How to Catch the Consumers' Heart: emotional effect of advertisements on buying behaviour

Xintong Zhu

Student Number: 2531549

Master thesis Cognitive Neuropsychology

August, 2015

VU University Amsterdam

Supervisors:

Dr. Pablo Cesar

Centrum Wiskunde & Informatica

Dr. Sander Los

VU University Amsterdam

CONTENTS

Abstract

Commercials are a type of marketing communication, of which mechanism theories have been widely studied. The studied mechanism theories mainly included two viewpoints: cognitive and affective. One of the affective effect viewpoints is the evaluative conditioning theory. According to this theory, the emotions invoked by advertisements will also present to the neutral logo of prosducts combined with advertisements, which influence the consumers' shopping behavior. The current study explored if the evaluative conditioning can provide new insights about the effectiveness of commercials, by conducting one experiment. The experiment included two phases: a conditioning acquiring phase in which emotional videos were matched to certain neutral logos, and a testing phase which consisted of a virtual online shopping experience. Both physiological data and self-report data of arousal and valence were recorded. The emotional arousal and valence's effect on the participants' choice in virtual online shopping were analyzed using a binary logistic model. The valence effect was not found neither with the physiological and the self-report data. The arousal effect was found based on the self-report data. Based on the results, it is suggested that evaluative conditioning does no influence people's shopping behavior, while the arousal level may affect people's shopping behavior.

Keywords: evaluative conditioning; galvanic skin response, facial expression, commercials, shopping behavior

Acknowledgement

First, I would like to express my sincere thanks to my supervisors: Dr. Pablo César and Dr. Sander Los, who have given their constant help: offered me helpful suggestions to the experimental design, labs' booking, data analysis, academic writing, and thesis processing schedule.

My thanks to Chen Wang, who really helped me a lot on Galvanic Skin Response sensors, running experiment and analyzing the data of GSR sensors. My thanks to Thomas Röggla, who developed the virtual online shopping website of the thesis experiment and the algorithm coding of GSR sensor data analysis, and helped me running the experiment. Thanks to Rosmi Rehman, who helped me to run the experiment. Thanks to Dr. Daniel Schreij, who helped me to solve the programming problems of OpenSesame. Thanks to Dr. Marie-Collete van Lieshout, who gave me the useful suggestions on data analysis.

My thanks also go to all the PhD students and staff in Distributed and Interactive Systems at Centrum Wiskunde & Informatica, who gave the useful suggestions on my thesis proposal.

Finally, I would like to thank to my family and friends, who support me a lot for my studying and doing research abroad.

Introduction

How often have you seen commercials on TV, the internet or while reading magazines? Have you ever bought products because of the commercials? Will you buy them again? If we try to recall and answer these questions, we will find that nowadays commercials are present everywhere. However, only effective commercials help companies to achieve their purpose (e.g., increasing sales). One core question remains: what kinds of commercials are effective? To explore the answer, it is important to understand how commercials affect the decision of potential customers. For this reason, the current study mainly focused on how commercials influence shopping behavior.

Commercials are a type of marketing communication, which is used to persuade consumers to buy products or services, or to take some other actions. It is an intrinsic form of communication in our society. Given the long history of the commercials, there have been many previous studies focused on them, including commercials' mechanism, which is the topic of the current study. Taking a panoramic view of these studies, previous researchers mainly had two viewpoints: cognition and emotional. One was that commercials influence consumers' cognition of the products (e.g. Strong, 1925; Vakratsas & Ambler, 1999), and the other one was that commercials have some emotional effect on customers (e.g. Zajonc, 2001; Stuart, Shimp, & Engle, 2002).

In the current study, one of the emotional effect viewpoints, called evaluative conditioning theory, was tested by experimental methodology. According to the evaluative conditioning theory, when advertisements combined with a neutral logo or a brand of products (conditioning process), the invoked emotions of advertisements will shift to the logo or brand, which influence the consumers' shopping behavior. To

explore if the evaluative conditioning can explain the effectiveness of commercials, in the current study, we performed one experiment including both a conditioning process that matched emotional videos and certain neutral logos, and a testing process, which involved a virtual online shopping experience.

Relevant Theories and Research Question

As early as 1898, E. St. Elmo Lewis proposed an advertising model called AIDA (Strong, 1925), which explained how advertising works on consumers' shopping behavior. This model defines four steps of engagement: attention, interest, desire and action. The model is considered as the first formal model for advertisements (Li & Leckenby, 2004). After E. St. Elmo Lewis's work, some researchers further developed AIDA by adding new steps, like satisfaction and confidence (Sheldon & McDowell,1923; Kitson, 1920). In addition, AIDA also edified a series of new models and theories, which are commonly called "hierarchy models of advertising". For example, Lavidage and Steiner in 1961 categorized different elements affecting the effectiveness of advertising into three categories: cognition, affection, and conation. Some researchers also had similar views and they suggested that the three categories are cognition, affection, and behavior (Palda, 1966; Barry, 1987). Besides the hierarchy models, some researchers only focused on one aspect of advertising. For instance, Naples (1979) believed that frequency of exposure influences the advertisements' effectiveness, while Brown and Stayman (1992) suggested that the attitude toward advertisements is the factor that affects their effectiveness.

To summarize, the cognitive and affective aspects of consumers were universally considered in the previous studies. For example, Vakratsas and Ambler (1999) summarized four crucial directions for understanding how advertising works: advertising input (e.g. message content and media scheduling), filters (e.g. motivation), consumer (e.g. cognition and affect) and consumer behavior (e.g.

choice).

Cognition and commercials In the early studies on the effect of commercials, some researchers believed that consumers' decisions are rational (Vakratsas & Ambler, 1999; Petty, Cacioppo, & Schumann, 1983). This means that commercials only can help consumers to understand the advantages of the products and persuade them to buy the products as an information process (Vakratsas & Ambler, 1999). To further explain how cognition influences the shopping behavior, there is a model: the elaboration likelihood model (ELM). This general model, based on persuasion, was applied in advertising. Petty and Cacioppo (1986) explained that there are two routes when people process information: central route and peripheral route. The central route is used when people have enough ability or motivation, and build their attitude by thoroughly evaluating the product. The peripheral route, however, happens when people have less interest or ability to process the information and the attitude will be affected by superficial characteristics of the advertisements. For example, Petty, Cacioppo, and Schumann (1983) found that when elaboration likelihood was low (i.e., that the participants had less interest), they had a more positive attitude towards advertisements with famous athletes which are the superficial characters. Whereas when elaboration likelihood was high, only the strength of the arguments within the advertisement can change the participants' attitude. Unfortunately, other researchers found that persuasion does not always work. For instance, Ohme, Reykowska, Wiener, and Choromanska (2010) indicated that some advertisements (e.g. Sony Bravia commercials), which only use videos with background music and without verbal introduction of advantages, were also effective. To explain this phenomenon, other factors should be considered.

Memory, which is another cognitive factor, was the other major focus (Zielske, 1959; Stewart, Pechmann, Ratneshwar, Stroud, & Bryant, 1985). According to these researchers' view, recall of advertising is a criterion of advertising effectiveness. In other words, only when consumers can recall the brands in the advertisements, they

are effective. However, the recall theory was challenged by some researchers (Haskins, 1964; Gibson, 1983). There were three main questioned views. First, Haskins (1964) concluded that recall of advertisements is not related to attitudes towards products. Additionally, Du Plessis (2005) pointed out that recall does not work until the consumers make a purchase decision, which means that it is hard to distinguish whether there was a causal relationship between memory of the advertisement and consumers' shopping behavior. Thirdly, recall theory cannot be applied to real life situation. Based on Zielske's investigation in 1959, advertisements will be quickly forgotten unless consumers were continually exposed to the advertisement, which is not possible in real life. Besides, high exposing frequency to one advertisement means more familiarity to the advertisement or the product. According to previous studies, familiarity causes positive affection (Zajonc, 2001; de Vries, Holland, Chenier, Starr, & Winkielman, 2010). So, we cannot simply conclude that the recall theory or other cognitive theory on their own can explain why advertisements are effective.

Affection and Commercials As we mentioned above, cognitive factors cannot on their own explain advertising effectiveness. There were plenty of studies which focused on the role of affection in commercials. For example, Zajonc (1980) argued that affection is the predominant evaluative system because it is faster and cruder than the cognitive process. In the 1980s, some researchers suggested that the emotion evoked by advertisements has a positive effect on consumers' attitudes to brands (Kroeber-Riel, 1979; Edell & Burke, 1987; Holbrook & Batra, 1987). For example, Holbrook and Batra (1987) asked the participants in lab studies to watch some selected commercials and then filled a self-report about affection. They found that feelings are important to predict the effectiveness of the advertisements and also contributed to the attitudes to brands. In order to explain the affective effect, one psychological definition was applied to the advertising field: evaluative conditioning.

Evaluative Conditioning (EC) De Houwer, Thomas and Baeyens (2001) pointed

that "the associative transfer of valence is commonly referred to as *evaluative conditioning* (EC) ". Evaluative conditioning is a kind of Pavlovian conditioning. In Pavlovian conditioning, the conditioned stimulus (CS) is a neutral stimulus (e.g., the sound of the ring), and the unconditioned stimulus (US) is a biological stimulus (e.g., taste of food). The US induces some unconditioned responses (UR) (e.g., salivation). After matching the CS and the US and repeating several times, the similar response to the UR called a conditioned response (CR) to the CS will happen when the CS appeared alone (e.g. salivate to the sound of the ring). Similarly, in the evaluative conditioning, the neutral stimulus (e.g. brand) is equivalent to the CS, the affective stimulus (e.g. advertisement) is equivalent to the US, the consumers' attitude to the affective stimulus is equivalent to the UR, and the consumers' attitude to the neutral stimulus is equivalent to the CR. For example, if the brand (CS) is always paired with an emotional advertising video (US), consumers' attitudes to the brand will change according to the evoked emotion of the advertising video.

As one of the first demonstrations of EC, Staats and his colleague (1957) matched nonsense words and emotional words (either positive words or negative ones), and they found that nonsense words acquired the same emotional valence of the paired emotional words. In subsequent studies, researchers also found the same phenomenon by using different kinds of stimuli (Levey & Martin, 1975; Baeyens, Eelen, & Bergh, 1990; Hammerl & Grabitz, 2000; Johnsrude, Owen, White, Zhao, & Bohbot, 2000). For example, Levey and Martin (1975) asked participants to first classify different pictures into different emotions (positive, negative, and neutral). Then they matched the most liked or most disliked pictures to neutral pictures. By using this kind of 'picture-picture' paradigm, they indicated that the participants' feelings about the emotional pictures shifted to the neutral pictures. Furthermore, many other stimuli have been considered: gustatory stimuli (e.g. artificial fruit flavours matched to sugar or harmful soap-like taste), haptic stimuli (e.g. touch of silk or sandpaper), and biological stimuli (e.g. neutral monochrome patterns combine with food reward or no food reward) (Baeyens et al., 1990; Hammerl & Grabitz, 2000; Johnsrude et al.,

2000).

In addition to those mono-stimuli experiments, researchers also demonstrated that EC occurred when US and CS were different kinds of stimuli, such as cross visual-auditory (Reekum, Marije, van den Berg, & Frijda, 1999; Zucco, 2012). For example, Reekum and his colleagues (1999) paired neutral visual stimuli of abstract paintings and emotional sounds or odors, and confirmed that the participants' emotions can shift from US to CS even if they are cross-domain. Eifert, Craill, Carey, and O'Connor (1988) established evaluative conditioning between neutral Greek letter (CS) and emotional music (US). Besides, researchers also found that the conditioned stimulus can be paired with more than one unconditioned stimuli which had the same valence (Stahl & Unkelbach, 2009). According to this experimental evidence of cross-domain EC, researchers found that EC can be applied to different fields, like consumer research, mental disorders, education, social psychology and so on (Loebnitz & Grunert, 2014;Stuart, Shimp, & Engle, 2002 ; Jones & Fazio, 2012).

EC and commercials In 1987, Stuart, Shimp, and Engle conducted several experiments of conditioning and advertising. They provided the experimental evidence of conditioning effect on advertising. In their study, emotional pictures were used as US while Brand L toothpaste was used as CS. The participants first viewed the Brand L Toothpaste for five seconds and then another five seconds of emotional pictures, which followed by another five seconds combination of the brand and emotional picture. Finally, there were two seconds called "down time". Four groups of participants were exposed to different numbers of trials (1, 3, 10, and 20 trials). The researchers indicated that all the four groups of participants felt more positive for the brand than the control group. Besides, the differences between the control group and the experimental group increased by the numbers of trials, which means that the strength of conditioning can influence the participants' attitudes to the brands. However, conditioning was not the only explanation of this phenomenon. More trials mean more familiarity. The familiarity may also be one of the reasons, which should

be verified later. In their study, the experimenters also changed the sequence of US and CS. They first presented the US and then the CS, which is called backward EC acquisition. They also found the same phenomenon according to the back forward sequence as the forward sequence, which meant that both forward EC and backward EC can influence the attitude of consumers. This finding inferred that with commercials, no matter when the logo is presented, the evaluative conditioning works.

After Stuart and her colleagues' study (1987), there was a series of studies which further proved EC's effect on advertising. For example, Gibson (2008) used two mature brands (Pepsi and Coca-Cola) to test if EC can change people's attitudes to recognize brands. The participants in the study were divided into three groups based on their preferences to the brands: strong preference for Coca-Cola, strong preference for Pepsi, and approximately equal preference. The brands were matched to emotional pictures or words during the EC acquisition procedure. After the acquiring procedure, both the implicit and explicit attitudes were tested. Results suggested that EC can only affect implicit attitudes of consumers on recognized brands.

Even though many previous studies supported that EC influences consumers' attitudes on brand, some other researchers hold the opposite opinion (Campbell & Keller, 2003; Rucker & Petty, 2004; Puccinelli, Wilcox, & Grewal, 2015). Campbell and Keller (2003) used television and Internet advertisements and found that repetition of advertisements had a negative effect on the participants' attitude, especially advertisements about unfamiliar brands. This finding was opposite to the evaluative conditioning theory, which predicts that repetition increases conditioning strength and makes people more favorable to the brands. In Rucker and Petty's (2004) study, participants were induced negative feelings with different arousal level, and then they were presented with the advertisements. They also found the improvement of shopping behavior by the increased arousal level. While negative emotion had no effect on the shopping behavior, which was inconsistent with the evaluative

conditioning theory's prediction. According to this research, induced arousal level is another important factor which should be considered.

Briefly, even though the EC explanation of advertising seems reasonable and had some experimental evidence, it still needs stronger evidence in view of those unsupported studies (Poels & Dewitte, 2006), which is our current study's purpose. We tried to overcome the limitations of the previous studies and used an experimental method to explore whether or not the evaluative conditioning influences the effect of advertisements.

Methodology

The limitations in the previous studies are the following. First, the experimental conditions were not strictly controlled. For example, the advertising stimuli they used was directly selected from real TV programs (Leonidas, Christina, & Yorgos, 2009; Pham, Geuens, & De Pelsmacker, 2013). Real advertisements always combine information of the real products, which is not neutral to the participants. Besides, the advertisements also contained different emotions. The combination of several emotions would mislead the participants' response to the brand or logo, which probably was not an adequate US. In addition to the combination of emotions, the unpredictable factors, such as the special person or actions in the advertisement (Ohme, Reykowska, Wiener, & Choromanska, 2009) and the background music (Zander, Apaolaza-Ibáñez, & Hartmann, 2010), can also influence the participants' behaviours. Second, surveys were mostly used to evaluate the participants' feelings towards the advertisement (Adalarasu, Felixia, Priyanga, & Jayaraman, 2014), which are too subjective. Surveys also have some other limitations besides subjectiveness. For example, the result of surveys can be easily influenced by many other factors, like social pressure, bandwagon effect, and so on. The results of surveys are conscious but the evaluative conditioning is unconscious which maybe cannot be reported by using

questionnaires (Yeh, Lin, Li, & Tsai, 2012). Third, the effectiveness of advertising was not well captured in the previous studies. The tools of testing advertising effectiveness in previous studies were questionnaires, like "Which product do you prefer?". The final goal of a company, increasing sales, is not necessarily equal to preference.

We solved those problems in the current study by applying the following methodology:

Emotion elicitation: in the EC acquiring process the most crucial part is to properly elicit participants' emotions as their unconditional response. How to elicit the proper emotion during the EC acquiring procedure should be carefully considered. According to previous studies, emotional words, pictures, music, or movies are used to induce emotions (Fossati et al., 2014; Dan-Glauser & Scherer, 2011; Vuoskoski & Eerola, 2012; Schaefer, Nils, Sanchez, & Philippot, 2010). Compared to other inducing stimuli, emotional movies have at least two advantages. First, it is demonstrated that emotional videos can elicit strong subjective and physiological changes (Kreibig, Samson, & Gross, 2013; Frazier, Strauss, & Steinhauer, 2004). Second, the emotions induced by emotional videos are stronger than the emotions induced by other kinds of stimuli, which is better for the acquiring conditioning procedure. For example, Westermann, Stahl, and Hesse (1996) compared 11 mood inducing procedures and suggested that emotional movies are most effective to induce both positive and negative emotions. Besides, emotional movies are auditory-vision stimuli, which are similar to TV advertisements. So the current study's result can be practical to real life. Thus, emotional videos selected by Schaefer and his colleagues (2010) were used to induce the participants' emotional response.

Autonomic measurements of emotions: As we mentioned above, in the previous studies questionnaires were used to explore the emotions of the consumers. However, surveys are too subjective and unstable. Instead of surveys, in some

previous studies automatic measurements were used to record people's emotions. Some researchers defined that emotion is a psycho-physiological process influenced by some situations or objects, no matter if it is conscious or not (Koelstra et al., 2012). Some previous studies demonstrated this definition. For example, Poels and Dewitte (2006) found that emotions can be detected by combining different physiological sensors which are able to detect small changes in individual's physiological status accurately with the development of technology.

Virtual online shopping: to directly test the participants' shopping behavior, a virtual online shopping website was developed. So the participants can shop online, which is similar to real life. The developed website better controlled other factors besides the conditioned logo, such as price, specification, and so on.

The automatic measurements include both facial expression and physiological reactions caused by autonomic changes on the nervous system (Winkielman, Berntson, and Cacioppo, 2001). For physiological reactions, some sensors, like Electroencephalography (EEG), Electromyography (EMG), galvanic skin response (GSR), and so on, were widely used in many fields including advertising research (Adalarasu et. al., 2014). For example, Ohme, Matukin, and Szczurko (2010) used EEG, EMG and GSR to measure the participants' emotional response to two versions of one TV commercial (Sony Bravia "Ball" commercial with the frog scene and without the frog scene). They confirmed that the physiological signals were significantly different between the two versions and suggested that measurements of physiological signals can be used to record consumers' emotional response to commercials. Dawson, Schell, and Filion (2000) suggested that galvanic skin response (GSR) can be used to measure emotional arousal, which is an important emotional dimension (Kensinger & Schacter, 2006). Even though the physiological sensors are valid to measure emotions, the sensitivity of the devices and accuracy of the algorithm to clean data should also be considered in order to get effective results (Poels & Dewitte, 2006). Based on these findings, in the current study the GSR sensor

was used to record the participants' emotional arousal in addition to the self-report questionnaires. Besides, to ensure that the results of the current study were reliable and valid, the raw data from the sensors were cleaned, low-pass filtered and smoothed.

Besides the physiological signals recordings, facial expression analysis is another method to understand individuals' autonomic emotional responses. As physiological signals cannot easily indicate emotional arousal or emotional valence (Matsumoto, Hwang, Harrington, Olsen, & King, 2011), facial expression is a helpful method to recognize individuals emotional status. Ekman, Davidson, and Friesen (1990) suggested that facial expressions are correlated to physiological status. Additionally, researchers also found that different emotions will be universally expressed on the face no matter the race (Matsumoto et al., 2011; Matsumoto, Keltner, Shiota, Frank & O'Sullivan, 2008). The previous studies inferred that researchers can use facial expression to analyse the participants' emotional response to different situations, including advertising. More and more systems can assess facial expression captured by a camera, such as iMotions system (De Lemos, 2007), Emotient Analytics system (Movellan et al., 2014), and so on. In the current study, a camera recorded the facial expressions of the participants and the Emotient Analytics system was used to analyse the facial expression in the videos. The system outputs the participants' emotional valence which is more objective than surveys.

Contributions

Compared to the previous studies, the current study had three main contributions:

- To explore the effectiveness of advertisements, the current study used an experimental method which strictly controlled irrelevant factors such as gender, age, advertising stimuli, and so on. Moreover, the experimental method helps us to better understand the causal relationship between induced

emotions and people's shopping behaviors.

- Due to the subjectivity of self-reporting, physiological measurements were used in the current study including facial expression recordings (emotional valence) and GSR sensor recordings (emotional arousal).
- In some of the previous studies, advertising effectiveness was hard to test. In the present study the advertising effectiveness was directly tested by using a virtual online shopping experience, instead of the questionnaires of the customers' attitude. So the consumers' shopping behavior, which is the intended result, was recorded and was considered as the dependent variable in the study.

Method

To test if the evaluative conditioning theory works for consumers' shopping behaviour, the current study first provided a conditioning procedure for the participants. Then, the participants' shopping behaviour was evaluated by a virtual shopping experience. In the conditioning procedure, some emotional videos used in a previous study (Schaefer, Nils, Sanchez, & Philippot, 2010) were used as the unconditioned stimuli, and some neutral logos were used as the conditioned stimuli. During the conditioning procedure, the participants' facial expression and their GSR were recorded, which can help us better understand their emotions. After the conditioning procedure, which combined emotional videos and certain logos, the participants took part in a virtual online shopping session. In the shopping procedure, all the products, which belonged to the same type, had different logos, but were otherwise identical (the same picture, the same specification, and the same price, etc.). The following sections detail the experimental methodology.

Participants

39 healthy volunteers participated in the study (20 males and 19 females). The mean age of the participants was 30.31 (SD = 13.47). Twenty participants were students from the VU University Amsterdam, and 19 participants were Ph.D. students or employees at Centrum Wiskunde & Informatica (CWI). All of the participants were paid by credits or by a 10 euros gift card (depending on their choice). Because of problems with the Galvanic Skin Response (GSR) sensors and cameras, the data of 8 participants were lost, and the data of 11 participants were not analyzed. In total, data from 20 participants were used in the analysis. Before they agreed to participate in the experiment, they were adequately informed about the experiment, explicitly mentioning that they were going to watch some really negative emotional videos and that their face would be recorded during the experiment. All the participants in the

experiments signed the informed consent before the experiment started.

Stimuli and Apparatus

The emotional videos used in the study were selected based on Schaefer, Nils, Sanchez, and Philippot's study (2010). Schaefer and his colleagues created a large database of 70 film clips including seven different emotions (anger, fear, disgust, sadness, amusement, tenderness, neutral). Before the experiment, all the emotional videos were selected according to the results of arousal and valence questionnaires included in the study by Schaefer and his colleagues (2010), which is openly available on their website (http://nemo.psp.ucl.ac.be/FilmStim/). In the present study, the English version of the film clips, cut by the experimenter, were used. The study only focused on the arousal and valence of the emotional film clips, which restricts the number of clips to 15: 6 positive film clips (three high arousal and three low arousal), 6 negative film clips (three high arousal and three low arousal), and three neutral film clips. Because the arousal of the neutral film clips in the Schaefer and his colleagues' study (2010) was almost the same, the present study included another three neutral videos (self-reported arousal: "While I was watching the film…" 1 = "I felt no emotions at all" to 7 = "I felt very intense emotions"; valence: Positive and Negative Affect Schedule (PANAS)). By using the same questionnaires as the pretest (Appendix A), all the neutral videos were also divided into two groups (high arousal and low arousal), as the positive and negative emotional videos. The length of the videos used in the study ranged from 40 to 186 seconds, with good image quality, and without any unclear part, or any logos. All the videos are described in Appendix B.

20 logos were downloaded from a free designing website and modified using Photoshop CS4 by the experimenter. The logos were professionally developed and did not contain any letters. There was a pretest to evaluate if the logos were neutral and unfamiliar. 8 of those logos, which were the most neutral and unfamiliar, were

selected (see Appendix C). 6 of them were matched to the different emotional videos according to the different arousal (high and low) and valence (positive, negative, and neutral) emotions. Two of them were not matched to emotional film clips, but only appeared during the online shopping phase as the unfamiliar logos. All the logos are listed in Appendix D.

A virtual online shopping website was used to test the participants' shopping behavior after the conditioning procedure. Thomas Röggla from the DIS research group at CWI developed the website (Figure 1). There were two kinds of products in each category in the virtual online shopping website (phones, tablets, MP3 players, and laptops). So there were 8 products in total in the website. For each product, there

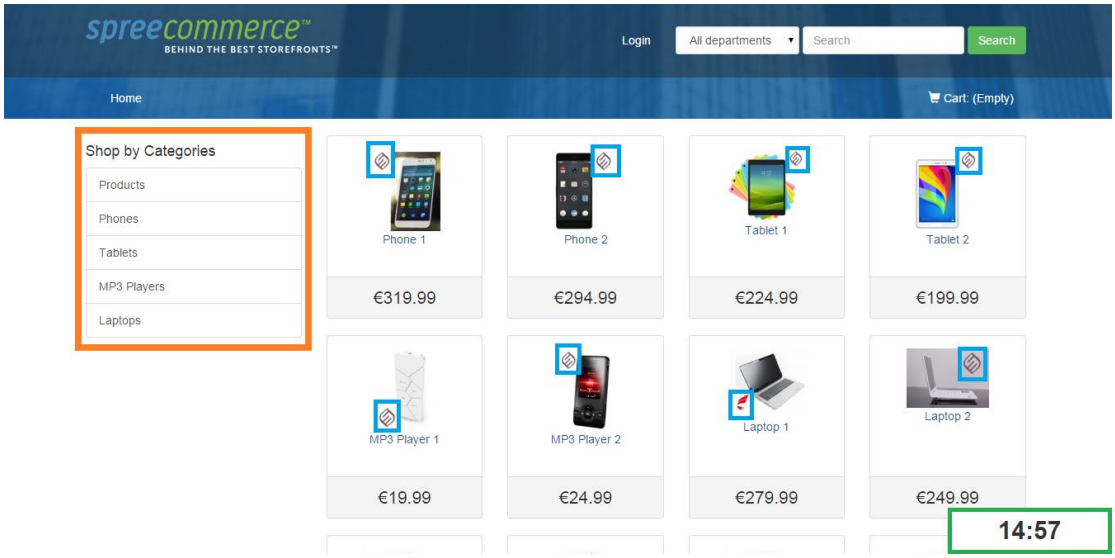


Figure 1 The webpage used in the virtual online shopping test. The number in the green square is the countdown. The blue square is the logo we want to test. The products in the orange square are the categories the participants can choose.

were 8 logos with the same price, product picture, and specifications for the participants to choose: 6 logos appeared during the conditioning procedure and the other two were unfamiliar never appeared to the participants. There was a countdown in the bottom right corner on the webpage, which reminded the participants how long time he or she can spend shopping. All the products were shown in one page, and the sequence of the products was random for different participants.

The data collection was done using GSR sensors and associated software**.** Because the experiment was conducted in different locations (VU University Amsterdam and CWI), two kinds of GSR sensors were used to record the participants' arousal while watching the emotional videos in different locations. One of the GSR sensors is developed by the commercial company BITalino (Figure 2). The range of the BITalino GSR sensor is from 0 to 1Mohm using a 3.3V virtual circuit connection (VCC). The bandwidth ranges from 0 to 3Hz, and the consumption is around 2mA. The sampling rate is 100Hz. During the experiment all the data were sent using Bluetooth to the connected laptop. The software OpenSignals, which was supplied by the company, was used to receive and visualize the data. Chen Wang from the DIS group at CWI developed the other kind of GSR sensor (Figure 3). The sensor node includes a radio frequency (RF) 12 wireless module. The sampling rate is 160Hz. During the experiment, all the data were sent to a sink node, which was connected to a laptop. All the data were collected by the Python code on the laptop. The GSR sensor attached the participants by connecting the Ag-AgCl electrodes to the patch (bottom left on Figure 2). The patches were attached the participants' two fingers of their non-dominant hand: index finger and middle finger.

Figure 2 Components of BITalino sensors (from bottom to top, from left to right: patch, battery, electrodes for GSR sensor, cable connected different sensor components to main board, different sensor components, main board, different sensor components, and electrodes for ECG sensor).

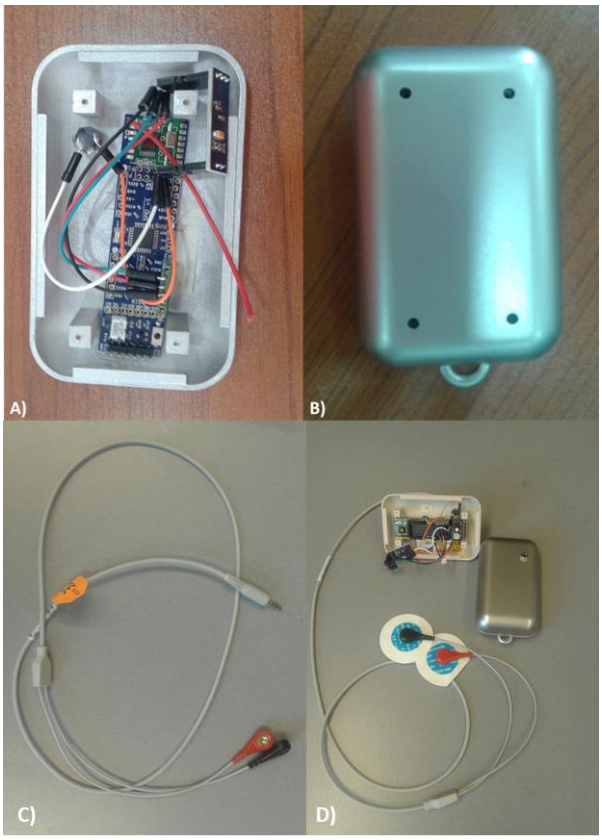


Figure 3 The GSR sensor developed by the DIS research group at CWI. A) The main part of the GSR sensor which includes Jeenode board and a low-pass filter. B) The cover of the GSR sensor. C) The cable connected the patch and the main part of the GSR sensor, which consists of the same materials as the BITalino GSR sensor. D) The whole GSR sensors.

To analyze the valence of the emotional videos, the facial expression of the participants was recorded by using a Logitech HD Pro Webcam C920 camera. The facial expression videos were then analyzed by Emotient Analytics (http://www.emotient.com/), which is a commercial company (see Figure 4). This software measures and detects people's facial muscle movements by the Facial Action Coding System (FACS), which specifies 9 action units in the upper face, 18 in the lower face, 14 head positions and movements, 9 eye positions and movements, 5 miscellaneous action units, 9 action descriptors, 9 gross behaviors, and 5 visibility

codes (Cohn, 2005). According to FACS, the software detects facial expression and divides them into 8 basic emotions (joy, surprise, anger, contempt, disgust, fear, sadness, and neutral). Of these, joy and surprise were categorized as positive emotional valence, while anger, contempt, disgust, fear, and sadness were categorized as negative emotional valence. Finally, the proportion of each emotional valence was output as the raw data. The sampling rate of the software is 1Hz. So every second, the software output one raw data point from the participants' facial expression. And this method's validity was confirmed by different studies (Rossi, 2012; Rossi, Fasel, & Sanfey, 2011).

Figure 4 The output of the Emotient Analytics software. The emotional face (left) and the valence of the whole video (right) are both output.

All the stimuli in the experiment were randomly shown in a 27 inches large, 1920*1080, 60 Hz Dell screen. The viewing distance was around 75 cm and all the participants were in a suitable position, which made them look at the middle of the screen. The emotional videos, which were matched to logos, were presented by Open Sesame 2.8.1. Besides, the GSR raw data were cleaned using Matlab 2013b.

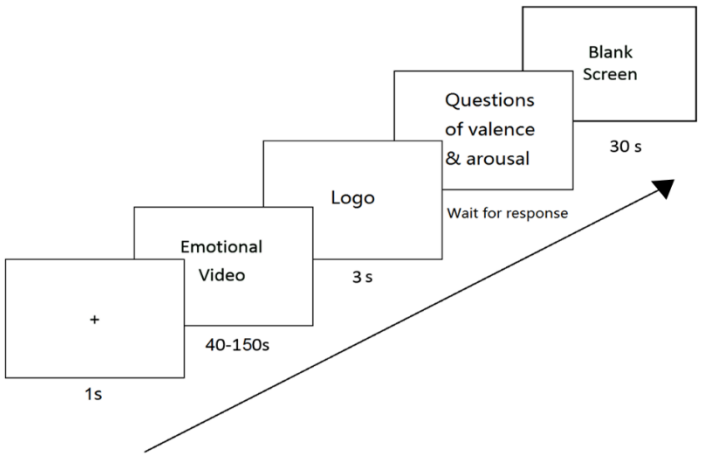


Figure 5 The procedure of every trial in the experiment.

Procedure

The participants sat in a quiet experimental room on a comfortable seat. After attaching the sensors, the experimenter provided verbal and written instructions to ensure that the participants understood the whole procedure. Then, the participants read and signed the informed consent. In the beginning of the experiment, there was a practice run so the participants got accustomed to the system. A 25 seconds' neutral movie was presented, which not only ensured that the participants were familiar with the experimental procedure, but also created a baseline condition that allowed the registration of autonomous physiological activation (valence and arousal) in the absence of external stimulation. This eliminated possible pre-emotional effects in the participants, and allowed the experimenter to test if the sensors worked well or not.

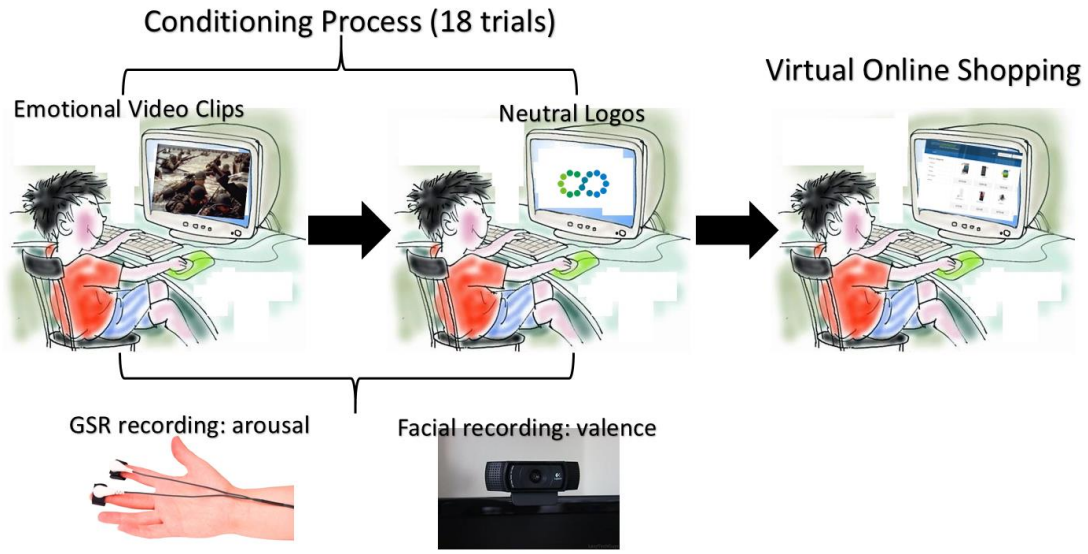


Figure 6 The procedure of the experiment. First is the conditioning process. When the participants watched the video clips and the logos, their arousal and valence were recorded by the GSR sensor and the camera. After watching 18 video clips (18 trials), they went to the virtual online shopping

test.

After the practice session, the actual experiment started. The participants first watched a fixation dot for one second followed by an emotional video. Immediately after watching the film clips, there was a matched logo that appeared for three seconds, which only matched to one emotion, allowing the formation of a strong conditioned response. In addition, the pairs of logo and emotional video clips were counter balanced among the participants. To better evaluate the valence and arousal of the video, the participants filled out two questions about the emotional film clip after the logo was shown: "This video made me feel: 1. extremely negative 2. negative 3. neutral 4. positive 5. extremely positive" and "When I watched the video, I was absorbed. 1. Totally disagree 2. Disagree 3. Maybe 4. Agree 5. Totally agree". These two questions' answers were recorded as the self-report data of valence and emotional arousal, respectively, and would be analysed later. Then, there was a blank screen for 30 seconds, so the participants return to the baseline condition again. There was an audio reminder at the end of the blank screen. The procedure for each trial is shown in Figure 5. There were two independent variables: arousal (high and low) and valence (positive, neutral, and negative). So there were 6 conditions in the experiment, and each condition was 3 times repeated during this phase, which means that the participants saw the corresponding logo 3 times, and 18 emotional video clips (18 trials) in total. After finishing watching the videos, the participants accessed a virtual online shopping webpage. The participants were requested to buy a product in the shopping website within 15 minutes. The virtual website was similar to the real online shopping website. So the participants can navigate through the website by controlling the mouse and decide which product he or she wanted to buy. When they made the decision, they can click the purchase botton and enter the fake email address (subject_ID@example.com). The whole procedure of the experiment is shown in Figure 6.

Data Analysis

Before analyzing the physiological sensors data (including both GSR data and facial expression data), the dataset was cut according to time stamps of the emotional videos. Then, the raw data of the GSR sensors were first cleaned from noise by a 2 Hz low-pass filter in Matlab 2013. Additionally, the smooth function in Matlab 2013 was used and the span of the moving windows of the function was 30 (data points), thus eliminating chance fluctuation and noise. Because of the different of sampling rate between different GSR sensors (100 Hz and 160 Hz) and facial expression analysis software (1 Hz), the clean data of GSR sensors were averaged from 100 Hz or 160 Hz to 1 Hz.

The facial expression recordings were upload to the website of Emotient Analytics (https://analytics.emotient.com/home). The software used algorithms to detected the facial expression according to FACS and return the raw data of their emotional valence (proportion of each emotional valance in the facial expression) (Sikka, Dykstra, Sathyanarayana, Littlewort, & Bartlett, 2013). The final data of the facial expression were the weighted results of the different emotional valences (Formula 1). In this formula, i stands for the valence (negative, neutral, or positive), f_i is the percentage of the valence according to the facial expression analysis software, and x_i is the weighing value (the weighing value of negative emotion is 1, the value of neutral emotion is 2, and the value of positive emotion is 3). F means the valence of the participants, and the higher F, the more positive of the emotions. Then, F was averaged according to each individual emotional movies of each participants. The final value was used in the data analysis as the valence value.

$$F= \sum(f_i * x_i) \quad \text{(Fomula 1)}$$

In the virtual online shopping, only the participants' choices of the logos were recorded. Because the logos were matched to the emotional videos, the choice of logos were also linked to the emotional videos. The dependent variable used in the

data analysis is if the participants chose the logo, which linked to videos. If they chose, the value was considered as 1. Otherwise, the value was considered as 0. Besides, the chosen frequency of each logo was also recorded as another dependent variable. And the familiarity of the logo was recorded as dichotomous data (Familiar: 1 / Unfamiliar: 0).

To better understand if we used the right unconditioned stimuli and established the evaluative conditioning correctly, one-way analysis of variance (ANOVA) was used separately for analysing the difference between different arousal groups, and between different valence groups. To analyse if the arousal grouping is right, the higher arousal group was considered as 2 and the lower one was considered as 1, which was the independent variable in the one-way ANOVA model. The self-report arousal level or physiological arousal data were considered as the dependent variable. To analyse if the valence grouping is right, the positive group was considered as 3, the neutral group was considered as 2, and the negative group was considered as 1, which was the independent variable. The self-report valence level or physiological valence data was considered as the dependent variable.

Binary logistic regression model was used for analysing the emotional valence and arousal effects on the people's shopping behaviors. In this model, gender and age were also considered as independent variables. Female was considered as 1 and male was considered as 0. Besides, because the range of age is large, we divided different age into three groups according to the Erikson's stages of psychosocial development (Erikson, 1959). Ages from 13 to 19 were recorded as 1, ages from 20 to 39 were recorded as 2, and 40 to 65 were recorded as 3. Both physiological data and self-report data were analysed by the binary logistic regression model. In the physiological data analysis, the GSR data represented the emotional arousal level and the facial expression data represented the emotional valence level, which were the continuous independent factors. The gender and the age were considered as the nominal independent variables. The choice value (Chose: 1 / Not chose: 0) of emotional video

selection linked to logos was considered as the dependent variable.

Beside, Person correlation was used for analysing the relationship between the familiarity of the logos and the participants' shopping choice. The dichotomous data of (Familiar: 1 / Unfamiliar: 0) was considered as the familiarity of the logos, and the chosen frequency of each logo was considered as the participants' shopping behavior.

Results

Self-report results

The average self-report arousal value of the participants is 3.56 (SD = 1.14). The average self-report emotional valence value of the participants is 2.94 (SD = 1.12). The self-report arousal value and emotional valence value for each movie are presented in Table 1.

One-way analysis of variance (ANOVA) was used to test if our division of arousal is valid according to the self-report data. The error variance of the dependent variable is equal across groups, $F(1, 365) = 1.92$, $p = .17$. So the ANOVA model can be used to test if the arousal grouping is valid. There is a significant main effect for arousal value, $F(1, 365) = 5.24$, $p < .01$. The mean value of the higher arousal group is 3.69 (SD = 1.11), and the mean value of the lower arousal group is 3.42 (SD = 1.15), which means that the higher arousal group is significant higher than the lower arousal group. So the arousal grouping is valid. To test if the grouping of valence is valid, a one-way ANOVA was also used. The self-report of emotional valence data was considered as the dependent variable. The homogeneity test result is significant, F (2, 364) = 18.83, p < .01, which means that the error variance of the dependent variable is not equal across groups. So ANOVA cannot be used. An independent samples Kruskal-Wallis test was used to compare different valence groups instead of one-way ANOVA. According to the result, there is a significant differences between different valence groups, $\chi^2(2) = 144.90$, $p < .05$. According to the Nemenyi test, which pairwise compares different valence level, there is a significant difference between negative group and neutral group, $\chi^2(1) = 126.14$, $p < .01$, and a significant difference between negative group and positive group, $\chi^2(1) = 144.91$, $p < .01$. So the negative group (mean = 1.93, SD = 0.87) is more negative than the neutral group (mean = 3.33, SD = 0.61), and the negative group is more negative than the positive group (mean = 3.53, SD = 1.07). However, there is no significant difference between positive group and neutral group, $\chi^2(1) = 18.76$, $p = .449$.

So the grouping of emotional valence is not too successful.

Binary logistic regression model analysis was used to describe the emotional valence and emotional arousal's effects on the participants' shopping behavior. In this model, gender and age of the participants were considered as the control variables and entered in block 1. The emotional valence value and the arousal value were considered as the independent factors, which entered in block 2. Besides, gender and age were considered as the categorical covariates, which were coded as the dummy variables in the model. Results showed that adding the valence and arousal factors significantly contributed to the predictive value of the model, $\chi^2(2) = 6.94$, $p < .05$. These four factors explained 5.0% of the variance in the participants'' shopping behavior (*Nagelkerke* $R^2 = .05$). According to the Hosmer and Lemeshow goodness-of-fit test, this model is good to fit the data, $\chi^2(8) = 7.37$, $p = .50$. In the model, arousal significantly predicted the participants' shopping choices, B = 0.32, *Wald* $\chi^2(1) = 3.94$, $p<.05$. This means that the higher arousal when the participants watch the advertisement, the higher the probability they will buy the product. The details of the model are presented in Table 2.

Table 1 Description of self-report arousal and valence value for each movie

Movie	**GSR**		**Facial Expression**	
	Mean	**SD**	**Mean**	**SD**
Benny Joone	3.80	1.11	3.85	0.75
Introduction of Bitalino	2.55	1.39	3.10	0.55
Blue	2.24	1.00	3.14	0.36
Introduction of car	2.45	1.28	3.05	0.39
Dead poets society	4.05	0.80	4.05	0.86
Hellraiser	3.79	0.79	2.21	0.85
Indiana Jones and the Last Crusade	3.95	0.97	2.67	0.86
The silence of the lambs	4.05	0.86	2.05	0.83
Life is beautiful	4.25	0.72	3.90	1.02
Lover	3.14	0.91	3.14	0.57

When a man loves a woman	3.33	1.02	3.43	0.75
Nature scene	2.80	0.95	3.80	0.62
The piano	4.21	0.63	1.45	0.69
The professional	4.30	0.86	2.05	0.69
Saving private Ryan	4.05	0.74	1.76	0.77
Sea scene	2.81	1.03	3.71	0.64
Schindler's list	4.20	0.52	1.45	0.60
A Fish Called Wanda	4.00	0.77	3.86	0.91
Total	**3.55**	**1.14**	**2.93**	**1.12**

Table 2 The binary logistic model of self-report data

Predictor	***B***	***Wald*** χ^2	***p***	**Odds Ratio**
Arousal	0.316	3.938	.047	1.372
Valence	-0.180	1.732	.188	0.835
Gender (compare male to female)	-0.218	0.344	.558	0.804
Age		1.686	.430	
Age (compare youth to aged)	-0.352	0.270	.603	0.703
Age (compare middle-age to aged)	0.361	0.829	.363	1.434

Physiological sensors results

The average GSR level of all the participants when they watched the emotion videos was 803.37 (SD = 150.29). The average facial expression value of all the participants was 1.76 (SD = 0.38). The average level of the GSR data and the emotional valence level of different emotional videos are presented in Table 3.

One-way analysis of variance (ANOVA) was used to test if our division of arousal is valid according to the physiological data. The grouping of the arousal level is the same as the self-report data. The homogeneity test result is significant, $F(1, 14239) =$

1.09, p = .30. The main effect of the arousal groups is not significant, F(1, 14239) = 81726.08, p = .07, which means that there is no significant difference between the lower arousal group and the higher arousal group according to the GSR sensor data. For valence group, we also used the same grouping method and same analysis method. The error variance of the dependent variable is equal across groups, F(1, 13318) =2.76, p = .06. So the ANOVA model can be used to test if the valence grouping is valid. There is a significant difference of the valence value between different valence groups, F(1, 13318) =20.44, p < .01. According to the post-hoc test with Tukey test, there is a significant difference between negative group (mean = 1.73, SD = 0.37) and neutral group (mean = 1.76, SD = 0.37) on the valence level, p < .01, and between negative group and positive group (mean = 1.78, SD = 0.40), p < .01. However, the difference between positive group and neutral group is not significant, p = .149. According to the result, the grouping of valence according to the physiological valence data is not too successful.

Binary logistic regression model analysis was used to describe the effects of emotional valence (the facial expression data) and emotional arousal (the GSR data) on the participants' shopping behavior. The physiological values for each movie for each participant were averaged, so the independent variables' value corresponds to the choice. Like the analysis of self-report data, in this model, gender and age of the participants were considered as the control variables and entered in block 1 of the model. The emotional valence value and the arousal value were considered as the independent factors, which entered in block 2. Besides, gender and age were considered as the categorical covariates, which were coded as the dummy variables in the model. Results showed that neither adding the control variables nor adding the independent variables contributed significantly to the predictive value of the model. Those four factors can explain 3.5% of the variance of participants'' shopping behavior (*Nagelkerke* R^2 = .035). According to the Hosmer and Lemeshow goodness-of-fit test, this model is good to fit the data, $\chi^2(8) = 9.164$, p = .330. In the model, all of the independent variables were insignificant to predict the participants' shopping choices. The details of the model were

presented in Table 4.

Table 3 Description of physiological arousal and valence data for each movie

Movie	GSR		Facial Expression	
	Mean	SD	Mean	SD
Benny Joone	832.43	110.08	1.82	0.36
Introduction of Bitalino	781.99	158.49	1.68	0.42
Blue	854.17	105.42	1.77	0.30
Introduction of car	847.15	76.38	1.80	0.36
Dead poets society	766.02	224.50	1.75	0.45
Hellraiser	809.26	125.11	1.78	0.35
Indiana Jones and the Last Crusade	775.83	172.99	1.78	0.35
The silence of the lambs	764.82	192.48	1.69	0.38
Life is beautiful	830.83	130.70	1.85	0.34
Lover	859.51	84.20	1.82	0.33
When a man loves a woman	771.43	185.58	1.81	0.37
Nature scene	811.34	131.18	1.75	0.39
The piano	859.86	82.32	1.74	0.31
The professional	767.03	169.75	1.70	0.39
Saving private Ryan	808.81	135.59	1.70	0.40
Sea scene	815.20	132.80	1.80	0.32
Schindler's list	821.31	102.77	1.77	0.35
A Fish Called Wanda	771.36	170.73	1.80	0.41
Total	**798.01**	**159.46**	**1.76**	**0.38**

Table 4 The binary logistic model of self-report data

Predictor	*B*	*Wald* χ^2	*p*	Odds Ratio
Arousal	-.001	0.078	.780	0.999
Valence	-0.657	0.559	.455	0.518

Gender (compare male to female)	-0.548	0.737	.391	0.578
Age		1.449	.485	
Age (compare youth to aged)	-0.804	1.108	.293	0.447
Age (compare middle-age to aged)	0.032	0.003	.960	0.968

Familiarity

A Person correlation between familiarity and the participants'' choices was analyzed. There is no significant correlation between familiarity and the participants' choices, $r(6) = .17$, $p = .68$.

Discussion

Physiological data

Based on the results of the sensor data, unfortunately, neither the whole binary regression model nor the independent variables (arousal level, valence level, gender and age) were significant. Those results mean that the arousal level, valence level, gender, and age cannot predict the participants' shopping choice in the virtual online shopping test.

To explain the results of the sensor data, the following reasons should be considered: 1) the validity of the physiological data, 2) experiment process, and 3) data analysis method.

First, regarding the validity of the physiological data, the arousal level was recorded by GSR sensor and the valence level was measured by using a facial expression analysis software, both of them rather using premature technology. According to the previous studies, GSR sensor is sensitive and influenced by many factors besides the emotional arousal such as body temperature, body movement, and so on (Westeyn, Presti, & Starner, 2006; Bakker, Pechenizkiy, & Sidorova, 2011). Even though we used a low-pass filter and smooth function to reduce possible effects of such factors, it is impossible to ensure that the data used in the data analysis part were clean. However, the validity of the GSR data was not the biggest problem, because it was supported many times based on the previous studies (e.g. Bakker et al., 2011; Latulipe, Carroll, & Lottridge, 2011). The most critical point in the present study was probably the validity of the facial expression data. In particular, we found that most movies failed to elicit strong facial expressions, such that many data values represented neutral emotion (The figures of facial expression raw data of each movie are shown on Appendix E). This finding means that the participants' facial expressions during the emotional video watching process were mostly neutral, which is not corresponding to

the actual situation and the self-report data. To test this viewpoint, Pearson correlation was used. We found that there is no significant correlation between the physiological data and the self-report data regarding valence, $p = .872$, which proved that the facial expressions data were not corresponding to the self-report data. So the validity of the facial expression data is in doubt. Why did this happen? Maybe the reason is that emotional valence is not always reliably expressed on people's face. Even if they feel happy or sad, the facial expression may be neutral at the same time. In addition, the experiment happened in the lab, which is not as relaxing as the home environment. And the participants knew that there was a camera to record their facial expression. Because of this, the participants may control their facial expression deliberately. Therefore, the lab environment influenced the participants' facial expression validity.

In addition to the physiological data's validity, the experimental process may also have influenced the final result. During the experiment, the participants only needed to watch the emotional videos and fill two questions about the emotional valence and arousal after every video. The whole process before the virtual online shopping experience lasted around 40 minutes. Because of the long duration, the participants may have lost their attention to the movie and its associated logo. So the GSR data recorded during the experiment may not reflect the arousal of the users.. To solve this problem, some tasks like memory test or searching task can be included in the experiment to ensure all the participants are engaged in the emotional videos. Another reason, which may influence the physiological data collection, is the sensor itself. During the experiment, all the participants were wearing the GSR sensors on their two fingers. This kind of wearing may make the participants feel uncomfortable and affect their performance during the experiment.

Thirdly, the insignificant result may be because of the data analysis method. In the data analysis, the binary regression model was used. To match the dependent variable (final choice) to the independent variables (GSR data and facial expression data), all the physiological data from each movie from each participant was averaged. By

averaging the physiological data, there was a lot of information lost, which may have influence the results. Moreover, according to the physiological result, the Nagelkerke R square is only 3.5% which means that the model explained little variance of the participants' shopping choices. The possible indication is that the regression model is not adequate. After consultation with experts on Statistics at CWI, they suggested that this type of research may require the creation of a new regression model, instead of the binary regression model. Nevertheless, that would require some extra research beyond the purpose of this thesis.

Self-report data

According to the results of the self-report data, only the arousal level of the participants significantly contributed to predict their shopping behavior. To explain it more precisely: the higher the arousal level induced by the advertisements, the higher probability that participants will buy the product. However, the evaluative conditioning theory of advertisement indicates that the emotional valence should significantly influence the participants' shopping behavior, which is different from the self-report result of the present study.

The effect of arousal on the consumers' shopping behavior was supported by previous studies (Rucker & Petty, 2004; Puccinelli, Wilcox, & Grewal, 2015). For example, Rucker and Petty (2004) found that increasing the arousal level of the participants can improve their shopping behavior, even though negative emotions were used to induce the participants' different emotional arousal level. In addition, Puccinelli and her colleagues (2015) also used a meta-analysis to indicate that both the arousal state of the consumers and the advertisements' arousal level influence the consumers' attitude to the commercials. The result of the self-report data is in accord with the results of some previous studies.

To explain the arousal effect on the shopping behavior, there are some possible explanations. First of all, high arousal level induced by the advertisement may also activate other brain circuits improving people's shopping motivation or decision-making. For example, in recent studies, some psychologists found that brain areas related to the motivation and decision making will be activated when emotional arousal stimuli provided to the participant (Sanfey, 2007; Critchley, Elliott, Mathias, & Dolan, 2000).

Second, higher arousal level improves consumers' memory of the advertisements, which may influence their shopping decision making. The relationship between advertising induced emotional arousal and memory was revealed by Bakalash and Riemer (2013). They found that there is a positive association between ad-elicited emotional arousal level and memory for the ad by using fMRI and self-reports measurements. Besides, according to Bettman (1979), the memory of commercials influence people's shopping choice. These previous results help us to infer that the higher the arousal level improves the consumers' shopping behavior. This explanation still needs some more experimental evidence. For instance, we can add the memory test in this experiment and analyze if there is any correlation between the memory test scores and the participants' choice. According to this, there may be a mediator between the advertising induced arousal level and the consumers' shopping behavior. We will discuss it in the Future Directions part.

As we mentioned above, the result of the self-report data is inconsistent with the evaluative conditioning theory. According to the evaluative conditioning theory, the consumers matched their emotional valence to the product, the logo, or the brand, and they transferred the emotional valence to the product, the logo, or the brand. So if the evaluative conditioning theory is true, the emotional valence of the advertisements should have a positive correlation to the participants' final choice, which is inconsistent with the present self-report result. Why does the result look different? There are several potential reasons. First, the previous studies demonstrating that evaluative conditioning

tested the advertising effectiveness by using questionnaires on attitude instead of real shopping behavior (Gibson, 2008; Stuart et al., 1987). In the present study, a virtual online shopping website was used to test the effectiveness of the advertisements. Because different dependent variable were used, the difference in the results is reasonable. The evaluative conditioning theory may only influence the consumers' attitudes to the products. In addition to the reasons above, we should consider weather the effective evaluative conditioning was established, which includes two points: right unconditioned stimuli and strength of the conditioning. According to the one-way ANOVA of valence grouping result, even though the grouping of valence are significantly different, there is no significant difference between positive group and neutral group, which cause problems for establishing the evaluative conditioning. And the strength of the evaluative conditioning may be the other problem. In the current study, because of the limitation of the duration, the strength of the evaluative conditioning may have been insufficient. The weak strength of the evaluative conditioning will also cause the insignificant effect of emotional valence.

Comparison between sensor data and self-report data

Comparing the physiological data and the self-report data, it is easy to conclude that the emotional valence of the participants does not influence their shopping behaviors. This indication is different to the evaluative conditioning, which suggests that the evaluative conditioning may not work on advertising effectiveness. Nevertheless, we cannot simply conclude that the evaluative conditioning theory does not work in the advertising field.

There are several issues we still need to check. The first issue is to check if we really tested evaluative conditioning. In the experiment, considering the whole duration of the experiment, each logo appeared only three times which is too little to establish evaluative conditioning. Besides, the backward conditioning schema may also

influence the effectiveness of the evaluative conditioning. A logo as the part of the emotional video may be better to create a stronger association. Furthermore, the CS used in the experimental conditioning building procedure may be not strictly suitable for every participant. Those emotional videos were selected according to the result of Schaefer and his colleagues' study (2010), and we divided those videos into different group based on their arousal and valence value. Then, each logo was matched to three emotional videos, which have the same level of arousal and valence. However, because of the individual differences, the classification of the CS may be not suitable to every participant. For some participants, the three emotional videos matched with the same logo may have different valence and arousal. According to the result of both the self-report data and physiological data, the standard error within group are larger. Besides, there is no significant difference of valence level between positive group and neutral group, which means that the evaluative conditioning did not set up as successfully as we expected.

Second, as we mentioned above, even though the evaluative conditioning was established successfully, the effect of this conditioning should be checked again. In the current study, to test the conditioning effect, the shopping behavior on the virtual shopping website was tested immediately after the conditioning acquiring procedure. This procedure caused three problems: 1) Online shopping is one type of shopping, which does not represent all kinds of shopping behavior; 2) Comparing to previous studies, the shopping behavior and not the attitude was tested. The evaluative conditioning may affect the attitude more significantly than the shopping behavior; 3) only the participants' choice was recorded, which may be an insensitive index to test the people's shopping behavior.

The conclusion of emotional arousal effect was hard to make, because the results of emotional arousal's effect is not the same between the self-report and the physiological data. Based on the self-report results, the emotional arousal has a significant positive effect on the participants' shopping behavior while there is no such

effect according to the physiological result. Why do the different results happen? Besides the physiological data validity and the data analysis problems mentioned above, another possible reason is that the different measures express different aspects of emotion and cognition. According to previous studies, the physiological data from different sensors represent the unconscious emotional state (Dienes, 2007; Bierman & Radin, 1998), while the self-report data represent the conscious emotional state of the participants. So it is not necessary to compare physiological data and self-report data. In addition, the arousal grouping is not valid for the physiological recording. According to the results of one-way ANOVA between different arousal groups, there is a significant difference of self-report data between the two arousal groups, while there is no significant difference of physiological data between those two groups.

Limitations

To conclude, there were at least three major limitations in the current study: experimental design problems, measuring problems, and application problems.

First, in the experimental design, the conditioning acquiring procedure was not so strict. During the establishing procedure, each logo combined three emotional videos with the same arousal level and valence level according. However, according to the one-way ANOVA results, this kind of combination may not work as we predicted. So the evaluative conditioning may not have been correctly established. To solve this problem, a real time analysis technique should be used during the experiment. By using this kind of technology, when the emotional video ends, the arousal and valence level will be known. So the right logo can be presented and adjusted to each particular participant. Additionally, the trials for establishing conditioning were not enough, which caused a weak conditioning. And the short duration between building conditioning and testing also caused other problems (e.g., insufficient time to move the evaluative effect to the shopping behavior). Last, the absence of a task during the

conditioning phase may have caused an engagement problem.

The second problem is related to measuring. The measuring problems include both the sensor recording and the online shopping testing. For facial expression, we found that the method was not valid. Many participants' faces were always neutral when they were watching the different emotional videos. The way the evaluative conditioning effect was tested also had some problems. For example, the online shopping website only tested the participants' behaviors instead of attitudes. Besides, the choice is not a sensitive dependent variable for testing the effectiveness, some more sensitive dependent factors (e.g. time spent in considering every product, time for making choice, and so on) should also be considered.

As the application problem, the stimuli used in the current study were emotional videos, which were better controlled but less practical than the real advertisements. How to apply the result into real life should be considered. Besides, buying product immediately after watching advertisements is not what normally happens.

Future directions

According to the discussion above, there are two topics that should be considered in future research.

First, according to the self-report result, the arousal level had a positive effect on the shopping behavior. One of the explanations of this result is that the arousal level influenced the memory of the advertisement, which then influenced the shopping behavior. This means that there is a mediator between the arousal level and the consumers' shopping behavior. The other future direction of the study is to explore what is the mediator between the physiological data and the shopping behavior.

The second direction is the combination between neuroscience and the shopping behavior. For example, in the current study, we suggested that the potential reason of the arousal effect is that arousal by the emotional videos activate the decision-making system in the brain. To explore this possibility, the method of neuroscience like the fMRI can be used.

Conclusion

According to the present study, there are three main conclusions, taking into account some known limitations:

1) The evaluative conditioning effect was not found neither according to the self-report data nor according to the physiological data.

2) Emotional arousal of the advertisement may have the effect on people's shopping behavior.

3) The familiarity of logos' effect was not found in the present study.

References

Adalarasu, K., Felixia, R. C., Priyanga, M. A., & Jayaraman, S. (2014). Evaluation of Entertainment Advertisement Using Different Techniques Review, *2*(3), 90–97.

Baeyens, F., Eelen, P., & Bergh, O. V. D. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective-evaluative learning.*Cognition and emotion*, *4*(1), 3-18.

Bakalash, T., & Riemer, H. (2013). Exploring Ad-Elicited Emotional Arousal and Memory for the Ad Using fMRI. *Journal of Advertising*, *42*(4), 275–291. http://doi.org/10.1080/00913367.2013.768065

Bakker, J., Pechenizkiy, M., & Sidorova, N. (2011, December). What's your current stress level? Detection of stress patterns from GSR sensor data. In*Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on* (pp. 573-580). IEEE.

Barry, T. E. (1987). The development of the hierarchy of effects: An historical perspective. *Current issues and Research in Advertising*, *10*(1-2), 251-295.

Bettman, J. R. (1979). Memory Factors in Consumer Choice: A Review. *Journal of Marketing*, *43*(2), 37–53. http://doi.org/10.2307/1250740

Bierman, D. J., & Radin, D. I. (1998). Anomalous Unconscious Emotional Responses : Evidence for a Reversal of the Arrow of Time 1.

Brown, S. P., & Stayman, D. M. (1992). Antecedents and Consequences of Attitude Toward the Ad: A Meta-Analysis. *Journal of Consumer Research*, *19*(1), 34. http://doi.org/10.1086/209284

Campbell, M. C., & Keller, K. L. (2003). Brand Familiarity and Advertising Repetition Effects. *Journal of Consumer Research*, *30*(2), 292–304. http://doi.org/10.1086/376800

Cohn, J. F. (2005). Observer-Based Measurement of Facial Expression With the Facial Action Coding System, 203–221.

Critchley, H. D., Elliott, R., Mathias, C. J., & Dolan, R. J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses:

a functional magnetic resonance imaging study. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *20*(8), 3033–3040. http://doi.org/http://dx.doi.org/20024039

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. *Behavior research methods*, *43*(2), 468-477.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2000). Handbook of psychophysiology. *Handbook of Psychophysiology. Cambridge University Press, Cambridge*, 200-223.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological bulletin*, *127*(6), 853.

De Lemos, J. (2007). *U.S. Patent Application 11/685,552*.

De Vries, M., Holland, R. W., Chenier, T., Starr, M. J., & Winkielman, P. (2010). Happiness cools the warm glow of familiarity: psychophysiological evidence that mood modulates the familiarity-affect link. *Psychological Science : A Journal of the American Psychological Society / APS*, *21*(3), 321–328. http://doi.org/10.1177/0956797609359878

Du Plessis, E. (2005). *The advertised mind: groundbreaking insights into how our brains respond to advertising*. Kogan Page Publishers.

Edell, J. A., & Burke, M. C. (1987). The power of feelings in understanding advertising effects. *Journal of Consumer Research, 14(3)*, 421–433.

Eifert, G. H., Craill, L., Carey, E., & O'Connor, C. (1988). Affect modification through evaluative conditioning with music. *Behaviour Research and Therapy*, *26*(4), 321–330. http://doi.org/10.1016/0005-7967(88)90084-8

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of personality and social psychology*, *58*(2), 342.

Erikson, E. H. (1959). Identity and the life cycle: Selected papers.*Psychological issues*.

Fossati, P., Hevenor, S. J., Graham, S. J., Grady, C., Keightley, M. L., Craik, F., & Mayberg, H. (2014). In search of the emotional self: an fMRI study using positive and negative emotional words. *American Journal of Psychiatry*.

Frazier, T. W., Strauss, M. E., & Steinhauer, S. R. (2004). Respiratory sinus arrhythmia as an index of emotional response in young adults. *Psychophysiology*, *41*(1), 75–83. http://doi.org/10.1046/j.1469-8986.2003.00131.x

Gibson, B. (2008). Can Evaluative Conditioning Change Attitudes toward Mature Brands? New Evidence from the Implicit Association Test. *Journal of Consumer Research*, *35*(1), 178–188. http://doi.org/10.1086/527341

Gibson, L. D. (1983). Not recall. *Journal of Advertising Research*.

Hammerl, M., & Grabitz, H. J. (2000). Affective-evaluative learning in humans: A form of associative learning or only an artifact?. *Learning and Motivation*, *31*(4), 345-363.

Haskins, J. B. (1964). Factual recall as a measure of advertising effectiveness. *Journal of Advertising Research*, *4*(1), 2-8.

Holbrook, M. B., & Batra, R. (1987). Assessing the role of emotions as mediators of consumer responses to advertising. *Journal of Consumer Research, 14(3),* 404–420.

Johnsrude, I. S., Owen, A. M., White, N. M., Zhao, W. V., & Bohbot, V. (2000). Impaired preference conditioning after anterior temporal lobe resection in humans. *The Journal of Neuroscience*, *20*(7), 2649-2656.

Jones, C. R., & Fazio, R. H. (2012). NIH Public Access, *2601*(10), 1–38. http://doi.org/10.1016/S0065-2601(10)43005-1.Evaluative

Kensinger, E. A., & Schacter, D. L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive, Affective, & Behavioral Neuroscience*, *6*(2), 110-126.

Kitson, H. D. (1920). *Manual for the Study of the Psychology of Advertising and Selling*. JB Lippincott.

Koelstra, S., Mühl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., … Patras, I. (2012). DEAP: A database for emotion analysis; Using physiological

signals. *IEEE Transactions on Affective Computing*, *3*(1), 18–31. http://doi.org/10.1109/T-AFFC.2011.15

Kreibig, S. D., Samson, A. C., & Gross, J. J. (2013). The psychophysiology of mixed emotional states. *Psychophysiology*, *50*(8), 799–811. http://doi.org/10.1111/psyp.12064

Kroeber-Riel, W. (1979). Activation research: Psychobiological approaches in consumer research. *Journal of Consumer Research*, 240-250.

Latulipe, C., Carroll, E. A., & Lottridge, D. (2011, May). Love, hate, arousal and engagement: exploring audience responses to performing arts. In*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1845-1854). ACM.

Lavidge, R. J., & Steiner, G. A. (1961). A model for predictive measurements of advertising effectiveness. Journal of Marketing. 25(10), 59-62.

Leonidas, H., Christina, B., & Yorgos, Z. (2009). The effects of culture and product type on the use of humor in greek tv advertising: An application of speck's humorous message taxonomy. *Journal of Current Issues & Research in Advertising*, *31*(1), 43-61.

Levey, A. B., & Martin, I. (1975). Classical conditioning of human 'evaluative' responses. *Behaviour Research and Therapy*, *13*(4), 221-226.

Li, H., & Leckenby, J. D. (2004). Internet Advertising Formats and Effectiveness. *An Invited Chapter for Thorson & Schumann*, (October), 1–31.

Matsumoto, D., Hwang, H. S., Harrington, N., Olsen, R., & King, M. (2011). No Title, *1*(3), 441–453.

Matsumoto, D., Keltner, D., Shiota, M. N., Frank, M. G., & O'Sullivan, M. (2008). Facial expressions of emotion. *Handbook of emotions*, New York: Guilford Press, 211-234.

Movellan, J., Bartlett, M. S., Fasel, I., Littlewort, G. F., Susskind, J., & Whitehill, J. (2014). *U.S. Patent Application 14/180,352*.

Naples, M. J. (1979). *Effective Frequency: The Relationship between Frequency and Advertising Effectiveness.* New York: Association of National Advertisers, Inc.

Ohme, R., Matukin, M., & Szczurko, T. (2010). Neurophysiology uncovers secrets of TV commercials. *Der Markt*, *49*(3-4), 133–142. http://doi.org/10.1007/s12642-010-0034-7

Ohme, R., Reykowska, D., Wiener, D., & Choromanska, A. (2009). Analysis of neurophysiological reactions to advertising stimuli by means of EEG and galvanic skin response measures. *Journal of Neuroscience, Psychology, and Economics*, *2*(1), 21–31. http://doi.org/10.1037/a0015462

Ohme, R., Reykowska, D., Wiener, D., & Choromanska, A. (2010). Application of frontal EEG asymmetry to advertising research. *Journal of Economic Psychology*, *31*(5), 785–793. http://doi.org/10.1016/j.joep.2010.03.008

Palda, K. S. (1966). The hypothesis of a hierarchy of effects: A partial evaluation. *Journal of Marketing Research*, 13-24.

Petty, R. ., & Cacioppo, J. . (1986). The elaboration likelihood model of persuasion, *19*, 1–24.

Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement. *Journal of Consumer Research*, *10*(2), 135. http://doi.org/10.1086/208954

Pham, M. T., Geuens, M., & De Pelsmacker, P. (2013). The influence of ad-evoked feelings on brand evaluations: Empirical generalizations from consumer responses to more than 1000 TV commercials. *International Journal of Research in Marketing*, *30*(4), 383-394.

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. http://doi.org/10.2501/S0021849906060041

Puccinelli, N. M., Wilcox, K., & Grewal, D. (2015). Consumers ' Response to Commercials : When the Energy Level in the Commercial Conflicts with the Media Context, *79*(March), 1–18.

Reekum, V., Marije, C., van den Berg, H., & Frijda, N. H. (1999). Cross-modal preference acquisition: evaluative conditioning of pictures by affective olfactory and auditory cues, *13*(6), 831–836. http://doi.org/10.1080/026999399379104

Rossi, F. (2012). Emotional Sophistication: Studies of Facial Expressions in Games.

Rossi, F., Fasel, I., & Sanfey, A. G. (2011). Inscrutable games? Facial expressions predict economic behavior. *BMC Neuroscience, 12*(Suppl 1), P281.

Rucker, D. D., & Petty, R. E. (2004). Emotion specificity and consumer behavior: Anger, sadness, and preference for activity. *Motivation and Emotion*, *28*(1), 3–21. http://doi.org/10.1023/B:MOEM.0000027275.95071.82

Sanfey, A. G. (2007). Social decision-making: insights from game theory and neuroscience. *Science (New York, N.Y.)*, *318*(5850), 598–602. http://doi.org/10.1126/science.1142996

Schaefer, A., Nils, F. F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition & Emotion*, *24*(June 2015), 1153–1172. http://doi.org/10.1080/02699930903274322

Sheldon, A. F., & McDowell, G. R. (1923). *The art of selling*. Sheldon School.

Sikka, K., Dykstra, K., Sathyanarayana, S., Littlewort, G., & Bartlett, M. (2013, December). Multiple kernel learning for emotion recognition in the wild. In *Proceedings of the 15th ACM on International conference on multimodal interaction* (pp. 517-524). ACM.

Stahl, C., & Unkelbach, C. (2009). Evaluative learning with single versus multiple unconditioned stimuli: the role of contingency awareness. *Journal of Experimental Psychology. Animal Behavior Processes*, *35*(2), 286–291. http://doi.org/10.1037/a0013255

Staats, C. K., & Staats, A. W. (1957). Meaning established by classical conditioning. *Journal of experimental psychology*, *54*(1), 74.

Stewart, D. W., Pechmann, C., Ratneshwar, S., Stroud, J., & Bryant, B. (1985). Methodological and theoretical foundations of advertising copytesting: a review. *Current issues and research in advertising*, *8*(2), 1-74.

Strong Jr, E. K. (1925). Theories of selling. *Journal of applied psychology*, *9*(1), 75.

Stuart, E. W., Shimp, T. A., & Engle, R. W. (1987). Classical conditioning of consumer attitudes: Four experiments in an advertising context. *Journal of*

Consumer Research, 334-349.

Stuart, E. W., Shimp, T. A., & Engle, R. W. (2002). Classical conditioning of consumer attitudes. *Consumer Behaviour Analysis.* …. Retrieved from http://scholar.google.com.au/scholar?q=Classical+conditioning+of+consumer+attitudes%3A+Four+experiments+in+an+advertising+context.+Journal&btnG=&hl=en&as_sdt=0%2C5#1

Vakratsas, D., & Ambler, T. (1999). How Advertising Works : What Do We Really Know ? *Journal of Marketing*, *63*(1), 26–43. http://doi.org/10.2307/1251999

Vuoskoski, J. K., & Eerola, T. (2012). Can sad music really make you sad? Indirect measures of affective states induced by music and autobiographical memories. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(3), 204.

Westermann, R., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity o f mood induction procedures : a meta- analysis. *European Journal of Social Psychology*, *26*(January 1995), 557–580. http://doi.org/10.1002/(SICI)1099-0992(199607)26:4<557::AID-EJSP769>3.0.CO;2-4

Westeyn, T., Presti, P., & Starner, T. (2006, October). ActionGSR: A combination galvanic skin response-accelerometer for physiological measurements in active environments. In *Wearable Computers, 2006 10th IEEE International Symposium on* (pp. 129-130). IEEE.

Winkielman, P., Berntson, G. G., & Cacioppo, J. T. (2001). The psychophysiological perspective on the social mind. *Blackwell handbook of social psychology: Intraindividual processes*, 89-108.

Yeh, S., Lin, H., Li, S., & Tsai, M. (2012). Unconscious evaluative conditioning: Attitude change induced by invisible words. Perception ECVP abstract, 41, 210-210.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American psychologist*, *35*(2), 151.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current directions in psychological science*, *10*(6), 224-228.

Zander, M. F., Apaolaza-Ibáñez, V., & Hartmann, P. (2010). Music in Advertising-

Effects on Brand and Endorser Perception. *Advances in Advertising Research (Vol. 1)*, *4*(17), 127–140.

Zielske, H. A. (1959). The remembering and forgetting of advertising. *The Journal of Marketing*, 239-243.

Zucco, G. M. (2012). The acquisition of odour preferences via evaluative olfactory conditioning. *Olfactory Cognition: From Perception and Memory to Environmental Odours and Neuroscience*, *85*, 269.

Appendix

Appendix A Pretest Questionnaires of Neutral Emotional Videos

Table 5 Positive and Negative Affect Schedule (PANAS) for Valence

After watching this video, I feel:

	not at all	*a little*	*moderately*	*quite a bit*	*extremely*
Active					
Afraid					
Alert					
Ashamed					
Attentive					
Determined					
Distressed					
Enthusiastic					
Excited					
Guilty					
Hostile					
Inspired					
Interested					
Irritable					
Jittery					
Nervous					
Proud					
Scared					
Strong					
Upset					

Table 6 Questionnaire for Arousal

	1	2	3	4	5	6	7
	No emotion at all	→					Very intense emotions
While I was watching the film I feel							

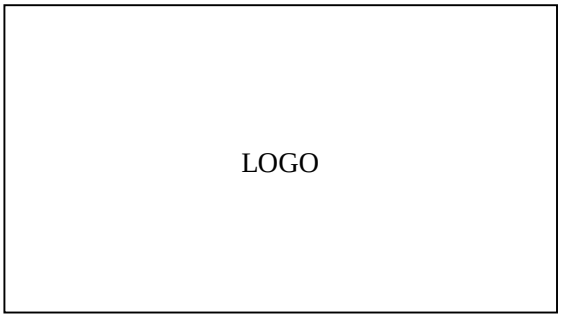
Appendix B Description of Emotional Videos

Table 7 The description of the emotional videos used in the experiment. The descriptions are based on Schaefer et al. (2010).

Movie	Duration (seconds)	Description
Benny Joone	73	Benny plays the fool
Introduction of Bitalino	89	A introduction of different sensors of Bitalino sensors
Blue	40	A man arranges some documents
Introduction of car	71	A introduction of accelerator, brake and clutch
Dead poets society	126	All the students climb on their desks to express their solidarity with their teacher
Hellraiser	90	Stains are growing and transforming to a monster
Indiana Jones and the last crusade	83	A man and a woman go through a sewage line
The silence of the lambs	186	A postmortem examination
Life is beautiful	93	A father and his son talk to the mother by a loud speaker
Lover	43	A girl go through the street
When a man loves a woman	97	Reconciliation between two lovers
Nature scene	56	Different waterfalls
The piano	60	A man cut off his wife's finger
The professional	151	The killer helps the girl to flight
Saving private Ryan	100	Fighting on the beaches
Sea scene	64	Ocean waving
Schindler's list	80	Dead bodies are being carried away
A fish called Wanda	155	A man was found naked

Appendix C Pretest Questionnaire of Logos

Table 8 The pretest questionnaire of the logos evaluation including both emotional valence and arousal questions. The logo was presented before the questions.



Please look at the logo and answer:

The logo makes me feel	**Extremely negative**	**Negative**	**Neutral**	**Positive**	**Extremely positive**
I am familiar with the logo	Totally disagree	Disagree	Maybe	Agree	Totally agree

Appendix D Logos Used in The Experiment

Figure 1 The logo matched to three video clips: Blue, Introduction of cars, and Lover.

Figure 2 The unfamiliar logo which did not match to any movies.

Figure 3 The logo matched to three movie clips: Benny Joone, When A Man Loves A Woman, and A Fish Called Wanda.

Figure 7 The logo matched to three video clips: The Professional, Life Is Beautiful, and Dead Poets Society.

Figure 5 The logo matched to three movie clips: Sea Scene, Natural Scene, and Introduction of Bitalino.

Figure 6 The logo matched to three movie clips: Saving Private Ryan, The Piano, and Schindler's List.

Figure 7 The unfamiliar logo which did not match to any movies.

Figure 6 logo matched to three movie clips: The Slience of The lambs, Indiana Jones and the Last Crusade, and Hellraiser.

Appendix E Facial Expression Raw Data of Each Movie

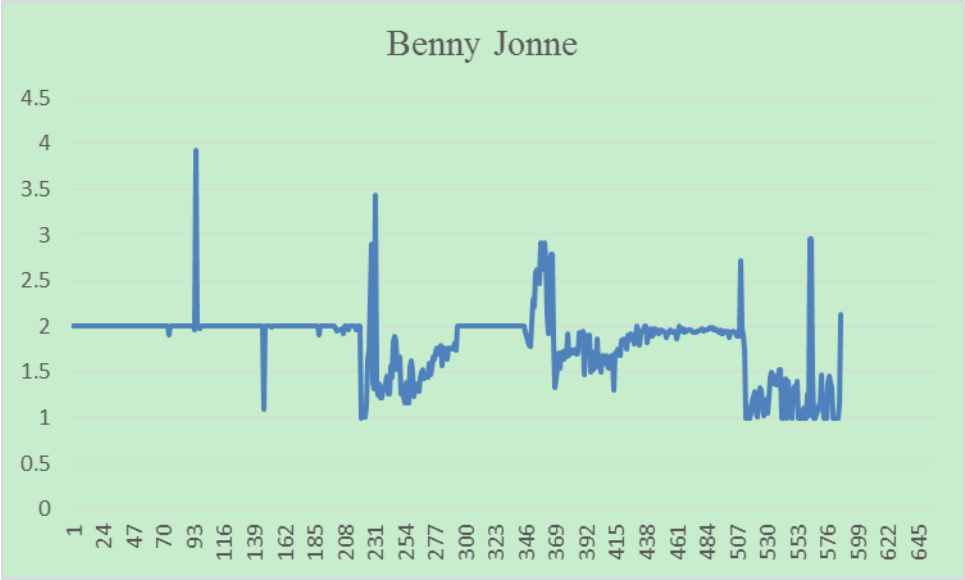


Figure 8 Facial expression raw data of Benny Joone

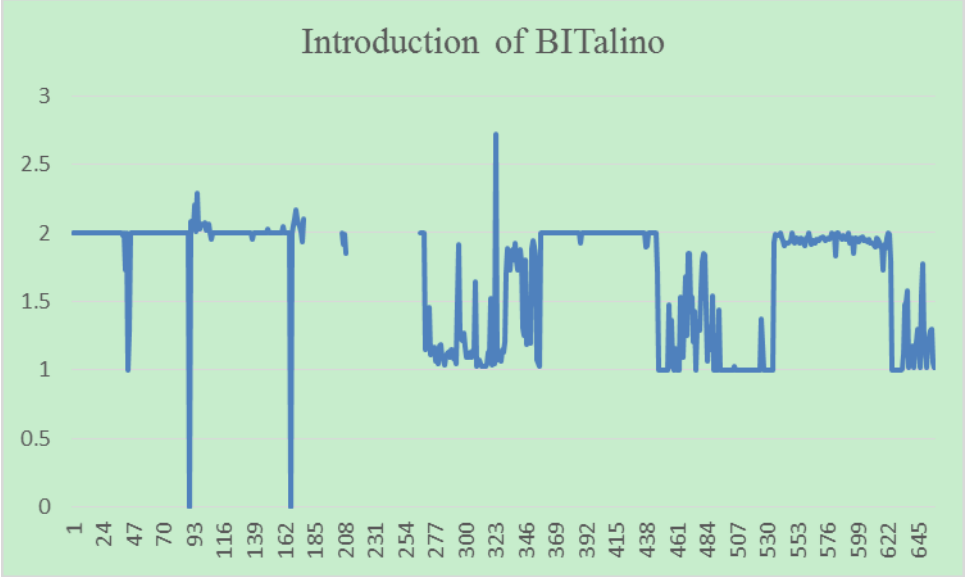


Figure 9 Facial expression raw data of Introduction of BITalino

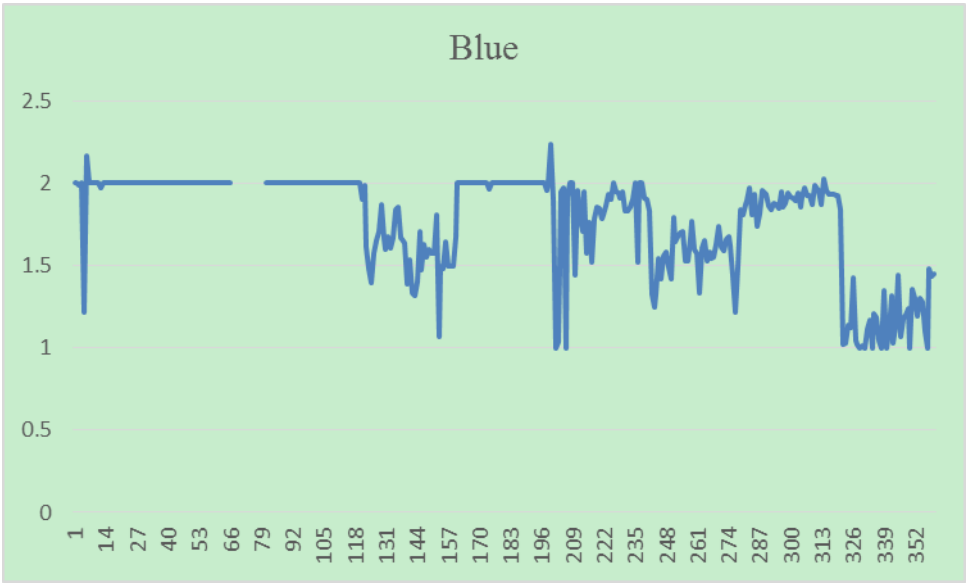


Figure 10 Facial expression raw data of Blue

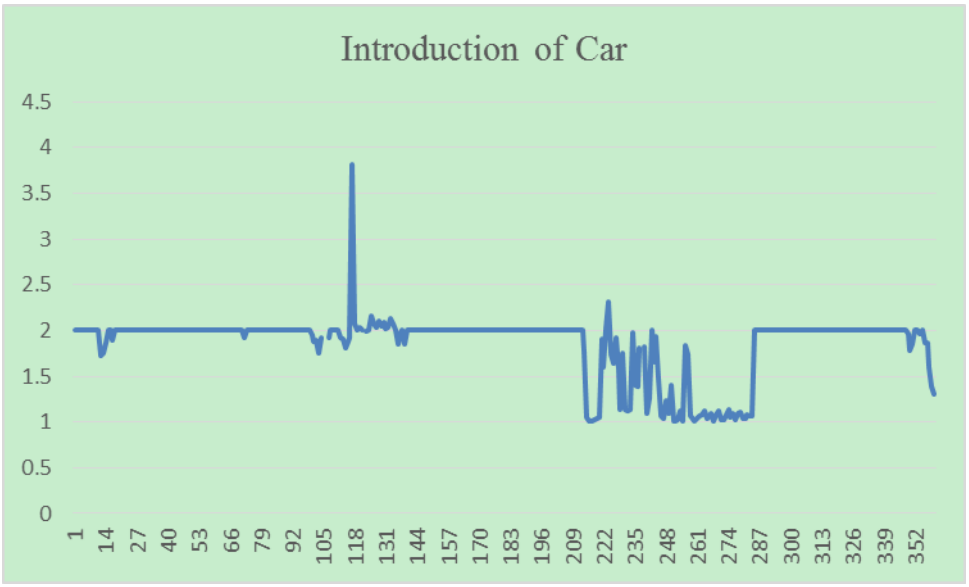


Figure 11 Facial expression raw data of Introduction of car

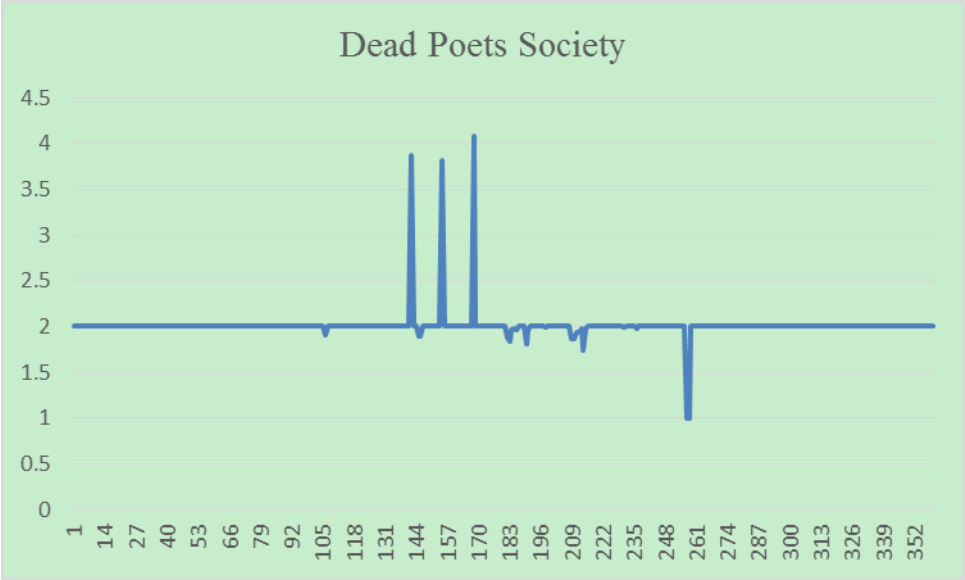


Figure 12 Facial expression raw data of Dead Poets Society

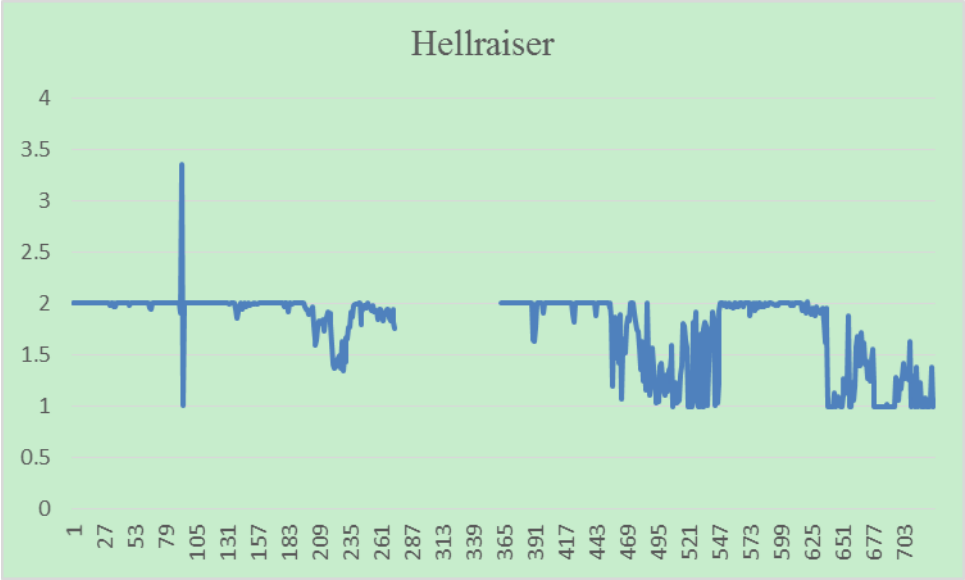


Figure 13 Facial expression raw data of Hellraiser

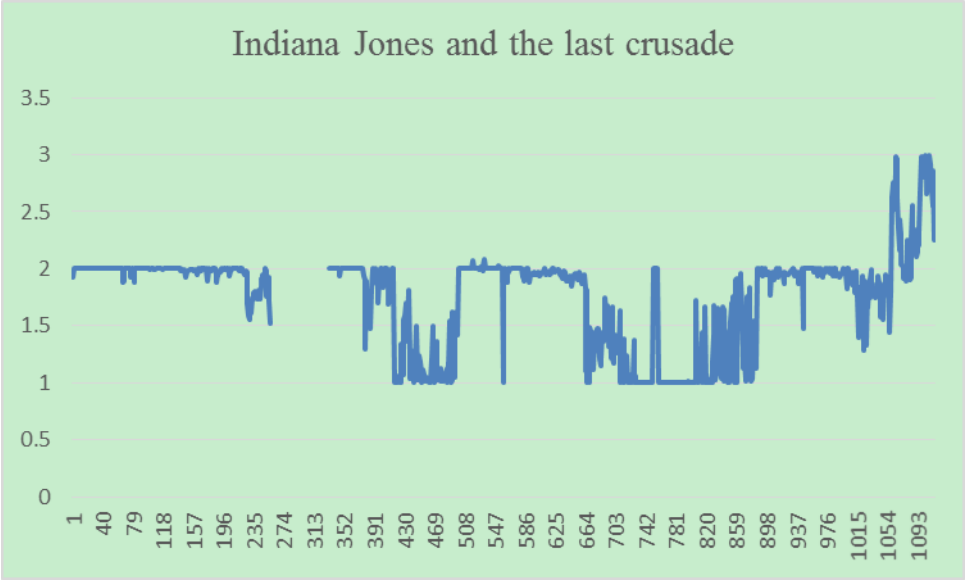


Figure 14 Facial expression raw data of Indiana Jones and the last crusade

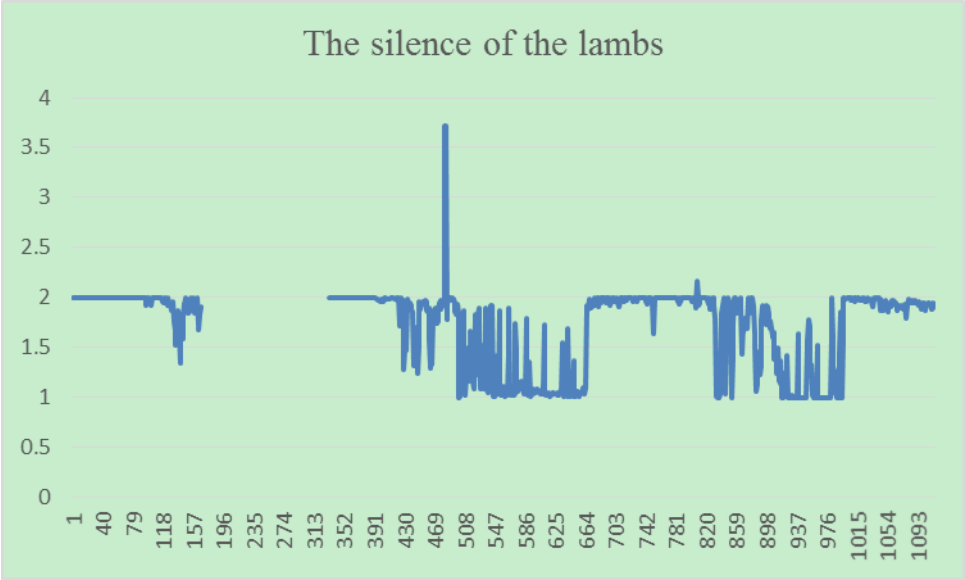


Figure 15 Facial expression raw data of The silence of the lambs

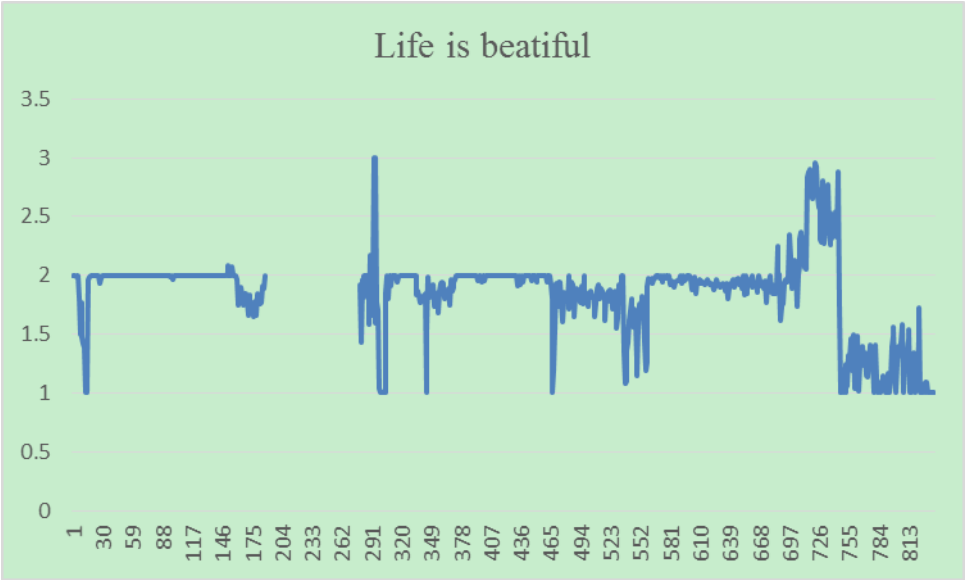


Figure 16 Facial expression raw data of Life is beautiful

Figure 17 Facial expression raw data of Lover

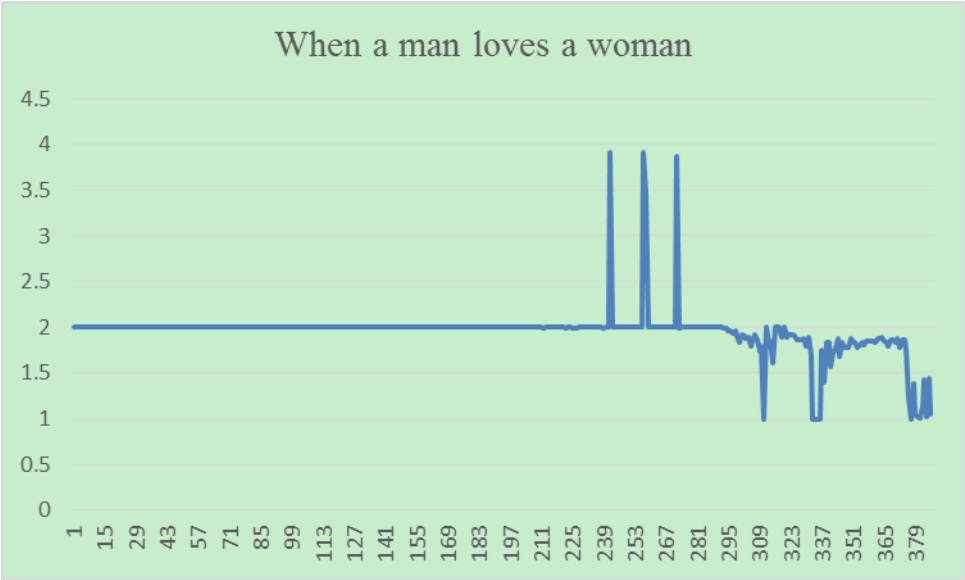


Figure 18 Facial expression raw data of When a man loves a woman

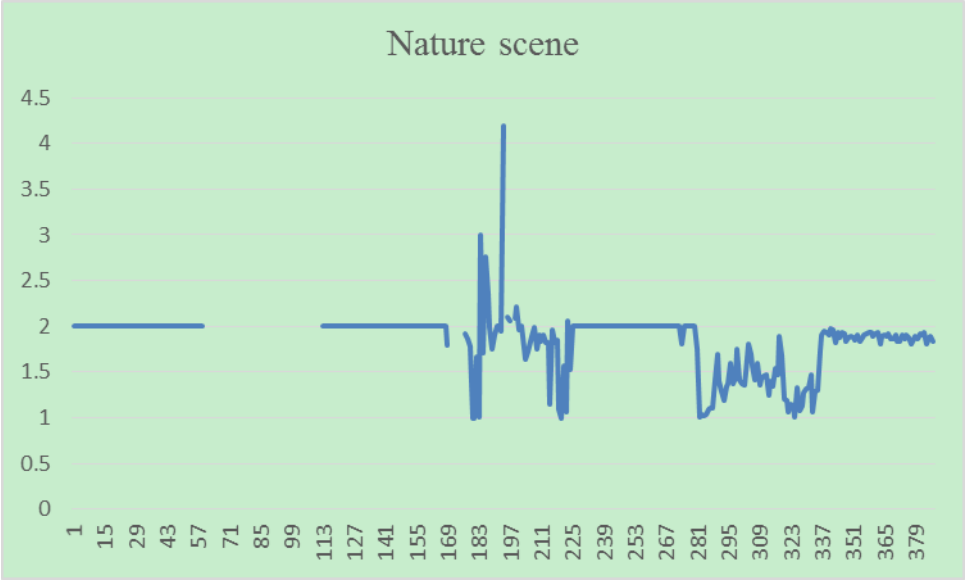


Figure 19 Facial expression raw data of Nature scene

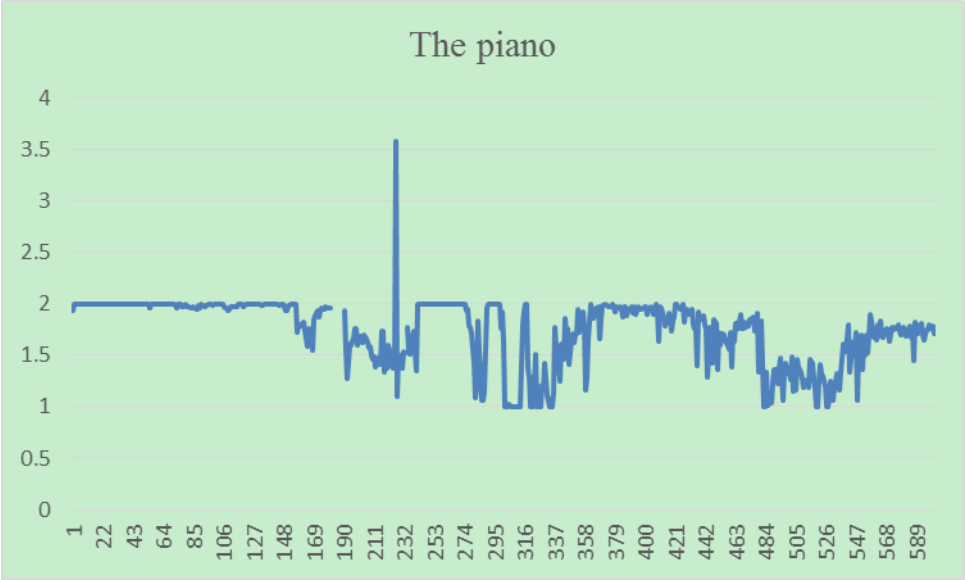


Figure 20 Facial expression raw data of The piano

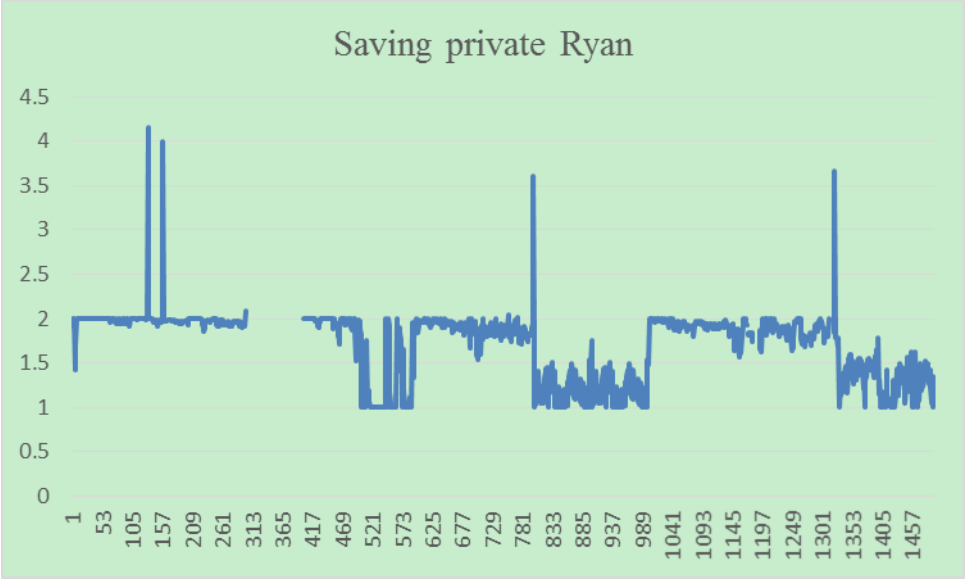


Figure 21 Facial expression raw data of Saving private Ryan

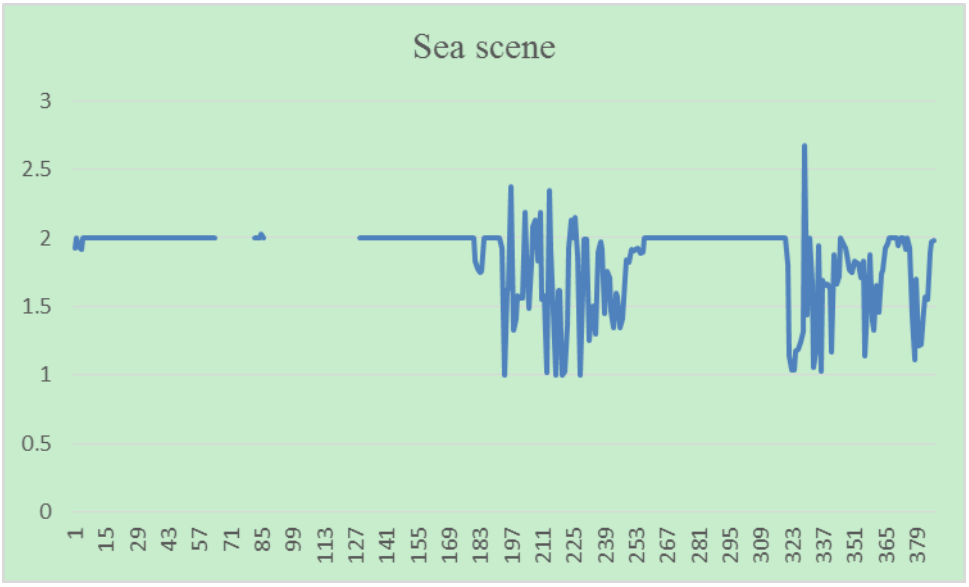


Figure 22 Facial expression raw data of Sea scene

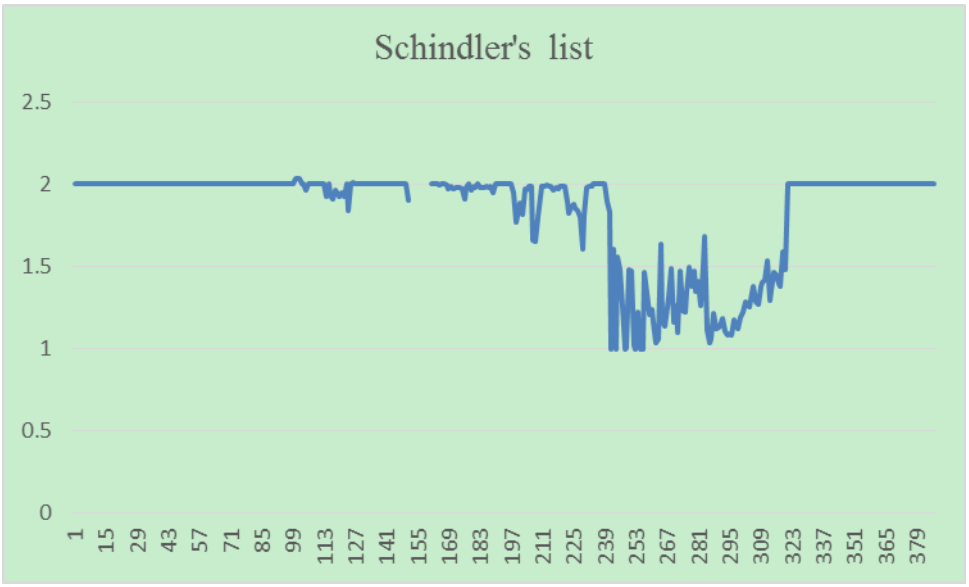


Figure 23 Facial expression raw data of Schindler's list

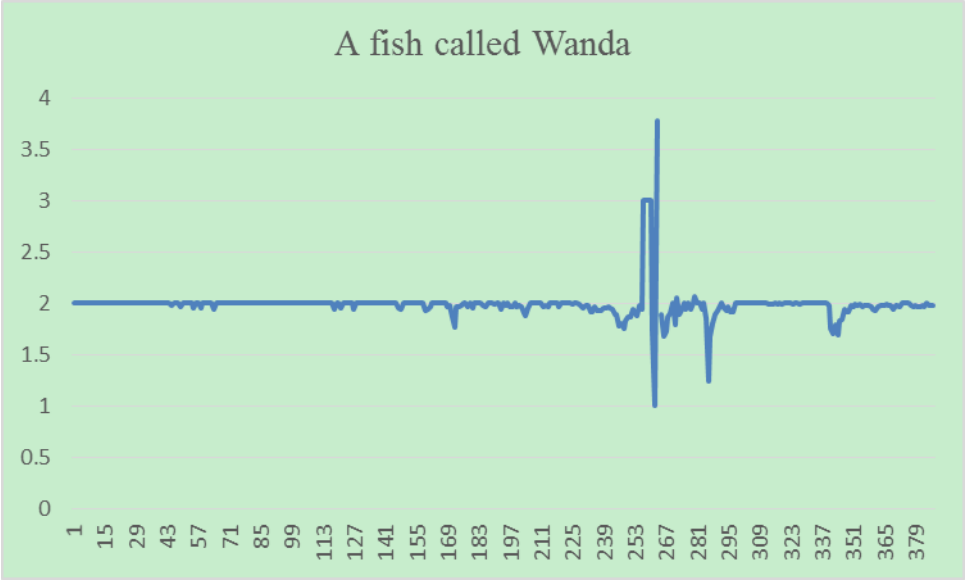


Figure 24 Facial expression raw data of A fish called Wanda