Measuring Audience Responses of Video Advertisements using Physiological Sensors

Chen Wang
Centrum Wiskunde & Informatica
Science Park 123
1098XG, Amsterdam
cw@cwi.nl

Pablo Cesar
Centrum Wiskunde & Informatica
Science Park 123
1098XG, Amsterdam
p.s.cesar@cwi.nl

ABSTRACT
The selection of the audio track, the best timing to overlay the logo, and the overall duration, all these issues affect the effectiveness of immersive media. Since traditional methods to evaluate the user experience of potential consumers (e.g., surveys or eye tracking) have severe limitations, we used data gathered from physiological sensor to measure the viewers' watching experiences. In this paper we report how we used our own Galvanic Skin Response (GSR) sensors to measure audience experience for the two different audio tracks of a commercial. Our results show that our GSR technology can play an important role for the advertisement community. In contrast with surveys, using GSR data relevant results can be obtained even with small number of participants, and the viewers' experiences are more vividly visualized. This enables advertisers to, for example, be able to decide the proper length of a commercial.

Categories and Subject Descriptors
H.5.m. Information interfaces and presentation (e.g., HCI):

General Terms
Human Factors; Design; Measurement

Keywords
Physiological computing; GSR sensors; audience experience; advertising videos

1. INTRODUCTION
Audio plays an essential role on the popularity and impact of products. The advertisement industry is fully aware that an appealing audio is as important as appealing visuals, and aims at producing engaging commercials that make costumers involved with their products. However, measuring the impact of the background audio of a commercial is a challenge. First, one should take into account the relationship between the audio track and the video content. Second, the available evaluation methods are constrained. For instance, traditional methods (e.g., surveys) fail to provide helpful timed information about the user experience. Other mechanisms, like eye tracking and facial expression have been used to observe the audience interests on video commercials [6]. But, eye tracking and facial expression data do not seem to be particularly useful for evaluating the effect of audio, since it is still unclear the relationship between auditory and visual attention.

Physiological sensors have been applied on audience research, e.g., user emotion [9] and user engagement [12]. Surprisingly, few studies used physiological sensors to measure audience experience during a commercial, and most of studies were related to the assessment of video content [3], or to investigate audience buying behavior [2]. None of them has used bio sensors to measure the impact of the audio track of a commercial.

In this paper, we used our own GSR sensors as an alternative tool to measure audience experience towards the audio tracks of commercial video. The reasons why we chose GSR sensors are because they are highly accurate for indicating the user internal state [4] compared to other bio sensors (e.g., respiration sensors). Besides, we could use the sensor data to monitor viewers' watching experience during the whole video. In such manner, we can analyze different perspectives for better understanding the audience experience. Moreover, based on the sensor results, we can discuss how can we help video designers reflect the proper length of videos, and the best timing to place the logo in commercials. In this paper, we are interested in the following research questions:

Q1: How can we use GSR data to visualize the viewers' experiences of the audio track of a commercial?

Q2: How can we use GSR data to help advertising designers to reflect about the length of videos and define the best timing for placing the logo of a commercial.

2. STATE OF THE ART
Audio impact of a commercial has invoked several research interests. For instance, Mandler [7]conducted surveys to conclude how different music types relate to personality. Neuroscientists found that commercials with an audio logo are more effective in the activating the areas of the brain that influence buying behavior [2]. Besides, psychologist Adrian North investigated the effects of playing either French music or German music in a supermarket: when French music played, the store sold five times the usual amount of French wine; when the soundtrack was German, twice as much German wine was bought [8].

The limitations of subjective methodologies sometime make physiological measures more attractive for empirical experiments. For instance, subjects may not remember how they actually felt,

ImmersiveME'15, October 30, 2015, Brisbane, Australia.

ACM 978-1-4503-3745-8/15/10…$15.00.
DOI: http://dx.doi.org/10.1145/2814347.2814352

and may not be able to articulate how they feel, or if they can articulate their feelings they often describe them in a non-extreme manner, therefore making statistical analysis difficult [10]. Therefore, physiological measures are extremely valued for being unobtrusive and therefore able to continually monitor the experience of the users without distracting him/her from their primary task.

Nevertheless, physiological measures in some cases are still dependent on subjective measurements. For instance, in order to understand the actual values from the sensors, we need some subjective user response (from interviews, questionnaires or from recordings of the session). The methodology is as follows: we first need to discover the patterns (e.g., GSR response to subjectively defined stimuli). Second, we use subjective reports to adequately classify their experience (or emotion) as positive or negative [13][5] .

3. METHODOLOGY

3.1 Apparatus

We built 15 GSR sensor nodes (sample rate: 50Hz) by using the open source Jeenode board (a clone of Arduino with a RF12 wireless module integrated) (Figure 1). The GSR sensors were built by using an operational trans conductance amplifier (OTA) and a low-pass filter (LPF), where an OTA circuit followed by a 2^{nd} order low pass Butterworth filter and the bandwidth of an LPF was cut at 0 -5Hz. After successful tested the GSR circuit in the lab experiments (e.g., watching videos), we asked the factory to produce the printed circuit boards (PCB) based on the interface requirements of the Jeenode boards (Figure 1: the right). In such manners, we can easily integrate the whole GSR system into a 3D printed box, which can be worn around the neck of the user. During the experiment, all the nodes sent the packets at different time slots to a sink node, which was connected to a laptop (Figure 1: the left).

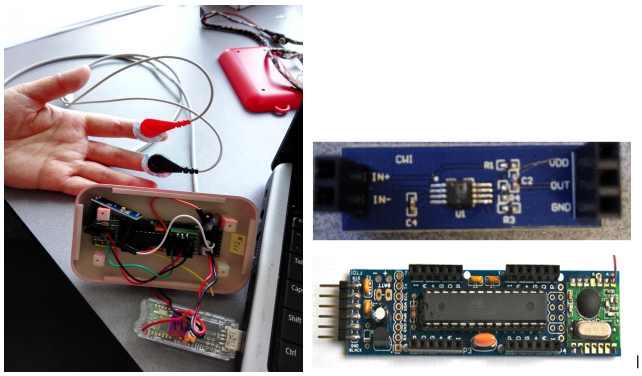

Figure 1: The GSR sensors (the left); The produced GSR sensors and the Jeenode board (the right)

3.2 Experimental Design

The commercial from Starbucks, called "What Do You Want 5 Minutes Of", was selected for the experiments (87 seconds). We replaced the original music (no narrative) by two different types of audio background: the up-tempo music (audio B) versus the ballad music (audio C). We presented the same video under three conditions (muted audio, audio B and audio C) to the same group users. The purpose of this design was to prevent the content of the commercial to affect the results. We first presented the video with muted music to all users, so that we could compare the results for the two different scenarios. After each video was played, there was a half hour break, so that we minimized the ordering influence on users (the video C was after the video B). There were pre (one time) and post questionnaires (two times) provided before and after each video.

3.3 Questionnaires

A pre-questionnaire and two post-questionnaires were provided before and after each video. Questions in the pre-questionnaire were mainly about the type and intensity of the emotions they had experienced during the day, and how much video design experience they had. The majority of the questions in the post-questionnaires dealt with their engagement, such as enjoyment, likeness, and the motivation to purchase the product. The questions were in the form of "Graphic Rating Scales". The line measured 100 mm and responses were measured to 1mm accurate.

3.4 Participants

We invited participants from one Chinese University: 8 females (Mean age = 21, SD = 2.08) and 7 males (Mean age = 21.17, SD = 1.47) attended the experiment. Before the experiment, all of them signed a consent form, and received a small gift after the experiment. None of them have seen the tested video before.

3.5 Methods

In the sensor data analysis, we used Analysis of Variance (ANOVA) to test whether there was a significant statistical difference on the SCL (skin conductance level) of the audience GSR data between the two test videos (B and C). All the assumptions related to ANOVA test were checked during the analysis of the data.

In the analysis of SCL, we normalized the GSR arousal level (see Figure 5) by using the first GSR reading as the base line [11]. This way we can calculate the level of skin arousal induced by watching the videos.

Figure 2 describes the different steps of the algorithm to compute the phasic changes of the SCR (skin conductance response), which is adapted from the paper [1]. The raw GSR signal was first processed by a 2Hz low-pass filter in order to eliminate misleading information and noise. After that, the GSR signal was derivated in terms of capture the phasic changes ($G'(t)$). However, the negative phasic changes were not our interests, so that the derivative signals was truncated into positive values (Thresholding

output: $G_+^{'}(t)$) in order to highlight the relevant phasic changes. In the next step, we apply a moving window to compute the mean arousal of the audience at a given window size (window size (W): 3 seconds) with an overlapping (overlapping window size: 2 seconds). Finally, in order to remove the user-dependent part related to the amplitude of the GSR derivative, as it may be varied from one subject to another, we normalized the GSR data by using the sum of subsampled skin response values as a denominator (i, j = 1……k, where k is the number of windows) to calculate the normalized individual mean arousal value (1) (n =1……N, N: the number of users).

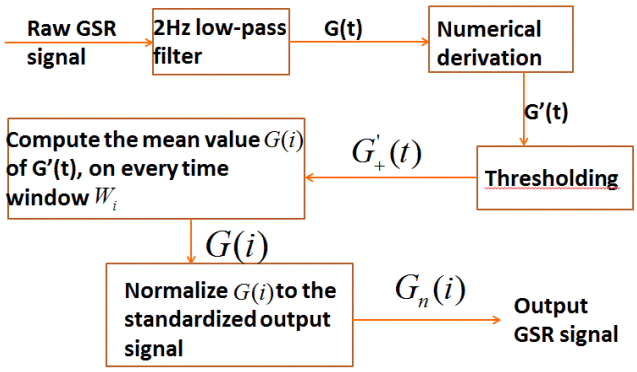

Figure 2: Description of the different steps of the algorithm to process the phasic changes of EDA signals.

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

In each window k, the mean p value is computed by averaging the p-value of the bilateral Mann-Whitney-Wilcoxon test performed between the latent unknown distribution of $G_n^i(k)$ (i = 1…..N) and the background noise. In our case, the 10% of computation results with the lowest mean are considered as the background noise [10]. Only an associated value lower than 5% is considered as significantly different from the background noise.

4. RESULTS

4.1 Survey Reports

All the participants rated high their enjoyment and likeness for both videos (B&C). Moreover, users reported similar attention values during watching, but they found difficult to specify which timing they had the highest attention level, although they understood both videos very well. Based on such results, we obtained general opinions from the viewers about their experience towards the videos: it seems they had a rather fair experience towards the two videos. But if we are unable to understand the user experience during the videos, it is unlikely to help video designers to make an improvement, e.g., Was the length of the videos right designed? Did the logo appear at the right moment? In addition, we could not find any correlations existed between the viewers' reports and the sensor data. This case of non-correlated results has happened in the past [14], and remains a topic for further discussion. How to best process data and how conscious and unconscious data correlate still needs to be investigated.

4.2 GSR Sensors

The algorithm performed on the SCR data showed some similarities for the two videos (Figure 3&4), which were consistent to the self-reports (e.g., enjoyment and likeness). The viewers were emotionally stimulated in the beginning, in the middle and towards the end. We labeled the computed significant moments by the red lines in Figure 3. However, In video B, significantly different skin responses appeared in seconds 45, 47, and 79. While in video C these happened in seconds 33, 47, and 75. Nevertheless, the appearance of the logo by the end of the commercial (second 84) did not induce any significantly different skin response. The associated p values related to the significant moments were visualized in Figure 4, and all were plotted with transferred logarithm values. The horizontal red line represents 0.05 significant level, thus the values under the red line can be considered as significant, which can be matched to the significant moments in Figure 3.

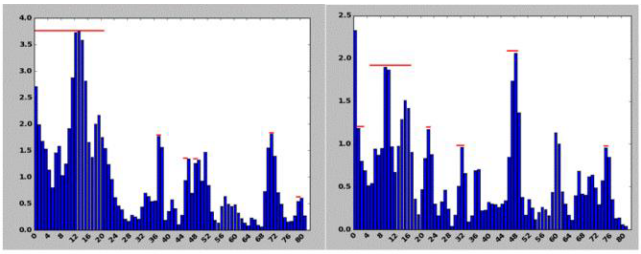

Figure 3: The computed SCR output value: only the moments labeled by the red lines are considered as significantly different SCR value (left: video B; right: video C). The x-axis is the time line of the video, and the y-axis is the normalized output mean GSR value.

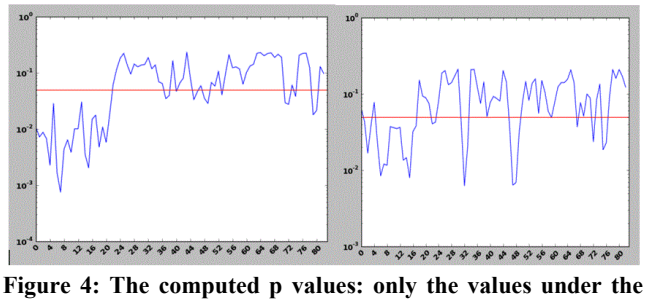

Figure 4: The computed p values: only the values under the red line are considered significant (<0.05) (left: video B; right: video C). The x-axis is time line of the video, and the y-axis is the transferred logarithm values of p.

The normalized algorithm on the SCL of GSR data was also coherent with the results of SCR. There is a statistical correlation between the GSR data distribution in the two videos: r =0.72 , p <0.01 , which means that there are some similarities regarding the GSR distribution, as we see in Figure 5: when users were aroused and when users' arousal

warned. In the first 25 seconds, the arousal of the users was higher than the rest of the videos. After that, the arousal of the users experienced a gradual decrease till the end. In addition, the arousal level invoked when watching video B and when watching video C was different. The viewers had a positive arousal value (normalized value: 24.8) when watching video B, while they had a negative arousal value (normalized value: -52.8) when watching video C. Thus, we can see that the different types of audio background invoked users in different manners: up-tempo audio track activated positive user arousal, while ballad audio decreased the arousal of the users. The ANOVA results show a significant difference on arousal levels between these two videos: $F(1, 86) = 364$, $p < 0.01$, $\eta_p^2 = 0.81$.

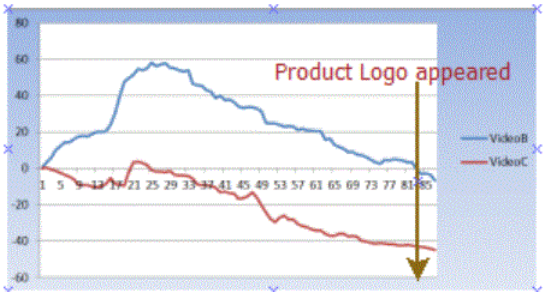

Figure 5: The GSR Data Distribution for the two videos.

Based on our experiments, we found that GSR sensors reported similar user experience as that found with the surveys. But, GSR sensors can better visualize such experiences. Thanks to the visualized results, it seems that the length of the video is rather long, as we already see the viewers' arousal warned at the end of the videos. If a commercial video is broadcasted into public, the viewers may have already changed the channel without waiting till the end. If that is the case, they will not even see the logo appearing at the end of the video. However, In order to make a concrete conclusion, more dedicated experiments are required, e.g., testing different timings for the logo and different durations.

5. CONCLUSION

We conducted a user experiment in which we used GSR sensors to measure the audience experiences for different audio tracks of commercials. Our studies showed that GSR sensors can report user experience than by using surveys. Nevertheless, with the visualized sensor data, video designers can reflect the decisions regarding the length of videos and the best timing of placing a logo. This information can be difficult obtained with surveys.

6. REFERENCES

[1] Julien Fleureau, Philippe Guillotel, and Izabela Orlac. 2013. Affective Benchmarking of Movies Based on the Physiological Responses of a Real Audience. In *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (ACII '13). IEEE Computer Society,Washington, DC, USA, 73-78. DOI=10.1109/ACII.2013.19

[2] Kenning PH, Plassmann H.2008. How Neuroscience Can Inform Consumer Research. *IEEE Transactions on Neural Systems and Rehabilitation Engineering. Dec*.

[3] Yoon, Kak, Paul Bolls, and Annie Lang. 1998."The effects of arousal on liking and believability of commercials." *Journal of Marketing Communications 4.2 : 101-11 Tavel, P. 2007. Modeling and Simulation Design*. AK Peters Ltd., Natick, MA.

[4] Krause, Andreas, Asim Smailagic, and Daniel P. Siewiorek. 2006."Context-aware mobile computing: Learning context-dependent personal preferences from a wearable sensor array." *Mobile Computing, IEEE Transactions on 5.2*: 113-127.

[5] P. Lang. 1995. The emotion probe: Studies of motivation and attention. *American Psychologist* . 50(5): 372–385.

[6] Krugman, Dean M., et al. 1994. "Do adolescents attend to warnings in cigarette advertising? An eye-tracking approach." *Journal of Advertising Research 34* : 39-39.

[7] George Mandler. 1987. Mind and body: psychology of stress and emotion. Motivation and emotion. *New York, Norton*.

[8] Adrian C. North. 1997, *Wine and song: the effect of background music on the taste of wine.* www.wineanorak.com/musicandwine.pdf

[9] Eva Oliveira, Mitchel Benovol, Nuno Rebeiro, and Teresa Chambel. 2011.Towards Emotional Interaction: Using Movies to Automatically Learn Users' Emotional States. *13th IFIP TC 13 International Conference,* Lisbon, Portugal, September 5-9, Proceedings, Part I.

[10] Picard, R., Daily, S.B., 2005 Evaluating affective interactions: Alternatives to asking what users feel. *CHI 2005 Workshop on Innovative Approach to Evaluating Affective Systems.*

[11] Boucsein, W. 2012. *Electrodermal activity* (2nd Ed). New York: Springer.

[12] Chen Wang, Erik N. Geelhoed, Phil P. Stenton, and Pablo Cesar. 2014. Sensing a live audience. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (CHI '14). ACM, New York, NY, USA, 1909-1912. DOI=10.1145/2556288.2557154

[13] C. Wang, Pablo Cesar. 2014. Do we react in the same manner?: comparing GSR patterns across scenarios. *Proceedings of Nordic Conference on Human Computer Interaction: Fun, Fast, Foundational* 2014 (NordiCHI 8), 501–510.

[14] C. Wang, Pablo Cesar.2015. Physiological Measurement on Students' Engagement In a Distributed Learning Environment. *Proceedings of International Conference on Physiological Computing System 2015* (PhyCS 2015), Angers, France, 2015