How Commercial Food Videos Affect Female Customers
Measuring Female Bio-response Towards Commercial Food Videos

Xintong Zhu, Yong Wang, ZhenZheng Wang, Xiguang Wang and Chen Wang
Future Media Convengence Institute, Xinhua Net, Jinyu Building, Xicheng District, Peking, China

Keywords: Physiological Computing, GSR Sensors, Audience Experiences, Advertising Videos.

Abstract: The concept design of a food television commercial (TVC) could affect the effectiveness of immersive media and user experience. Traditional methods (e.g., surveys or eye tracking) for evaluating the targeted consumer's responses are serious limited in several aspects. In this paper, through closely working with the TVC designers, we used the data gathered from physiological sensor to measure viewers' watching experiences. We thereby conclude how we used our own Galvanic Skin Response (GSR) sensors to measure audience responses to the three clips of food TVCs. The results demonstrate that GSR sensors can provide fine-grained information for the advertisement community. Compared to subjective evaluation methods, the continuous user experience can be vividly visualized, and this enables designers to efficiently evaluate the impact of a TVC.

1 INTRODUCTION

The advertisement industry has been fully aware that a content design for a commercial campaign plays a vital role in spreading the name and the impact of products. Different food promotion strategies shall be specified with different concept designs in a campaign. In particular, to attract the targeted customers, advertising companies always seek for the optimal way to design a television commercial (TVC) that could seize the attention and further bringing the tangible revenues.

However, it is a great challenge to measure the impact of a TVC. First, obtaining such information is rather expensive and time consuming. For instance, the commercial companies generally hire agencies to run market researches that normally take several months. One of the feasible approaches is to carry out online questionnaires to collect the answers from the target customers and analyse the data. Second, traditional methods (e.g., questionnaires) are not the most appropriate one to measure the impact of a TVC, since questionnaire data is discrete, which means the data is hard to be connected to the whole duration of a TVC. In addition, subjective reports can easily collect information from the questions that users are aware of, but it is unlikely to ask them to answer some questions related to attention or engagement (Latulipe et al., 2011). Last, recently, eye movements

Figure 1: The experimental session.

(Krugman et al., 1994)and facial expressions (Joho et al., 2009)have risen as popular methods to measure the impact of a TVC. We doubt the credibility of such methods as a watcher might be completely absent-minded when they are watching a video. If so, eye tracking and facial expression data seem to be particularly meaningless for evaluating the influence of a TVC, since the relationship between user inner attention and visual attention still remains unclear.

Physiological measurements have many advantages as an alternative tool to monitor, in real-time, audience responses in a video consumption. The physiological data are continuous, providing time series that can be mapped to events (e.g., the moment that a brand logo appears) of a TVC. Second, the non-intrusive nature of the physiological measurements guarantees a seamless experience of the audience. In

Zhu, X., Wang, Y., Wang, Z., Wang, X. and Wang, C.
How Commercial Food Videos Affect Female Customers - Measuring Female Bio-response Towards Commercial Food Videos.
DOI: 10.5220/0006394300350044
In *Proceedings of the 4th International Conference on Physiological Computing Systems (PhyCS 2017)*, pages 35-44
ISBN: 978-989-758-268-4

addition, physiological sensors have been applied in audience research, e.g., user emotion (Oliveira et al., 2011)and user engagement(Picard and Daily, 2005). Although some studies use them to assess the impact of a video (Yoon et al., 1998), the audio effect (Kenning and Plassmann, 2008), or use questionnaires to investigate audience's purchase behaviour. Surprisingly, no studies have ever used physiological sensors to measure women's experience when watching a commercial. None of those studies has paid particular attention to female customers, who are most of the time the actual decision maker in a household's shopping journey.

In this paper, requested by the producers who have rich experiences in designing commercial videos, we dive into female customers' experience in commercial video consumption. Specifically, we first interviewed the producers and collected their research interests. Then, we designed the experiment and chose the appropriate methods to answer the research questions. In our case, we chose GSR sensors to measure user experience, as they are highly accurate for indicating the user arousal and the emotional intensities compared to other bio sensors. Taking advantage of the sensor data, we could map their response to specific events (e.g., the moment that the food brand appears).

Our contribution has twofold. On the one hand, we focus our research interests on female customers, since their watching experience has never been investigated in the previous studies. On the other hand, we use GSR sensors, instead of questionnaires, to continuously record user responses in a food TVC, a research area that has not been explored before.

1.1 Research Questions

The research questions were generalized from the interviews with the producers. They are both very experienced in commercial video production. They believe that there are many ways to produce a good TVC, but how to demonstrate that one concept design is superior than the other one is a challenge. In particular, the producers consider that how female customers are influenced by these TVCs is extremely important, because they are major buyers in households. In other words, if a TVC hits their heart, the chance is high that they will buy the product in the future. The producers proposed three TVCs to run the experiments, in which the different types of food are involved. They had particular interests in female reactions, i.e., how female would react when the food close-up shot first appears, how do they response differently when watching the three videos, and how their bio-responses are related to their subjective reports.

The interview was semi-conducted and coded based on the notes. Four researchers participated and concluded three research questions:

- R1: What are the differences in female consumers' responses when watching the three videos?
- R2: How do the female consumers react to the close-up of food first appearing in the videos?
- R3: How is the female bio-response correlated to the subjective reports?

The remainder of this paper is organized as follows. Related work will be reported first and the design alternatives for the hardware and the software are introduced next. Then the results are exhibited. followed with the discussion and the conclusion at last.

2 RELATED WORK

There are various definitions for audience response. For instance, some studies use different synonyms, such as audience engagement, audience feedback, audience interests, audience interaction and audience (or user) experience. Christopher Peters et al. (Peters et al., 2009) described audience response as a combination of focus, interest, perception, cognition, experience and action. On the other hand, Heather L. O'Brien et al. (O'Brien and Mclean, 2009) pointed out that audience engagement (response) can facilitate users with more enriching interactions in computer applications: engaged users tend to recommend the products (or service) to others. In addition, in affective computing, users' emotional response (Wang, Prendinger, and Igarashi, 2004) is obtained as an evaluation tool to define user engagement. In game application studies (Fischer and Benford, 2009), audience engagement (response) refers to players' state awareness and synchronization. Audience biofeedback, e.g., arousal, is also used as an indicator of the levels of players' engagement(Chanel et al., 2008).

Audience response can be measured in two ways: explicitly and implicitly. Explicit methods normally require users' intentional inputs, like surveys or ratings, whereas implicit measurements generally capture audience feedback through physiological sensors, such as GSR sensors. During the process of data collection, sensor readings are collected in real time without interrupting audience's watching journey.

Visual analysis methods, e.g., eye movements, are also classified into the scope of explicit methods of measuring audience response. In such studies, eye-gaze, eye movements and head movement trackers are installed to define users' interested spots during a video consumption or other applications like game studies. Some other methods, e.g., physiological sensors, are also combined with visual analysis methods to capture audience response, in which audience inner state are characterized, e.g., boredom or fatigue. However, these studies could not avoid intentional requiring audience annotations constantly, by which the labels of the attributes of the data are generated(Kapoor et al., 2007).

Audience response has a significant impact on many applications. There are plenty of quantitative studies executed in game researches based on the reliability and suitability of physiological sensors. Pejam et al.(Mirza-Babaei et al., 2013) have explored the possibilities of improving games design by providing user biofeedback, and the results showed that a combination of user GSR feedback would help designers choose proper design strategy for higher game play quality. In the game industries, Sony, for example, recently has decided to add GSR sensor into their new game controller, DualShock 4 (DualShock, 2017), where users' GSR-response is detected as interest level towards the game. Another example (Sakurazawa et al., 2003) is that users' GSR-response is used as the users' agitation in a game, and the more often a user feels agitated, the more enemies will appear in a game.

Previous studies have also shown that GSR sensor is one of the indicators for users' cognitive and emotional status. Lin et al (Lin et al., 2008) successfully investigated how audience GSR-response differs in the different movie sections. In particular, the fluctuations of GSR data were linked to events during a 3D movie experience.

Matthew et al. (Pan et al., 2011)demonstrated a novel interaction by using GSR sensors in an audio stream bookmarking, where users' GSR-response was monitored as a response to the external interruptions. GSR sensors were also applied in a wearable system in order to help users select the high arousal photos, which were the most relevant to the users' ordinary daily life(Sas et al., 2013).

Web applications are also benefited from audience response research. In the paper(Lunn et al., 2010), users' biofeedback was used to distinguish audience response within the different age scopes, in terms of a web 2.0 application. Audience affective states were also employed to investigate what kind of interaction technique on the web has a significant impact on users(Hart et al., 2012). Transforming audience physiological signals (audience response) into a smile icon was implemented in an online chat(Wang et al., 2004). Users' biofeedback on preference, e.g., like or dislike, was used as the input data in order to improve the accuracy of online recommendation system(Madan et al., 2004).

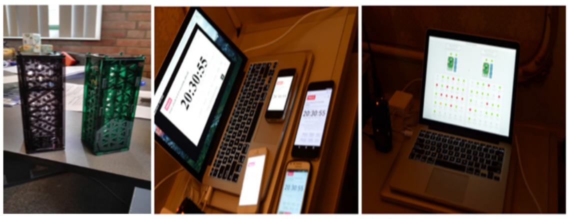

Figure 2: The GSR sensors and the monitoring system.

Audience response also plays an important role in some other applications. For instance， in Olympic Games, audience clapping frequency was visualized on the display screen to encourage athletes' performance(Barkhuus, 2008); audience 'cheering meter" was measured to aid voting at rap competitions(Aigner et al., 2004).

GSR, is also known as galvanic skin response, electro dermal response (EDR), psych galvanic reflex (PGR), skin conductance response (SCR), or skin conductance level (SCL). GSR sensors measure the users' electrical conductance of the skin, where users' sweat glands are varied and controlled by the sympathetic nervous system. Therefore, GSR sensors are normally considered as an indicator of psychological or physiological arousal or stress. When users are highly aroused, users' skin conductance increases in turn. GSR sensors can be either purchased from commercial companies or self-developed. Commercial companies, such as BioPac, Thought technology and Q sensors, offer this type of GSR sensor with a high price. Although such commercial sensors allow researchers to start experimenting immediately, they do not provide functions to measure groups of users simultaneously. The reason is that the communication protocol normally is Bluetooth, which has limitation to connect cell nodes in wireless network, e.g., a master and up to 7 slave piconet networks.

3 METHODOLOGY

3.1 Participants

All the participants were recruited from our institute per the requirements (from 25 to 34) of the producers, and they had not any visual or acoustic problems (Figure 1). All participants were divided into 4 teams with 4 persons in each.

3.2 Stimuli and Apparatus

The stimuli consisted of 3 different short videos (video A, B, and C). There was a short pause left between video clips (32 seconds) with grey screen to let users jump out of the previous watching experience. The video A is intended to prompt the organic blueberry, the video B advertises the ham product, and the video C addresses on the beer chicken. The three videos were played in different sequence for each team in order to minimize the influence of order. The duration of the experiments was 12 minutes 34 seconds in total.

We built the GSR sensors using a Jeenode board with a RF12 wireless module, a low pass filter and some accessories (e.g., electrodes) (Figure 2). The wireless function of the RF12 module makes it possible for simultaneously measuring audience at a large scale.

The sensors have been validated through a number of experiments (anonymous). All the sensor subordinates simultaneously send GSR data packets back

to the master sink node, which is connected to a laptop. The master node communicates with all the subordinate nodes by using a polling mechanism. In our case, we set up the sampling rate at 2Hz, which is not optimal, but it is sufficient to recover the sensor data. The sensors are robust against noise because of the circuit design and the data smoothing procedure in the software (Figure 3). In addition, all the sensor data were synchronized with the time stamp of the videos.

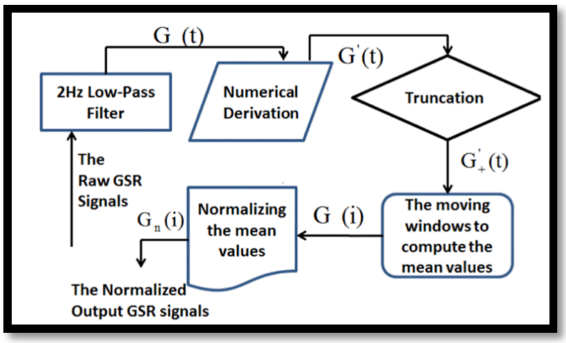

Figure 3: the diagram of the feature extraction of the GSR data.

3.3 Interviews and Questionnaires

The purpose of the interviews and questionnaires is to better interpret the sensor data. All participants were interviewed before and after they watched the videos. The interviews mainly focused on the following five aspects: the title of their job, their favourite food, their favourite video，which specific scenes in the videos that they were interested in, and reasons for likes and dislikes.

The questionnaires were designed at seven scales. The pre-questionnaires include the following questions:

1. *What is your job and responsibilities?*
2. *what kind of food do you like to eat?*
3. *what kind of food you do not like to eat?*
4. *what taste of food do you like?*
5. *what taste of food you do not like?*
6. *what kind of video do you like?*
7. *what kind of video you do not like?*
8. *Do you often watch video by smart phone?*
9. *Do you often buy food online?*

The post-questionnaires have three questions:

1. *What do you want to eat after watching?*
2. *Which video do you think the best?*
3. *Which style of the video do you like? And the reason?*

3.4 Software

All the videos were played on the four same-type smart phones, and the participant hold the phone in their most comfortable gesture. The whole experiment was video-recorded by using the facilities installed in our user lab (Figure 4). The recording video streams were collected by the software written in Python, which is installed in the observation room. All the sensor data was analysed by SPSS and MathWorks (Matlab).

3.5 Experimental Procedures

Before the experiment started, all the participants were asked to fill an informed consent form and the pre-questionnaires. Then the interview began and oral instructions were provided. After that, the experimenters helped the participants to wear the sensors, and then they opened the video play program, and clicked the “play” button. When the experiment was finished, the interviews and the post-questionnaires were conducted after the sensors were taken off.

3.6 Data Analysis

To explore the participant’s GSR response, both the event-related skin conductance response (SCR) and

Figure 4: the user lab and the observation room.

the skin conductance level (SCL) were analysed and normalized by Z-score. In addition to that, we also used analysis of variance (ANOVA) and Least Significant Difference (LSD) to test the significance of physiological data in three videos.

There are several steps to analyse Electro Dermal Activity (EDA) to extract the event-related SCR data based on Fleureau (Fleureau, Julien, Philippe Guillotel, and Izabela Orlac. 2013) and his colleagues' work.

Normally, when participants receive an engaging stimulus, their GSR value will increase quickly with a latency of 1-3 seconds, and after reaching a maximum value, it will recover to a value around the baseline level. Firstly, a 2Hz (*G(t)*) low-pass filter was applied to remove noise from the raw data, such as other physiological signals and electrical noise. Then a derivation (from $G(t)$ to $G'(t)$) was applied to calculate the rate of change of the GSR data. In doing so, we could know if the GSR value is ascending (positive values) or descending (negative values). After that all the negative values were discarded, with only the positive ones being kept (from $G'(t)$ to $G'_{+}(t)$), which means that we only focus on the increasing phases of the GSR signal, because the negative phases only reflect the recovery of the signal to the baseline. These steps helped us to extract the SCR data from the raw electrophysiological data.

$$G_n(i) = \frac{G(i)}{\sum_{j=1}^{k} G(j)} = \frac{\int_{W_i} G'_{+}(t)dt}{\sum_{j=1}^{k} \int_{W_j} G'_{+}(t)dt} \tag{1}$$

To temporally analyse emotional flow, we applied an overlapping time moving window with a window size of 30 samples (30 seconds), and an overlap of 15 samples (15 seconds). This step helped us to smooth the data and remove the users' GSR latency. So the mean values of $G'_{+}(t)$ were converted into $G(i)$ ($1 \le i \le k$, k is the number of the moving windows). $G(i)$ is the mean derivative value of one subsample in one specific moving window.

Since each individual may have a different amplitude for the derivative GSR signal when exposed to the same stimulus, $G(i)$ was divided by the sum of the subsampled skin response values (Formula (1)), and the output was $G_n(i)$ ($1 \le n \le N$, N is the number of the sample; $1 \le i \le k$, k is the number of the moving windows).

$G_n(i)$ is the individual value in a moving window, which cannot represent the whole group's response, considering some outliers, differences from person to person, noise (e.g. body movements) and so on. To define whether the group had a significant arousal or not, a statistical test called the bilateral Mann-Whitney-Wilcoxon (MWW) test was used. This test detects whether there is a significant difference between the audience arousal response ($G_n(i)$ and the background noise. We took the lowest 10% of the values in $G_n(i)$ as background noise. $G_n(i)$ of a single time sample was compared to the background noise of each time sample, which means that we used MWW test to compare k times and obtain k p-values for each time sample. The final p-value of each time sample is the averaged value of those k p-values. For final p-values lower than 5%, we considered the response during that time sample to be significantly different from the background noise.

The mean value of EDA of the first 32s blank video (grey screen) were used as the baseline, which was then subtracted from the raw electrophysiological data, to remove individual differences.

Z-score is the number of standard deviations from the mean a data point is. Z-scores range from -3 standard deviations (which would fall to the far left of the normal distribution curve) up to +3 standard deviations (which would fall to the far right of the normal distribution curve). The basic z score formula for a sample is:

$$z = (x - \mu) / \sigma \tag{2}$$

where, μ is the mean of the population, σ is the standard deviation of the population. The absolute value of z represents the distance between the raw score and the population mean in units of the standard deviation. z is negative when the raw score is below the mean, positive when above.

Here, we use Z-score to standardize the EDA in order to compare the changes of EDA when different short videos were played.

The ANOVA is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups. When ANOVA gives a significant result, this indicates that at least one group differs from the other group. Yet, the omnibus test does not indicate which group differs. In order to analyse the pattern of difference between means, the ANOVA is often followed by specific comparisons, and the most commonly used method is to compare two means. Here, we used least significant difference (LSD) to compute the smallest significant difference between two means as if these means had been the only means to be compared (*i.e.*, with a *t* test) and to declare any significant difference larger than the LSD. So, we firstly used ANOVA to analyse weather the EDA is different between some close-up shot and normal frame, and then the LSD analysis was used to compare them to see whose difference is significant.

4 RESULTS

In the Results section, there are four parts that help to answer the research questions mentioned before. First, the overview of the participants' physiological response is reported by showing the mean z-score of the 16 participants. So we could build a general idea of the participants' response of the three videos according to the first part. After that, the results of the SCR of the participants are described in the second part, and the results of the SCL are reported in the third part. Both the general response to different videos and the responses to specific scenes are analysed, enabling us to understand the participants' reaction to both the entire videos and the specific moments in the videos like the food close-up. And then, the results of the questionnaires are shown in the last part of the section, helping us to compare the results of the objective measurement (GSR sensor) and the subjective measurement (questionnaire).

4.1 Overview of the Participants' Physiological Response (R1)

First, the mean Z score levels corresponding to the time, analysed from the data of 16 participants during the whole experiment, are shown in Figure 5. An overview of the participants' physiological response to the three videos could be thus built.

In Figure 5, there are three lines representing the Z score of the participants' galvanic skin response to each video. Some descriptions of the moments in the video corresponding to the peaks are also exhibited in the figure. The peaks mean that the participants had the emotional arousals at this moment.

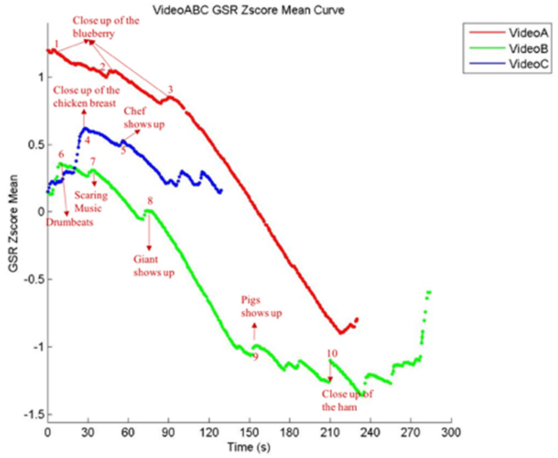

Figure 5: The mean of Z score for each video. The red line represents the Video A, which promotes the blueberries. The green line represents the Video B, which addresses the story of making ham. The blue line shows the Video C, which advertises the beer chicken. The X-axis is the time and the Y-axis is the GSR Z score. The red annotations are corresponded to the peaks, indicating that the participants had the emotional arousals.

From Figure 5, we can safely conclude that the participants' physiological response has three trends: 1) The participants' response decreased during all three videos, indicating that the participants were losing their interests while watching the videos, especially during the Video A. 2) It is easy to find that the participants had the emotional arousal to the close-up of the food when watching all videos. 3) Besides, the arousal level in the three videos varies. According to the mean Z score, the arousal level in the Video A is higher than the Video C, and the arousal level in the Video B remains the lowest.

4.2 SCR Related Video Events (R1&R2)

Based on the algorithm mentioned in the software section, event related significant emotional arousal were detected as shown in Figure 6, 7, and 8.

The SCR results from the three videos are rather different. The participants were significantly aroused in almost the entire watching process of the video C (Figure 8). As a contrast, only two moments caught the significant arousal in the video B (Figure 7) and no significant arousal appeared in the video A (Figure 6).

In addition, the SCR analysis also reveals which events significantly aroused the viewers. For instance, in the Video B, the participants were highly engaged in the scenes where the giant showed the

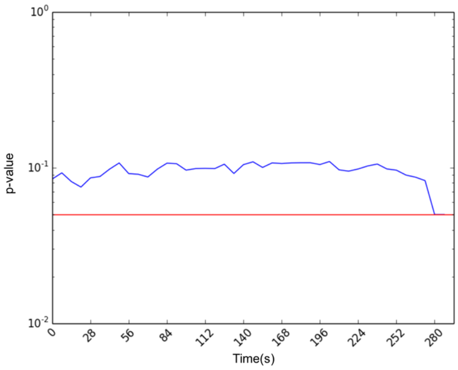

Figure 6: The extracted SCR signals of the participants during the Video A. The x-axis is the time in seconds, and the y-axis is the mean p value of the bilateral MWW test. The blue line is the participants' mean p value, the red line represents the critical value (p = .05). When the blue line is below the red line, it indicates that the participants had a significant emotional arousal response at 0.05 p value level.

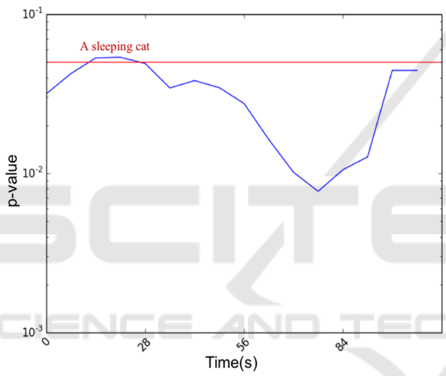

Figure 7: The extracted SCR signals of the participants during the Video B. The x-axis is the time in seconds, and the y-axis is the mean p value of the bilateral MWW test. The blue line is the participants' mean p value, the red line represents the critical value (p = .05). When the blue line is below the red line, it indicates that the participants had a significant emotional arousal response at 0.05 p value level.

ham to his friend, and the slogan and logo shown up at the end of the video, except for the moment (around 10 seconds) that they saw a sleeping cat. For the video A, the participants did not show any significant emotional moments. While for the video C, only two events significantly stimulated the participants' emotions.

4.3 SCL Levels in Three Videos (R1&R2)

The user SCL level differences in the videos are shown in Figure 9 and Figure 10, and the significant bio-responses were observed in them.

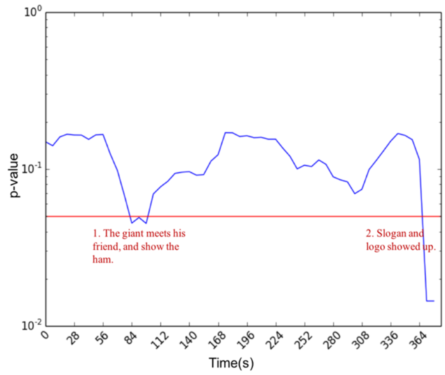

Figure 8: The extracted SCR signals of the participants during the Video C. The x-axis is the time in seconds, and the y-axis is the mean p value of the bilateral MWW test. The blue line is the participants' mean p value, the red line represents the critical value (p = .05). When the blue line is below the red line, it indicates that the participants had a significant emotional arousal response at 0.05 p value level.

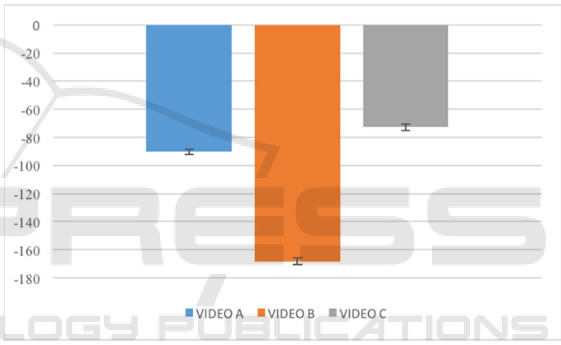

Figure 9: The mean and standard error of the SCL data of each video.

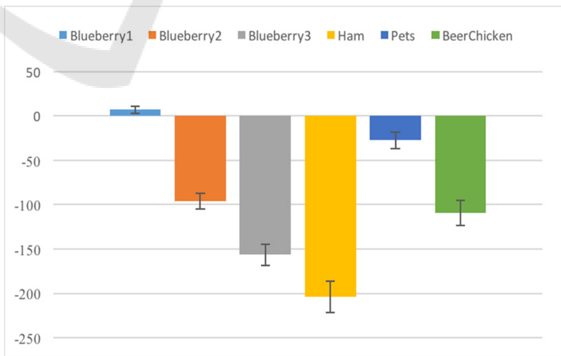

Figure 10: The mean and standard error of the SCL data of each scene.

The SCL of the participants in the three videos are significantly different (F (2, 14) = 393.36, p < .05). By using the LSD method to do the pairwise comparison, the participants' SCL data is found to be significantly different from each other. It shows that the SCL is the highest in the video C, while it is the lowest in the video B. It is also spotted that the viewers' SCL showed significant differences among the six scenes where the close-up scenes occurred (three from the blueberry video, and only one close-up from the other two videos respectively): F (5, 11) = 53.62, p < .05). The LSD method shows that the SCL data is significantly different from each other, except for the second close-up of the blueberry and the one of the beer chicken. In particular, the SCL of the participants during the first close-up of the blueberry is significant higher than the other scenes.

4.4 Questionnaires (R3)

There is no significant difference among the evaluation of the videos: F (2, 14) = 5.20, p = .11 based on the repeated measuring ANOVA test of the questionnaire.

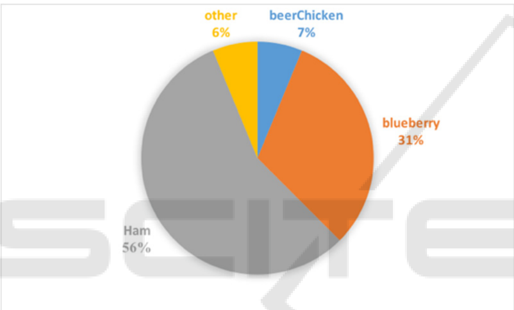

Figure 11: The disliking ratio reported from the questionnaires.

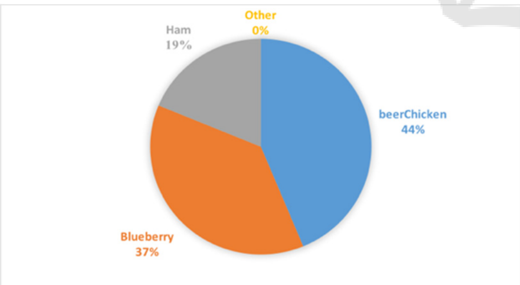

Figure 12: The liking ratio reported from the questionnaires.

The ratio on the likes and dislikes videos shows in Figure 11 and 12. More than a half of the participants disliked the video B because of the dark image and scary background music. For the most favourite video, 44% of the participants chose the video C, 37% of the participants chose the video A, and only 19% of the participants chose the video B.

The results of the questionnaire are similar to the results of the SCR and SCL results. For example, both the SCR and SCL results indicate that the video C is the most engaging one while the video B is the least engaging one. To sum up, the results obtained from the sensors are consistent with the ones obtained from the questionnaires.

5 DISCUSSION

In this paper, we conducted an experiment by using GSR sensors, where user bio responses towards the three food TVCs were analyzed. Through a comparison with the questionnaires, it has been confirmed that the sensor measurement is consistent with their subjective reports.

The study has answered the research questions clearly. First, according to both the GSR results and the questionnaires, the participants preferred the video C (Beer Chicken) the most, and showed the least preference to the video B (Ham). Second, some specific scenes, like the close-up of the food, could induce the participants' emotional arousal. Third, the results of the GSR data were grounded by the questionnaires.

There are some interesting findings discovered from the results of the study. First, the attention in all videos is decreasing from the start to the end of each video according to the z-score figure, which means that the participants were losing their interest during all the videos, especially after 90 seconds. It suggests that the length of a TVC is crucial. This information is extremely helpful for producers, especially when it comes to the effectiveness of a TVC. Second, some interesting phenomena were observed among the female consumers. One of them is that they prefer some specific moments in a video, e.g., the close-up of the food, the good-looking figure of the actors, and the appearance pets. These reasons may explain why the video C is the most popular one. In addition, it seems that women are sensitive to the bright scene and the joyful music background, which were reported from the post-interviews.

Nevertheless, there are some differences found between the SCL and the SCR. According to the results of the SCL, the arousal level of the video C is the highest, while the video B is the lowest, which is consistent with the results of the questionnaire. However, according to the results of the SCR, there is no significant arousal moments during the video A, which means the video A did not significantly arouse the viewers. We assume that the two patterns of the EDA may reflect the different user experiences, and we need further investigation for such phenomenon.

This study proves again that our GSR sensor system is robust for such studies. In particular, the monitoring system has helped us learn the actual status of sensors. Even if the sensor gets broken, we can easily replace it with a new one. However, the sensor prototype is a bit of bulky for users to wear. A half size of the current one will be an optimal shape for an unobtrusive watching experience.

In this study, we did not categorize which specific emotions are induced in viewers, considering the fact that we only have one type of sensors in our case. Therefore, more sensors are required if user emotional states need to be further classified. In the future work, we can add other types of sensor (e.g., ECG sensors) to obtain more sensor data, which may help us better define emotions elicited during video consumption.

6 CONCLUSION

In this paper, we have reported the experiment on female bio-response towards the three types of food TVCs. The results have exhibited how those videos could affect female reactions and their watching experiences. Our study presents that physiological data does have superior advantages on measuring user experiences compared to subjective reports. By following our method, researchers can design an experiment with their own research purposes. Furthermore, other similar studies, e.g., new media design, can be also benefited from our learning experience.

Besides, our work also demonstrates that the combination of the hardware and software solution can be rather helpful for commercial companies. By using our method, they can pre-assess the effects of TVCs, especially among targeted consumers who are particularly interested in investigation. In such a manner, it can reduce the risk before the launch of products, and pre-sampling test method can help them to adjust the marketing strategy.

In addition, it has been fully demonstrated that our GSR sensor system is robust and can simultaneously and accurately capture the GSR signals from users. The system allows us to quantify user experience, and at the same time keeps the confidentiality with user. Currently we are working on the process of scaling up the system and attempting with the integration of different sensors (e.g., ECG sensor and acceleration sensors). The other types of sensor can be integrated into our sensor network to provide a more complete representation of user experience.

REFERENCES

DualShock. (2017). [online] En.wikipedia.org. Available at: http://en.wikipedia.org/wiki/DualShock. Factors in Computing Systems (CHI '13). ACM, New York, NY, USA, 1499-1508.

Smith, J., 1998. *The book*, The publishing company. London, 2nd edition.

Aigner, W., Tomitsch, M., Stroe, M. & Rzepa, R. Be a judge!: wearable wireless motion sensors for audience participation. CHI '04 Extended Abstracts on Human Factors in Computing Systems, 2004. 1617-1621.

Barkhuus, L. Engaging the crowd:studies of audience-performer interaction. Extended Abstracts Proceedings of the 2008 Conference on Human Factors in Computing Systems, CHI 2008, Florence, Italy, April, 2008. 2925-2930.

Chanel, G., Rebetez, C. & Pun, T. Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. International Conference on Entertainment and Media in the Ubiquitous Era, Mindtrek 2008, Tampere, Finland, October, 2008. 13-17.

Fischer, J. E. & Benford, S. Inferring player engagement in a pervasive experience. Sigchi Conference on Human Factors in Computing Systems, 2009. 1903-1906.

Fleureau, Julien, Philippe Guillotel, and Izabela Orlac. 2013. "Affective benchmarking of movies based on the physiological responses of a real audience." Affective Computing and Intelligent Interaction (ACII), Humaine Association Conference on. IEEE.

Hart, J., Sutcliffe, A. & De Angeli, A. Using affect to evaluate user engagement. CHI '12 Extended Abstracts on Human Factors in Computing Systems, 2012. 1811-1834.

Joho, H., Jose, J. M., Valenti, R. & Sebe, N. Exploiting facial expressions for affective video summarisation. ACM International Conference on Image and Video Retrieval, Civr 2009, Santorini Island, Greece, July, 2009. 1-8.

Kapoor, A., Burleson, W. & Picard, R. W. 2007. Automatic prediction of frustration. *International Journal of Human-Computer Studies,* 65**,** 724-736.

Kenning, P. H. & Plassmann, H. 2008. How neuroscience can inform consumer research. *IEEE Transactions on Neural Systems & Rehabilitation Engineering A Publication of the IEEE Engineering in Medicine & Biology Society,* 16**,** 532.

Krugman, D. M., Fox, R. J., Fletcher, J. E., Fischer, P. M. & Al, E. 1994. Do adolescents attend to warnings in cigarette advertising? An eye-tracking approach. *Journal of Advertising Research,* 34**,** 39-52.

Latulipe, C., Carroll, E. A. & Lottridge, D. Love, hate, arousal and engagement:exploring audience responses to performing arts. International Conference on Human Factors in Computing Systems, CHI 2011, Vancouver, Bc, Canada, May, 2011. 1845-1854.

Lin, T., Maejima, A. & Morishima, S. Using subjective and physiological measures to evaluate audience-participating movie experience. Working Conference

on Advanced Visual Interfaces, AVI 2008, Napoli, Italy, May, 2008. 49-56.
Lunn, Darren, Harper & Simon 2010. Using galvanic skin response measures to identify areas of frustration for older web 2.0 users.
Madan, A., Caneel, R. & Pentland, A. S. GroupMedia: distributed multi-modal interfaces. International Conference on Multimodal Interfaces, 2004. 309-316.
Mirza-Babaei, P., Nacke, L. E., Gregory, J., Collins, N. & Fitzpatrick, G. How does it play better?:exploring user testing and biometric storyboards in games user research. Sigchi Conference on Human Factors in Computing Systems, 2013. 1499-1508.
O'Brien, H. L. & Mclean, K. 2009. Measuring the User Engagement Process.
Oliveira, E., Benovoy, M., Ribeiro, N. & Chambe, T. Towards emotional interaction: using movies to automatically learn users' emotional states. Ifip Tc 13 International Conference on Human-Computer Interaction, 2011. 152-161.
Pan, M. K. X. J., Chang, J. S., Himmetoglu, G. H., Moon, A., Hazelton, T. W., Maclean, K. E. & Croft, E. A. Galvanic skin response-derived bookmarking of an audio stream. International Conference on Human Factors in Computing Systems, CHI 2011, Extended Abstracts Volume, Vancouver, Bc, Canada, May, 2011. 1135-1140.
Peters, C., Castellano, G. & Freitas, S. D. 2009. *An exploration of user engagement in HCI*, ACM.
Picard, R. W. & Daily, S. B. 2005. Evaluating affective interactions: Alternatives to asking what users feel. *Chi Workshop on Evaluating Affective Interfaces Innovative Approaches*.
Sakurazawa, S., Yoshida, N., Munekata, N., Omi, A., Takeshima, H., Koto, H., Gentsu, K., Kimura, K., Kawamura, K. & Miyamoto, M. A computer game using galvanic skin response. International Conference on Entertainment Computing, 2003. 1-3.
Sas, C., Fratczak, T., Rees, M., Gellersen, H., Kalnikaite, V. & Coman, A. Affectcam: Arousal-augmented sensecam for richer recall of episodic memories. CHI '13 Extended Abstracts on Human Factors in Computing Systems, 2013. 1041-1046.
Wang, H., Prendinger, H. & Igarashi, T. Communicating emotions in online chat using physiological sensors and animated text. CHI '04 Extended Abstracts on Human Factors in Computing Systems, 2004. 1171-1174.
Yoon, K., Bolls, P. & Lang, A. 1998. The Effects of arousal on liking and believability of commercials. *Journal of Marketing Communications,* 4**,** 101-114.