jacobs, J. J. (2000). When a Car Makes You Smile: Development and Application of an Instrument to Measure Product Emotions. *Advances in Consumer Research, 27*, 111-117.

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proc Natl Acad Sci U S A, 111*(15), E1454-1462.

Ekman, P. (1982). Emotions in the Human Face. *Cambridge University Press*, 353-395.

Ekman, P., & Frisen, W. V. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement *Consulting Psychologists Press, 22*.

Ekman, P., & Frisesen, W. V. (1986). A New Pan-Cultural Facial Expression of Emotion *Motivation and Emotion, 10*(2).

Frisesen, W. V., & Ekman, P. (1983). EMFACS-7: Emotional facial action coding system.

Galmarini, M. V., Symoneaux, R., Visalli, M., Zamora, M. C., & Schlich, P. (2016). Could Time–Intensity by a trained panel be replaced with a progressive profile done by consumers? A case on chewing-gum. *Food Quality and Preference, 48*, 274-282.

Garcia-Burgos, D., & Zamora, M. C. (2013). Facial affective reactions to bitter-tasting foods and body mass index in adults. *Appetite, 71*, 178-186.

Gutjar, S., de Graaf, C., Kooijman, V., de Wijk, R. A., Nys, A., ter Horst, G. J., et al. (2015). The role of emotions in food choice and liking. *Food Research International, 76*, 216-223.

He, W., Boesveldt, S., Delplanque, S., de Graaf, C., & de Wijk, R. A. (2017). Sensory-specific satiety: Added insights from autonomic nervous system responses and facial expressions. *Physiol Behav, 170*, 12-18.

Howard, A., Zhang, C., & Horvitz, E. (2017). Addressing bias in machine learning algorithms: A pilot study on emotion recognition for intelligent systems. In, 2017 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO).

Jager, G., Schlich, P., Tijssen, I., Yao, J., Visalli, M., de Graaf, C., et al. (2014). Temporal dominance of emotions: Measuring dynamics of food-related emotions during consumption. *Food Quality and Preference, 37*, 87-99.

Kang, Y., Wang, J., Wang, Y., Angsuesser, S., & Fei, T. (2017). MAPPING THE SENSITIVITY OF THE PUBLIC EMOTION TO THE MOVEMENT OF STOCK MARKET VALUE: A CASE STUDY OF MANHATTAN. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W7, 1213-1221.

King, S. C., & Meiselman, H. L. (2010). Development of a method to measure consumer emotions associated with foods. *Food Quality and Preference, 21*(2), 168-177.

Kuilenburg, H. v., Wiering, M., & Uyl, M. d. (2005). A model based method for automatic facial expression recognition. Lecture notes in computer science. *Machine Learning: ECML, 3720*, 194-205.

Laros, F. J. M., & Steenkamp, J.-B. E. M. (2005). Emotions in consumer behavior: a hierarchical approach. *Journal of Business Research, 58*(10), 1437-1445.

Leitch, K. A., Duncan, S. E., O'Keefe, S., Rudd, R., & Gallagher, D. L. (2015). Characterizing consumer emotional response to sweeteners using an emotion terminology questionnaire and facial expression analysis. *Food Research International, 76*, 283-292.

Lewinski, P., Fransen, M. L., & Tan, E. S. H. (2014). Predicting advertising effectiveness by facial expressions in response to amusing persuasive stimuli. *Journal of Neuroscience, Psychology, and Economics, 7*(1), 1-14.

Mateus, A., Weber, H., & Rodriguez, A. (2016). Emotion and Mood in Design Thinking.

Meiselman, H. L. (2015). A review of the current state of emotion research in product development. *Food Research International, 76*, 192-199.

Peltier, C., Visalli, M., & Thomas, A. (2019). Using temporal dominance of emotions at home. Impact of coffee advertisements on consumers' behavior and methodological perspectives. *Food Quality and Preference, 71*, 311-319.

Pineau, N., Schlich, P., Cordelle, S., Mathonnière, C., Issanchou, S., Imbert, A., et al. (2009). Temporal Dominance of Sensations: Construction of the TDS curves and comparison with time–intensity. *Food Quality and Preference, 20*(6), 450-455.

Poels, K., & Dewitte, S. (2006). How to Capture the Heart? Reviewing 20 Years of Emotion Measurement in Advertising. *Journal of Advertising Research, 46*(1), 18-37.

R Core Team. (2018). R: A language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between Emotion and Facial Expression: Evidence from Laboratory Experiments. *Emotion Review, 5*(1), 16-23.

Richins, Marsha L. (1997). Measuring Emotions in the Consumption Experience. *Journal of Consumer Research, 24*(2), 127-146.

Rocha-Parra, D., García-Burgos, D., Munsch, S., Chirife, J., & Zamora, M. C. (2016). Application of hedonic dynamics using multiple-sip temporal-liking and facial expression for evaluation of a new beverage. *Food Quality and Preference, 52*, 153-159.

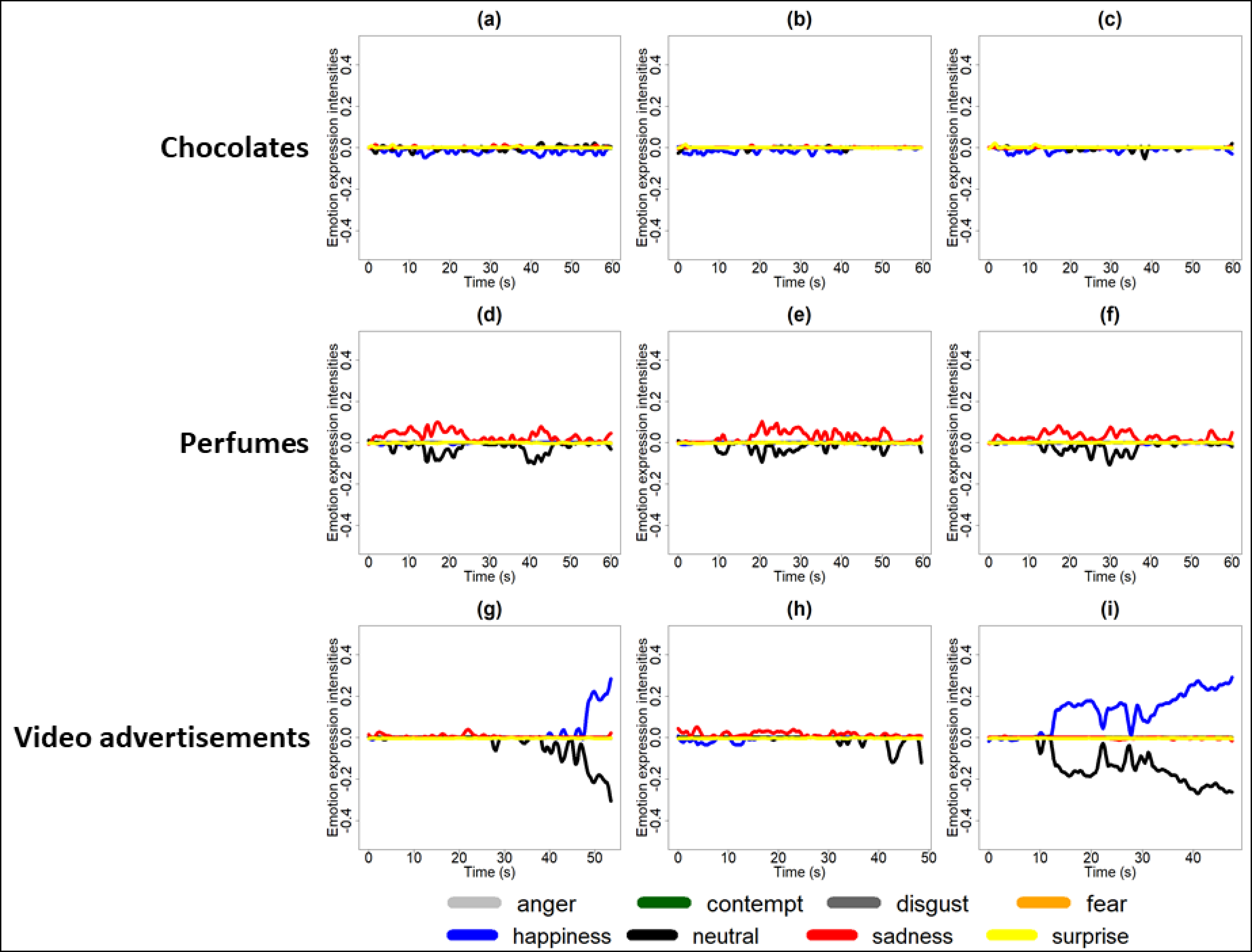
Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Network*, *18*(4), 317-352.

Scherer, K. R., Schorr, A., & Johnstone, T. (2001). Appraisal process in emotion: Theory, methods, research. 92-120.

Takáč, P., Mach, M., & Sinčák, P. (2016). Cloud-based facial emotion recognition for real-time emotional atmosphere assessment during a lecture. In 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC).

Zhi, R., Wan, J., Zhang, D., & Li, W. (2018). Correlation between hedonic liking and facial expression measurement using dynamic affective response representation. *Food Res Int*, *108*, 237-245.

Fig.1: Emotions temporal curves for Choc 1 (a), Choc 2 (b), Choc 3 (c), Perf 1 (d), Perf 2 (e), Perf 3 (f), Ad 1 (g), Ad 2 (h) and Ad 3 (i).


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
Chocolates
Perfumes
Video advertisements
Emotion expression intensities
Time (s)
anger
contempt
disgust
fear
happiness
neutral
sadness
surprise