POLITECNICO DI MILANO

Faculty of Industrial Engineering

Master of Science in Management Engineering

Emotion Impact on Ad Effectiveness and Purchase Intent by Facial Reactions and Self-report Measures

Supervisor: Prof. Lucio Lamberti

Supervisor: Ing. Elena Magri

TUGBA AYDOGAN

Matricola: 899344

April/2020

Acknowledgements

I would like to express my endless thanks to Alessandro A. Parla, who always had faith in me and supported me to complete my masters in the most productive way. I will always be thankful for his patience, his faith and his genuine help.

INDEX

List of Figures

List of Tables

ABSTRACT

This thesis investigates the impact of ad-evoked emotions on the important variables of interest in order to measure advertising effectiveness and predict purchase intent: attitude toward the advertisement, attitude toward the brand, customer satisfaction and purchase intention. In this thesis, the review of the progress in the fields of affective computing, advertising and consumer behaviour is elaborated in order to shed light on the association between emotion and ad effectiveness & purchase intent. A combination of self-assessment and facial emotion recognition software tool were employed in laboratory settings. 24 subjects were exposed to the six short video commercials which evoke happy, sad and neutral emotions and asked to fill the questionnaire while their real-time facial responses and survey data were collected. In online environment, a larger group of subjects, 226, watched the six commercials and completed the more detailed self-report.

The results of both methods show that emotion has an important role on ad effectiveness and purchase intent, particularly ‘happiness’. While ad effectiveness can be read real-time via autonomous measures, alas purchase intent require to be supported by other factors which can be measured through self-reports.

EXECUTIVE SUMMARY

My thesis pursues the goal to study the impact of emotion in video stimuli on the effectiveness of advertising. It questions whether happy, sad and neutral emotional stimuli impress the watcher differently in a way that the effectiveness of advertising will be higher, or the buy decision of human being will be affected in a persuasive context. The study focuses on how emotions affect and ultimately support the marketing initiatives.

Reading consumer opinion and understand behaviour patterns have appealed many scholars for a very long time. However, today the attention spent by marketers and advertisers on this field are much higher comparing to decades ago due to the advent of technology. The main motives of incremental increase of interest in affective computing in the fields of marketing and advertising are the exponential increase of human-computer interaction and the feasibility of reading affections real-time.

With the question on mind regarding the emotion impact in marketing and advertising disciplines, an extensive literature review is started investigating what emotion means in different domains such as psychology, computer science, neuromarketing, affective computing, recognition studies, so on so forth. Based on reviews, emotions were not clearly distinguished from moods, attitudes and feelings. However, emotions are referring to the subconscious part of the human brain while for instance attitudes largely belong to cognitive parts. Then, the review continued asking what other factors affect ad effectiveness and purchase decisions. The factors are mainly clustered under the name of tripartite of cognitive, conative and affective states; nonetheless, these factors namely are brand awareness, attitude toward advertisement, attitude toward brand, customer satisfaction, to name a few. Followingly, it is analysed how emotion is detected and what traditional and state-of-the-art methods are used in order to gauge affect accurately in literature. The academic studies focusing on emotion recognition methods group the techniques in implicit and explicit measures. Implicit measure requires a direct communication with the subjects whilst the explicit techniques collect emotion data non-invasively. As the core of this study, in this thesis, an extensive review of face emotion recognition is analysed with the process the method passes through; feature extraction, feature selection, classification. In addition to the face emotion recognition, subjective measures are extensively described as a method to measure emotion traditionally.

Obtaining knowledge from the previous studies, this work concentrated on the practical use of affective state information on prediction of buying behavior and measuring advertising effectiveness.

Thereafter reviewing existing studies in this field and setting research questions, the factors affecting ad effectiveness and purchase intentions are identified in order to design the experimental part of the thesis.

The empirical study of this thesis is conceived in three parts; starting with stimuli selection, continues with experiment design for the laboratory environment and ending with experiment design for the online settings. In the stimuli selection phase, over 60 stimuli candidates were evaluated considering the video characteristics which are semantic, tone of voice, music and general emotion of the video commercial. All videos were chosen with the length of 1-minute and promoting only product, not service. The first selection of stimuli candidates is completed iteratively due not having a consensus on videos in terms of the emotion the commercials evoke. The emotion types of the videos are happiness, sadness, neutral and anger. The latter emotion type was discarded from the stimuli list, for the commercials never fully evoke pure anger and indeed angry commercials involve mostly amusement. Therefore, three dominant basic emotion videos were the stimuli candidates. In the second step of evaluation stimuli, a validation of stimuli selection was supported by a group of subjects who are 12 Italian individuals, speak English and understand English language sufficiently. The subjects were directed a survey including the questions regarding the general perception of emotion revealed by the ads, the emotion of the music the video has, the emotion of tone of voice and emotion in the semantics of the video. The survey included the selected videos which were uploaded to Youtube with a private account with the purpose of following analytics of video watch time. The commercials achieve at least 80% of the consensus by subjects were selected with one exception which is one sad video (HP, printer commercial, 58%). This video was chosen by experts working at the Design and Communication department of Politecnico di Milano because of the reason that other sad emotion video candidate did not perform better than HP video at conveying sadness.

Following to choosing stimuli, the survey items were selected from the existing similar researches. The questions in the surveys were well structured including 5-point Likert scale verbal questions and Self-Assessment Manikin visual questions. The questionnaire was designed intending to collect responses of watchers about their purchase intent, their attitude towards brand and advertisement, their previous experience with the brand if there is any, their attitude change towards the brand. Alongside the affective, cognitive and conative attitude questions, the questionnaire also includes the demographic questions such as age, sex and education level. The questionnaire designed for the laboratory environment is relatively less intense than the questionnaire designed for online settings. In the questionnaire which is intended to spread online the number of questions were higher to better understand the customers' attitudes towards the video commercials remotely.

A group of subjects (18) recruitment were conducted by Pheel Lab at Politecnico di Milano and the other group of subjects (8) attended the experiment voluntarily. The subjects were chosen with the criteria of being Italian native speaker. The subjects were taken one by one, requested their consent to be videotaped and warned not to cover their faces with their hands. The FaceReader developed by Noldus (version 7.0) is used to record the watchers' facial reactions and the questionnaire was conducted using Surveymonkey platform. The extensive other survey conducted online was spread using the platform of Amazon Mechanical Turk and filled by 226 test participants. In both surveys, the gender, sex and educational background were purposely balanced.

The laboratory experiments lasted three days. The experiments started with the exposure of 1-minute gray screen and ended with the same gray screen in order to record the neutral facial expressions of the participants. The sequence of video exposure was randomized in order to avoid ordering effects. Each video was followed by a questionnaire designed including the question items regarding the promoted product and brand. The survey spread online was filled by 350 subjects from numerous parts of the world. On the Amazon's Mturk platform, the subjects were supposed to paste a code at the end of the survey proving that they complete the task in exchange of some allowance. Also, in this phase, the Youtube account helped me to track whether the subjects watched the video or not.

Subsequent to completing of experiments in both environments, the data collected were subjected to data cleaning and reliability phases. Before that, two subjects attended lab experiment were needed to be discarded because of having too much missing data. The facial data was recalibrated considering the incremental changes in subjects' faces after the exposure of the video. The data collected online was also subject to discard a group of subjects who complete the task less than 12 minutes and the number of subjects were reduced to 226. The following analysis of data phase was completed using Microsoft Excel and SPSS Statistical software (version 26). The reliability analysis was made using Principal Component Analysis and the high number of questions were separated into different dimensions.

In the analysis of experiments, first of all, the manipulation was somewhat successful given the fact that FaceReader measures correlated sufficiently high with participants' self-reports for happy videos. Alas, sad and neutral videos were not significantly distinguished using facial values of sadness and neutral. The nonparametric tests were chosen to run data (Paerson correlation). The reason of facing this result could be that the stimuli evoke other emotions. The valence values of facial reactions are relatively better at identifying affective states of both cognitive and behavioural measures. Because valence values are calibrated considering all positive and negative emotions. Another implication from the study is that the emotional communication has a greater impact on viewer. The advertising effectiveness is highest when the exposure of the stimuli evokes happiness. On the contrary, the neutral videos have negative impact on viewer based on survey responses which delivered high disliking scores for the neutral ads. In addition to this, the ad effectiveness is dependent on attitude towards brand more than affective states. When the subjects have already a positive attitude towards the brand promoted in the ad, they are more prone to like the commercial than the ones who are, for instance, neutral. Hence, here it is seen that the predictive power of attitude towards brand (52% of explaining variance in ad effectiveness) is higher than the impact of emotions (36% of explaining variance in ad effectiveness). Moreover, this study indicates that purchase intent is also affected by the impact of emotions. The ads evoked happiness has blatantly greater impact on customer in a convincing context. However, in order to predict purchase intent, it is needed to gather other information such as customer satisfaction. Because, this study shows that people who had an unpleasant experience before do not opt to change their mind to make a repurchase in future regardless of the emotion evoked by the commercial they are exposed to. Nonetheless, this does not necessarily mean that gauging emotions are secondary. Because emotion does play a certain role to convince people for making a buy where there is not a very unpleasant experience.

Very importantly, in this thesis, it is proven that the FaceReader is a tool that can measure ad effectiveness real-time. The association between the frequency and intensity of facial reactions of happiness and liking the commercial is statistically significant. Therefore, measuring ad effectiveness is feasible with non-invasive tools. On the other hand, prediction of purchase intent with the FaceReader did not yield significant results. The reason for this can be lack of sufficient data or stimuli selection. The stimuli chosen for this study promote products like computer and car which are not the products easily decided to be bought by the impact of one short advertisement.

To the best of my knowledge, sad and neutral emotions were not analysed as predictor factors of ad effectiveness and purchase intent and the study on these emotions show the potential importance of further studies focused on other types of emotions evoked by the commercials.

The study has some certain limitations such as selection stimuli in a more detailed way. For instance, engineering on other characteristics of video could lead a clearer understanding on what makes the dominant emotion of stimuli dominant and to what extent the other factors affect the general perception of emotion in the video (e.g. semantic analysis). Furthermore, this study conducted a self-report to receive the responses of customers regarding conative and cognitive attitudes. The self-reports pose some disadvantages; nevertheless, specific to this study, it is seen that answering questions caused fatigue in subjects. This can be avoided using other explicit measures to fathom these aspects.

CHAPTER 1: INTRODUCTION

Throughout the history, communication has been a vital means in human life in a variety of ways. Emotions, feelings, intentions and human behavior contributed to a large extent to convey information one another. However, today, affects are of utmost importance not only for human-human communication but play a crucial role to interacting with computers too.

By reason of this impact in life, affective science represents an intriguing topic to study for many scholars within various academic fields, ranging from psychology, neuroscience, and cognitive science but also in many interdisciplinary domains like affective computing, computer vision. In different disciplines, emotions have been investigated differently. In the instance of computer science, the scholars contemplated over the affective states of human interactions with computers. On the other hand, marketing researchers have attempted to probe human behavior via emotion recognition to better understand pleasurable and desirable consumer experiences (Lee and Kwon, 2010). In particular, in marketing and consumer behavior, emotions have been perceived in a way that they enable marketers to “read” human behavior as they convey significant information about consumers’ opinions. Reaching the information about end users through ultimate tools seems an effective manner to increase the success of delivering persuasive communication and to achieve measuring and predicting the consumer’s current and future opinions (Lewinski, Fransen & Tan, 2014).

However, studies in affective computing also show the possibility to augment the quality of interaction via gauging emotions. As one example of emotion recognition applications, it can be given the study of Park and Gates (2009) proving that real-time proper customer satisfaction measurement is applicable near real-time which makes it feasible to control the conversation immediately and build a desirable relationship accordingly. The use of affect detection can be varied depending on different study fields. In addition, the affective computing confronts various number of challenges including objective annotation of affects, cognitive bias and social desirability, uncontrollable environment of subject, just to name a few.

The core objective of this study is to focus on the impact of general perception of emotion on advertisement effectiveness and on purchase intent. I will hereby conduct a literature review of the importance of emotion in marketing and advertising fields, what methods and techniques were employed to detect affect in experimental studies published in journals during recent periods. Subsequently, it will be introduced the empirical study conducted using the FaceReader software by Noldus for a part of this thesis. Alongside measuring the measuring appropriateness of this tool to detect the affect from facial reactions, the self-assessment methods are employed to gauge the effect of perceived emotion on buying intention and ad effectiveness.

This thesis is planned in six chapters. In this section, it is covered the importance of this study area and the objective of this thesis. The rest of the study is organised as follows. The second chapter focuses on the motivation of this study, the research questions that intended to be answered by this thesis and the methodology designed. The third chapter first introduces the emotion from the point of view of marketing and affective computing world with such details like the terminology used for emotions, the contributions of emotion in marketing and advertising fields, what to code as emotion types and how to code emotions objectively so that human-centred computer systems can grasp the affect unobtrusively and interact with humans. Besides, , it will be detailed the components of the purchase intent and ad effectiveness. In the same chapter, I will present the emotion analysis of facial expressions and verbal self-report in detail and a brief overview of other emotion recognition methods in affective computing. The chapter ends with several research studies which also focus on the emotion detection so as to measure purchase intent and advertising effectiveness.

The fourth chapter provides information regarding experimental design which is composed of three parts: stimuli selection, survey design and experiment design. This chapter first details the criteria of choosing the adequate stimuli candidates and validation of selected stimuli. Subsequently, questionnaire items, instruments (survey and FaceReader) and experiment environments (laboratory and online settings) are explained.

The analysis of facial expressions and survey responds as an approach to predict ad effectiveness and purchase intent will be presented in the fifth chapter of the work.

Five main chapters of this thesis are to be pursued by the general discussion of findings, limitations of the study and further research suggestions in the last chapter.

CHAPTER 2: RESEARCH OBJECTIVES AND METHODOLOGIES

The motivation of this research comes from the importance of measuring affective states of individuals and also the inquiry of predicting purchase intent and advertising effectiveness employing emotion measuring tools.

The desire of recognizing affective states and understand decision making mechanism date back to a long time ago. In past, as a traditional method, the consumers were asked their opinions and feelings towards various stimuli using surveys. However, with the advent of technology, last couple of decades, measuring emotion real-time has become feasible and the interest in this field escalated by academicians and business professionals.

Benefiting from the previous studies, this thesis aims to find out the practical use of affective state information on prediction of buying behavior and advertising effectiveness. Based on the problems stumbled upon, the objective set of the research is identified, and the research objectives are as follows:

i. To gauge the contribution of factors that predict the purchase intent and advertising effectiveness

ii. To validate the usability of FaceReader tool to read the facial emotion

iii. To validate the usability of FaceReader to measure the ad effectiveness and buying behavior

iv. To assess the relationships among and between affective, cognitive processing and ad effectiveness

v. To assess the relationships among and between affective and behavioral intentions

Setting objectives, the research questions are formulated in the following fashion:

Research Question 1: Do emotion type of the video stimuli, ad liking, rewatchability and attitude towards the brand predict the advertising effectiveness?

Research Question 2: Do emotion type of the video advertisement, customer satisfaction, pre-purchase intent predict the post-purchase intent?

Research Question 3: Do facial expressions predict the ad effectiveness?

Research Question 4: Do facial expressions predict the post-purchase intent?

The reviews, empirical and conceptual studies conducted in affective computing and consumer behaviour fields are of large range. In the extensive review of Poels & Dewitte (2006), what emotion is in marketing, how emotional state is measured and methods of emotion recognition, impact of emotion on other predictors of ad effectiveness are explained in detail from the perspective of the advertising field. However, in order to deepen in consumer behaviour and emotion recognition methods separately and jointly, I analysed the articles in two ways: one with the search keywords to comprehend the feasibility of automatic emotion reading in computer science, the other with search terms to reach social studies conducted in this field . Although for a detailed research I have used many search words using the search engines Scopus, Sciencedirect and Google Scholar, I hereby list only the most relevant keywords used for this research in table X. The articles and conference papers published in last two decades (from 2000 to 2020) were considered only.

Affective Computing Science	Consumer Behaviour & Marketing
• Facial emotion recognition	• Cognitive, conative and affective attitudes
• Facereader	• Market research; advertising; marketing
• Facial coding AND affective studies	• Attitude towards brand AND emotion
• Automatic facial image analysis	• Attitude towards advertisement
• Human-computer interaction	• Purchase intent AND emotion
• Physiological measures	• Satisfaction measurement
• Nonverbal behaviour	• Cognitive emotions
• Facial expressions of emotion	• Predicting user emotion
• Automatic detection of emotions	• Consumer behaviour
• Valence-arousal-dominance	• Role of emotions in marketing
• Affect analysis of facial responses	• Emotion impact in advertising
• Feature extraction AND classification AND face emotion recognition	• Neuromarketing
	• Emotion measurement in marketing
• Human-centred computing	• Affective states of consumer

Table 1: The keywords used for literature review part of the thesis

A mixed method has been designed in order to address the research questions. The data collected for this research are both primary data and secondary data. Secondary data provides essential academic support to the thesis which is collected from published or unpublished sources and played a crucial role for designing surveys and experiments. The methods of primary data collection are broadly autonomic measures and quantitative measures. The autonomic measures are the face emotion recognition software values and the quantitative data is visual and verbal self-reports. The FaceReader and survey instruments chosen in order to answer the research questions are based on the studies in similar fields that pursue the emotion impact on purchase intent and ad effectiveness (McDuff et. al, 2015; McDuff et. al,2012; Soleymani & McDuff, 2017; Lewinski et. al, 2014; Hamelin et. al, 2017; Li, Walters, Packer & Scott, 2017). The experiments and questionnaires were conducted in both real-world setting and laboratory environments. In the literature, both methods are preferred for different reasons and avoided for some limitations which are explained in detail in the following sections. The subjects were recruited by Politecnico di Milano Pheel Lab for the experiments and through Amazon Mechanical Turk platform for the web-based surveys. The laboratory environment experiments were performed with only Italian individuals in order to avoid cultural differences in the facial values (Anagnostopoulos et. al, 2012). The online setting survey study was completed by participants from various countries. For both environments, the gender and age distribution of the subjects were balanced due to the considerable effect on the perception (Loui et. al, 2013). The stimuli chosen for this thesis are six video commercials that evoke three basic emotions: happiness, sadness and neutral. From the article of Horvat et. al (2015) inferred that the video stimuli have a greater influence and induce more affective reaction than image stimuli. The video stimuli selection is validated by the help of experts and using a pilot survey study. This method of annotation of stimuli is common in literature (Chen et. al, 2018; Lewinski et. al, 2014; Hamelin et. al, 2017). The questionnaire is designed in a way to receive the affective, cognitive and conative attitudes of participants. The questions are conceived to measure the general emotion of video, purchase intent, the attitude towards brand and advertisement. The well-structured survey includes close ended 5-point Likert scale questions. In the visual self-report, the participants were shown the dimensions of emotions employing Self-Assessment Manikin instrument which bares less burden for the subjects to understand the dimensions (Morris et. al,2002). Following to the experiment part, the data was processed using Add-ins for Microsoft Excel, SPSS statistical software(version 26) and FaceReader developed by Noldus (version 7.0). The reliability of the questions groups that are designed to measure the purchase intent, attitude towards ad, the like was measured first computing PCA to fathom the components produced, afterwards ran Cronbach’s alpha in order to see the consistency of the items. The nonparametric tests were run to process facial recording values. Spearman rank order correlation and Paerson correlation was computed in order to assess the degree of association between facial emotional states and self-reported measures of ad effectiveness and purchase intent predictors. Friedman test was employed for the assessment of usability of FaceReader to gauge the emotion from face reactions. The test chosen for difference between groups was Chi square test followed by post-hoc and Bonferroni correction test.

CHAPTER 3: LITERATURE REVIEW

To go beyond the postulation of importance of affections, the affective computing is reviewed from the studies conducted in social science and computer science. In the following sections, first it will be analyzed the meaning of emotion in marketing, advertising and consumer behavior disciplines, the worth of detecting emotion of individuals for the same disciplines and what the other components to that emotion is related to, the like. In the second part of the review analysis, it will be detailed facial emotion recognition and self-report methods and then followingly presented other autonomic methods to gauge the emotion.

3.1 Emotion

The importance of emotion has been accepted by a vast majority of researchers from many disciplines and sub-disciplines. The scholars from various research fields built many studies over this topic with numerous perspectives. Whilst sociologists, as an instance, overlook the emotion as an important metric to draw social interactions and patterns, communication scientists analyze affections to better understand the perspectives of managers, employees and consumers. In criminology, the emotion reading is of great potential to detect danger beforehand (Barreto, 2017). However, in particular, in marketing, advertising and consumer behavior fields, the importance of emotion is always emphasized, and the inquiry of reading emotion is of great value due to its potential (Poels & Dewitte,2006). According to the review of Barreto (2017), the marketers approach the role of emotion from both consumer and brands' perspectives. In marketing field, the real propensity is to predict the human's next action towards a collusion which means, in this case, an engagement between two parties.

Some marketers work on emotional situation in the moment the experience or product is consumed. In the study of Chen et. al (2018), the authors examine the emotion induced during the consumption of product and compare the methods of measuring the emotion during the product consumption. They concluded that emotion does have an impact to choose the product however the method of measuring the affective state influences the performance of measurement. Another group of marketers focus on emotional reaction and satisfaction associations. Different studies investigated the emotion between loyalty, shopping behaviors, etc. (Yu & Dean, 2001; Zhang, Cheung, Lee; 2014). Advertising academicians and business professionals never ignore the affection impact in the advertising process (Poels & Dewitte, 2006). There are a number of studies testing the impact of emotion on attitude towards advertisement, brand and purchase intent (Derbaix, 1995, Lewinski et. al, 2014, McDuff et. al, 2015, McDuff & Soleymani, 2017; Hamelin et. al, 2017; Poels and Dewitte, 2006; Teixeira et. al, 2012) Many studies in this field show that the ad effectiveness is associated with the emotions of the stimuli and purchase intent requires a more exhaustive work. Consumer behavior is a very close discipline to marketing and advertising in terms of giving importance to the affective science. In the review of Barreto (2017) is seen that people react more positively towards the product when there is a smiling face on the package. About the role of emotion in consumer behavior, LeBlanc et. al(2012) states that emotion has a very strong correlation with attention. Hamelin et. al(2017) gives the example of people pay more attention on negative stimuli comparing to any other emotional stimuli.

On the other hand, the emotions are subjective feelings and therefore are subjected to interpretations. Reading emotion requires the preexisting knowledge of psychology science in order to comprehend what the emotion is and how to categorize and judge in objective manners and also affective computing knowledge regarding the questions that are what to code as emotion and how to code the emotion. In computer science, the affective states of individuals are perceived in the same way of a human-being does (Butler & Lewinski, 2014). According to the study of Butler and Lewinski (2014), the human-being mostly recognizes the emotion on the face at 60 to 80% accurately. Barreto (2017) also express the difference of interpretation accuracy between humans and computers. The author states the computer can outperform the human observers at detecting emotion and reasons this difference in a way that humans may not perceive the slight changes in face whereas computer realizes even minor movements. There are plenty of techniques to read the emotion in literature: face emotion recognition, speech emotion recognition, self-reports, heart rate, skin conductance, eye tracking, so on. The methods have benefits and limitations comparing one another. Nonetheless, in this study, it will be detailed the method of facial emotion recognition, self-report and the studies conducted these techniques in order to figure out emotion impact on advertising effectiveness and purchase intent.

3.1.1. Emotion in Marketing

The affections in marketing and consumer behaviour are used interchangeably with moods, action tendencies, attitudes, subjective feelings, intentions, thoughts, etc. (Bagozzi et. al, 1999; Cohn, 2006). As Cohn states in his article (2006), although the emotions are of different components in terms of terminology, the pursuit of reading emotions comes from the inquiry regarding analysing human behaviour patterns. On the other hand, Bagozzi et. al (1999) conducted an exhaustive study regarding the difference between these terms used in marketing field. They take the emotion as central to the actions of individuals. However, moods, attitudes and others result from the existence of different stimuli. For example, the authors describe that attitudes are evaluative judgments, not affective states. In the article of Taylor (2000), it is asserted that having little consistency on terminology in this field poses a problem and limits the increase of knowledge in theory. In the literature review of this thesis, the studies presented overlook the emotional state as a feeling state of individuals not cognitive responses when they refer to emotion.

As Bagozzi et. al (1999) claimed traditionally the focus point of the role of emotions in marketing was mostly on salespeople and consumer relationship (Kidwell et. al, 2011; Verbeke & Bagozzi, 1998). The authors perceive the emotion more than the relationship of salespeople and consumers. Taylor (2000) states the concern in emotion should enhance. He wrote in his article that managers are focused on 'facts' more than power of emotion although emotion is one of the most significant tools to obtain 'facts' about consumers. Bogazzi et. al (1999) indicated the role of emotion in marketing as an indicator, mediator, moderator of end user responses.

The affections have a direct influence on the decision mechanism of individuals (Small and Verrochi, 2009; Poels & Dewitte, 2006). From the example presented in Taylor's paper (2000), it is scientifically proven that in cases in which the agents are expected to make the most rational decision (e.g. a manager level individual that choose a luxury car), the consumers behave emotionally. In the experiments conducted by Edell & Burke (1987), it is seen that the end user makes decision about picking the most appealing car among many, however not the most functional one. This behaviour can also be explained by the scientific argument which is provided in the paper of Walla et. al (2011) that human brain stops functional specialisation when emotional and cognitive attitudes jointly affect the thought or behaviour of individual.

Emotions, as Bagozzi et. al (1999) stated are the key players in goal setting processing. Furthermore, recognizing the emotional state of the customers is a way to understand his reactions due to the role of emotion mediating responses towards stimuli. From the example given by Taylor (2000), it is implied that consumers who feel threatened by stimuli in previous experiences feel closer to the products providing protection. In salesperson-consumer relationship, customer seeking protection must be welcome with more protective, safe and secure strategic move. As

Plutchik (1980) provides a circular model (fig. 3-1), the strategies to approach the end user can be modelled using this structure. For instance, in order to plan the relationship with the customer at the very beginning phase, the decision to how to move towards customer (e.g. a quick sale or more intense relationship) can be modelled depending on how customers feel. The ones feel stressed would be approached on the opposite side of the circle, with more calming effect. Nonetheless, as published in the article of (Kellerman, 1980), customers opt to develop emotional defence mechanism and understanding the affective states of customer is of greater importance in this regard.

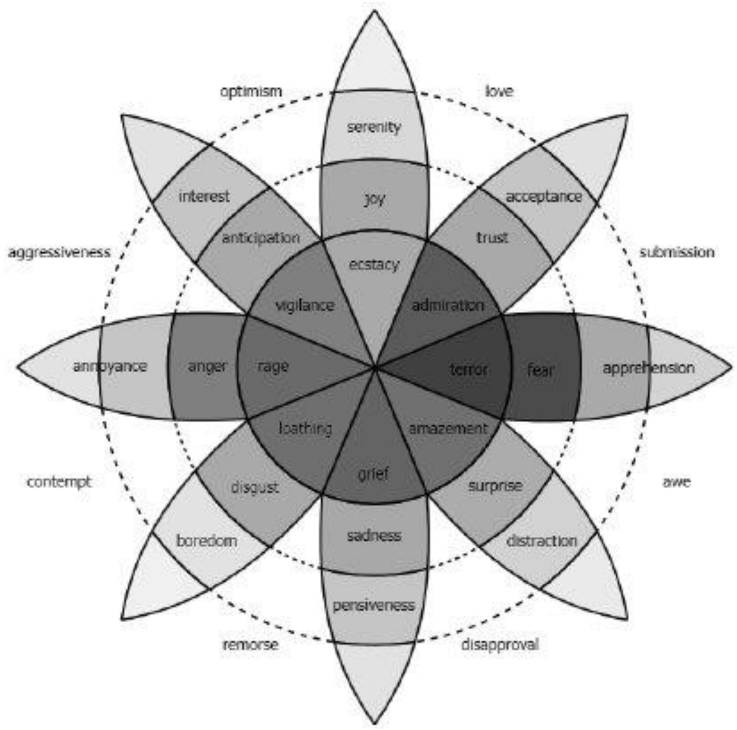


Figure 3 1 : Wheel of emotions of Plutchik (Plutchik, 1980)

As Bagozzi et. al (1999) declare affections are particularly important in marketing to interpret both consumer opinion and managers. The use of information obtained through emotion recognition methods could support marketing activities (e.g. market research, monitoring marketing campaigns) in plenty of ways (Eyben et. al, 2017). From the article of Taylor (2000), it is seen that

Chance (1980) conducts an empirical study showing the difficulty of segmenting patients by therapists. While some patients require an intense relationship and rather hard to communicate comparing to others, the other patients are more open to get help by counsellor. To deal with both types of patients effectively, the strategic moves could be planned differently. The same logic as Taylor (2000) explains is valid between advertisers, marketers and end users. In this sense, emotions support marketing activities from the managerial perspective.

More importantly, as asserted before, emotion is a marker of advertising success (Bagozzi et. al, 1999). Providing scalable and automated are especially important in this field. The details of this topic will be given in the next part.

3.1.2. Emotion in Advertising

With the advent of technology and exponentially increasing intensity of using social media, the dynamics of advertising have been changed dramatically in plenty ways. As reported in NY Times (2015), from 1985 to 2016 the number of advertisements people stumble upon in their daily life has increased from 2000 to 5000. In the article of McDuff et. al (2015), the authors highlighted the growing interest in video pointing out that the time spent for video watching online has increased three times between 2012 and 2013. Realizing this difference in consumer life, the companies changed their advertising strategies and invested more in digital content. The impact of the ads is always aimed to be explained with more concrete and reliable predictors. However, the measuring advertising effectiveness is not quite an easy task to explain with quantitative terms (Corvi and Bonera, 2010). Traditionally, advertisers and academicians measure the behavior and attitude of customers employing self-assessments. Nonetheless, measuring the effect of advertising and predicting the factors influencing customer response are also subjected to the change and be more dynamic.

The components in order to be measured for ad effectiveness are namely attitude towards brand (Haley, 1990), attitude towards ad (Lewinski et. al, 2014), brand familiarity, ad familiarity, purchase intention, recall (Poels and Dewitte, 2006) , brand interest, brand awareness (Sharifi, 2014) and particularly emotion. In literature, the predictors of ad effectiveness and purchase intent are clustered as the tripartite or trilogy: affective, cognitive and conative processing. This tri-partite representation of attitudes are mostly evaluative measures rather than emotion (Andrade et. al, 2008). The cognitive attitude is composed of attitude toward ad and brand, in other words knowledge and belief toward an object. Actually, the attitude toward brand and ad is the belief phase of attitude.

Conative attitude refers to the behavioural actions of consumers. Performing an action, having an intention and behavioural tendency by an individual towards an object are associated with the attitude of conation. Since conative processing is perceived as willingness to perform a given behaviour, the degree of its association with affective and cognitive attitudes remained as a mystery for scholars and business professionals. Several methods have been applied to discover the correlation between these approaches. For instance, Morris et. al (2002) studied on the relationship among predictors using the trilogy and the conceptual model presented by the study is as in figure 3-2.

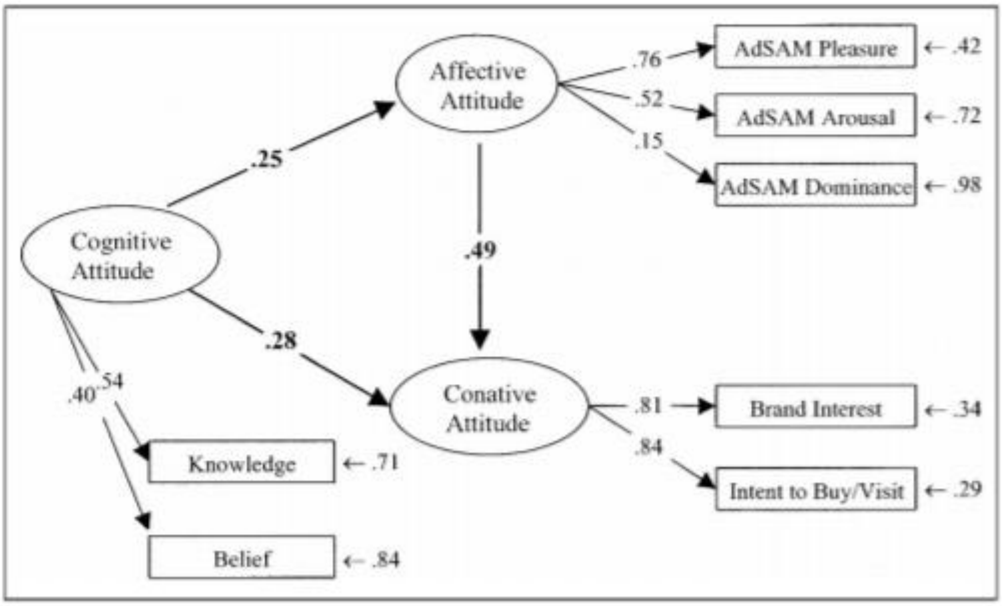


Figure 3 2: SEM Path DIagram showing Affective, Attitude and Conative Attitudes (Morris et. al, 2002)

Their study assessed the extent to which the cognitive, affective and conative attitudes influence each other and how accurate they measure purchase intent. In the study, it is concluded that affective attitudes dominate cognitive attitudes to measure the conative attitude. To put it simple, valence and arousal dimensions of emotion explained that emotion has a greater effect on behavior than thoughts. The findings of this research are contradicted with the results of the study conducted by Li et. al (2017) According to Li et. al, 2017, arousal was not as much good predictor as valence. In their study, the authors proposed a framework to assess the relationship the emotion evoked by the tourism ads and the effectiveness of advertising on customers. The framework is figured in the Fig. 3-3:

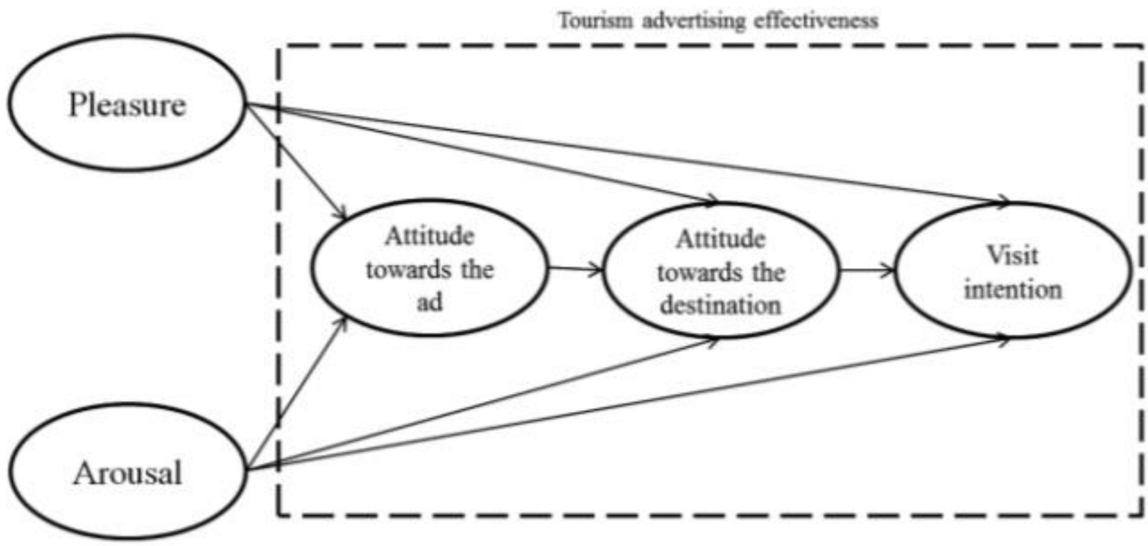


Figure 3 3: The conceptual framework for ad-evoked emotional responses and advertising effectiveness (Li et. al, 2017)

However, in order to compute ad effectiveness and purchase intent, ignoring the importance of emotion and concentrating only cognitive process could make marketers and advertisers end up not understanding the consumer behaviours at all (Allen, Machleit, and Kleine, 1992; as cited in Morris et. al, 2002).

Apart from the studies concentrated on the trilogy of attitudes, in literature it is found other researches focusing on the emotion impact on ad effectiveness and purchase intent employing facial emotion recognition techniques. These studies measure attitude towards brand and ad, brand interest, recall, customer satisfaction, so on to compute their relationship with emotion and ad effectiveness.

<u>Attitude towards ad:</u>

Attitude towards ad is of special importance to evaluate ad effectiveness (Haley, 1990). 'Likability' of the advertisement is a common way to fathom the attitude towards ad (Haley, 1990; Morris et. al, 2002; Lewinski et. al, 2014). The liking denotes the elements such as enjoying and valence of the ad. In the articles measuring 'liking' is captured through explicitly asking the participants how they feel towards the ad (Teixeira et. al, 2013; Li et. al, 2017; McDuff et. al, 2015; Lewinski et. al, 2014). The examples of explicit measuring as follows:

- "How much did you LIKE the AD that you just watched?". (McDuff et. al, 2015)
- "I like/dislike this ad very much." (From 0 to 7 – likert scale) (Debaix, 1995)
- "Did you like the video?" (McDuff et. al, 2012)

In the research of McDuff et. al (2015), the likeability of the commercial is compared using the tools facial coding and surveys. The researchers concluded their study asserting that liking of the ad is possible to measure with the increase of smile and intensity of facial reaction toward the stimuli. This way, the ad effectiveness can be gauged not needing the explicitly intervening the experience of the viewers. Still, as Morris et. al (2002) stated, ad effectiveness can be measured through likeability of the ad, but for more insightful attitudinal information the expansion of quest is needed.

The relationship of attitude toward ad with affective state does exist and emotion has an impact on attitude towards ad (Poels & Dewitte, 2006). It is proven using methods of visual self-report (Morris, 2002), the autonomic facial coding (Lewinski et. al, 2014) and verbal self-report (Derbaix, 1995). However, not in all studies the correlation between two variable is found significant. In the study of Li et. al (2017), while pleasure dimension of emotion is correlated with attitude towards ad, arousal does not yield significant results. Also, in the same study of Derbaix (1995), the facial emotion reading and attitude towards brands are not associated. Although the existence of such findings which does not prove the association, the direct impact of emotion on ad effectiveness using this component is accepted in reviews papers (Poels & Dewitte, 2006; Barreto, 2017).

Attitude towards Brand:

Attitude towards brand has a strong connection with the emotional attitude of individuals. As Walla et. al indicated in their study (2011), attitude towards brand is shaped by both cognitive and affective attitudes; they measured the attitude towards brand using objective measures such as EMG signals and proved the connection between subjects' emotions and subjects' attitude towards brands. In the same study, it is also stated that the importance of this processing is come from the fact that it is one of the key drivers for a firm's economical performance.

Also, as Mitchell and Olson (1981) stated the existence of attitude or the change in the attitude towards brand are perceived as a factor that can also predict how ad is effective (cited in Teixeira, 2013). However, in the article of Lewinski et. al, 2014, it is seen that the attitude towards brand is of same concept with the attitude towards ad except that in the latter, the questions are forwarded regarding and in the brand related one the questions are about the brand promoted in the commercial. The items of attitude towards brand questions in self-reports are like:

- "I recognize the values transmitted by this brand"
- "In the past, I have had positive experience with this brand"
- "I have a positive consideration for this brand"
- "This brand offers me good value for the price" (Han et. al, 2011)
- "I am very committed to this brand" (Han et. al, 2011)
- "I think this brand deserves my effort to maintain a relationship" (Han et. al, 2011)

The attitude towards brand is possible to be measured by affective processing (Derbaix, 1995; Morris et. al, 2002; Poels & Dewitte, 2006;). In Lewinski's article (2014), screening happiness level of facial expressions can weakly explain the variance in attitude toward brand (7%) which ratio is higher when it comes to attitude towards brand (25%). This result can be explained by the paper published by Morris et. al (2002) saying that attitude towards brand measures and emotions are both cognitive responses. However, in the study of Walla et. al (2011), as stated before, the connection between emotional attitude and brand attitude are already proven and on top of that, these authors outlook self-report as a weak tool in this regard. Because, the cognitive attitudes are polluted by individual biases.

Purchase Intent:

Purchase intent is perceived as a predictor of ad effectiveness in some studies (e.g. Lewinski et. al, 2014; Li et. al, 2017) and computed independent of ad effectiveness in some other studies (e.g. McDuff et. al, 2015). However, as inferred from the study of Tan and Chia (2007), the change in belief about the ad leads the change in attitude towards the advertisement and this results in influence in purchase behaviour (as cited in Hamelin et. al, 2017). In literature, the association of purchase intent and affective state is not as much strong as that of ad effectiveness. It is found in the article of Teixeira et. al (2013), purchase intent is measurable through facial emotions recognition under some condition such as when the brand is shown after the ad. In the study of Lewinski et. al (2014), purchase intent is of more complexity to be gauged. Although in the article of Li et. al (2017) purchase intent (visit intent) can be measured and the performance of measurement is subject to the choice of instrument, there is no clear consensus made by academicians to the best of my knowledge. The examples of questions directed in the surveys to collect data regarding buying behavior are as following:

- “Next time you are buying [product category] how likely are you TO PURCHASE products from each of these brands?” (McDuff et. al, 2015)
- “The product shown in the video impressed me particularly”
- “Does this ad make you want to visit/buy the destination/computer (advertised in the ad)?” (Li, Walters, Packer and Scott, 2017)
- “The probability I would consider buying this luxury brand is high.” (Bian & Forsythe, 2012)

Customer Satisfactions:

Satisfaction is not counted as basic emotion or any dimension of emotion however it has mutual features with positive affects (Bagozzi et. al, 1999). According to Sharifi 2014), the purchase intent is under the influence of past purchase experience in decision-making process. Sharifi in his article (2014) emphasized the importance of post purchase process for a consumer to consider a future purchase and stating that a dissatisfied experience could even cause a remorseful thought. The questions regarding previous experiences of individuals:

- “I would buy again products of this brand”
- "How satisfied are you with_?") anchored with "very satisfied"/"completely satisfied" and "very dissatisfied"/"completely dissatisfied." (Peterson & Wilson, 1992)
- “I was satisfied with the product I bought”

Brand Awareness:

From the article of Sharifi (2014), the existence of brand awareness is of significance for making a purchase. Citing from the article of Aaker (1996), Sharifi claims that the consumers opt to make a buy from a brand they have any slight idea. This idea can be supported by the paper of Walla et. al (2011). The researchers highlighted the scientific fact between brand awareness and emotion: culturally known brands prompt prefrontal cortex activity which is related to emotional information processing. In the literature, in order to gauge if subjects are familiar with the brands, subjects are directed some questions through self-assessment like:

- When I think of a laptop, brand … comes to my mind (Buil et. Al, 2011)
- … is a brand of a laptop I am really familiar with (Buil et. Al, 2011)
- I am aware of brand … (Buil et. Al, 2011)

3.1.3. Affect Measuring Methods

Detection of affectional state can be done in multimodal forms. The expression of emotion is done mostly by face, gestures, body language, posture (Bagozzi et. al, 1999). In the research conducted by Lewinski et al. (2014) the ways of measuring emotion are categorized in two ways: self-report and autonomic measures such as vocal emotion recognition, facial expressions, heart rate, respiration system. In this research, as an empirical study, the facial reactions will be analysed and for this reason, the face emotion recognition steps will be explained in greater detail. Followingly, as the other instrument of the study, the verbal and visual self-assessment methods are analysed. Alongside the reviewing these techniques in detail, an overview of other emotion detection methods will be introduced.

3.2.3. Face Emotion Recognition

Facial Emotion Recognition (FER) has been the first study area for affective computing communities since face, especially the eyes, has been perceived the window to the soul as William Shakespeare once said. However, how is that possible to read the face to see the emotion of the person? The answer of this question lies in three different fields of academic world: psychological theories, principles of computer algorithms and recognition studies (Butler & Lewinski, 2015).

To begin with what emotion is in the affective computing world:

Emotion recognition process starts with objectively describe what the emotion is. In the science of affection computing, classification and description of human emotions have always been quite a controversial topic for academicians to build a consensus on. These issues are mainly about how to decide the kinds and range of emotions, how to judge or label emotions in an objective manner, so on. According to Ekman & Power (1999), there are six basic emotions which are internationally accepted and do not depend on cultural differences: anger, disgust, fear, joy, sadness and surprise. These emotions are stated as discrete categories. On the other hand, there is another way of classifying the affects that is called continuous and includes three dimensions: pleasure, dominance and arousal (W. Dai, Han, Y. Dai & Zu, 2015). Pleasure is also named as valence in literature. Pleasure dimension implies the interval of the emotion range between positive or negative. Arousal dimension refers the magnitude of the emotion, mainly gauges intensity of the emotion. The dominance scale of the emotions is computed relying on submissiveness of emotion. To explain the dominance scale better, the distinct could be shown between anger and fear which have same characteristics for the valence and arousal dimensions, nonetheless in dominance scale is seen that anger is dominant whilst fear is quite submissive (Albert, 1980). Nonetheless, there has been a disagreement among researchers regarding whether discrete emotional states are universally expressed or not (McDuff, 2017).

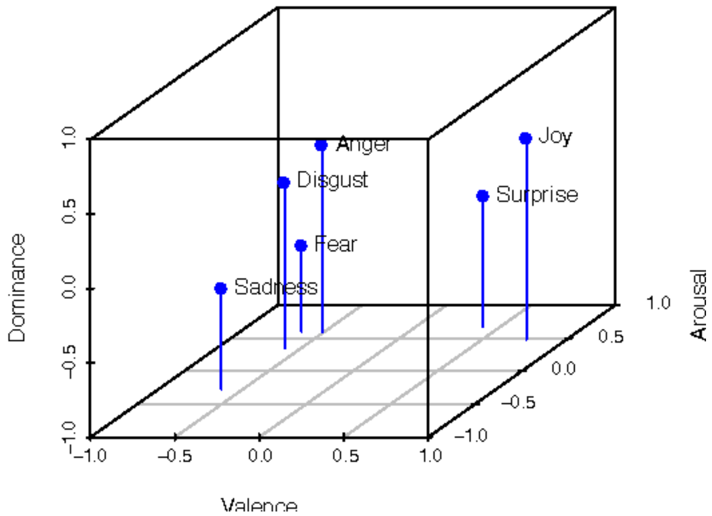


Figure 3 4 The three-dimensional emotion classification model of the valence-arousal-dominance dimensions. (Ekman, 1992; Russell and Mehrabian, 1977)

As much important as describing the emotion, annotation plays a crucial role in affective science.

How to code emotion?

In the emotion recognition field, another harsh problem which is remained unsolved is how to judge the emotions objectively. Simply put, which variable explains what emotion to what extent.

To overcome such issue, the main methods employed in the empirical studies are professional human judges and automatic emotion recognition system.

The categorizing of judging affect methods was made little differently by McDuff & Soleymani (2017) like implicit tagging which means analyzing responses of participants who are exposed to the stimuli and explicit tagging which means to ask the participant tagging the emotion directly. However, Lutfi et. al (2013) emphasized the second problem arises from this section is that not only tagging the emotion is problematic but tagging the right emotion that can be agreed on by others creates a much bigger problem. Firstly, crowdsourcing requires to use the automated facial/vocal recognition technology and Internet. Using crowdsourcing, the participant is exposed to the stimuli and reactions of participants are collected to make an objective decision about the emotion of the stimuli. Exploiting crowdsourcing could be with the aim of asking to label of emotion or collecting their facial responses or both. Secondly, automatic emotion recognition uses the technique of tracking voice/facial signal changes over time and decides what the emotion is coming from the numerical indicator of the degree of emotional signals. However, this way it is avoided to come across speaker dependency issue (Jones & Jonnson, 2015). In some academic studies, the researchers tend benefiting from professional human experts to judge the emotion (Jones & Jonnson. 2005; Lutfi et. al, 2013; Parks and Gates, 2009; McDuff et. al, 2015) However Lutfi et. al states that human annotation does not contribute much to label the emotion comparing to the machine prediction. In the study of Butler & Lewinski (2014), the authors compared the performance of FaceReader comparing the results of human annotators. And in that paper, it is also inferred that software for automatic emotion recognition works in the same that a human perceives the emotion when it comes to basic emotions detection. However, in both ways, the accuracy can never be 100% due to the fact that the recognition of emotion is based on human judgment and human performs maximum 90% to classify the emotion correctly. As asserted in the research study of Lugovic & Horvat (2017) and the review study of Barreto (2017), the computer(55% to 95%) could annotate the emotions better than humans (60%). Barreto (2017) reasons this difference by pointing out the fact that the computers could perceive even minor changes in the face whereas humans judge the overall facial expression. Also, Barreto adds that computers outperform human annotators in terms of fast change of facial reactions.

There are two prominent methods in literature to measure facial expression of emotion. These methods namely are manual facial coding and facial electromyography (EMG) method (Harrigan, Rosenthal and Scherer, 2005). In manual facial expression coding, the face can be detected through eye, thus observation of the face. The other technique is more about the detecting the movements occurring under the skin and closer to the field of physiological measures.

Starting from the computer algorithms of the facial expression reading, recognizing the emotion on the face is a very challenging task on so many levels. First and foremost, the data required to be collected for this purpose is quite complex and is exposed to objective labelling issues and also bringing the issues regarding the environment. The data dealt with in this field suffers from the limitations such as lightning, resolution, head pose, orientation, registration errors and identity bias, just to name a few (Kuilenburg et. al, 2005; Sariyanidi, Gunes & Cavallaro, 2015).

To overcome the issues mentioned above, there are various techniques to be used. However, before that, looking at the figure 3-5 shown below simply pictures all the steps that need to be through for categorizing one facial expression as one of the emotions.

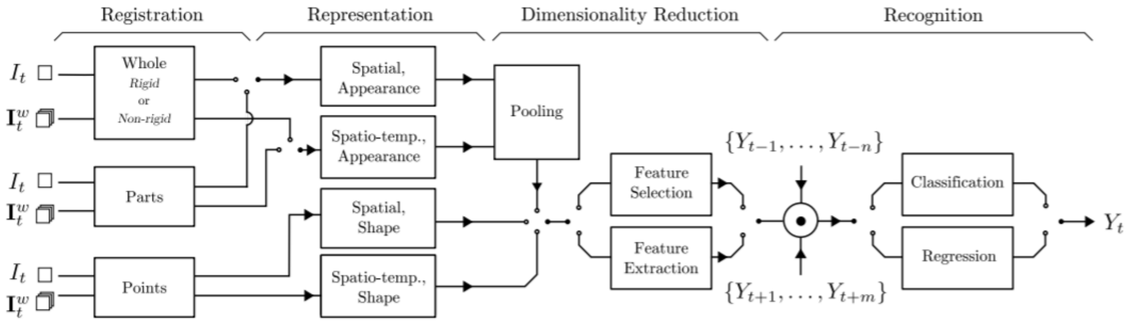


Figure 3 5: The framework of face emotion recognition processes (Sariyanidi, Gunes and Cavallaro, 2015)

Face emotion recognition as behavioral measures register the face first. This step could be in three ways: whole rigid, parts or points of the facial image. In order to register the face, the most popular technique used is Active Appearance Model (AAM). AAM is built in a training phase and this technique uses global facial information. Because though local facial information provides good results, they fail at detecting the emotion in the whole facial expression. Subsequently, as seen in the framework, appearance or shape models are applied to create the feature vector which is used to detect the emotion. The feature vector is the vector containing related information to comprehend input data and classify one of the emotions as output data. In that sense, the AAM is a representation technique to deal with the limitations of environment. AAM is performed pointing landmarks of the face which then compose a shape vector. Shape vector is a vector created in a way that can measure any face to detect the emotion. AAM also creates the texture vector. For more information, see Kuilenburg et. al (2005). Furthermore, the feature vector created is passed through a dimension reduction step to rid of the data that does not contribute much to detect the affect. For that, PCA is the widely used one and proven to perform sufficiently. As the last step of the FER process, the output data can either be a discrete emotion or a continuous data that shows the intensity and frequency of facial values (Sariyanidi et. al, 2015). The last step, classification phase, mostly employs neural networks in literature. For the basic emotions' classification, in practice 7 different classifiers could have been trained and testes however this could lead the problem of “mutual responsiveness”. The reason is that classifiers are supposed to yield the outputs for six basic emotions and neutral state and the highest output is supposed to categorize one of the emotions, but no network could yield a high output in this scenario. On the other hand, employing one classification network for seven distinct emotional states could have been a way to avoid the mutual responsiveness issue (Kuilenburg et. al, 2005).

Many affect models regarding facial emotion coding are examined to measure their performance of achieving the goal of affect detection and recognizing the facial actions. Among them, the most popular choice of FER is to use manual coding in literature, namely Facial Action Coding System (FACS).

Facial Action Coding System

Manual coding is a field that has been worked on for a very long time and first framework was created by Landin in 1924. Since then there are plenty of methods developed in this field. However, Cohn (2006) reviewed many methods that manually label human emotion on the face and stated that Facial Action Code System (FACS) outperforms the other methods in terms of differentiating same anatomically discrete movements. The FACS is a system that works on a human face decomposing it into specific action units (AUs) depending on visually noticeable movement on the face and/or combines the action units to read the affective state appeared on the face. Basically, the FACS separates face into action units, say components, looking at muscle movement and then interprets the emotion on face appropriately. As Sariyadini et. al(2015) explains, the system in this method interprets the facial action unit using the temporal evolution of face. The temporal evolution of face is basically composition of four steps which are neutral, onset, apex and offset. The face first takes the appearance of a neutral version of its own which means nonexistence of any emotional state. Then, it evolves to the onset phase in which facial muscles begin to contract and intensity level escalates. On the third phase, apex, it reaches a maximum and stable level. Offset phase converts the face to the neutral state by relaxing the facial muscles. Through the AUs, it is doable to detect the emotion. However, as Kuilenburg, Wiering & den Uyl state in their paper (2005), there occur numerous facial expressions to be interpreted as emotions. Besides, a great range of emotion is also a challenge here in order to be assigned to various facial expressions. The most used way to interpret facial expressions and categorize them into emotions through FACS is to benefit from Ekman's universal emotional expressions. Furthermore, facial muscle actions are more objective descriptors comparing to facial expressions; as Ekman et al. (1972) advocated facial emotions are inborn and do not depend on cultural differences which make them internationally recognized and also provides researchers the system as in high-level decision making process according to the emotional FACS rules (EMFAC). The manual system of action coding is quite well described in a +700- page guide to acquire objective labelling of the affect state (Ekman & Friesen, 1978; Ekman, Friesen & Hager, 2002; cited in Butler & Lewinski, 2014). This system at a large scale employs discrete emotions to decode the emotion. There are also studies to detect the non-basic emotions which can either be limited range of emotions such as "relief" or continuous emotion affect dimensions like valence, arousal and dominance (Sariyadini et. al, 2015).

Over this topic, in the article of Cohn (2006), he categorizes the action units in a way that they are additive and non-additive. Additive AUs are like the features in speech emotion recognition, each AU explains the emotion to a considerable degree. In non-additive AUs, they provide complementary information. The joint production of two AUs could mean a whole different thing than their separate interpretations. Also, Zeng et. al (2009) stated that the combination of 27 basic AUs and a number of AU descriptors could explain the other emotional states, complex ones, different from basic emotions (e.g. pain). A list of AUs is shown in Fig. 3-6.

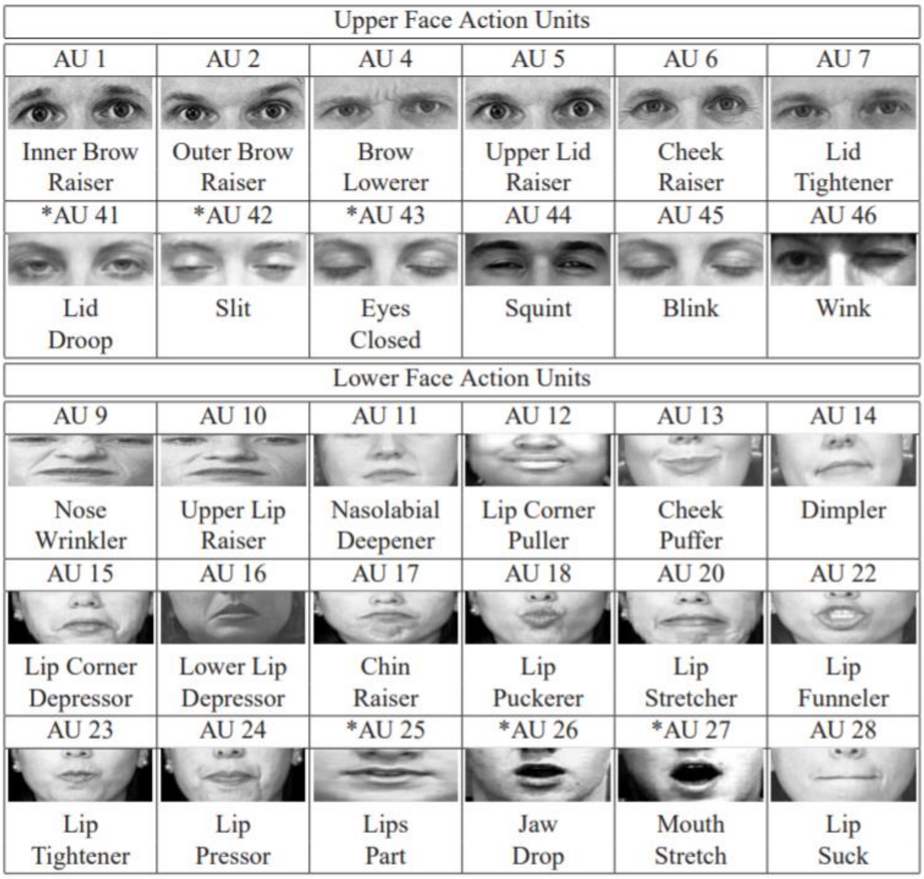


Figure 3 6: 28 Action Units of Face Emotion Recognition

In the study of Butler and Lewinski (2015), it is seen the comparison of performance to detect the affect state between basic emotions and FACS using two different emotional databases. Through a software for automated facial coding it is computed action units and also categorized them into one of the affections. From the study, it is inferred that the facial action units are far more exhaustive work comparing to assigning the face expression an emotion category. By means of software which is a FaceReader developed by Noldus, each AU is due to contribute to the classification of facial expression and the combination of AUs are required to do it correctly. To do this, it is obligatory to compute the scores of probability and intensity of facial reactions of 17 different action units. This, in contrast, happens among six basic emotions when it comes to categorizing the emotion from facial expressions. Apart from the higher number of computing tasks bringing higher risk problem, the objectivity of labeling emotion also poses an issue when it has to do with FACS. In basic emotions and FACS module of FaceReader, the validity is measured in comparison to the human coders labelling performance.

And about the distinction between two approaches, simply, human coders perform better at detecting the facial affective state with six basic emotions. For a human coder to code the action unit accurately takes a very long time to acquire the skill of detecting emotion this way. Since the learning curve of software follows a similar way to gauge the emotion, this makes the process more complex in that sense. However, the only manner to handle face expression detection is to compare the results of human coders and software and pursue the higher performance. But also, as Cohn (2006) states the overall emotion perception could be completely different with the help of additive information. Even though the result of Butler & Lewinski's study (2015) was proving the basic emotions as a more robust approach, they also pointed out that the psychology research is more deep regarding basic emotions and the algorithm to compute the emotion is built on the past psychology studies. From their conclusion, the performance of basic emotions(88% accuracy) and FACS module(0.69) implies that the FACS is promising albeit it is not a proven approach to be used to gauge the affective state. Also, they compared their study to a very similar study which was conducted using an older version of FaceReader and employing only one database for both test and train in 2005. They have come to a conclusion that automated facial coding software, a FaceReader, makes the emotion recognition feasible and make good progress.

Apart from the comparison of basic emotions and action units' emotion detection values obtained through FACS module, how is a FaceReader works to compute the affect state? Also, in the experimental part of this thesis, it is employed a FaceReader developed by Noldus (version 7.0). A FaceReader first detect a human face and creates a 3D Active Appearance Model(AAM). AAM is a powerful technique to overcome the issues caused by environmental effects such as lightning.

Followingly, in the last step, the software computes the scores of probability and intensity of face reactions on a continuous scale which is from 0 to 1. This software is a commercial tool to be used in FER fields and easily adopted with Qualtrics and crowdsourcing platforms (Butler & Lewinski, 2015). The software is also robust towards different types of faces which can be age differences, gender distinct, nationality-based groups, so on. Its performance is measured comparing a human coder detecting performance of an emotion as explained before. However, to confirm the validity of its usability, even with human coders there not 100% success to detect the affect. Opposite, it is often 60-80% accuracy for human coders. Since basic emotions is a relatively more naïve system, this issue has been softened for the objective labelling issues.

3.2.5 Subjective Measure

Self-assessment has been the most widely method used in marketing researches and marketing business applications to inspect the opinion of the consumers. This is an explicit and direct method to learn what the consumer thinks about the stimuli in terms of the emotion evoked by the stimuli, the intensity of the emotion and the whether the content is intriguing for him/her, to name several. The reasons for self-reports to be popular is numerous; however, the major ones are that this explicit measurement method is cheap to apply regarding time and cost, ease of use even in large amount of questionnaire samples, providing the possibility of asking question that could not be answered with behavioral and autonomic measures(e.g., personality questions) and also ease of interpreting the results emerged from them (Hoskin, 2012). The importance comes from this method that it represents the scientific research to the public. On the other hand, this popular approach to grasp the emotion poses drawbacks. As Hoskin (2012) claims in his study, many questionnaires fail at checking the validity of the questionnaire measuring what the study demands to assess. As an instance from this study, among depressed and non-depressed people, even though the measure is valid, the subtle differences could matter, and this self-report method might fail at measuring it in every aspect.

Another disadvantage of self-assessment is that the technique is leaning on the honesty of subjects which is highly dubious because it may be subjects not unwillingly biased (e.g. laboratory environment effects, social desirability) or subjects blending the truth purposefully. Nevertheless, owing to this handicap the behavioral and autonomic measures presented in theory. Also as Hoskin (2012) indicated that this would be avoided with the reverse questions to eliminate the bias effect. Moreover, the questions directed this way are prone to be understood differently or not understood at all. For instance, due to not self-measuring own emotions, the subjects could come up with random answers. In addition, forwarding scale-based questions are also opted to be varying by individual.

Furthermore, with the advent of technology, the self-report seemed to be a very powerful tool in terms of gathering big size of data however, this also drew on the issue of not having control over the sample in case they need help with some questions, they could not have the chance to clarify the doubts (Hoskin, 2012). Additionally, Hamelin et. al(2017) proved that the findings gathered by surveys could have reached with a smaller number of participants using behavioral measures methods.

Hence, this method has advantages and disadvantages and in case this method used or intended to be used in a research, the advantages and drawbacks are ought to be paid attention. After having compiled the general features of the explicit method, in the following part I will section the approaches to apply self-report: verbal self-report, vision self-report and moment to moment report.

3.2.5.1 Verbal Self-report

The verbal form of the subjective measures generally comprises open-ended questions and/or rating the emotions on a range. Emotion measurement in this fashion is applicable either asking the dimensions of emotions or direct emotion types which is termed basic emotions. The dimensions are pleasure, arousal and dominance, whereas the basic emotions are happiness, sadness, surprise, fear, anger and disgust (Lewinski, Fransen and Tan, 2014). These methods of categorizing emotions are well-explained before. Also, it is stated in the study of Havlena and Holbrook, 1986 that the dimensional approach is a more useful method to capture the information regarding the emotion felt by stimuli rather than basic emotions do.

Apart from the drawbacks and benefits the subjective measures hold, one of the challenges caused by verbal self-reports is that the respondents are exposed to stimuli first and the questionnaire after; this might cause the issue of examining the perception of emotions rather than emotion themselves (Poels & Dewitte, 2016). Though, this might be overcome with the fusion of other emotion recognition methodologies.

3.2.5.2. Visual Self-report

Visual self-reports are another way of asking the respondents what the stimulus evokes by means of a questionnaire. In this sort of self-report, the subjective confronts less tedious questions in the survey directed to them after the stimuli shown (Lewinski et. al, 2014). Besides, as Morris et. al(2002) claimed the verbal self-reports are not a great choice to provide all the emotions as visual one can due to the variability of meaning of each emotion for each individual. The most prominent instruments of this questionnaire are AdSAM® (Morris et. al, 2002) and PrEmo (Desmet, 2002) and the examples are demonstrated in Figure (3-7) and Figure (3-8). As seen in the figure (3-7), the respondents have the cartoon character as options to rate the emotion they feel (Lewinski et. al, 2014).

AdSAM® is a modified version of Self-Assessment Manikin and it basically measures the emotion with the visualized version of questions. This approach benefits from Pleasure, Arousal, Dominance bipolar dimensions (PAD) theory to capture all sorts of emotions and PAD enumerates the affects into the three independent dimensions. This tool has been very often utilized in advertising and marketing practices and researches. Moreover, this tool is a problem-solver in terms of cognitive processing since the verbal self-report, especially open-ended questions, is onerous.

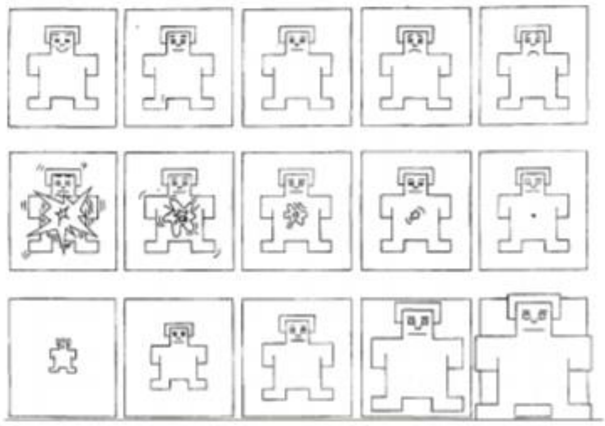


Figure 3 7 The Self-Assessment Manikin visualizing PAD dimensions of Emotion (Morris et. al, 2002)

The other way of visualizing self-reports is Product Emotion Measurement Instruments (PrEmo). This approach provides 14 emotions that are half pleasant emotions half unpleasant emotions. In the figure (3-8) shown below, 14 animations of emotions are accompanied by three other options which are basically asking the subject to what extent they agree on the emotion chosen. For example, in the case of amusement, the respondent confronts the options of totally agreeing on the emotion, to some extent feeling the emotion or not feeling the emotion at all. The hidden three scale and fourteen visualized emotions do present a wider choice range than SAM could. Likewise, the animations are displayed with facial, vocal and bodily expressions. when the subject clicks on the emotion type. The more detail is presented in the study of Desmet P. (2005).

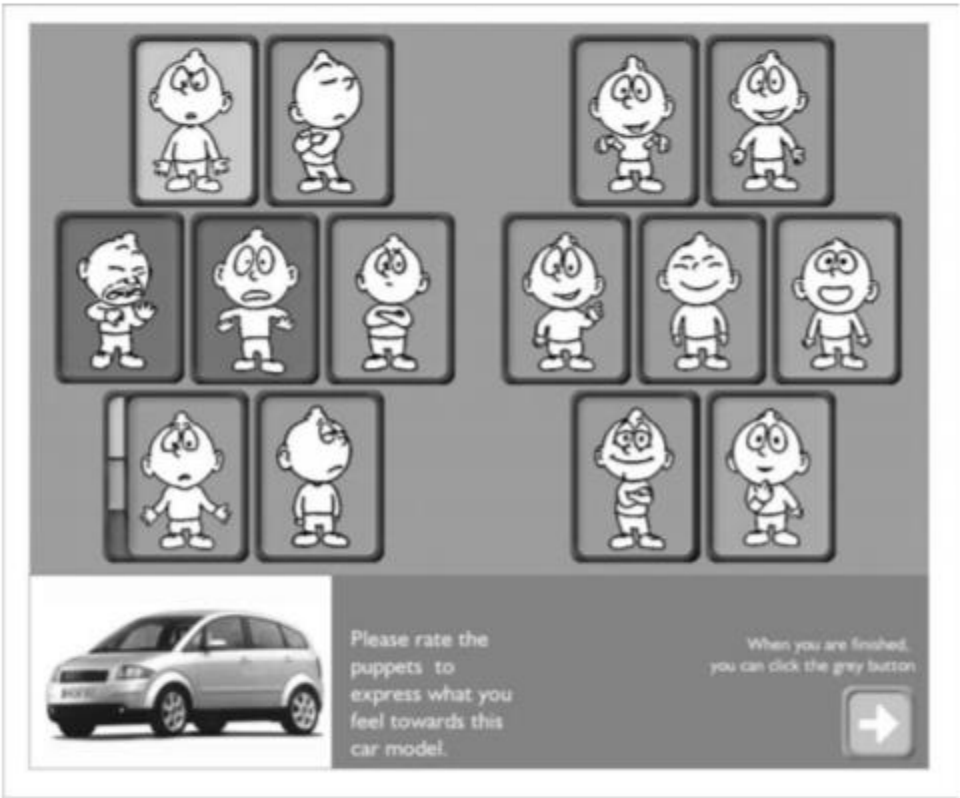


Figure 3 8: PrEmo Product Emotion Measurement Instrument (Desmet, 2005)

Albeit visual self-report seems to be a relief for the problems emerged from verbal self-report to some extent, this method also bears the issues of general subjective measures.

3.2.5.3 Moment to Moment Report

This is the third approach used in self-report emotion measurements. The major benefit of choosing this method over the verbal and visual self-report, it handles the problem of receiving the emotion responses after the stimuli exposure. Simultaneously, in this method, the participant is supposed to respond the questionnaire and be exposed to the stimuli.

3.2.4. Overview of Other Emotion Measurement Methods

In literature, there exist other autonomic measures that aim to capture the emotion using different sources: speech emotion recognition (Ramaksrishnan, 2012; Ayadi et. al, 2011; Park and Gates, 2009; Schuller et. al, 2005; Zeng et. al; 2009; Yang and Lugger, 2010; Ren and Quan, 2012; Morrison, Wang & Silva, 2007), heart rate, skin conductance, brain imaging (Maaoui, Abdat and Pruski, 2014; Vecchiato et. al, 2011; Han et. al, 2015; Vroomen & de Gelder, 2000; Zeng et. al, 2009; Anagnostopoulos et. al, 2012). Lee and Kwon, 2010; Horvat et. al, 2015), to name a few.

These methods have the similar steps such as feature extraction, feature selection, classification at macro level. Differently, speech emotion recognition and face emotion recognition are more alike in the sense that both measures face the challenge that the subjects could behave differently in the laboratory settings and that could reason biases, manipulate the effects. For this reason, both techniques are also known as behavioral measures of emotion. However, the rest of the other measurement methods of emotion are uncontrolled or beyond the control by subjects (Poels and Dewitte, 2006). As some instances from literature, heart rate technique follows the patterns of heartbeat to detect level of attention, arousal, cognitive and physical effort (Lang, 1990). Reading heart rate is possible using only a small device on top of a finger, but the technique is advised to be used with the combination other methods in order to get more reliable results (Lang, 1990). In the experimental study of Maaoui, Abdat and Pruski (2014), heart rate features were combined with skin conductance, muscular tension and facial expressions features in different processes of emotion recognition with the aim of achieving better accuracy. In this way, there is no intervention of subjects. Moreover, The skin conductance, on the other hand, measures the level of sweat via electrodes so as to detect the emotion of people towards a stimulus (Poels & Dewitte, 2006). Brain-imaging is another implicit measure of emotion and does not require the intervene of subjects either.

However, still, in micro level, the advantages and limitations of employing each method varies greatly depending on the goal of the study. For example, heart rate poses a problem in which the researchers might stumble upon some changes in heartbeat that is affected some other factors (Lang, 1990). Another example is that the use of skin conductance technique brings about special equipment and high level of knowledge to interpret the results (Poels and Dewitte, 2006). The most obvious drawback of employing Brain imaging is the cost the method incurs.

3.1.4. Empirical Studies of Emotion Recognition in Marketing and Advertising Literature

Thus far it has been shown what emotion means in marketing and advertising. It has been also detailed the facial emotion method. Since there are various methods to gauge emotion and dissimilar approaches towards the emotion impact in marketing and advertising, in order to limit my search, I analysed the empirical studies in literature using the keywords 'facial emotion recognition AND ad effectiveness AND purchase intent' in order to see the similar studies done in this field. The articles presented in last decade were considered. In this part, a part of these studies will be presented as a review.

In the article of Lewinski et. al (2014), the effectiveness of advertising was tested through using FaceReader and self-assessment instruments. They conducted the psychophysiological experiments collecting data on Amazon Mturk platform which means that the subjects were in natural settings. However, firstly the stimuli were chosen after performing a pre-testing study in which a number of subjects criticized the type and intensity of emotion evoked by the stimuli and agreed upon. The test participants were then shown six stimuli with three levels of amusement as high, medium and low. The research questions of the study are to distinguish the face expressions intensity level towards different amusement level of the video stimuli and also to measure the feasibility of advertising effectiveness by monitoring facial reactions. They had worked on only one type of emotion which is happiness which is a limitation of this paper because as authors also state that the stimuli could evoke other sorts of emotions too. The reason to choose both self-report and FaceReader (developed by Noldus) for this study was that the surveys are challenging to capture the advertising effectiveness as a tool and FaceReader is a proven method to gauge the emotion. They have tested the relevant hypothesis using nonparametric Friedman test and Spearman's rank-order correlation because the data was not normally distributed. And the results of the study proved that people react significantly happier when the stimuli is high level of amusement video. Cognitive responses correlate with FaceReader results. Therefore, advertising effectiveness is measurable employing a automated emotion coding software.

In this study McDuff, Cohn and Picard (2015), the authors sought to clarify the relationship of facial responses with ad liking and change in purchase intent. They aimed to model ad liking and purchase intention of the viewers using the facial expressions towards the ads. In addition, they wanted to find the relationship of these three metrics. An enormous dataset has been used for this study and the data have been collected in natural environment (Amazon Mechanical Turk). Ultimately, the researchers had the propensity to identify the features of aggregate emotional responses that make an ad effective such as where to use brand name. Previous studies claimed that psychological and self-reported measurements capture different information. Also, including emotion in an ad can keep the viewer paying more attention. The participants were recruited under the constraint that they were supposed to use the promoted product before. Followingly, they were asked to attend the experiment and answer related questions. The researchers of this study have used the-state-of-art methods like HOG for the feature extraction and SVM. The frames were examined looking at eyebrow raises, smiles, disgust and valence expressions. For all have been used binary measurements. As an instance, existence of smile is 1 and nonexistence of smile is 0. The images from these datasets were labelled for the presence of an expression by human coders. The result arises from this work is that employing the face expressions gave only 17% significant response from the viewers. Although responses are sparse, different people respond to the ads differently and the ads elicited a range of expressions (from strong positive valence to negative valence). So, the researcher created the aggregate metrics and this way gave rich information about the effectiveness of the ads. Liking score and purchase intent scores have been obtained from the self-reported responses of viewers. Training and test analysis are that using leave one ad out method is chosen for the test and training part. SVM is chosen as a classification method. In the end the findings, more liked videos generated higher valence values and created more smiles upon subjects' faces. However, the prediction of intention yielded less strong outcomes than liking model. On the other hand, there is a significant difference recorded for the purchase intent if the brand appears right before the peak positive emotion.

An article by Chen et. al (2018) concentrates on measuring emotion evoked by oral care products' flavours using self-assessment, EEG, EMG and cardiovascular measures. They conducted the experiments in the laboratory environment and asked participants to taste five different flavours. In this field, they employed to state-of-the-art instruments as they claimed. Before the main study, the stimuli were chosen based on a pilot study in which the participants rated the emotion evoked by the stimuli and had a compromise. In their study, the authors directed subjects the questions related the valence, arousal and dominance, basic emotions, overall attitude and purchase intent. Meanwhile the subjects were monitored with physiological instruments. They used the test Friedman and Wilcoxon due to the reason of violating the normality assumption. The significance distinction was recorded between five flavours in valence and arousal and they stated that the subjects had no clear understanding of what dominance explicitly means. Additionally, they asserted that asking directly emotion dimensions through survey is not wise because of unclarity. They suggested to use self-assessment tool rather for overall attitude and buying intention questions. To test EEG, EMG and HRV values, they used repeated measures ANOVA and proved that EEG and HR tools are better predictors for VAD dimensions of affect than surveys. They had no significant record for HRV and EMG. The limitations of this research is that they divided the subjects into two groups because of the hardness to use EEG and EMG tools concurrently.

Another study conducted regarding emotion and advertising effectiveness is that of Hamelin, Thaichon and Moujahid (2017). The researchers aimed to measure the long-term impact of advertising upon attitudes using the safe driving stimuli. For this purpose, they employed a face emotion recognition software (Gfk-EMO) and self-assessment. The surveys were filled by participants subsequent to the exposure of the stimuli and also 2 weeks after in order to see the impact driven by the ads for two different type of video ads. One was a more instructional and the other stimulus was more harsh and performing a car accident; low and high emotional commercials. They concluded the study showing the distinct effect of high emotional ad on driving attitudes and recorded the recall of ad higher for the more intense commercial. The implication of the study is that emotion can control the attitudes and actions.

Another study in literature respect to measure the emotion influence evoked by the commercials on advertising effectiveness and purchase(visit) intent is of Li, Walters, Packer and Scott (2017). They aimed to enlighten this search with the ads from tourism industry and computed the records of attitude towards ad, visit intent and postexposure destination attitude. The authors employed the visual self-assessment (SAM Manikin) instrument, verbal self-assessment facial EMG and SC tools with the purpose of investigating the consistency among the instruments alongside answering the advertising effectiveness questions. The experiments were held in the laboratory environment. The stimuli range is humour, romance, adventure, youth, family and rationality. The researchers used the non-parametric tests to examine the results of hypothesis testing. What is concluded in this study is that valence is a better predictor than arousal to understand the customer propensity according to the survey results. The outcomes of this paper are contradicted with the results obtained from Morris et. al (2002) which was telling that there is a direct impact of emotional responses on purchase intent with survey results. However, there are other studies also agree with that of Li et. al (2017) and denied with that of Morris et. al (2002) and say that there is no direct impact of emotion on purchase intent when verbal self assessment is used. About the physio-psychological experiments findings, the authors noted the significant results existed for pleasure but with a weaker impact comparing to self-report outcomes. Regarding skin conductance, no significant results were recorded at all. All in all, depending on the tool, the results would have been obtained slightly different.

The limitations of this study are a few. One is that the people had higher propensity to visit unfamiliar places but the stimuli chosen was promoting the famous destinations. Another one is that many subjects were mostly university students and therefore the sample distribution in terms of age was not balanced. Also the researchers recommend a future study benefiting from a natural settings.

CHAPER 4: EXPERIMENTAL DESIGN

In this part of thesis, I elaborate the methods employed to respond the research questions. As a first phase of the research, the stimuli candidates and the constraints of stimuli selection are well-detailed (Section 4.1). Afterwards, a pre-survey study was conducted, which was designed with the aim of capturing baseline measures of perceived emotions in the stimuli. Using pilot studies to have a consensus on adequate stimuli selection was applied in past academic studies (e.g. Hamelin et. al (2017), Chen et. al, (2018)). This phase is explained in the section 4.2. Followingly, two well-structured questionnaires were designed to gauge the impact of stimuli on advertising effectiveness, purchase intent and, as the core part of the study, emotion of the stimuli. The survey spread over Internet was designed in more detail due to capture the impact in a better way. This part is covered in the section 4.3.

Subsequent to the design of self-assessment studies and selection of stimuli steps, the experiments were staged using the FaceReader software by Noldus (version 7.0) in the Pheel Lab at Politecnico di Milano. In the lab part of the experiments, the process was involving the exposure of stimuli to subjects, recording their facial reactions for emotion recognition analysis and requesting the subjects answer a self-report questionnaire. The framework is shown in section 4.2. In the same section, the expanded version of survey designed to be applied in online settings will be explained.

4.1 Stimuli Selection

In the very beginning of the selection of stimuli, the main constraints regarding stimuli selection were that the stimuli were intended to promote a product with the average length of 1-minute and standard deviation of 10 seconds. The reason of choosing a product to promote in the ad instead of service is that the product is relatively a better choice to measure purchase intent after exposure of the video on the consumer and expecting the immediate answer of potential purchase decision. The reason of making the video length a minute is that creating engagement with the viewer in order to deliver the message, convince the consumer to buy and promote the brand requires long time and on the opposite, the long exposure time could cause boredom in the viewer. Relatively the average length of 1-minute stimuli could be a good fit to gauge the emotion output of engagement. On this basis, a big number of videos were screened considered the emotion the videos elicit. The categorization was made depending on the general emotion of the video which were restricted to sad, happy, neutral and anger out of basic emotion lists and the emotion in the music, semantic and tone of voice. In the selection of emotion types of stimuli were mainly the result of the software employed to measure the facial reaction, which is Facereader by Noldus. This FaceReader software can measure six basic emotions: happy, sad, angry, surprised, scared, and disgusted. Considering the research questions, it is more suitable to direct a happy, sad, neutral and angry ads to measure the ad effectiveness on purchase intention via facial expressions. The logic to decide the semantic emotion is fundamentally linked to linguistic emotions features. The salient words are sufficient to explain the affect state to some extent (Anagnostopoulos et. al, 2012; Ren & Quan, 2012). The tone of voice is mostly related to the voice emotion features. As an instance, the high pitch usually refers to the emotions requiring high arousal like anger (Gong et. al,2015).

As a result of looking at these constraints, 22 videos were chosen as good potential for the stimuli in the first round. However, for the 4 emotion categories, some videos failed at the second selection phase because of not fulfilling the requirements regarding the semantic, music and voice characteristics features coherently. Therefore, 24 new videos were added to the list to be reexamined in order to see whether they elicit the emotion in a clearer way and are coherent in terms of the emotion evoked in the semantic, music and via tone of voice. After this round, a third one was needed to be sure to have enough number of stimuli candidate before going through another examination which was including expert investigation such as one video maker and one user experience designer.

Subsequent to processing the videos for the pre-survey part, 5 commercials were selected for each aforementioned affect sorts. The ads were including some characteristics which are not in line with the constraints; however, they were thought as potential due to the difficulty of gathering all requested features. The commercials are shown below:

HAPPY	**Length**	**General Emotion**	**Semantic**	**Music**	**Voice Characteristics**
Ikea	01:01	Happiness	Neutral Speech	Happy Music	Happy Voice Features
Kellogs's	01:02	Happiness	Happy Speech	Happy Music	Happy Voice Features
McDonalds	00:30	Happiness	Happy Speech	Happy Music	Happy Voice Features
Knorr	00:58	Happiness	Happy Speech	Normal Tempo	Happy Voice Features
Jollibee	01:00	Happiness	Happy Speech	Happy Music	Happy Voice Features
SAD					
Subaru	00:59	Sadness	Sad Speech	Sad Music	Sad Voice Features
HP	01:10	NO Sad	Neutral-Sad Speech	Sad Music	Sad Voice Features
Heineken	01:00	Sadness	Sad Speech	Sad Music	Sad Voice Features
Origami Extra Gum	01:01	Emotional	Neutral Speech	Sad Music	Sad Voice Features
Canon	01:00	NO Sad	Neutral Speech	Sad Music	Sad Voice Features
ANGRY					
Hyundai	00:46	Angry	Anger Speech	Angry Music	Angry Voice Features
BMW	00:30	Angry	Anger Speech	No Music	Angry Voice Features
Adidas	01:10	Angry	Angry Speech	Angry Music	Angry Voice Features

OralB	00:30	Angry	Angry Speech	Neutral Music	Angry Voice Features
Bridgestone	00:32	Angry-Funny	Neutral Speech	No Music	Angry Voice Features
NEUTRAL					
The Glenlivet	00:54	Neutral	Happy Speech	Neutral-Happy Music	Neutral Voice Features
Kindle	01:25	Neutral	Neutral Speech	Neutral Music	Neutral Voice Features
Lenovo	00:59	Neutral	Neutral Speech	Happy Music	Neutral Voice Features
Tide	00:52	Neutral	Neutral Speech	No Music	Neutral Voice Features
OralB	00:51	Neutral	Neutral Speech	Neutral -Happy Music	Neutral Voice Features

Table 2: The stimuli candidate for Pilot Study (The Commercial List, Duration, General Perceived Emotion, Semantic, Music and Voice Characteristics)

4.2. Validation of Stimuli

In the pilot study, the main goal is to select adequate stimuli for the experiment part. To put it simply, in addition to selection of the stimuli considering the emotion in semantic, tone of voice and music, the emotion in the ad is supposed to be perceived in a way that a sample group could agree on the type of affect which evokes by the stimuli.

The survey was created through “Google forms” as separate 12 different forms for 12 participants. The difference in the forms was only the sequence of the stimuli so as to eliminate the ordering effects. The survey started with the personal questions as regards to the obtain demographic structure of participants such as age, gender, educational status with the intent of balancing the demographic features of viewers. The participants were given the basic information concerning the study objective. The participants were chosen with the constrain of their nationality. They were all native Italian speakers who speak English well. Because it was intent to avoid the effect of cultural background in perceiving the emotion.

The questions directed to the survey respondents were designed to assess the general emotion of the video, emotion in the music of the video, emotion in the semantic and emotion conveyed via tone of speaker. They were also asked to rank the comprehensibility range of the video with 5-point scale. The questions measuring the emotion of the video from different perspectives had the options of "happiness, sadness, anger, neutral" as explained before.

The videos were uploaded to a brand-new YouTube channel with the captions of different Greek alphabet letters. The reason behind to use a different alphabet letters was to neutralize the effect of the brand name on the viewer and keep the videos anonymous. For, when they watch the video, they were exposed to see the caption as a first impression. These videos were embedded into the Google forms following the randomized list of videos. The forms including the advertisements and questions were shared online and the results were collected and processed through Google documents.

The pilot study survey results:

The demographic structure of the participants is shown below:

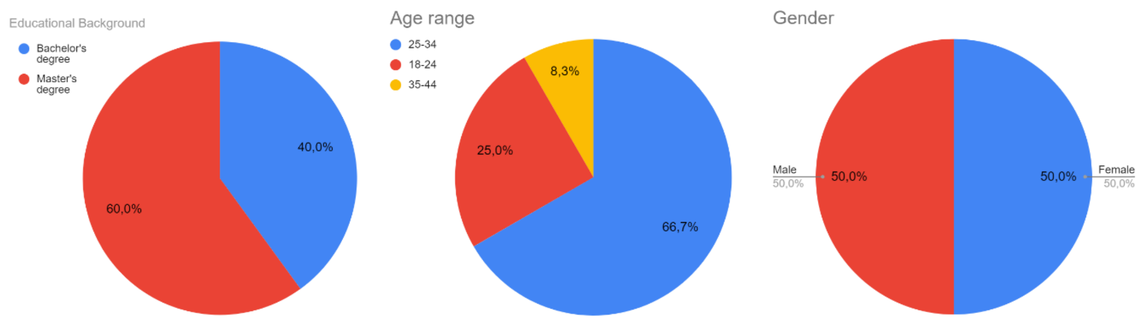


Figure 4 1: Educational Background, Age and Gender distribution of Pilot Study Participants (12 subjects)

From the participants, the pre-survey study has been concluded with the result is that they perceived all videos which were aimed to convey happy affect as happy setting the lower threshold 70%. The column chart demonstrates the results separately for each happy advert.

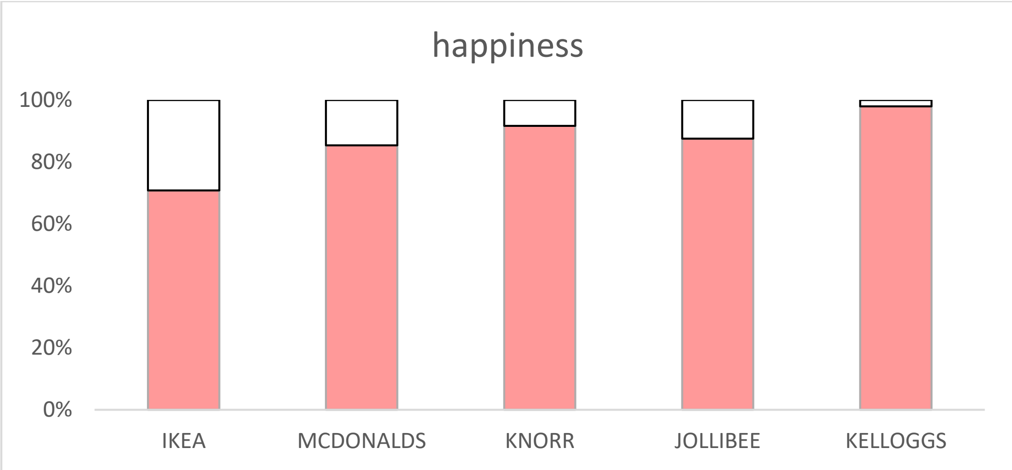


Figure 4 2: The self-assessment results of Pilot Study regarding the general perception of emotion in happy stimuli

The happy commercials were examined from different aspects as shown in the table below:

Happiness	**General Emotion**	**Music**	**Tone of Voice**	**Semantic**	**Overall**
IKEA	75%	100%	58%	42%	69%
MCDONALDS	100%	100%	75%	67%	85%
KNORR	100%	92%	83%	92%	92%
JOLLIBEE	92%	92%	92%	75%	88%
KELLOGS	100%	100%	100%	92%	98%

Table 3: The results obtained through pilot study for happy stimuli

As regards the table of happiness stimuli performance for this pre-survey study, it is clearly seen that overall, the participants agree on perceiving the happiness affect conveyed through the ads. The results are shown in percentage of people who chose "happiness" option over the other emotion types options. Whilst, general emotion of the video and the music are more obvious to them, semantic and the tone of voice are more questionable in terms of the emotion elicited. However, the ads give good results for this emotion type. Moreover, even though the ads were investigated from few perspectives, the main goal was to achieve good result in terms of the "general emotion of the video" part.

The results reflected from the participants for the neutral state of emotion were similar to those of happy one. In the following graph, the general emotion of the Lenovo, Tide and Glenlivet commercials were perceived as neutral affect commercials according to this self-report study.

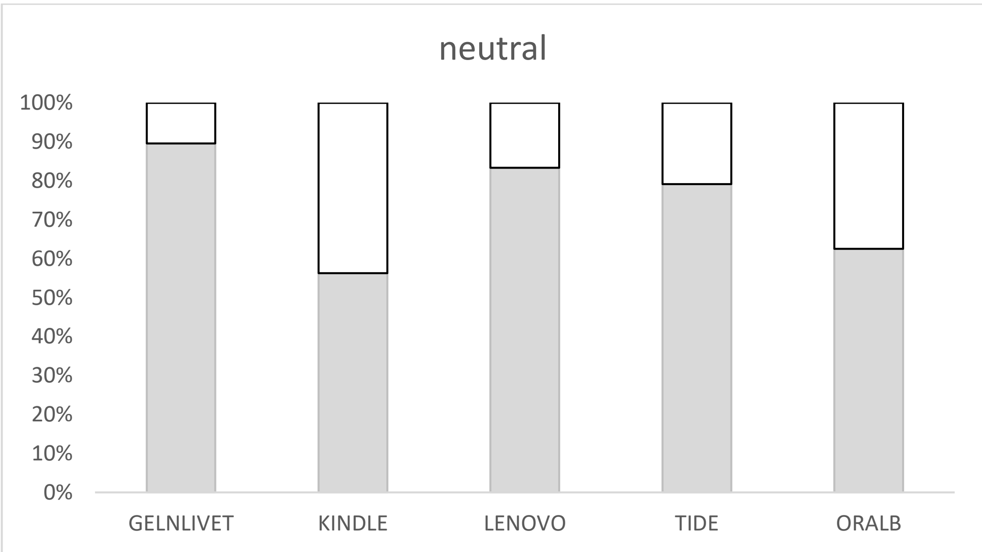


Figure 4 3: The self-assessment results of Pilot Study regarding the general perception of emotion in neutral stimuli

Neutral	General Emotion	Music	Tone of Voice	Semantic	Overall
Glenlivet	100%	92%	83%	83%	90%
Kindle	58%	50%	67%	58%	58%
Lenovo	92%	67%	83%	92%	83%
Tide	75%	100%	67%	75%	79%
OralB	83%	25%	58%	83%	63%

Table 4: The results obtained through pilot study for neutral stimuli

As regards the emotion evoked through the music, semantic and the tone of voice of the speaker in the video, the results were promising for Glenlivet and Lenovo.

For the sadness and anger videos, the outcomes were more challenging. People were perceived the ads closer to the neutral state for some of them.

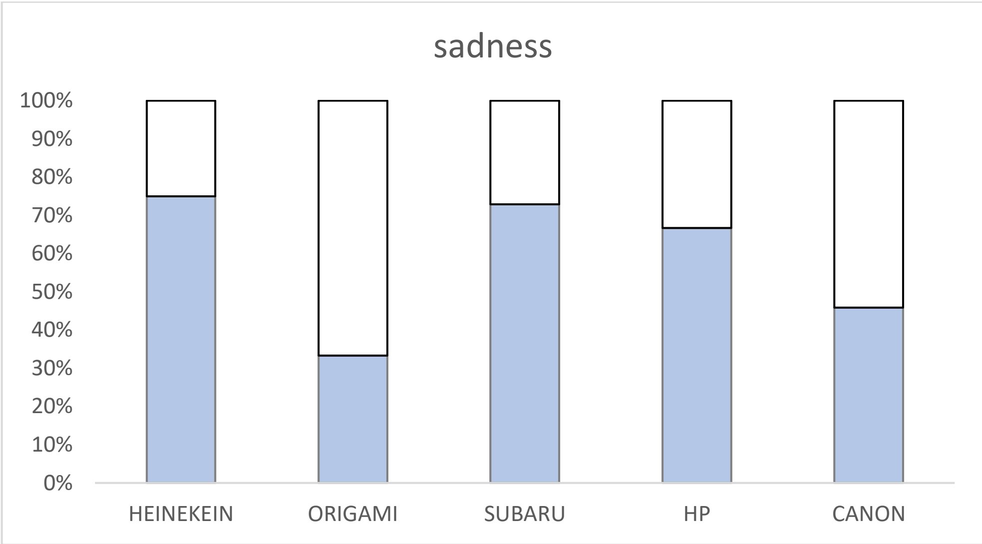


Figure 4 4: The self-assessment results of Pilot Study regarding the general perception of emotion in sad stimuli

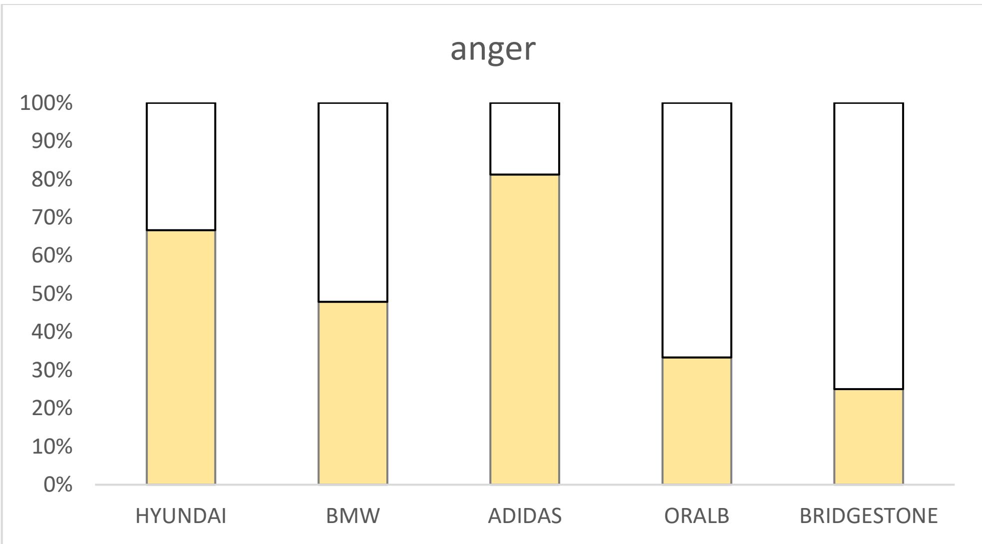


Figure 4 5: The self-assessment results of Pilot Study regarding the general perception of emotion in happy stimuli

Regarding the sad videos, participants came to a consensus about Heineken and Subaru ads being sad, while HP ad was the following one in the ranking. Concerning the angry ads, Adidas and Hyundai ads were flagrantly thought the only anger advertisements among these 20 ads they were exposed to.

Sadness	**General Emotion**	**Music**	**Tone of Voice**	**Semantic**	**Overall**
Heineken	100%	100%	50%	50%	75%
Origami	33%	75%	17%	8%	33%
Subaru	83%	92%	50%	67%	73%
HP	58%	58%	67%	67%	63%
Canon	58%	42%	58%	8%	42%

Table 5: The results obtained through pilot study for sad stimuli

Anger	**General Emotion**	**Music**	**Tone of Voice**	**Semantic**	**Overall**
Hyundai	83%	75%	58%	50%	67%
BMW	75%	8%	50%	58%	48%
Adidas	67%	83%	83%	92%	81%
Oral-B	67%	8%	33%	25%	33%
Bridgestone	42%	17%	25%	25%	27%

Table 6: The results obtained through pilot study for angry stimuli

The consequents of this study for “sadness” and “anger” emotion stimuli in the presurvey phase were not as much bright as “happiness” and “neutral” ones. Nonetheless, it was still good to have the consensus over two videos for each emotion at least for this phase. The reason behind this was to stimulate the participants in the laboratory environment with two videos for each emotion and compare the results coming from the self-report study and the facial expressions toward the stimuli.

Expert Analysis

Concluding this presurvey step with the information shared led the process to the next step which is to consult the experts. In this phase, the two colleagues from Design and Communication department of Politecnico di Milano were reworked on the selected stimuli. Apart from their competences in their field, one had the special interest in video-making field. After screening the videos with them, the final selection was made choosing Glenlivet and Lenovo commercials as the neutral affect stimuli, Kellogg's and Knorr advertisements as happiness videos, Subaru and HP adverts as the sadness emotion stimuli. In this step, it was decided to eliminate the adverts conveyed anger affect due to the reason that the ads in general which carries anger feelings were mostly turn out to be funny angry adverts. The inherent nature of the advertisement industry does not transmit the feeling as it is and that causes confusion in the viewer' side. In other respects, HP ad outperformed at the conveying the emotion to the perceiver than Heineken did according to the design and communication experts.

After having the approval of experts, the process has been finalized for the stimuli selection.

4.3. Survey Design

The reactions towards video advertisements are planned to register via FaceReader and Self-Assessment instruments. Following the exposure of stimuli, the respondents are requested to fill the close-ended questions in a structured survey.

The questionnaire includes 25 questions. Starting with demographic questions, after each video the subjects are directed 22 questions in the laboratory survey and 33 questions in survey spread online. Following the video, the survey continues asking the affective attitudes of the subject first. The items of affective attitude questions were borrowed from the study conducted by Horvat et. al (2015) McDuff & Soleymani (2017) and Li, Walters, Packer, Scott (2017). The affective attitude questions were designed in two sorts: discrete emotion (McDuff et. al, 2017; Li, Walters, Packer, Scott, 2017) and dimensions of emotion which is "valence-arousal-dominance dimensions" (Horvat et. al,2015). For the VAD dimension questions, Self-Assessment Manikin (SAM) was chosen in a visual form. The subjects were shown the dimensions by means of a series of figures which was explained before (see 3.2.5). This method has been used in many advertising researches as detailed in the related section (Morris, Woo, Geason and Kim, 2002; Chen, Gao, Lv, Qie and Ma, 2018; Li, Walters, Packer and Scott, 2017).

About the purchase intent and cognitive processing, the survey involved several questions for each attitude. For instance, in the questionnaire filled in the lab settings the number of 5-point scale questions regarding "attitude towards brand" is four and in the online environment it is eight. Furthermore, the questions of questionnaires are selected based on the studies measuring the same aspects: purchase intent (Derbaix, 1995; McDuff et. al, 2015), attitude towards brand (Li, Walters, Packer and Scott, 2017; Bian and Forsythe, 2012), attitude towards ad (Derbaix, 1995; McDuff et. al, 2012; McDuff et. al,2015; Lewinski et. al, 2014), customer satisfaction (Peterson & Wilson, 1992), brand awareness (Bui et. al, 2011; Han et. al, 2011).

Appendix B and C show the both questionnaires involving all questions.

4.2 Experiment Design

Completing the stimuli selection and questionnaire design, the experiments were held in both laboratory and internet environments. Next chapters cover the details regarding methods used in this part of the empirical study.

Data Collection in Laboratory Settings

The experiments were performed in Politecnico di Milano, the Pheel Lab (Physiology, Emotion, Experience Lab).

A group of participants (18) were recruited by Pheel Lab and the other group of participants attended the experiments voluntarily (8). The subjects with the average age group of 25-34 years were native Italian speakers. The total number of subjects were 26; however, 2 participants had to be excluded from the analysis because of too many missing facial values. Of 24 subjects, the gender distribution was 13 Female and 11 Male with the large wide of educational background. 11 out of 24 were able to comprehend the content of the advertisement due to speaking the English language. The group of subjects that can understand English are relatively younger, possessing a higher level of education.

After arriving at the lab, the subjects were asked to give their consent to be videotaped during the study, warned about not covering their faces with hands for the software to perceive the emotional status of face and the participant were also informed about the approximate duration of the study alongside basic information how the experiment was about to take place as a basic information set.

Before the experiments took place, each subject was placed on a comfortable seat viewing a computer monitor which was set in front of them on a table. The door was kept close for each subject to feel comfortable and not disturbed.). The subjects were requested to use the headsets for the stimuli. During the study, the facial reactions towards the stimuli shown were monitored through the camera “Logitech HD Pro Webcam C920” and were recorded by FaceReader developed by Noldus (version 7.0 The subjects were taken one by one and the experiments lasted three days.

The experiments followed three stages (Fig. 4-6), namely in the beginning the viewers were recorded watching a gray screen in order to get their neutral facial expression. As the second stage, the viewers were asked to fill the personal questions that measured the demographic features of the subjects. In the same phase, following the personal questions they were requested to watch the stimuli and answer the related questions. The sequence of videos was randomized to avoid ordering effects. The experiment for each individual followed the same structure which starts with demographic questions, follows with the exposure of the first random stimulus, continues with the survey including the questions regarding the video shown and then the experiment goes on with second random video, questions related to the product promoted in second video, so on so forth. After the viewers were exposed the six videos and the full questionnaire, in the third phase, they were again asked to look at the gray screen for another minute with the same reason in the first stage. During the time the subjects were exposed the study, the timing for each stimuli and blank screens were gauged with the aim of tracking the change of emotional state in their facial expressions. Each stimulus included 12 questions intending to measure change of purchase intention, brand awareness, change of the attitude towards brand, advertisement likeability, advertisement effectiveness, to name a few. The experiment took approximately 15-20 minutes for each viewer.

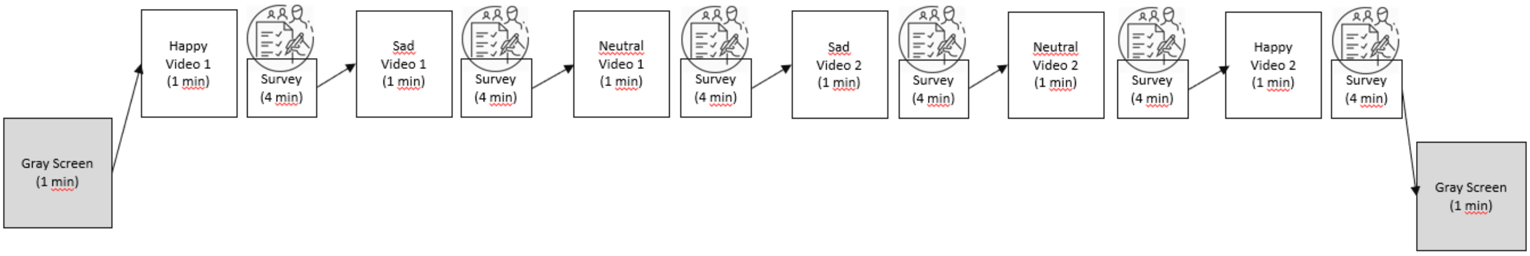


Figure 4 6 : The timeline of experiment for each subject. The sequence of dominant emotion (happiness/neutral/sadness) were randomized. In the figure, it is shown one of the randomized combinations of exposure stimuli.

Data Collection over Internet

Crowdsourcing is a means to reach a larger group of people from diverse populations and for this purpose, the expanded version of questionnaire was conducted on Amazon's Mechanical Turk platform. Amazon Mturk is a platform used by some scholars for the kind of study I conducted (Lewinski et. al, 2014; McDuff & Soleymani, 2017; McDuff et. al, 2012). The main advantage of using crowdsourcing platforms for data collection, it is more efficient in terms of both time and cost than in laboratory environment. For this study, the test participants were recruited in exchange of paying 0.35$ for completing the task.

Using an online platform to reach a big number of people poses few problems. One is to guarantee that the subjects complete the task. This was avoided by the help of platform in the beginning phase. The chosen platform asks test participants paste a code which appears after completing the survey. The second is the problem of being sure of all participants watching the commercials. To overcome this issue, the analytics of the Youtube account that was created for this study was monitored with filtering the experiment date and average watching duration. However, it was not certain whether the participants watch some ads more than once and the others not. This would be avoided to some extent if the crowdsourcing platform and/or Youtube was providing the IP information of the participants. Other doubts regarding if they paid attention to the videos or not do also exist. But, on the other hand, this would be an advantage considering that test participants answered the questions in natural settings. It is a way to avoid social desirability biases.

CHAPTER 5: ANALYSIS OF RESULTS

This chapter represents the data analysis based on three sources: output of facial reactions obtained through the FaceReader, the results of 24 self-assessment report of the stimuli gathered in the laboratory environment and the results of the 226 questionnaire distributed online to answer the research questions.

First, the raw data of the FaceReader's output was in a form that gives information regarding six basic emotions and the emotional dimensions of valence and arousal. The outcome obtained through FaceReader of Noldus is shown in the Fig. 5-1. The tool can record over 30 frames per second, labelling the type of emotion detected on the participant's face real-time. Different colors shown bottom side of the figure imply the different values for basic emotions.

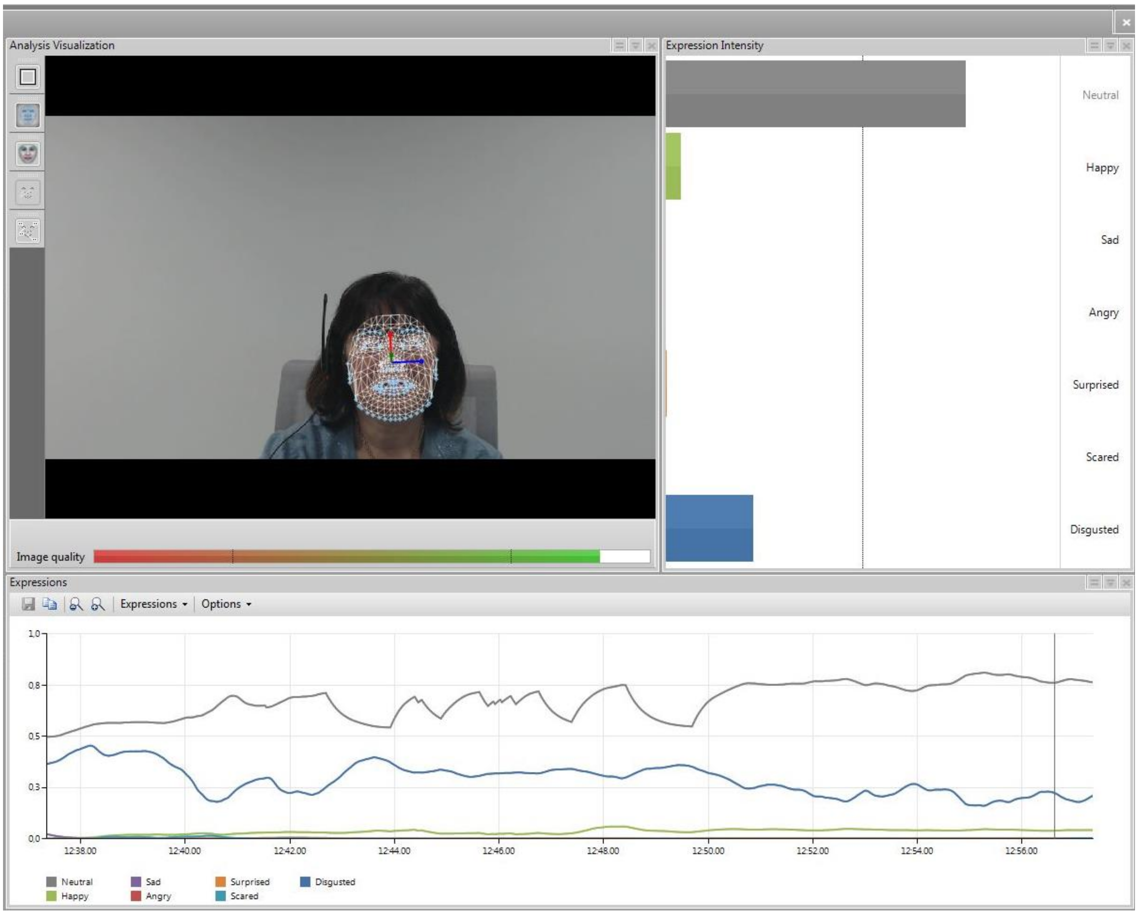


Figure 5 1 : The graphic output of FaceReader for Participant 3. Distinct colours correspond to basic emotion

5.1 Data Cleaning

The data obtained for this study was immense due to collecting six basic emotion recordings for 30 times each second. For this reason, the average value of each second was calculated and processed the data accordingly. Additionally, just like advised in the articles of the FaceReader by Noldus, the emotional state changes were calibrated by taking into account the neutral states of the participant. For instance, a person was detected with a high level (0.4) of sadness in the most neutral image. This signifies that when the value of sadness exceeds 0.4 for the subject, the participant was perceived as ‘sad’.

The data collected through questionnaires were also subjected to data cleaning process. The number of test participants reached over 350 in the beginning; however, given the work in progress information from platform, around 100 participants work were excluded owing to completing the task in less than 12 minutes. 226 test subjects completed a big part of the survey (more than 80% of it).

5.2. Descriptive Analysis

After computing the incremental change in the facial expressions, the survey results were processed through SPSS statistical package.

- The gender, age and education level distribution of the subjects joined the experiments in the laboratory are as following:

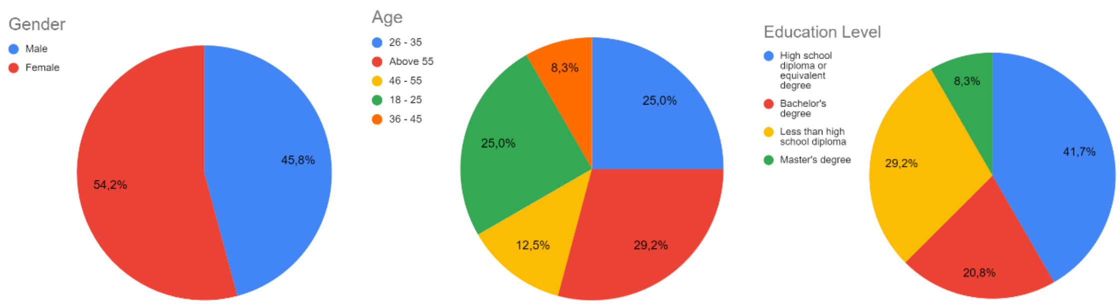


Figure 5 2: Educational Background, Age and Gender distribution of Experiment Participants (24 subjects)

- The gender, age and education level distribution of the respondents recruited through online channel are as following:

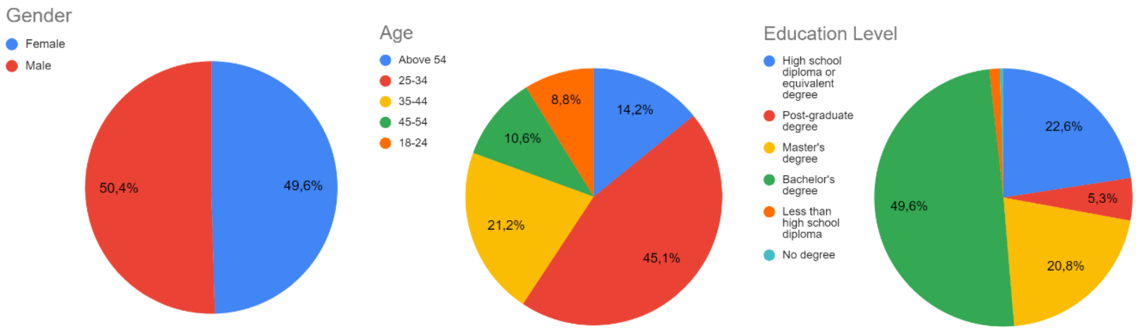


Figure 5 3: Educational Background, Age and Gender distribution of Online Survey Study Participants (226 subjects)

The gender distribution of the questionnaire collected through online sources was 112 female 114 male with average age group of 26-35. The test participants of this platform were from numerous parts of the world (Fig. 4-7) which was detected using IP information to make the statistics. The survey took approx. 15 minutes to complete. The average time spent for the survey changes for each subject from 12 to 30 minutes as both platforms provide the "work-in-progress" information (Surveymonkey and Amazon Mturk). On the other hand, the average completion of task in laboratory is between 25-30 mins which is almost the same with internet settings.

Figure 4 7: The map indicates 226 test participants' geolocation distribution, based on their IP address.

The participants of both environments are shown in absolute numbers as follows:

		Lab. Settings		Online Environment	
Variable	Item	Frequency	%	Frequency	%
Gender	Female	13	54%	112	49,6%
	Male	11	46%	114	50,4%
Age	18 – 25	6	25%	20	9%
	26 – 35	6	25%	102	45%
	36 – 45	2	8%	48	21%
	46 – 55	3	13%	24	11%
	Above 55	7	29%	32	14%
Education Level	No degree	0	0%	1	0%
	Less than high school degree or equivalent	7	29%	3	1%
	Highschool degree	10	42%	51	23%
	Bachelor's degree	5	21%	112	50%
	Master's degree	2	8%	47	21%
	Postdoctoral degree	0	0%	12	5%
Total		24	100%	226	100%

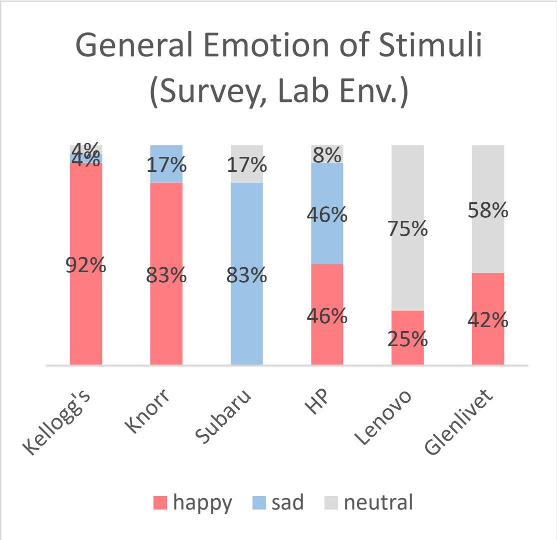
Table 7: The table indicates the distribution of demographic characteristics of the subjects who attended the experiments held in lab and internet settings.

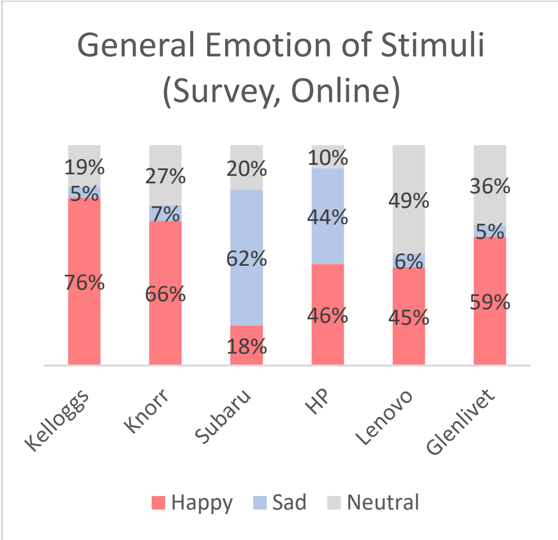
5.3. General Perception of Emotion Prediction from Facial Reactions and Self-report measures

In this part of analysis of experiments, it was aimed to find out how the subjects perceive general emotion of the stimuli. As expressed before, the subjects were asked to fill the survey which includes questions assessing their affective attitudes towards the videos. The affective attitudes were asked in two forms: basic emotion and VAD dimensions of emotions. Alongside receiving the cognitive answers of subjects via surveys in lab and online environments, the subjects were also recorded via the FaceReader in lab settings. Using two instruments enabled us to collect the data regarding basic emotions and VAD dimension of emotions of the video commercials. In the following two sections, the outcome of this research will be detailed and interpreted.

5.3.1. Specific Emotions

Following the stimuli, the participants classify the general perceived emotion of stimuli as one of the three basic emotions which are happiness, sadness and neutral. The figures in Fig. 5-4 visualize the outcomes of this question. Whilst the first and second bar charts demonstrate the results of surveys, the facial reactions are indicated in the last bar chart.





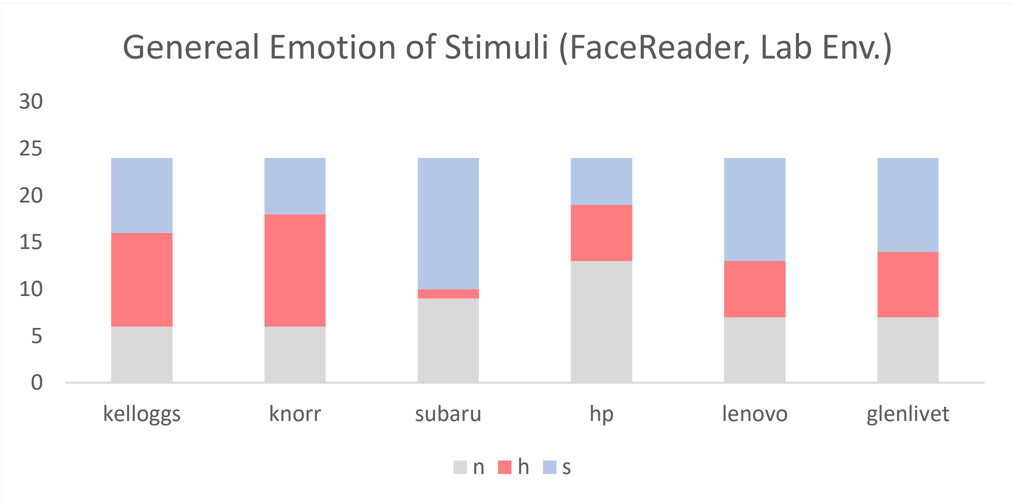


Figure 5 4: The column charts display the self-reported measures of perceived happy/sad/neutral ads in the lab settings (up-left) and internet(up-right). The chart in the bottom show the basic emotions revealed by facial reactions

The data collected from FaceReader are values of six basic emotions and valence-arousal-dominance dimensions of facial expressions. The most dominant basic emotion among happy, sad and neutral facial reactions values were chosen as the emotion of face towards the stimuli shown for each individual.

Apart from the facial reactions, in surveys normally the ideal version of these charts could have revealed the sort of result that the bars for Kellogg's and Knorr ads are fully pink color, Subaru and HP are completely blue and Lenovo and Glenlivet are entirely gray color, this means that the ad groups are perceived completely different from each other by anyone. Nonetheless, as described in the section 3.1.3, in general human observers are detecting the emotion successfully maximum at 90% and very much often 60-80%. Knowing this, the outcome is as expected for laboratory settings and less successful for online environment. Facial reactions are good enough only to distinguish the dominant emotion of the video when the videos are compared to each other. For instance, even though first two videos are not fully perceived happy, these ads are happier comparing to sad and neutral ones. Neutral ads do not give a good result if we only look at the bar charts of neutral ads and decide the emotion of the ad. It should be not ignored that during the exposure to the stimuli, the subjects could not perform fully sad or neutral or happy emotions. The fluctuations are also represented in the bar charts of face values. Comparing the three chart figures, the overall perception of success is acceptable in this regard.

Another interpretation is that the charts obtained through the experiments show that people are more prone to choose the emotion of the ads as happy albeit some were intended to evoke sad and neutral emotions. This is no big surprise due to having harmony of emotions also in the validation of stimuli part. The results are in line in comparison of pre-survey and main surveys in that sense except “the Glenlivet” ad.

As bottom line, what captured my attention here is that whilst in the laboratory settings, the results are closer to the ideal version, in the online environment, this is slightly further. The reason for this could be that in natural settings people are exposed more stimulants than in a lab environment. Another reason could be the subjects are of different nationalities. While in the lab environment all subjects were Italian speakers, in online settings the test participants were from all around the world.

However, in order to prove the difference between the ad groups statistically, it was computed the Chi-square tests and followingly post-hoc and Bonferroni correction on SPSS. The results yielded show that the difference between ads is significant. The table 6 for more detail is shown below:

	Kelloggs		Knorr		Subaru		HP		Lenovo		Glenlivet		P value
Lab Survey - Happy	22	**(4,7)**	20	**(3,8)**	0	**(-5,15)**	11	(-0,22)	6	(-2,46)	10	(-0,67)	0,00
Lab Survey – Neutral	1	**(-3,01)**	4	(-1,55)	4	(-1,55)	2	(-2,52)	18	**(5,29)**	14	**(3,34)**	
Lab Survey – Sad	1	(-2,33)	0	(-2,87)	20	**(7,89)**	11	**(3,05)**	0	(-2,87)	0	(-2,87)	
Online Survey – Happy	108	**(6,34)**	101	**(3,8)**	26	**(-8,74)**	71	(-1,52)	66	(-1,88)	92	(1,91)	0,00
Online Survey – Sad	7	**(-5,16)**	11	**(-4,68)**	88	**(12,89)**	67	**(7,46)**	9	**(-4,91)**	8	**(-5,42)**	
Online Survey – Neutral	27	(-2,38)	42	(0,03)	29	(-2,02)	16	**(-5,14)**	73	**(6,63)**	57	(2,83)	

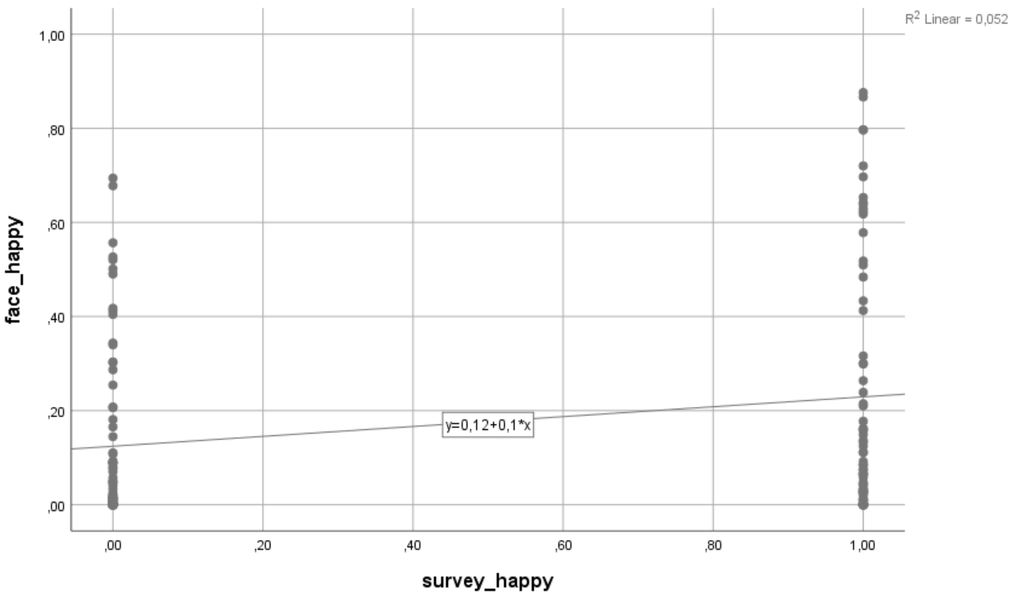
Table 8: Contingency table, Count, Adjusted residual, results of Chi-Square Tests

In addition to have a p value (0.000) that shows the difference is significant considering the significance level is ,05 , the Bonferroni correction indicates that the emotion perception of stimuli is significantly different for six commercials. The adjusted residuals reveal where the difference is coming from. As a way to interpret this table, in the laboratory environment, Kellogg’s commercial is perceived happy by a larger group of people in the sample than Lenovo or Glenlivet ads. Looking at the table this way, it is seen that there is a distinction among commercials in terms of the emotions the commercials evoke.

The similar result is obtained for the data obtained through online sources and with only a bit smaller difference between the groups. However, the test is significant for separating the ads according to basic emotions.

About the comparison of facial reaction and self-reported perceived emotion, the basic emotions can be read from the facial reactions if the emotion evoked the ad is happy. In the figures below, it can be seen that when the stimulus is “happy” stimulus, the “happiness” level of facial reaction is significantly higher. The association is moderate between happy expressions and happy videos variables. In the table 7, the associations between facial reactions and the stimuli are shown as output of Pearson’s correlation test. There is also a statistically significant result between sad videos and sad facial reactions. Nonetheless, this association does not yield any result because the sadness level of neutral advertisements is quite high. Even though the subjects perceive the neutral ads either happy or neutral according to survey results, the sadness level of face expressions is surprisingly much different than expected. Consequently, this contradiction does not tell that the FaceReader is not a tool to measure the sadness level of face; nevertheless, the stimuli reveals mix of emotions for subjects.

Overall, it can be noted that the FaceReader seems promising to read the emotion of happiness.



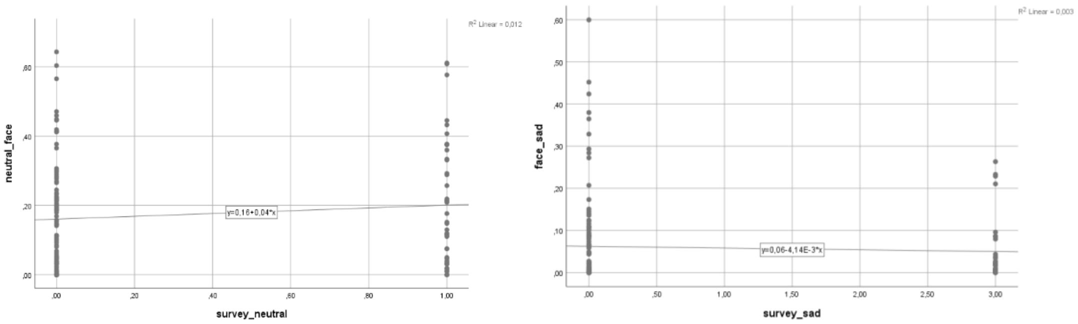


Figure 5 5 : The scatter plots of the correlation between facial reactions and perceived emotion of stimuli for happy, sad and neutral, respectively (Point biserial correlation analysis)

	Perceived Emotion of Stimuli (Self Report measures)						
	Happy		Sad		Neutral		
Facial Exp.	R_S	P value	R_S	P value	R_S	P value	N
Happy	,334	**,000**					144
Sad			-,261	**,002**			144
Neutral					,071	,396	143

Table 9: The correlation between facial reactions and emotion of stimuli for happy, sad and neutral, respectively

5.3.2 Valence and Arousal Dimensions of Emotion

The valence depicts if the emotional state of the subject is positive or negative and that is calculated as the intensity of 'happy' minus the intensity of 'sad ' expression. The valence values extracted from the FaceReader for three type emotional commercials are visualized as follows.

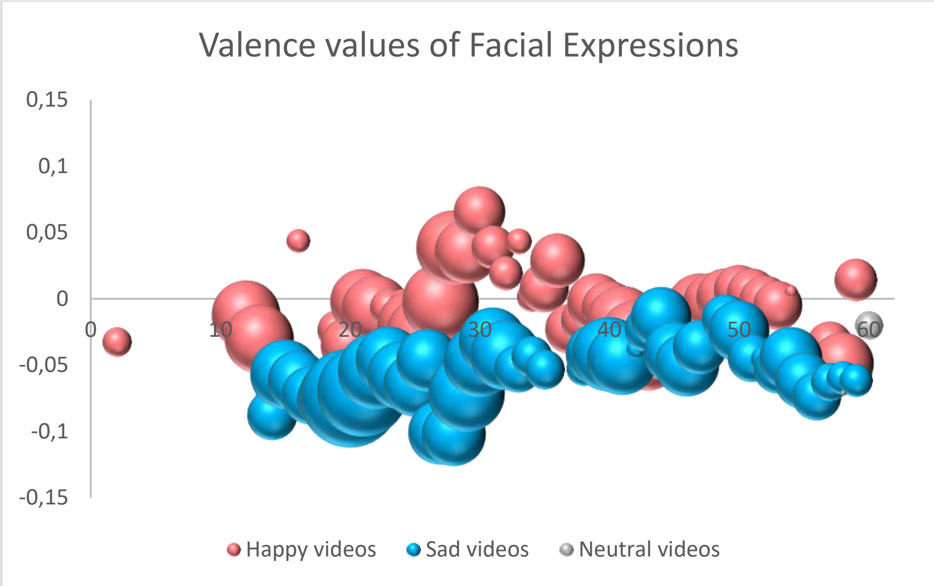


Figure 5 6: The scatter chart shows that the happy ads caught positive valence values of facial expressions while sad ads captured negative valence and neutral ones gained neutral valence in facial expressions. Horizontal axis shows the seconds of a minute video. Vertical axis is the value of the facial reactions.

This graph shows the clear distinction between happy, sad and neutral commercials in terms of valence values of facial expressions. On the other hand, the subjects declared the degree to which the each video is positive or negative with Self-Assessment Manikin (SAM) 5-point question. The usability of FaceReader was measured comparing the autonomic and self-report values.

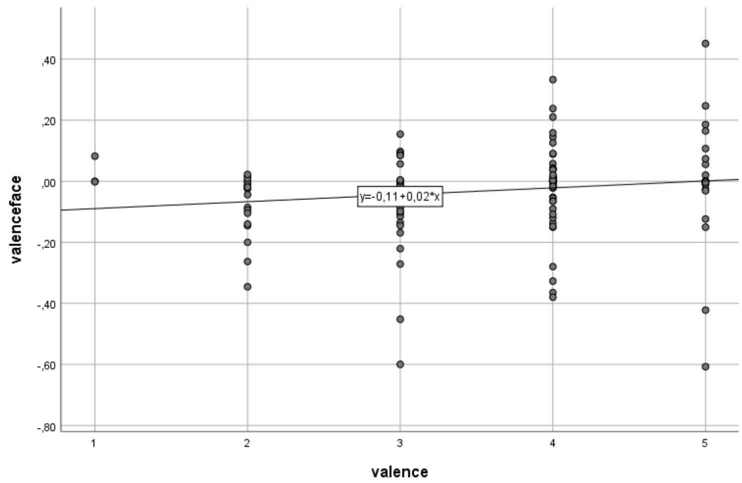


Figure 5 7: The scatter plots of the correlation between valence values of facial reactions and self+reported valence values for all commercials

The Spearman rank correlation was run using the values of both tools. It can be deducted that the FaceReader can be used instead of self-report to gauge the valence value of the stimuli from facial reactions. The association between two measures are at moderate level (R_S=,281 and p value=0,01)

The survey results collected from both environments also give the similar results for commercials designed to evoke similar emotions. For happy perceived ads, the valence values highest and for sad perceived ads, the valence values the lowest. However, what has captured my attention here is that people opt to perceive the videos more positive than their actual level of positivity. 1 to 5, 1 is very negative and 3 is neutral and in the survey results, it is seen that the lowest level of valence is anchored around 3, while neutral videos are little higher than 3. This might be expected considering that commercials rarely evoke extreme sad feelings. The results in a greater detail is visualized in the table 9.

	Kelloggs		Knorr		Subaru		HP		Lenovo		Glenlivet		χ2	P value
Face Arousal in SAM	-0,03	(0,16)	-0,05	(0,16)	-0,03	(0,13)	-0,01	(0,13)	0,01	(0,15)	0,00	(0,13)		0,71
Face Valence in SAM	0,02	(0,16)	0,02	(0,11)	-0,07	(0,1)	-0,00	(0,13)	-0,03	(0,18)	-0,02	(0,12)	14,43	0,019
Lab Survey Arousal in SAM	3,42	(0,18)	3,08	(1,02)	3,79	(0,78)	4,25	(0,85)	3,29	(0,95)	3,17	(1,01)	29,91	0,00
Lab Survey Valence in SAM	3,92	(0,88)	3,79	(0,59)	2,75	(1,03)	4,04	(1,12)	3,42	(0,65)	3,50	(0,88)	28,20	0,00
Online Survey Arousal in SAM	3,74	(1,15)	3,45	(1,19)	3,34	(1,18)	3,26	(1,28)	3,75	(1,03)	3,21	(1,14)	31,87	0,00
Online Survey Valence in SAM	4,27	(0,85)	3,92	(0,96)	3,04	(1,22)	3,82	(1,07)	3,73	(1,03)	3,66	(1,04)	85,57	0,00

Table 10: Mean values, standard deviations, results of Chi Square and Friedman Tests for facial values and both questionnaires. Arousal and valence are scales from 1 to 5 for surveys and -1 to 1 for facial reactions

To prove the statistical distinction between the valence values of facial expressions, it was run Friedman test (because the data violated the normality assumption). The test produced p value as 0.019 which is lower than the level of significance that is taken 0.05 and so it can be deducted that the valence values obtained through facial signals are significantly different from each other (Table 9).

Overall, in comparison of valence values to basic emotions, this study finds Valence values are more reliable when the stimuli reveal mix emotions.

In the second question of Self-Assessment Manikin (SAM), the subjects were asked to rate the arousal level of the stimuli alongside being recorded arousal level of their facial reactions. The results for the facial signals did not yield a significant result considering not having sufficient data. On the other hand, the outcome coming from survey parts for arousal values do not coincide with. The reason could be that the subjects do not comprehend the meaning of arousal or each subject interpret differently about what it really means. However, in literature this was reported as a disadvantage of using survey for VAD dimensions of emotion due to the fact that people do not comprehend valence, arousal and dominance terms albeit they are explained the meanings or shown the drawings regarding dimensions (Li et. al, 2017).

5.4. Reliability

The questions for measuring affective, conative and cognitive attitudes of subjects were multiple. To separate the dimensions, Principal Component Analysis dimension reduction technique was employed using SPSS. The PCA was run following a varimax rotation and taking into account each factor with eigenvalues greater than 1.0. It was found distinct dimensions for conative, cognitive and affective attitudes, namely Purchase Intent, Emotional states and Change in Attitude towards Brand. The multiple questions of each item are listed below:

Following each commercial, the subjects were asked to answer the following 'valence-arousal-dominance' questions (Horvat et. al, 2015):

Valence

Q: "Please evaluate the valence of the video(Valence describes the extent to which an emotion is positive or negative)"

Very positive		Neutral		Very negative
1	2	3	4	5

Arousal

"How much were you aroused with the ad?(Arousal means the extent to which you were evoked by the video)Very Active refers to the states like "excitement, euphoria, excitation, rage, agitation, anger"Very Quite refers to the experiences such as "relaxation, tranquility, idleness, meditation, boredom"

Very Active		Neutral		Very Quiet
1	2	3	4	5

Dominance

"Please evaluate the dominance of the video.(Dominance refers to the power of the emotion, extent to which the video denotes something that is weak or strong.)Very Dominant picture refers to states like "feeling lack of control, withdrawal, resignation, submission, subordination, intimidation" Very Weak refers to the states such as "feeling in control of situation, being important, recognized, decisive, influence"

Very Weak		Neutral		Very Dominant
1	2	3	4	5

Following each commercial, the subjects were asked to answer the following 'attitude towards brand' questions which were computed as a mean variable:

Q: "I recognize the values transmitted by this brand"
Q: "In the past, I have had positive experience with this brand"
Q: "This brand is my favourite brand"
Q: "I have a positive consideration for this brand"
Q: "This brand offers me good value for the price"
Q: "I am very committed to this brand"
Q: "I think this brand deserves my effort to maintain a relationship"

Strongly disagree		Neutral		Strongly agree
1	2	3	4	5

The 'purchase intent' question forwarded to the viewer subsequent to having watched each ad is:

Q: "If I needed to buy this product, I would probably buy the product of the brand shown in the video."

Q: "The product shown in the video impressed me particularly"

Strongly disagree		Neutral		Strongly agree
1	2	3	4	5

The ‘customer satisfaction’ question forwarded to the viewer subsequent to having watched each ad is:

Q: “I was satisfied with the product I bought”				
Q: “I would buy again products of this brand”				
Strongly disagree		Neutral		Strongly agree
1	2	3	4	5

Subsequently, in order to measure the consistency of questions belonging to these dimensions, the Cronbach’s a coefficient was calculated. Table 5-8 reveal the Croanbach’s a coefficient values. As seen in the table, their reliability is acceptable considering the general accepted level of reliability (.70). Afterwards, the common variables were described with taking the median values of the responses for the questions. The ad effectiveness and purchase intent scores were measured using the new common variables.

Measures	LAB Cronbach's Alpha (a)	ONLINE Cronbach's Alpha (a)
Attitude Towards Brand	,843	,937
Customer Satisfaction	,886	,869
Purchase Intent	,748	,819
Attitude Change Toward Brand	,772	,868

Table 5 8: Cronbach's a coefficient of measures

5.5. Ad Effectiveness Prediction

After assessing the performance of the software and survey instruments so as to see the general perceived emotion of the stimuli, it is moved on with the association of ad effectiveness and its predictors. In order to answer research question regarding the impact of the predictors on advertising effectiveness, I have analyzed the advertisement liking, re-watchability of the commercial and the attitude towards the brand as components of the advertising effectiveness. In other words, the analysis here is about how ad effectiveness is affected by cognitive and affective attitudes of individuals.

In the first part of this subsection, it was assessed whether it is feasible to measure the advertising effectiveness observing the face emotion reactions towards the stimuli. In this case, the higher level of liking the commercial is formulated as higher ad effectiveness.

Continuing with the affective answers from the survey questions, it was assessed the relationship the perceived emotion of the ad and ad liking with the purpose of seeing the impact of perceived emotion type on ad liking.

Moreover, it was analyzed the degree of predicting ad effectiveness by looking at the association between ad liking and the independent variables such as attitude towards brand, rewatchability.

In order to measure the ad liking and rewatchability, the questions forwarded to the viewers are as follows:

Following each commercial, the subjects were asked to answer the following 'liking' question (McDuff et.al, 2015):

Q: “How much did you LIKE the ad you just watched?”				
Very High 1	2	Average 3	4	Very Low 5

Following each commercial, the subjects were asked to answer the following ‘rewatchability’ question (McDuff et.al, 2012):

Q: “Would you like to watch this video again?”				
Very highly likely would watch it 1	2	Might watch it 3	4	Very highly unlikely would watch it 5

5.5.1 General Perception of Emotion and Ad Liking

In this sub-section, the association between affective attitudes and ad effectiveness is assessed through observing facial expressions and self-report answers, respectively.

5.5.1.1 Ad Liking Prediction from Facial Expressions

In order to fathom the usability of face emotion recognition software to read the ad liking in the face, Spearman-s rank correlation was run using SPSS. The face values of the participants are calculated with the average value of happiness, sadness and neutral states. However, for sad commercials the sad level of facial reactions is captured when the message of the commercial is transmitted. Due to the fact that, a sad advertising is rarely revealing sad emotion for the whole duration; rationally thinking, the sad commercial values were taken into account for the time the viewers had emotional moment and comprehended the message of the videos.

The hypothesis to test the association between two variables is as following:

Hypothesis 1: There is no monotonic association between frequency and intensity of facial reactions of happiness, neutrality, sadness and self-assessed measures of likeability of happy, sad and neutral perceived ads.

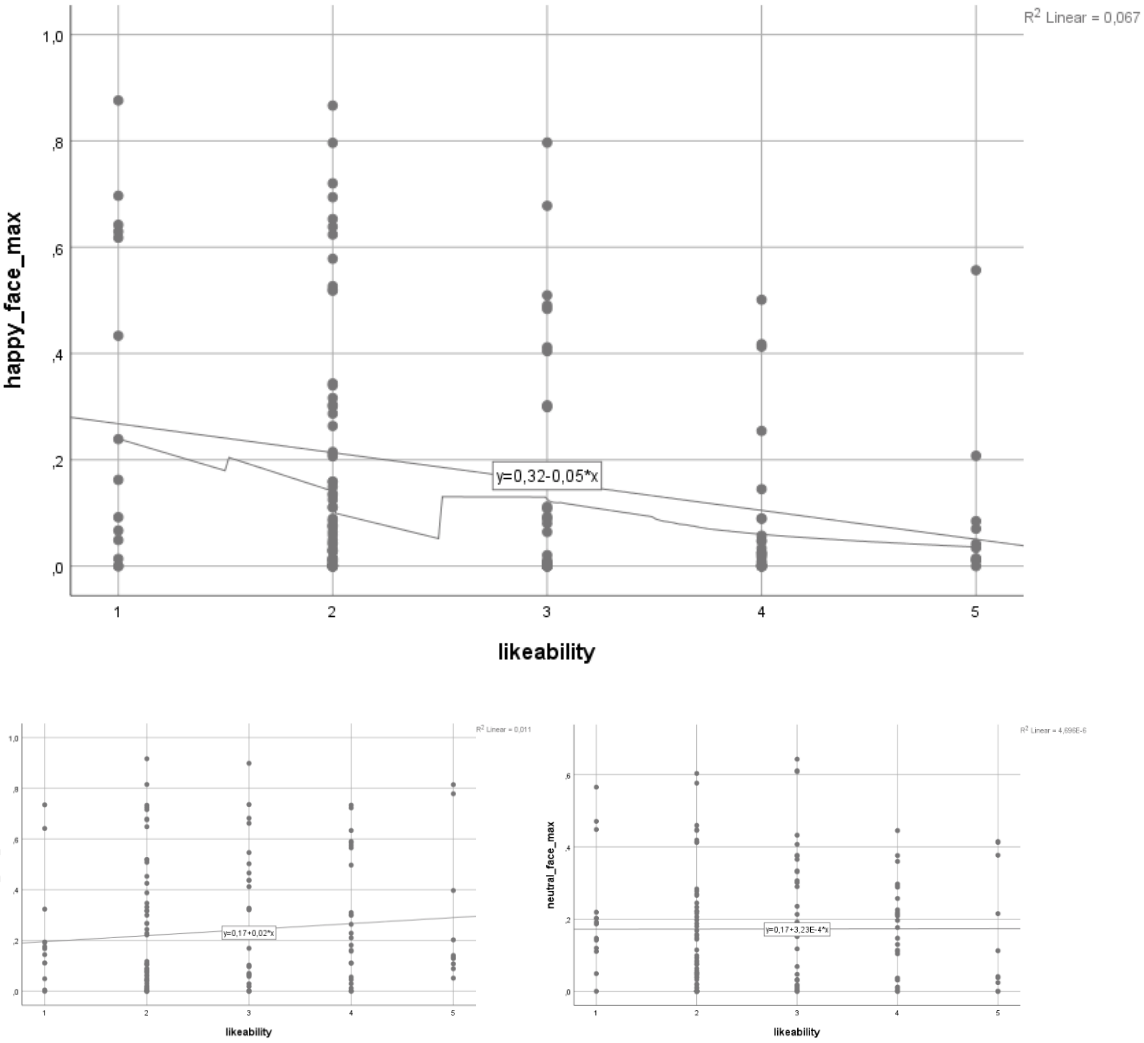


Figure 5 9: Figures show the scatterplots of the correlation between facial expressions of happiness/sadness/neutral and advertising likeability, respectively. (1, I like the ad very much & 5, I like the ad very little)

The correlation which is at a weak and moderate level is significant for both happy and sad face expressions and likeability, respectively. Since the p values (0.026 and 0.001) are less than level of significance (0.05), the null hypotheses is rejected for the happy and sad emotion values. This means that there is a positive correlation between frequency and intensity of facial reactions of happiness and self-assessed measures of likeability of the commercials. There is a negative correlation between likeability of ad and sad emotion values, hence when the sad facial reaction increases, the likability of the commercial is statistically lower. The real reason behind the negative correlation, the subjects like the neutral ads far less than happy and sad ads. However, the sad level of facial reactions towards the neutral ads is high. In consequence, the correlation yields significantly negative. Furthermore, it is p value of 0.980 as insufficient evidence against the null hypothesis that two variables emotion and ad liking are not associated for the neutral emotion values.

	Advertisement Likebility		
Stimuli	R_S	P value	N
Happy	-,187	**,026**	144
Sad	,266	**,001**	144
Neutral	-,060	,980	143

*Table 11: Correlation between facial reactions and Advertisement Likeability *Two-tailed*

5.5.1.2. Self-reported Measure of General Emotion Perception and Ad Liking

In this part of advertising effectiveness analysis, the association was tested between the self-reported general emotion perception of stimuli and the ad liking. To put it simple, it was aimed to observe that does the subject like the commercial significantly more depending on perceived emotion type or not?

Hypothesis 2: There is no statistically significant association between the self-reported measures of liking the ad and self-reported happy, sad and neutral perceived emotion in the stimuli.

To prove this statement, Chi square test was run on SPSS statistics. The bar charts and results are as follows for the lab survey:

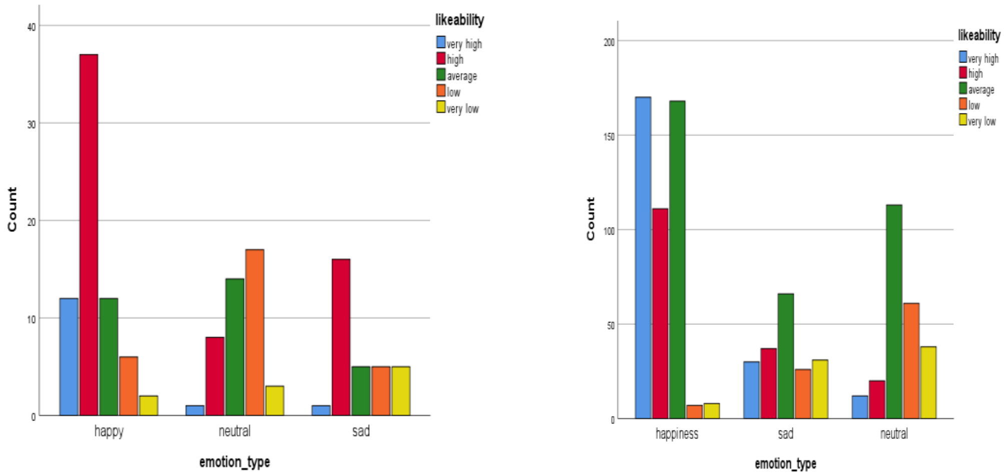


Figure 5 10 The clustered bar charts show the group categories in three emotion states and the frequency of liking counts in these groups

		LAB				ONLINE			
		Happy	Sad	neutral	total	happy	Sad	neutral	total
Likeability	very high	**17%**	**3%**	2%	10%	**37%**	**16%**	5%	24%
	High	**54%**	**50%**	19%	42%	**24%**	**19%**	8%	19%
	Average	17%	16%	33%	22%	36%	35%	46%	39%
	Low	9%	16%	**40%**	19%	2%	14%	**25%**	10%
	very low	3%	16%	**7%**	7%	2%	16%	**16%**	9%
	Total	100%	100%	100%	100%	100%	100%	100%	100%

LAB.	ONLINE

	N	Phi	Cramer's V	Sig.	N	Phi	Cramer's V	Sig.
Emotion - Liking	144	,515	,364	,000	898	,517	,366	,000

Table 12: The association between Self-reported measures of General Emotion Perception of Stimuli and the level of advertisement liking. (Chi Square)

Hypothesis test summary table proves that the null hypotheses must be rejected . From the bar charts and tables clearly seen that when the ad is stated a happy ad, it is more likely to be enjoyed. Followingly, the sad perceived stimuli also have a positive impact on the advertising effectiveness although not as much as happy ads. On the contrary, the neutral perception of emotion in the video commercials are more likely to be less liked comparing to the others. The details of the test are shared in the table. Cramer's V outcome indicate that the strength of association between the perception of emotion and ad liking is moderate which is statistically significant.

5.5.3 Attitude Towards Brand and Ad Liking

The other predictor of ad effectiveness is the attitude of subjects toward the brands promoted. The association between these variables is tested using Goodman and Kruskal's Gamma measures on SPSS statistics. The results are as following:

	LAB.				ONLINE			
	Gamma (γ)	Asymptotic Standard Errora	Approximate Tb	Aprox. Sig.	Gamma (γ)	Asymptotic Standard Errora	Approximate Tb	Aprox. Sig.
Happy	-0,485	,122	-3,499	0,000	-0,371	0,052	-6,893	0,000
Sad	-0,448	,207	-2,229	0,000	-0,491	0,064	-6,872	0,000
Neutral	-0,500	,193	-2,374	0,000	-0,478	0,063	-6,874	0,000
Total	-0,520	,081	-5,945	0,000	-0,522	0,029	-16,379	0,000
N of Valid cases	112				840			

Table 13: The Gamma measures of association between attitude towards brand and ad effectiveness (in the detail of partial distinction of emotions)

The tests run using two surveys collected in lab and natural settings and both resulted in statistically significant. The attitude towards brand explains the variance of ad liking at around 52%. However partial association is also significant in the perception of emotion of video stimuli. When the video is perceived happy, the predictor power of the attitude toward brand variable drops. This means that, the variable " perceived emotion type of stimuli" also might pose an impact to explain the ad effectiveness.

5.5.4. Rewatchability and Ad Liking

A close relationship exists between rewatchability of the ad. The association between the attitude towards brand and ad liking is tested using Goodman and Kruskal's Gamma measures on SPSS statistics. The results reveal that when subjects consider rewatch the commercial, it is because the higher possibility of "liking ad" is statistically evident. In the partial distinction of emotions evoked by the stimuli, it is seen that when the stimuli are perceived happy, the rewatchability variable explains the variance of ad effectiveness at 72% which would indicate a strong effect.

	Gamma (γ)	Asymptotic Standard Errora	Approximate Tb	Aprox. Sig.
Happy	-0,722	0,034	-17,496	0,000
Sad	-0,664	0,054	-10,356	0,000
Neutral	-0,76	0,043	-13,199	0,000
Total	-0,765	0,02	-30,722	0,000
N of Valid Cases	898			

Table 14: The Gamma measures of association between rewatchability of ad and ad effectiveness (in the detail of partial distinction of emotions

Regression:

In the last part of ad effectiveness, it was run a regression analysis to see the impact of cognitive and affective processing. With this object in mind, the ad liking as an independent variable and the attitude toward brand, general emotion perception of stimuli and rewatchability as dependent variables, Ordered Probit regression was run through SPSS statistics tool.

Hypothesis 3: There will be no significant prediction of Advertisement Effectiveness by Attitude towards brand, rewatchability and general perceived emotion type of the ad

Since the beginning, the statistics tests were run on two separate data: lab environment findings and online questionnaire findings. However, estimating the lack of data for this analysis, the output of the test was not significant to reject or not reject the hypothesis with the data gathered in lab. On the other hand, with a bigger sample obtained through internet, the test results are as following:

The dependent variable is: Advertisement Liking

	Laboratory Settings					Online Settings				
Variables	Estimate	Std. Error	Wald	df	Sig.	Estimate	Std. Error	Wald	df	Sig.
						-1,293	0,075	297,279	1,000	0,000
Attitude towards Brand	-1,048	0,285	13,523	1,000	0,00024	-0,522	0,082	40,260	1,000	0,000
Happy	-0,938	0,451	4,325	1,000	0,038	-1,383	0,185	56,037	1,000	0,000
Sad	0a			0,000		-0,756	0,201	14,115	1,000	0,000
Neutral	0,576	0,507	1,289	1,000	0,256	0a			0,000	

Table 15: The output of ordered probit regression between the dependent variable 'ad effectiveness' and its predictors: attitude towards brand, rewatchability and emotion. The impact of factors are sequenced from biggest to smallest as rewatchability, attitude towards brand and emotion.

The regression test is significant due to p value (,000) of the tests are lower than the level of significance (0,05). The interpretation of the regression shown in the table is that attitude towards brand, rewatchability and the type of emotion perceived through the stimuli does matter for the subject to like the commercial. Thus, the independent variables impact the advertisement effectiveness. Apparently, while the happy ads are significantly more important, the other types do not influence the ad effectiveness that much. The odds of happy perceived commercials considering ad liking to be too high was 0.351 (95% CI, 0.162 to 0.947) times that of sad perceived commercials, a statistically significant effect, Wald $\chi2(1) = 4.325$, $p = .038$. Having a negative estimate for happy perceived videos means that the subjects enjoy it more due to the fact that '1' is assigned as 'I like the ad a lot'.

As a whole study of impact on emotion type on advertisement effectiveness, the implication is that from facial reactions it can be read whether the subject would like the ad significantly. Also, liking the advertisement is affected by having a positive attitude towards brand, the positive approach to re-watch the ad, being exposed to a happy stimulus. Looking at the predictors in detail, it is seen that while rewatchability(76%) and attitude towards brand(52%) explain the ad effectiveness at a relatively better scale, the emotion(36% for cognitive responses, 19% facial reactions) has a little less impact to explain the variance of the ad liking.

5.6. Purchase Intention Prediction

Knowing the importance of emotion on consumer behaviour in an influential context, in this part of analysis of results, in order to answer the research question regarding the emotion impact on purchase intent,

- the relationship between facial reactions and future purchase intent
- The relationship between previous purchase intent and post purchase intent
- The relationship between customer satisfaction and future purchase intent
- The regression analysis of future purchase intent taking the independent variables self-reported perceived emotion of stimuli, customer satisfaction, previous purchase intent

The listed tests were run on SPSS. This analysis was computed in a similar way to advertisement effectiveness. The questions forwarded to the subjects are shown below. The median values of the questions were computed as new variables. The reliability results were depicted in the section of Reliability (5.2).

The 'prior purchase intent' question forwarded to the viewer subsequent to having watched each ad is:

Q: "Prior to this study, were you planning to purchase the product shown in the ad?"	
Yes	No
1	2

5.6.1. General Perception of Emotion and Purchase Intent

5.6.1.1 Facial Expressions and Purchase Intent

In order to measure the usability of face emotion recognition software for the purpose of predicting the buying intent automatically, the hypothesis 4 was tested using Spearman-s rank correlation. The nonparametric test was due to be used because the data violated the normality assumption. The facial expressions values are calculated the same way with the advertising effectiveness's values.

Hypothesis 4: There is no correlation between frequency and intensity of facial reactions of happiness, neutrality, sadness and self-assessed measures of purchase intent

	Purchase Intent Change		
Stimuli	R_S	P value	N
Happy	,094	,264	144
Sad	-,051	,547	144
Neutral	,11	,188	143

*Table 16: Correlation between facial reactions and Purchase Intent change *Two-tailed*

The output of correlation analysis between facial reactions and purchase intention did not give a significant correlation result. Also, in the study of Li et. al (2017) there was not proof regarding an association between purchase intent and facial reactions. However, it was run another association test between likeability of the ad and purchase intention, and it is found that there are significant and positive correlation. It is known that there is a correlation between facial expression and likeability as well. Based on this, assuming that there might be a correlation between two variables if it is obtained more facial data. The following table are the Chi-square test results for lab and online settings that show the correlation among purchase intent and likeability of the commercial.

			Laboratory Purchase Intent3					Total	ONLINE					Total
			strongly disagree	disagree	Undecided	agree	Strongly agree		strongly disagree	disagree	Undecided	Agree	Strongly agree	
Likeability	very high	Count	0	1	1	8	4	14	4	7	17	90	94	212
		% within likeability	0	0,071	0,071	0,571	0,286	1	0,019	0,033	0,08	0,425	0,443	1
		Adjusted Residual	-0,7	-1,1	-1,9	1,3	2,7		-3,5	-4,5	-5,4	0,6	10,9	
	High	Count	1	3	15	34	8	61	17	45	84	161	40	347
		% within likeability	0,016	0,049	0,246	0,557	0,131	1	0,049	0,13	0,242	0,464	0,115	1
		Adjusted Residual	-1	-3,5	-0,9	3,1	1,5		-2,2	0,7	1,7	2,8	-4,4	
	average	Count	0	3	14	13	1	31	18	20	22	13	4	77
		% within likeability	0	0,097	0,452	0,419	0,032	1	0,234	0,26	0,286	0,169	0,052	1
		Adjusted Residual	-1,2	-1,4	2,3	0,1	-1,3		5,6	3,9	1,6	-4,4	-3,2	
	Low	Count	2	16	8	2	0	28	5	11	39	85	28	168
		% within likeability	0,071	0,571	0,286	0,071	0	1	0,03	0,065	0,232	0,506	0,167	1
		Adjusted Residual	1,2	6	0	-4,1	-1,9		-2,4	-2,4	0,7	2,9	-0,8	
	very low	Count	2	3	3	2	0	10	22	25	29	16	2	94
		% within likeability	0,2	0,3	0,3	0,2	0	1	0,234	0,266	0,309	0,17	0,021	1
		Adjusted Residual	3	1	0,1	-1,4	-1		6,3	4,6	2,4	-4,9	-4,4	
Total		Count	5	26	41	59	13	144	66	108	191	365	168	898
		% within likeability	0,035	0,181	0,285	0,41	0,09	1	0,073	0,12	0,213	0,406	0,187	1

Chi-Square Tests

	Value	Df	Asymptotic Significance (2-sided)
Pearson Chi-Square	73,829a	16	0
Likelihood Ratio	71,157	16	0
Linear-by-Linear Association	46,622	1	0
N of Valid Cases	144		

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	284,204a	16	0
Likelihood Ratio	269,696	16	0
Linear-by-Linear Association	101,461	1	0
N of Valid Cases	898		

Table 17: The contingency table of ad liking and purchase intent (Chi square tests)

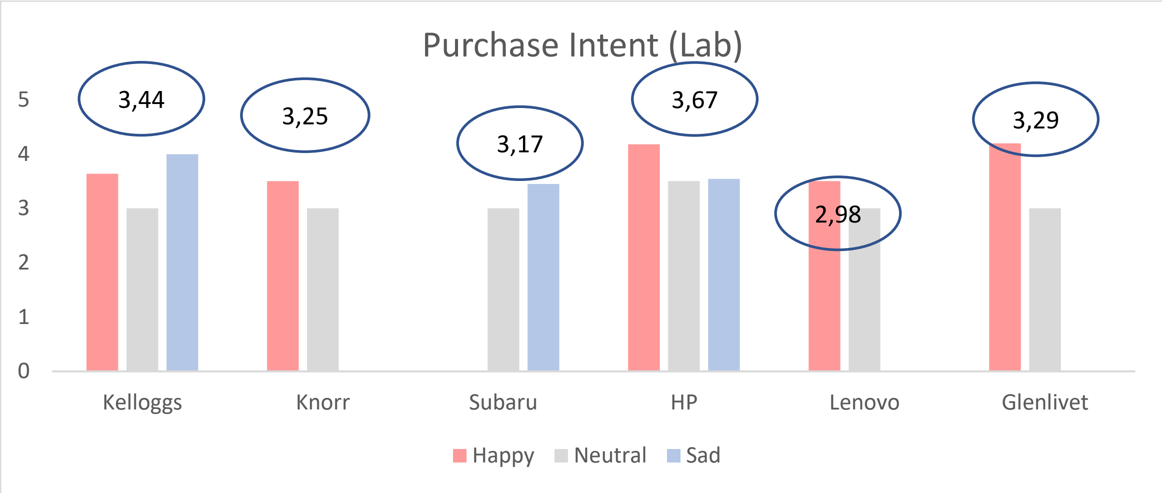
From the table, it can be seen that there is a tendency to make a purchase with higher probability when the likeability of the commercial is higher.

5.6.1.2. Self-reported Measure of General Emotion Perception and Purchase Intent

The purchase intent of viewers was measured through self-report questionnaires. The important point in this part is that the purchase intent does not depend on the emotion of the commercial or the commercial itself only. The need of product depends on some other parameters which will be detailed in the following sections.

Thus, in the graphs and figures the purchase intent scores for each advertisement are visualized. The outcome of the interpretation of data is that the commercials stimulated happy emotions could have a slightly higher positive impact on the viewers. However, some commercials evoke different kinds of emotions for different groups of subjects. Knowing this, it is also graphed the distinction of purchase intent of subjects depending on their perception of emotion in the videos. Seemingly, the subjects opt to see the commercial as a happy ad approach more positively to make a purchase in future comparing to the other subjects that state the ad is sad or neutral.

For both environments, it is first observed that when the stimuli were perceived happy, it is higher likely to persuade the viewer to make a purchase regardless of product or brand type. Second observation is that rarely sad perceived videos could pose an impact on purchase intent positively but the neutral perceived ads in general do not convince subjects about a possible future purchase. However these graphs do not give the weighted average of purchase intent. This means that for example Kellogg’s ad has a remarkable purchase intent score when it is perceived sad but number of subjects perceived it sad are not as much high as subjects who found it happy. The importance of distinction between emotion impact is analysed through Chi Square Test for Association using SPSS tools. The results are shown in the table 17.



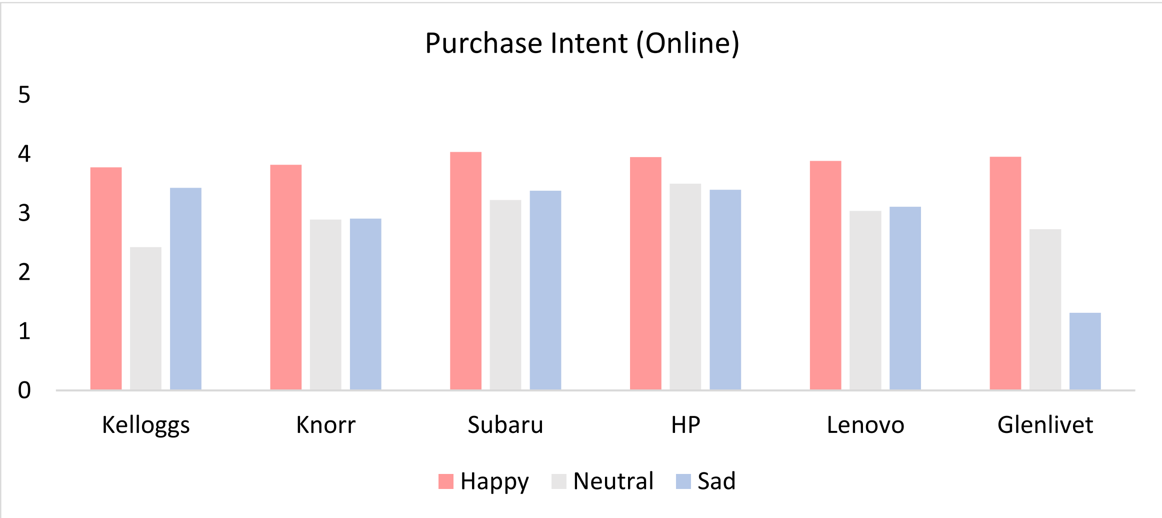


Figure 5 11: The purchase intent of subjects is shown in the bar charts for lab and online env. The emotion distinction of each commerciasl points out the significant difference in purchase intent depending on the perception of emotion of the video.

Purchase Intent	Lab				Online			
	Happy	Sad	neutral	total	happy	sad	neutral	total
1	1%	6%	**9%**	5%				
1,5	0%	4%	**3%**	2%	0%	5%	3%	2%
2	5%	10%	**18%**	10%	1%	0%	6%	2%
2,5	5%	6%	**10%**	6%	6%	19%	0%	8%
3	10%	19%	21%	15%	10%	12%	13%	11%
3,5	**15%**	**20%**	13%	15%	**14%**	**37%**	**22%**	23%
4	**29%**	**20%**	14%	23%	**16%**	**19%**	**22%**	18%
4,5	**19%**	**8%**	7%	13%	**38%**	**7%**	**28%**	26%
5	**17%**	**7%**	4%	11%	**7%**	**0%**	**6%**	5%

5,5					**7%**	**2%**	**0%**	4%
Total	100%	100%	100%	100%	100%	100%	100%	100%

	LAB.				ONLINE			
	N	Phi	Cramer's V	Sig.	N	Phi	Cramer's V	Sig.
Emotion – Purchase Intent	144	,511	,362	,002	898	,17	,295	,000

Table 18: The association between Self-reported measures of General Emotion Perception of Stimuli and the level of purchase intent. The purchase intent is scaled from 1 to 5 in which 5 shows the highest propensity of possess the product. (Chi Square)

The information regarding the tables are compiled taking the association between perceived emotion of stimuli and purchase intent score of individuals. The Chi Square test of association yielded statistically significant results which indicate that there is an association between two variables, however the strength of the association is moderate for both experiments held in lab and online environments.

5.6.2 Future Purchase Intent and Prior Purchase Intent

Another independent variable that influence the future purchase intent is prior purchase intent. The subjects were asked if they had planned to buy the product promoted in the video commercial before the study. Regarding answering "Yes" and "No" to this question, the subjects were divided into two groups and their post purchase intent score are demonstrated below (Fig. 5-12). It is clear that prior purchase intent has an observable impact on post purchase intent. People already planned to buy the product are closer to make this purchase from the brand promoted in the ad.

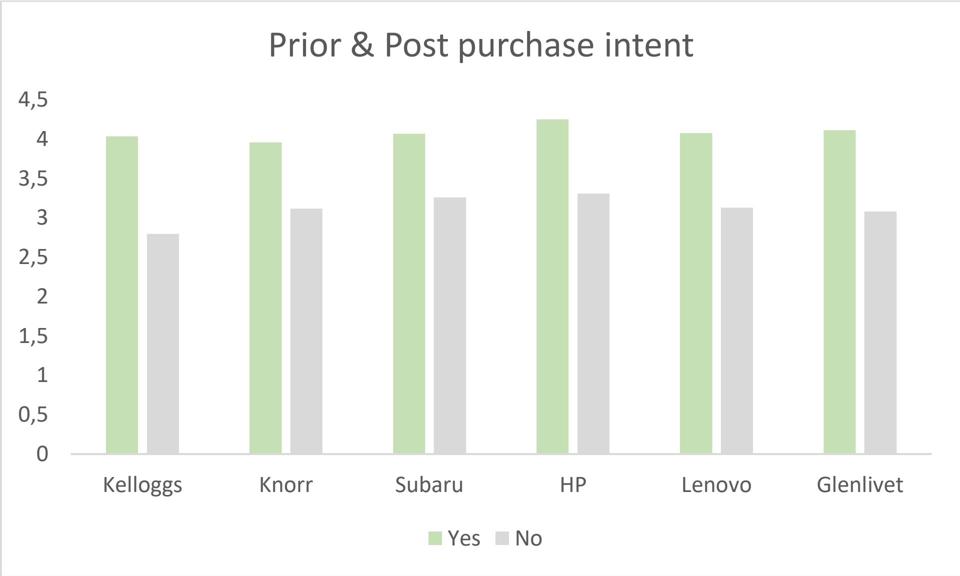


Figure 5 12: The post purchase intent is charted for six commercials. "Yes" signifies people that planned to purchase the product in the video before. "No" means they had no buying plan before.

The figures were created using the data collected from online sources because the data gathered in the lab was not relatively much less to make a reliable interpretation since there is many divisions in this analysis. Yet, the results were similar and the test results of tests for both environments are as following:

	Kellogg's	Knorr	Subaru	HP	Lenovo	Glenlivet
Online						
Yes	4,04	3,96	4,07	4,25	4,08	4,11
No	2,80	3,12	3,26	3,31	3,13	3,08
Lab.						
Yes	3,88	3,63	4,00	4,08	3,20	3,75
No	3,00	3,31	3,21	3,58	3,11	3,45

Figure 5 13: The average purchase intent scores of experiments conducted in both environments. Yes and No signify the existence of purchase intent prior to the study

	LAB.				ONLINE			
	N	Phi	Cramer's V	Sig.	N	Phi	Cramer's V	Sig.
Kellogg's	24	,814	,814	,014	142	,625	,625	,000
Knorr	24	,521	,521	,367	154	,446	,446	,000
Subaru	24	,772	,772	,026	143	,401	,401	,000
HP	24	,520	,520	,371	154	,548	,548	,000
Lenovo	24	,395	,395	,442	148	,468	,468	,000
Glenlivet	24	,469	,469	,626	157	,398	,398	,002
Total	144	,394	,394	,004	898	,451	,451	,000

Table 19: The association between existence of prior purchase intent and the post purchase intent. (Chi Square – Cramer's V)

The Chi-square for association test was run in order to fathom the degree of association between purchase intent scores post and prior to the study. In consequence, albeit it is proven the observable impact of prior purchase intent, in the test analysis results it is seen the degree of association between two variables. The prior purchase intent is associated to the post purchase intent by explaining the 45% of variance. In the lab environment, the test is significant enough to claim the distinction however it is not said directly for the impact on each video stimuli.

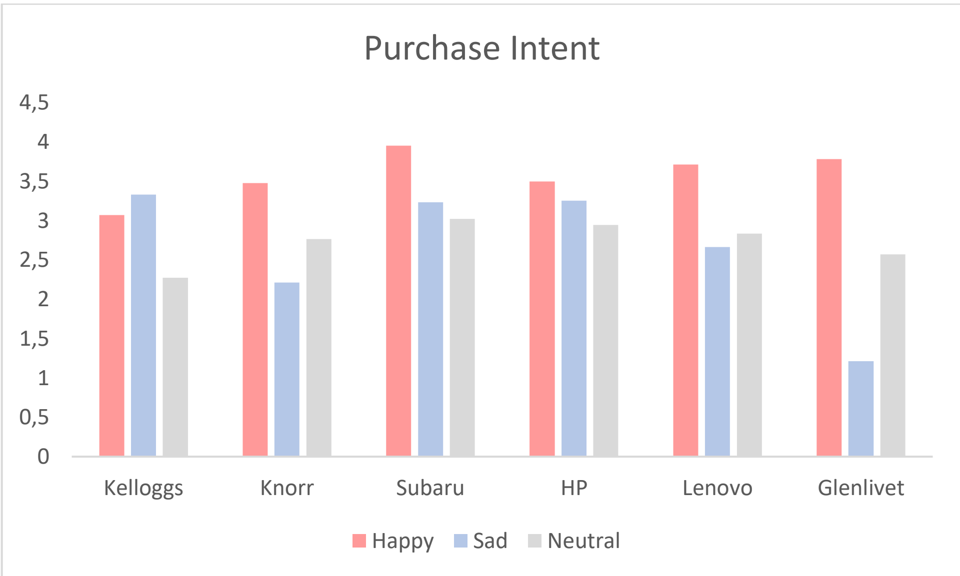


Figure 5 11: The purchase intent scores of subjects who had no previous intent to buy the promoted product are shown in the bar chart. The happy, sad and neutral columns signify the difference of purchase intent depending on how it was perceived.

Looking in the greater detail to the subjects who had no intent of buying the promoted product before the study, it is seen that happy perceived videos are better to convince the viewer to make future purchase. The exception is the Kellogg's ad in this sense. It is visualized in the Figure 5-11 whether people show propensity to buy the product or not. But this figure shows the average purchase intent score and the number of people found this commercial as a sad ad is quite low. In the table, the Cramer's V measures demonstrate that purchase intent is different depending on the type of emotion perceived in the videos. Albeit it is not distinguishable for some ads, from the figure it can be seen that happy emotion dominates in the sense of purchase intent anyway.

	LAB.				ONLINE			
	N	Phi	Cramer's V	Sig.	N	Phi	Cramer's V	Sig.
Kellogg's	7	1.095	,775	,395	61	,529	,374	,252
Knorr	16	,462	,462	,755	84	,668	,472	,002
Subaru	19	,637	,637	,259	106	,405	,286	,360
HP	12	,824	,582	,615	96	,558	,394	,019
Lenovo	19	,407	,407	,533	102	,543	,384	,017
Glenlivet	20	,629	,629	,340	112	,673	,476	,000
Total	93	,464	,328	,218	561	,372	,263	,000

Table 20: The table shows the significant difference in purchase intent occurred to the viewer depending on how the commercial is perceived. The table gives the results of the subjects who did not have previous intent to buy the promoted product and declare the intent to buy by the sheer effect of commercials.

To sum up, apparently prior purchase makes a difference for people to make a purchase decision. Prior purchase intent decision explains the variance of post purchase intent at 45% *(Phi test=,451 , p<0,05)*. The strength of non-existence of purchase intent shows the predictor power of prior purchase intent better. Also, the impact on purchase decision is dependent on the emotion of stimuli they perceive when there is no prior intent to buy the product.

5.6.3 Purchase Intent and Customer Satisfaction

The reason to take customer satisfaction as an independent variable on purchase intent is that the customers who were satisfied with the product they bought before would have approached with more positive attitude to make a repurchase from the brand promoted in the ad than the ones who had negative experience (Sharifi, 2014). From the following bar chart, it is understood that while satisfied people are more prone to have a repurchase intent, the ones who are unsatisfied were not affected by the commercial regardless of the emotion type (< 3,0).

Figure 5 14: The figure demonstrates the difference in the future purchase intent of the subjects who were satisfied with prior purchase and the ones that were dissatisfied. The degree of purchase intention of people who are satisfied with the product shown in the video are the green columns and unsatisfied ones' intention levels are the green ones.

Running the Goodman and Kruskal's Gamma measures on SPSS tool, it is obtained that the association between two variables purchase intent and customer satisfaction is relatively strong and statistically significant *(G=.554, p<.05)*.

Apart from the people who had customer satisfaction scores regarding their previous purchases, it was also analysed the impact of commercials for participants that never made a purchase before. The following figure especially marked the impact of emotion on buying intent depending on how it was perceived. The three commercials chosen for this figure are the ones which were perceived differently by many subjects about the emotion type they evoke. Seemingly, when the commercials are approached as happy videos, they create a considerably higher impact on people to think a next purchase. In this part of the study, it was aimed to see the sheer difference of emotion perception on post purchase intent excluding the impact of previous purchase.

On the other hand, the other three stimuli are perceived as expected. For instance, The Kellogg's and Knorr videos were perceived happy and for this reason, these ads created a positive influence on future purchase intent. In the last part of purchase intent analysis, all will be explained as an output of regression analysis in a greater detail.

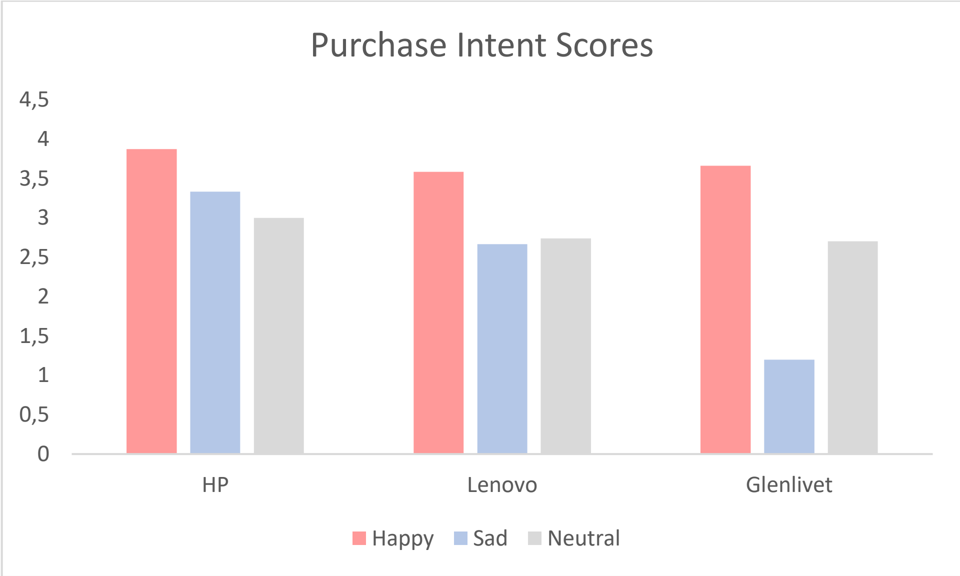


Figure 5 2: The figure indicates the purchase intent of subjects who had no customer satisfaction score due not having a purchase regarding the product from promoted brand in videos. The commercials in the bar chart are subject to different opinions in terms of emotion they evoke and the difference of purchase intent for each commercial

ONLINE	N	Gamma	Sig.
HP	18	,525	,025
Lenovo	63	,475	,002
Glenlivet	101	,490	,000

Table 21: The gamma results of emotion impact difference in purchase intent for the subjects who had no prior purchase of the product promoted.

Ordered Probit Regression Analysis:

Having three aforementioned components of purchase intention prediction as independent variables, I have regressed these variables on future purchase intent. The hypothesis and the result table are as follows:

Hypothesis 5: There will be no significant prediction of Purchase Intent by Customer satisfaction, prior purchase intent and general perceived emotion type of the ad

The dependent variable is: Post Purchase Intent

	Laboratory Settings					Online Settings				
Variables	Estimate	Std. Error	Wald	df	Sig.	Estimate	Std. Error	Wald	df	Sig.
Customer Satisfaction	1,0658167	0,3331379	10,235687	1	0,0013775	1,400	0,108	167,824	1,000	0,000
Happy	-0,470	0,534	0,775	1,000	0,378	0,754	0,197	14,641	1,000	0,000
Sad	0a			0,000		0,514	0,237	4,707	1,000	0,030
Neutral	-0,486	0,646	0,566	1,000	0,452	0a			0,000	
Existence of Prior Purchase Intent	1,161	0,480	5,846	1,000	0,016	1,218	0,165	54,596	1,000	0,000
No Existence of Prior Purchase Intent	0a			0,000		0a			0,000	

Table 22: The output of ordered probit regression between the dependent variable 'purchase intent' and its predictors: customer satisfaction, existence of prior purchase intent and emotion. The impacts of factors are sequenced from biggest to smallest as customer satisfaction, prior purchase intent and emotion.

The p value of both tests (0.00) are lower than the level of significance (0.05) that demonstrates the regression analyses are meaningful and the models explain the data significantly (the variation explanation is 0,598 and 0,448 for both environments, Nagelkerke values). It was found that the purchase intent does depend on customer satisfaction from previous engagement and previous purchase intent. Post purchase intention might also have a significant association with the emotion type of stimuli exposed towards the viewer. However, looking at the results gathered in two different environments, this deduction is questionable at some point. It should be noted the difference in terms of number of subjects in two settings is quite large. Therefore, it would be more rational to consider the online settings' findings. This means, emotion type could pose an impact on purchase intent significantly. Observing estimates, happy is slightly more effective emotion type in a persuasive context. The odds of happy perceived commercials considering purchase intent to be too high was 2.124 (95% CI, 1.444 to 3.125) times that of non-business owners, a statistically significant effect, Wald $\chi 2(1) = 14.641$, $p = .000$.

In the detail of components influencing purchase intent, the power of explaining the variance in purchase intent is quite high for customer satisfaction (52%) comparing to the existence of prior purchase intent (around 45%) and emotion of the stimuli (36%). However, in the non-existence of customer satisfaction and prior purchase intent, the emotion impact on purchase is remarkably higher (50%). But the overall interpretation of measuring purchase intent is that the will to buy the product is more complex than having a positive approach towards the ad. For instance, the will of buying a car (e.g. Subaru ad) does depend on various factors like income status of the individual. However, as expressed before, the change in the attitude towards the ad may lead a propensity to make a purchase in future.

CHAPTER 6: CONCLUSIONS

This thesis was designed to answer three research questions indicated at the beginning:

Research Question 1: Do emotion type of the video stimuli, ad liking, rewatchability and attitude towards the brand predict advertising effectiveness?

Research Question 2: Do emotion type of the video advertisement, customer satisfaction, pre-purchase intent predict post-purchase intent?

Research Question 3: Do facial expressions predict ad the effectiveness?

Research Question 4: Do facial expressions predict the post-purchase intent?

To begin with the usability of FaceReader in order to measure the affective states, it can be concluded that FaceReader results are significantly correlated with the happiness level of facial records and self-reported measures. However, the sadness level and neutral level of facial expressions of individuals do not correlate with the self-reported results. However, ads evoking sadness are not purely sad content commercials. The sad advertisements have ups and downs in a sense that sad ads are mostly subjected to convey mixed emotions. Considering this fact, the FaceReader is able to detect the emotion of face when the conditions are met. However, when there is no obviously dominant emotion, it is harder to detect emotion real-time using a face emotion recognition software. On the other hand, FaceReader was somewhat more successful at detecting cognitive valence values from the facial reactions. The reason of this difference between valence values and discrete emotions is that valence values are calibrated considering other negative emotions than only sadness. Therefore, the impact of other emotions on subjects' faces were considered to some extent.

The findings of first research question reveal that ad effectiveness can be measured through the information regarding consumers' cognitive, conative and affective attitudes. In this study, it is implied that emotional state as a predictor has the lowest power to explaining the variance of ad effectiveness (36%). However, the exposure of emotional content instead of neutral commercial will make higher the effectiveness of advertising. On the contrary, neutral video commercials are particularly disliked by the subjects and the implication is that marketers should engineer on the emotional content of the stimuli in order to create a positive engagement.

Moreover, the importance of emotional advertising is asserted in literature review part; the change in belief could lead the change in attitude and therefore a change in purchase intent (Ton & Chia, 2007). Thus, emotion might also influence the other predictors and in overall create a bigger impact.

As a response to the research question regarding purchase intent; the purchase intent is at a large-scale dependent on the previous experience of customer. In the results of the study, the individuals had bad experience with the product promoted in the video, they were not particularly impressed to change their buy behaviour in future. Nonetheless, the subjects who had never consumed the product of the brand promoted in the video were affected by the emotion of the video. While the persuasive impact of happy perceived videos were the highest, neutral ones had no contribution significantly. All in all, the significant impact of emotion in purchase intent is relatively lower than in ad effectiveness. The similar results were obtained in previous studies (Lewinski et. al, 2014; McDuff et. al, 2015) The reason was explained by researchers in a way that purchase intent is of more complexity (e.g. buying a car depends on other factors like income). The same reason is valid also in this study because some of videos promote the products like computer and car.

Third and fourth research questions are related to the use of FaceReader and the feasibility of automatically measure the ad effectiveness and purchase intentions. The FaceReader can automatically analyze the ad effectiveness as seen in happy videos, which is valuable in terms of having the chance to read emotion real-time and in a non-invasive way. Nevertheless, for purchase intent prediction, there is need of having more information regarding other attitudes and demographic features of individuals. Employing only FaceReader did not yield significant results at predicting post purchase intent.

More importantly, the novel contribution of this study to the fields of affective computing, advertising and marketing is that not only ad-evoked happiness will likely increase the ad effectiveness, but sad commercials too build positive customer engagement. Quite the opposite, neutral videos have negative impact on ad effectiveness, subjects disliked the neutral interaction. The impact of 'sadvertising' also exist on purchase decisions of the individuals in addition to the impact of happy videos.

In this study, it is deducted that an emotional communication is perceived better by potential customers than neutral interaction. Therefore, it is suggested to marketers and advertisers to focus on emotions while designing their advertisement and communication strategies.

To the best of my knowledge, sad and neutral emotions were not analysed as predictor factors of ad effectiveness and purchase intent and the study on these emotions show the potential importance of further studies focused on other types of emotions evoked by the commercials.

Limitations and Further Research

The study conducted in this thesis faces some limitations which should be avoided in future work.

In this thesis, the stimuli were chosen depending on the most dominant emotion the commercials evoke. However, the video advertisements most probably include the minor contradictions in terms of emotions they evoke. Considering this fact, this study focused on three criteria in the beginning: semantic, music, tone of voice. Even so, some commercial stimuli used in this study were subjected to be perceived differently by a considerable amount of people. Therefore, the marginal contributions of each factor should be engineered because of the importance of context. For instance, for semantics chosen in the video commercial could be examined by speech emotion recognition software and presented as stimuli. This way, the limitation of human observers which play crucial role in the validation of stimuli phase could have been avoided. This limitation should be eliminated in future work to better figure out the usability/ reliability of face emotion recognition software.

In the beginning phase of this thesis, it was aimed to measure the impact of angry video commercials. In the validation of stimuli phase, it is understood that the angry commercials were not completely evoke the anger itself, oppositely they may result in evoking happiness. However, maybe not anger but other types of emotion in future studies should be analysed (e.g. disgust). Because, in contrast to the study of Lewinski et. al (2014), it is found in this work that happiness is not only emotion type which could affect ad effectiveness. Therefore, different discrete emotions' impact could be an interesting analysis.

This study employed self-report instrument to measure cognitive, conative and affective attitudes of the subjects. The performance of FaceReader was gauged making the comparison with the results of self-reports. However, as stated in the literature review, self-assessment technique suffers from individual biases which means that the responses of self-reports are subjected to cognitive processing and can be controlled. For the further work, the questions forwarded to the subjects can be minimized using the other tools for example EMG signals like in the study of Walla et. al (2011). The authors measured Attitude towards Brand using physio-psychological measures. Moreover, this should be seen as a must due to the fact that in the experimental phases, it was seen that answering questions cause fatigue in subjects which affect the facial reactions as a byproduct.

Another limitation of this work is that it was unknown if the subjects were familiar with the advertisement they were exposed in the experiments. However, although there is an argument in research over this topic: Repeated stimuli exposure caused evoking neutral valence or caused higher familiarity which brings about liking (Andrade et. al, 2008). Employing less-known stimuli could give the possibility to measure impact of “mere exposure”.

References

Aaker, David A. *Building Strong Brands*. London, Simon And Schuster, 1996.

Anagnostopoulos, Christos-Nikolaos, et al. "Features and Classifiers for Emotion Recognition from Speech: A Survey from 2000 to 2011." *Artificial Intelligence Review*, vol. 43, no. 2, 9 Nov. 2012, pp. 155–177, 10.1007/s10462-012-9368-5.

Andrade, Eduardo, et al. *The Nature and Role of Affect in Consumer Behavior The Seven Sins of Consumer Psychology View Project Behavior and Clinical Change View Project*. 2014.

Bagozzi, Richard, et al. "The Role of Emotions in Marketing." *THE ROLE OF EMOTIONS Journal of the Academy of Marketing Science*, vol. 27, no. 2, 1999, pp. 184–206.

Barreto, Ana. *Application of Facial Expression Studies on the Field of Marketing*. 2017.

Bian, Qin, and Sandra Forsythe. "Purchase Intention for Luxury Brands: A Cross Cultural Comparison." *Journal of Business Research*, vol. 65, no. 10, Oct. 2012, pp. 1443–1451, 10.1016/j.jbusres.2011.10.010.

Buil, Isabel, et al. "Examining the Role of Advertising and Sales Promotions in Brand Equity Creation." *Journal of Business Research*, 2011, pp. 1–8.

Chen, Yue, et al. "Comparing Measurements for Emotion Evoked by Oral Care Products." *International Journal of Industrial Ergonomics*, vol. 66, July 2018, pp. 119–129, 10.1016/j.ergon.2018.02.013.

Cohn, Jeffrey. "Foundations of Human Computing: Facial Expression and Emotion Automated Analysis of Nonverbal Behavior in Depression View Project 3D Facial Expression Understanding View Project Foundations of Human Computing: Facial Expression and Emotion." *Proceedings of the 8th International Conference on*

Multimodal Interfaces, 2006.

Corvi, E, and M Bonera. “The Effectiveness of Advertising: A Literature Review.” *10th Global Conference on Business & Economics*, 2010, pp. 1–11.

Dai, Weihui, et al. “Emotion Recognition and Affective Computing on Vocal Social Media.” *Information & Management*, vol. 52, no. 7, Nov. 2015, pp. 777–788, 10.1016/j.im.2015.02.003.

D'errico, Francesca, et al. “Cognitive Emotions in E-Learning Processes and Their Potential Relationship with Students’ Academic Adjustment.” *International Journal of Emotional Education*, vol. 10, no. 1, 2018, pp. 89–111.

de Gelder, Beatrice, and Jean Vroomen. “The Perception of Emotions by Ear and by Eye.” *Cognition & Emotion*, vol. 14, no. 3, May 2000, pp. 289–311, 10.1080/026999300378824.

Derbaix, Christian M. “The Impact of Affective Reactions on Attitudes toward the Advertisement and the Brand: A Step toward Ecological Validity.” *Journal of Marketing Research*, vol. 32, no. 4, Nov. 1995, p. 470, 10.2307/3152182.

Desmet, P. *Designing Emotions*. 2002.

Desmet, Pieter. “Measuring Emotion: Development and Application of an Instrument to Measure Emotional Responses to Products.” *Funology*, Springer, Dordrecht, 2003, pp. 111–123. Accessed 2 Mar. 2020.

Edell, Julie A., and Marian Chapman Burke. “The Power of Feelings in Understanding Advertising Effects.” *Journal of Consumer Research*, vol. 14, no. 3, Dec. 1987, p. 421, 10.1086/209124.

EKMAN, PAUL. “Are the Basic Emotions?” *Psychological Review*, vol. 99, no. 3, 1992.

Ekman, Paul, and Erika Rosenberg. *What the Face Reveals: Basic and Applied Studies of*

Spontaneous Expression Using the Facial Action Coding System (FACS), Second Edition. 1997.

El Ayadi, Moataz, et al. "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases." *Pattern Recognition*, vol. 44, no. 3, Mar. 2011, pp. 572–587, 10.1016/j.patcog.2010.09.020.

Eyben, Florian, et al. "Emotion on the Road—Necessity, Acceptance, and Feasibility of Affective Computing in the Car." *Advances in Human-Computer Interaction*, vol. 2010, 2010, pp. 1–17, 10.1155/2010/263593.

Haley, Russell I., and Allan L. Baldinger. "The ARF Copy Research Validity Project." *Journal of Advertising Research*, vol. 40, no. 6, Nov. 1990, pp. 114–135, 10.2501/jar-40-6-114-135.

Hamelin, Nicolas, et al. "Emotion and Advertising Effectiveness: A Novel Facial Expression Analysis Approach." *Journal of Retailing and Consumer Services*, vol. 36, May 2017, pp. 103–111, 10.1016/j.jretconser.2017.01.001.

Han, Heesup, et al. "Cognitive, Affective, Conative, and Action Loyalty: Testing the Impact of Inertia." *International Journal of Hospitality Management*, vol. 30, no. 4, Dec. 2011, pp. 1008–1019, 10.1016/j.ijhm.2011.03.006.

Han, Junwei, et al. "Arousal Recognition Using Audio-Visual Features and FMRI-Based Brain Response." *IEEE Transactions on Affective Computing*, vol. 6, no. 4, 1 Oct. 2015, pp. 337–347, 10.1109/taffc.2015.2411280.

Harrigan, Jinni A, et al. *The New Handbook of Methods in Nonverbal Behavior Research*. Oxford, Oxford University Press, 2012.

Havlena, William J., and Morris B. Holbrook. "The Varieties of Consumption Experience: Comparing Two Typologies of Emotion in Consumer Behavior." *Journal of*

Consumer Research, vol. 13, no. 3, Dec. 1986, p. 394, 10.1086/209078.

---. *The Varieties of Consumption Experience: Comparing Two Typologies of Emotion in Consumer Behavior*. Dec. 1986. ", in Peterson, R.A., Hoyer, W.D. and Wilson, W.R. (Eds), The Role ofAffect in Consumer Behavior: Merging Theories and Applications, Lexington Books,Lexington, MA, pp. 17-52.

Heni, N, and H Haman. "Design of Emotional Educational System Mobile Games for Autistic Children." *2nd International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2016*, 2016, pp. 631–637.

Horvat, Marko, et al. *Comparing Affective Responses to Standardized Pictures and Videos: A Study Report*. 2015.

Hoskin, Rob. "The Dangers of Self-Report – Science Brainwaves." *Sciencebrainwaves.Com*, 3 Mar. 2012, www.sciencebrainwaves.com/the-dangers-of-self-report/.

Husnjak, Sinisa, et al. "Possibilities of Using Speech Recognition Systems of Smart Terminal Devices in Traffic Environment." *Procedia Engineering*, vol. 69, 2014, pp. 778–787, 10.1016/j.proeng.2014.03.054. Accessed 2 Mar. 2020.

Jones, Christian, and Ing-Marie Jonsson. "AUTOMATIC RECOGNITION OF AFFECTIVE CUES IN THE SPEECH OF CAR DRIVERS TO ALLOW APPROPRIATE RESPONSES." *DBLP*, 2005, pp. 1–10.

Kamaruddin, Norhaslinda, et al. "Measuring Customer Satisfaction through Speech Using Valence-Arousal Approach." *6th International Conference on Information and Communication Technology for the Muslim World*, 2017, pp. 298–303.

Kidwell, Blair, et al. "Emotional Intelligence in Marketing Exchanges." *Journal of Marketing*, vol. 75, no. 1, Jan. 2011, pp. 78–95, 10.1509/jmkg.75.1.78.

LANG, ANNIE. “Involuntary Attention and Physiological Arousal Evoked by Structural Features and Emotional Content in TV Commercials.” *Communication Research*, vol. 17, no. 3, June 1990, pp. 275–299, 10.1177/009365090017003001.

Lebai Lutfi, Syaheerah, et al. “A Satisfaction-Based Model for Affect Recognition from Conversational Features in Spoken Dialog Systems.” *Speech Communication*, vol. 55, no. 7–8, Sept. 2013, pp. 825–840, 10.1016/j.specom.2013.04.005.

LeBlanc, Vicki R., et al. “Predictable Chaos: A Review of the Effects of Emotions on Attention, Memory and Decision Making.” *Advances in Health Sciences Education*, vol. 20, no. 1, 6 June 2014, pp. 265–282, 10.1007/s10459-014-9516-6.

Lee, Dongkeon, et al. “The ChatBot Feels You -A Counseling Service Using Emotional Response Generation.” *International Conference On Big Data And Smart Computing Bigcomp 2017*, 2017, pp. 437–440.

Lee, Haesung, and Joonhee Kwon. “Combining Context-Awareness with Wearable Computing for Emotion-Based Contents Service.” *International Journal International Journal International Journal International Journal of Advanced Science and Technology of Advanced Science and Technology of Advanced Science and Technology of Advanced Science and Technology*, vol. 22, 2010, pp. 13–25.

Lewinski, Peter, et al. “Automated Facial Coding: Validation of Basic Emotions and FACS AUs in FaceReader.” *Journal of Neuroscience, Psychology, and Economics*, vol. 7, no. 4, Dec. 2014, pp. 227–236, 10.1037/npe0000028.

---. “Predicting Advertising Effectiveness by Facial Expressions in Response to Amusing Persuasive Stimuli.” *Journal of Neuroscience, Psychology, and Economics*, vol. 7, no. 1, 2014, pp. 1–14, 10.1037/npe0000012.

Li, Shanshi, et al. “A Comparative Analysis of Self-Report and Psychophysiological Measures of Emotion in the Context of Tourism Advertising.” *Journal of Travel Research*, vol. 57, no. 8, 12 Oct. 2017, pp. 1078–1092, 10.1177/0047287517733555.

Liu, Zhen-Tao, et al. “Speech Emotion Recognition Based on Feature Selection and Extreme Learning Machine Decision Tree.” *Neurocomputing*, vol. 273, Jan. 2018, pp. 271–280, 10.1016/j.neucom.2017.07.050.

Loui, Psyche, et al. “Effects of Voice on Emotional Arousal.” *Frontiers in Psychology*, vol. 4, 2013, 10.3389/fpsyg.2013.00675.

Maaoui, C., et al. “Physio-Visual Data Fusion for Emotion Recognition.” *IRBM*, vol. 35, no. 3, June 2014, pp. 109–118, 10.1016/j.irbm.2014.03.001.

McDuff, Daniel, et al. “Crowdsourcing Facial Responses to Online Videos.” *IEEE Transactions on Affective Computing*, vol. 3, no. 4, 2012, pp. 456–468, 10.1109/t-affc.2012.19.

---. “Predicting Ad Liking and Purchase Intent: Large-Scale Analysis of Facial Responses to Ads.” *IEEE Transactions on Affective Computing*, vol. 6, no. 3, 1 July 2015, pp. 223–235, 10.1109/taffc.2014.2384198.

McDuff, Daniel, and Mohammad Soleymani. “Large-Scale Affective Content Analysis: Combining Media Content Features and Facial Reactions.” *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 1 May 2017, pp. 339–345, ieeexplore.ieee.org/document/7961761. Accessed 2 Mar. 2020.

Mitchell, Andrew A., and Jerry C. (Jerry Corrie) Olson. “Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attitude?” *Advertising & Society*

Review, vol. 1, no. 1, 2000, 10.1353/asr.2000.0010.

Morris, Jon D, et al. “The Power of Affect: Predicting Intention.” *Journal of Advertising Research*, vol. 42, no. 3, May 2002, pp. 7–17, 10.2501/jar-42-3-7-17.

Morrison, Donn, et al. “Ensemble Methods for Spoken Emotion Recognition in Call-Centres.” *Speech Communication*, vol. 49, no. 2, Feb. 2007, pp. 98–112, 10.1016/j.specom.2006.11.004.

Nass, Clifford, et al. “Improving Automotive Safety by Pairing Driver Emotion and Car Voice Emotion.” *CHI ’05 Extended Abstracts on Human Factors in Computing Systems - CHI ’05*, 2005, pp. 1973–1976.

Park, Youngja, and Stephen Gates. “Towards Real-Time Measurement of Customer Satisfaction Using Automatically Generated Call Transcripts General Terms.” *International Conference on Information and Knowledge Management, Proceedings*, 2009, pp. 1387–1396.

Plutchik, Robert. “Emotion: A Psychoevolutionary Synthesis.” *The American Journal of Psychology*, vol. 93, no. 4, Dec. 1980, p. 751, 10.2307/1422394.

Plutchik, Robert, and Henry Kellerman. *Emotion. Vol. 1, Theories of Emotion*. Boston, Academic Press, 1980.

Poels, Karolien, and Siegfried Dewitte. *How to Capture the Heart? Reviewing 20 Years of Emotion Measurement in Advertising*. *SSRN Electronic Journal*, utomatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition, 2006, pp. 1–47.

Ramakrishnan, S. *Recognition of Emotion from Speech: A Review*. 2012. *Speech Enhancement, Modeling and Recognition- Algorithms and Applications,* Speech Enhancement, Modeling and Recognition- Algorithms and Applications, 0 2012, pp. 121–138.

Ren, Fuji, and Changqin Quan. “Linguistic-Based Emotion Analysis and Recognition for Measuring Consumer Satisfaction: An Application of Affective Computing.” *Information Technology and Management*, vol. 13, no. 4, 22 Aug. 2012, pp. 321–332, 10.1007/s10799-012-0138-5.

Russell, James A, and Albert Mehrabian. “Evidence for a Three-Factor Theory of Emotions.” *Journal of Research in Personality*, vol. 11, no. 3, Sept. 1977, pp. 273–294, 10.1016/0092-6566(77)90037-x.

Sariyanidi, Evangelos, et al. “Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, 1 June 2015, pp. 1113–1133, 10.1109/tpami.2014.2366127.

Schuller, Björn, et al. “META-CLASSIFIERS IN ACOUSTIC AND LINGUISTIC FEATURE FUSION-BASED AFFECT RECOGNITION.” *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2005, pp. 1325–1328.

Shahin Sharifi, Seyed. “Impacts of the Trilogy of Emotion on Future Purchase Intentions in Products of High Involvement under the Mediating Role of Brand Awareness.” *European Business Review*, vol. 26, no. 1, 7 Jan. 2014, pp. 43–63, 10.1108/ebr-12-2012-0072.

Small, Deborah, and Nicole Verrochi. “The Face of Need: Facial Emotion Expression on Charity Advertisements.” *Journal of Marketing Research*, vol. XLVI, 2009, p. 777.

Spirina, Anastasiia, et al. “Could Emotions Be Beneficial for Interaction Quality

Modelling in Human-Human Conversations?” *Ecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics*, 2017, pp. 447–455.

Story, Louise. “Anywhere the Eye Can See, It’s Likely to See an Ad.” *The New York Times*, 15 Jan. 2007.

Tan, Soo Jiuan, and Lily Chia. “Are We Measuring the Same Attitude? Understanding Media Effects on Attitude towards Advertising.” *Marketing Theory*, vol. 7, no. 4, Dec. 2007, pp. 353–377, 10.1177/1470593107083162.

Tawari, Ashish, and Mohan Trivedi. “Speech Based Emotion Classification Framework for Driver Assistance System.” *2010 IEEE Intelligent Vehicles Symposium*, 2010, pp. 174–178.

Taylor, John G., and Nickolaos F. Fragopanagos. “The Interaction of Attention and Emotion.” *Neural Networks*, vol. 18, no. 4, May 2005, pp. 353–369, 10.1016/j.neunet.2005.03.005.

Taylor, Ronald K. “MARKETING STRATEGIES: GAINING A COMPETITIVE ADVANTAGE THROUGH THE USE OF EMOTION.” *Competitiveness Review*, vol. 10, no. 2, Feb. 2000, pp. 146–152, 10.1108/eb046407.

Teixeira, Thales, et al. “Emotion-Induced Engagement in Internet Video Advertisements.” *Journal of Marketing Research*, vol. 49, no. 2, Apr. 2012, pp. 144–159, 10.1509/jmr.10.0207.

Tian, Ying-Li, et al. “Facial Expression Analysis.” *Handbook of Face Recognition*, 2005, pp. 247–275, 10.1007/0-387-27257-7_12.

Tokuno, Shinichi, et al. “Usage of Emotion Recognition in Military Health Care Detecting Emotional Change under Stress.” *Defense Science Research Conference and Expo*

(DSR), 2011.

Turley, L.W., and J. Richard Shannon. "The Impact and Effectiveness of Advertisements in a Sports Arena." *Journal of Services Marketing*, vol. 14, no. 4, July 2000, pp. 323–336, 10.1108/08876040010334547.

Van Kuilenburg, Hans, et al. *A Model Based Method for Automatic Facial Expression Recognition*. *Lecture Notes in Computer Science*, Springer, 2005, pp. 194–205.

Vecchiato, Giovanni, et al. "On the Use of EEG or MEG Brain Imaging Tools in Neuromarketing Research." *Computational Intelligence and Neuroscience*, vol. 2011, 2011, pp. 1–12, 10.1155/2011/643489.

Verbeke, W. "The Adaptive Consequences of Pride in Personal Selling." *Journal of the Academy of Marketing Science*, vol. 32, no. 4, 1 Oct. 2004, pp. 386–402, 10.1177/0092070304267105.

Walla, Peter, et al. "Objective Measures of Emotion Related to Brand Attitude: A New Way to Quantify Emotion-Related Aspects Relevant to Marketing." *PLoS ONE*, vol. 6, no. 11, 2 Nov. 2011, p. e26782, 10.1371/journal.pone.0026782.

Weninger, Felix, et al. "On the Acoustics of Emotion in Audio: What Speech, Music, and Sound Have in Common." *Frontiers in Psychology*, vol. 4, 2013, 10.3389/fpsyg.2013.00292.

Yang, B., and M. Lugger. "Emotion Recognition from Speech Signals Using New Harmony Features." *Signal Processing*, vol. 90, no. 5, May 2010, pp. 1415–1423, 10.1016/j.sigpro.2009.09.009.

Yu, Yi-Ting, and Alison Dean. "The Contribution of Emotional Satisfaction to Consumer Loyalty." *International Journal of Service Industry Management*, vol. 12, no. 3, Aug. 2001, pp. 234–250, 10.1108/09564230110393239.

Zhang, Kem Z.K., et al. “Examining the Moderating Effect of Inconsistent Reviews and Its Gender Differences on Consumers’ Online Shopping Decision.” *International Journal of Information Management*, vol. 34, no. 2, Apr. 2014, pp. 89–98, 10.1016/j.ijinfomgt.2013.12.001.

Zhihong Zeng, et al. “A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, Jan. 2009, pp. 39–58, 10.1109/tpami.2008.52.

APPENDIX A

Pilot Study (Googleforms)

2/3/2020 Emotion Recognition Questionnaire

Emotion Recognition Questionnaire

In this survey, you are going to be watching 20 videos.
Afterwards you will answer 5 question for each video in terms of what kind of emotions they evoke.
*Required

1. What is your gender? *

Mark only one oval.

- Female
- Male
- Prefer not to say

2. What is your age? *

Mark only one oval.

- Under 18
- 18-24
- 25-34
- 35-44
- 45-54
- Above 54

3. What is your age? *

Mark only one oval.

- Under 18
- 18-24
- 25-34
- 35-44
- 45-54
- Above 54

https://docs.google.com/forms/d/18x1zSz7qBlzTZz__SXX2_Z1SBs7EC7KXbsMXVq5v_0Q/edit 1/32

For whole questionnaire: Adobe Acrobat Document

APPENDIX B

Laboratory Environment Survey

The questions are the same for 6 commercials which are listed in the Stimuli Selection Section

Questionario (Versione italiana)

Vorremmo chiedere il tuo aiuto in questo sondaggio per rispondere in modo onesto al seguente questionario. La vostra collaborazione sarà molto apprezzata. Grazie per il vostro tempo e aiuto.

1. Qual è il tuo genere?

A. Donna
B. Uomo
C. Preferisco non dirlo

2. Quanti hanni ha?

A. 18-25
B. 26-35
C. 36-45
D. 46-55
E. Superiore a 55

3. Qual è il tuo titolo di studio?

A. Meno del diploma di scuola superiore
B. Diploma di scuola superiore o equivalente
C. Laurea breve
D. Laurea magistrale
E. Laurea post-laurea

Si prega di rispondere alle seguenti domande riferendosi al video che hai appena guardato.

4. Secondo te quale tra le seguenti descrive meglio l'emozione generale del video?

A. Felicità
B. Neutro
C. Tristezza

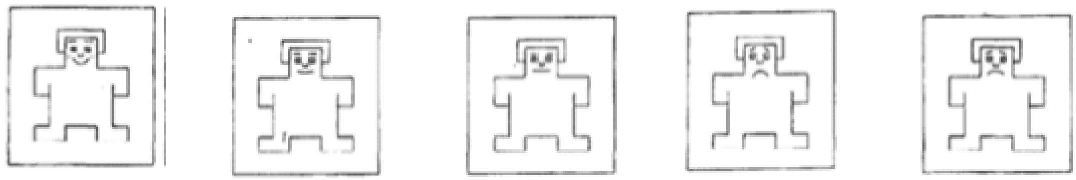
5. Si prega di valutare l'intensità dell'emozione selezionata

A. Molto intenso
B. Intenso
C. Abbastanza intenso
D. Leggermente intenso
E. Non intenso

6. Valuta il livello di valenza suscitata da questo video. La valenza descrive la misura in cui un'emozione è positiva o negativa.

(Valenza descrive la misura in cui un'emozione è positiva o negativa)

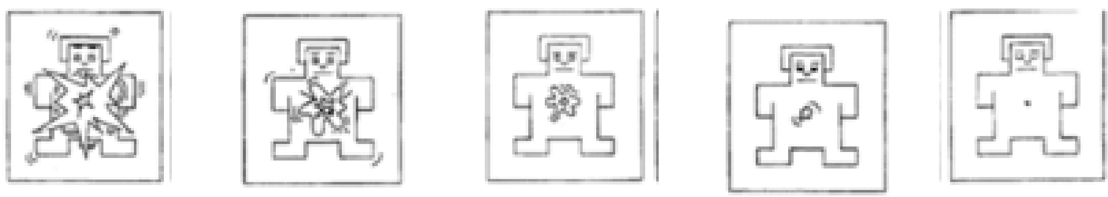
Molto positiva	Positiva	Neutra	Negativa	Molto negativa



7. Valuta il livello di AROUSAL suscitato da questo video. L'arousal descrive quanto sei stato stimolato/colpito.

(per stimolato, si intende che punto sei stato colpito dal video)

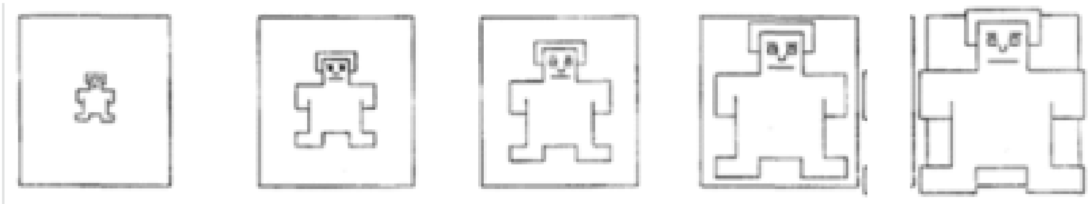
Molto stimolato	Stimolato	Neutro	Calmo	Molto calmo



8. Valuta il livello di DOMINANCE suscitata da questo video. La dominance descrive quanto tu ti sia sentito sopraffatto dallo stimolo o di avere lo stimolo sotto controllo.

(La dominanza si riferisce al potere dell'emozione, fino a che punto il video denota qualcosa che è debole o forte.)

Molto dominato dallo stimolo	Dominante dallo stimolo	Neutro	Debole	Molto Debole



9. Quanto ti è piaciuta l'inserzione pubblicitaria che hai appena visto?

A. Moltissimo
B. Molto
C. Nella media
D. Poco
E. Pochissimo

10. Conoscevi già la marca mostrata nel video?
A. Sì
B. No

11. Se conoscevi già la marca mostrata nel video, esprimi il tuo grado di accordo con la seguente affermazione

	Fortemente d’accordo	D’accordo	Né d’accordo né disaccordi	Non sono d’accordo	Fortemente in disaccordo
In passato ho avuto esperienze positive con questa marca					
Ho una considerazione positiva di questa marca					
Mi riconosco nei valori trasmessi da questa marca					
Questa marca è la mia marca preferita					

12. Hai mai acquistato un prodotto di questo marca?

A. Sì
B. No

13. Se hai acquistato un prodotto di questa marca, esprimi il tuo grado di accordo con le seguenti affermazioni

	Fortemente d’accordo	D’accordo	Né d’accordo né disaccordi	Non sono d’accordo	Fortemente in disaccordo
Sono rimasto soddisfatto del prodotto acquistato					
Acquisterei nuovamente prodotti di questa marca					
Sono disposto a diffondere pareri positivi a proposito di questa marca					

14. Prima di questo studio, avevi intenzione di acquistare il prodotto mostrato nell'annuncio?

A. Sì, quel prodotto di quella marca
B. Sì, quel prodotto ma di un'altra marca
C. No, ma volevo acquistare un altro prodotto di quella marca
D. No, non ero interessato a quel prodotto nè a quella marca

15. Alla luce di quanto hai visto nel video, esprimi il tuo grado di accordo con le seguenti affermazioni

	Fortemente d’accordo	D’accordo	Né d’accordo né disaccordi	Non sono d’accordo	Fortemente in disaccordo
La mia considerazione di questa marca è cambiata positivamente					
La mia considerazione di questa marca è cambiata negativamente					
Sono venuto a conoscenza di valori di questa marca di cui non ero consapevole					
Mi sento più propenso ad acquistare un prodotto di questa marca					
Il prodotto mostrato nel video mi ha colpito particolarmente					
Se avessi bisogno di acquistare questo prodotto, acquisterei probabilmente il prodotto della marca mostrata nel video					

APPENDIX C

Online settings Survey (Surveymonkey)

The questions are the same for 6 commercials which are listed in the Stimuli Selection Section

Affect Detection

1) To which gender you belong to?

A. Female

B. Male

C. Prefer not to say

2) To which age group you belong to?

A. Under 18

B. 18-24

C. 25-34

D. 35-44

E. 45-54

F. Above 54

3) What is your highest qualification?

A. No degree

B. Less than high school diploma

C. High school diploma or equivalent degree

D. Bachelor’s degree

E. Master’s degree

F. Post graduate degree

4) Please watch the 1-minute video first, then continue with the questions.

A. I watched this 1-minute video.

5) How would you describe your overall attitude toward the ad you have just watched?

A. Happiness

B. Sadness

C. Neutral

6) Please evaluate the intensity of emotions for the selected emotion

A. Very intense

B. Intense

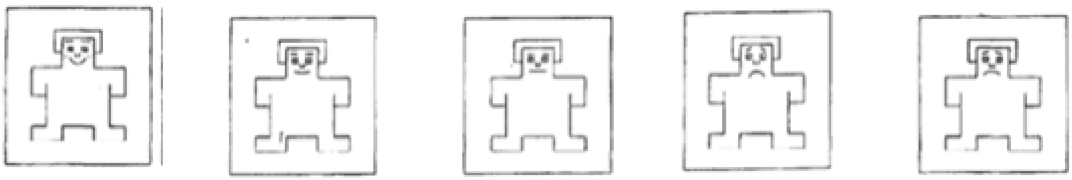
C. Fairly intense

D. Slightly intense

E. Not intense

7) Please evaluate the valence of the video
(Valence describes the extent to which an emotion is positive or negative))

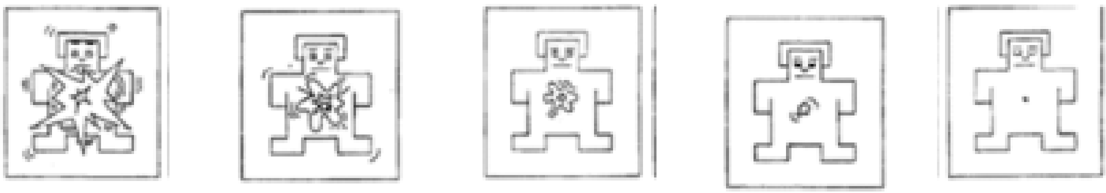
Very positive	Positive	Neutral	Negative	Very negative



8) Valuta How much were you aroused with the ad?
(Arousal means the extent to which you were evoked by the video)

Very Active refers to the states like "excitement, euphoria, excitation, rage, agitation, anger"
Very Quite refers to the experiences such as "relaxation, tranquility, idleness, meditation, boredom"

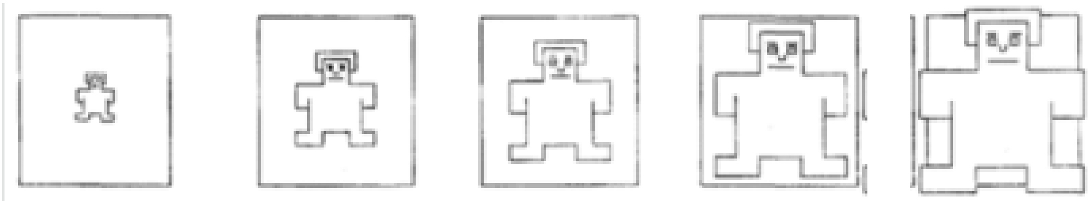
Very active Active Neutral Quite Very quite



9) Please evaluate the dominance of the video.
(Dominance refers to the power of the emotion, extent to which the video denotes something that is weak or strong.)

Very Dominant picture refers to states like "feeling lack of control, withdrawal, resignation, submission, subordination, intimidation"
Very Weak refers to the states such as "feeling in control of situation, being important, recognized, decisive, influence"

Stimulus very dominant Stimulus dominant Stimulus Neutral Stimulus weak Stimulus very weak



10) How much did you like the commercial that you just watched?

A. Very high
B. High
C. Average
D. Low
E. Very low

11) Would you watch this video again?

A. Highly likely would watch
B. Likely would watch
C. Might watch
D. Unlikely would watch
E. Highly unlikely would watch

12) Do you remember the brand promoted in the ad?

A. Yes

B. No

13) If you remember the brand, please express your degree of agreement with the following statements

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
I am aware of brand Kellogg's					
Kellogg's is a brand of breakfast cereal I am really familiar with					
When I think of breakfast cereal, Kellogg's comes to my mind					

14) If you knew the brand shown in the video, please indicate your level of agreement or disagreement with the following statement

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
This brand is my favorite brand					
I think this brand deserves my effort to maintain a relationship					
I have a positive consideration for this brand					
In the past I have had positive experiences with this brand					
I am very committed to this brand					
The product I gained worth the price I paid					
I recognize the values transmitted by this brand					
This brand offers me good value for the price					

15) Have you ever purchased a product of this brand?

A. Yes

B. No

16) If you have purchased a product of this brand, please indicate your level of agreement or disagreement with the following statement

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
I am willing to spread positive opinions about this brand					
I was satisfied with the product I bought					

I would buy again products of this brand					

17) Prior to this study, were you planning to purchase the product shown in the ad?

A. Yes

B. No

18) Considering what you have watched in the video, please indicate your level of agreement or disagreement with the following statement

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
I became aware of the values of this brand of which I was not aware					
My consideration of this brand has changed negatively					
My consideration of this brand has changed positively					
I feel more inclined to buy a product of this brand					
If I needed to buy this product, I would probably buy the product of the brand shown in the video					
I will recommend this supplier to my family and friends					
The product shown in the video impressed me particularly					

APPENDIX D

It is reported below a part of the summary sheets of all the articles, conference paper and books I have analysed.

AUTHORS: D. Morrison, R. Wang, L. C. D. Silva
TITLE: Ensemble methods for spoken emotion recognition in call-centres
SOURCE: Speech Communication
YEAR: 2007
NO OF QUOTES: 212
VOLUME: 49
ISSUE: 2
PAGES: 98-112
AFFILIATION: New Zealand
KEYWORDS: affect recognition; emotion recognition; ensemble methods; speech processing; speech databases
CATEGORY: Article
TYPE: Empirical Research
OBJECTIVE: Comparison of portrayed and natural (call-center) emotion data sets and comparison of some base-level classifiers and two ensemble classification schemes

SUMMARY: In this research, the aim is to compare two different databases: one is a natural data set obtained from a CC and including two emotions (neutral, angry speech; the others are eliminated due to low distributions), the other emotion data set(ESMBS) acquired from another study and includes a wider set of emotion(anger, disgust, fear, joy, sadness, surprise).
Prosodic and contextual features were used to identify emotion. Of these, pitch, energy, vocal tract,rhythm, etc determine emotion in speech to some extent and the results are consistent for some of them and the others still contradictory.
Classification techniques:
There are many developed classification methods in literature and some of them proved to make good progress. In this study, 2 database were examined using SVM, Random tree, KNN, etc. Among these methods, only top 5 are retained: SVM, KNN, Multi-layer perceptron, Random forest, K*. SVM is the best one in terms of accuracy.
In order to optimise the classification time and accuracy, it is possible to apply feature selection algorithms. In this paper, Stepwise forward selection, PCA and Genetic search algorithms are used.
Apart from the base classification methods, ensembles of classifiers were aimed to use in this study with the aim of improving accuracy. These ensemble classification methods are unweighted vote and stacked generalisation. The unweighted vote sums the predictions of each class from the base classifiers and picks the most popular class. The stacked generalisation combines predictions from multiple classifiers, followingly use these to train a meta-learner that gives the last result based on level-0 predictions and then determine which ones are correct. In the end, there is a minimized error rate. The results of the experiments based on these new classification methods show that on Natural dataset, both ensemble classification methods show improvement over the base classifiers. On acted dataset, StackingC performs better than all the base classifiers.
The results of testing of feature selection methods indicate that Forward selection works best for the Natural data and the genetic search proves more accurate on the ESMBS.
All in all, in this research, one of the aims was to see if there is a superior result of using combined classification methods and the results give a better accuracy. The other goal was to examine the natural dataset(call-center) comparing to the portrayed speech and to see the differences. It is found that the acted speech yields the researcher a high amount of control over the emotion expressed. However, for the natural dataset, variance throughout the emotions is higher.

AUTHORS: W. Dai, D. Han, Y. Dai, D. Xu
TITLE: Emotion recognition and affective computing on vocal social media
SOURCE: Information & Management

YEAR: 2015
NO OF QUOTES: 49
VOLUME: 52
ISSUE: 7
PAGES: 777-788
AFFILIATION: China, United States
KEYWORDS: social media; social network; voice instant messaging; vocal data mining; emotion recognition; affective computing
CATEGORY:
TYPE: Empirical Research
OBJECTIVE: To develop an effective computational method for processing complex and dynamic emotions from the speech signals of vocal social media
SUMMARY: The research is about using the real-world data obtained from a real application and recognizing the emotion from the speech. It proposes a method to follow and a research schema to realize. Based on the suggested system, experiments and applications are realized and results shared for a further study. The difference of vocal social media from a real-world speech database is that listeners can be affected by the speaker in the group. It happens effectively because of the sense of belonging to a particular social group, continuance intentions toward these groups etc. Accordingly to change in emotions reflects behavioral reactions as well. At the first step of the proposed method, there is voice waves to be converted speech signals. Afterwards to get the proper feature set, it is a must to pre-process the signals. Machine learning(LS-SVR Model) comes in the 3rd phase and followingly we can estimate PAD values(pleasure, arousal, dominance). PAD is a proven method to estimate emotions. As a last step, it is acquired the percentage of typical emotions. The experiment here; during the ML phase, the training data is used from different databases (Chinese Database CASIA, vocal social media database). To get an adequate recognizement of emotions, it is important to train the ML with the relevant training samples. This needs to be generalized to apply in other vocal social media. So, the performance of the proposed method must be tested and verified through large samples.In the study, 25 acoustic parameters used and these parameters created a feature vector for the modelling part. After modelling, estimated PAD values collected. Going deeper, to have an objective outcome, some methods are applied to avoid subjectivity in interpretation like evaluations of psychologists. For the experiment, it is also considered to focus only vocal emotion rather than semantic information.

Experiment has been applied by reporting an actual serious incident in food safety in the media and few vocal social media groups were observed. In the end, it is deducted that emotions are perceived with high accuracy and also the changes in emotions caused by familiar people can be observed. The experiment result has also good generalizability for different social media group.

AUTHORS: C.N. Anagnostopoulos, T. Iliou, I. Giannoukos
TITLE: Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011
SOURCE: Artificial Intelligence Review
YEAR: 2012
NO OF QUOTES: 134
VOLUME: 43
ISSUE: 2
PAGES: 155-177

AFFILIATION: Greece

KEYWORDS: speech features; emotion recognition; classifiers

CATEGORY: Journal

TYPE: conceptual (?) -> theoretical

OBJECTIVE: The review of all SER

SUMMARY: The review of all SER and compilation of the improvements done in recent years. The consensus about how to distinguish emotions: 4 Dimensions (Evaluation-Pleasantness, Potency-Valence, Activation-Arousal, Unpredictability). ASR is not well-developed yet.
No need to compare classification algorithms due to the fact that there is a lack of uniformity in the way these methods are evaluated (different test sets, feature vectors, evaluation framework).
Feature Vector (Evolution during years):

LLDs Functionals(mean)
-Prosodic f. (pitch, energy,..)
-Spectral f. (MFCC)
Firstly (2003) prosodic features were used, (2007) LLDs started to be used widely. Followingly, alongside these, rhythm and sentence duration were added (2009). (2010) functionals were included in the feature set.
The improvement made related to feature selection was to use more features and eliminating the ones indicating low performance. Also, there was a consensus about suprasegmental features outperformed segmental features and MFCC were transferred to supra-segment. Apart from paralinguistic features, a list of "salient words" like "no", exactly provided an improvement to recognize the emotions.
For feature selection and classification, the methods used so far have been listed with their pros and cons. Obviously, there is potential to use ensemble classification schemas.

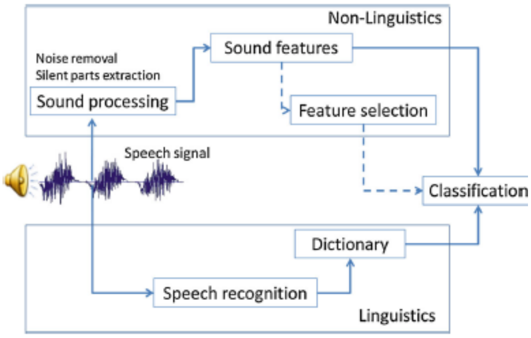


Fig. 1 Speech emotion recognition processing pipeline

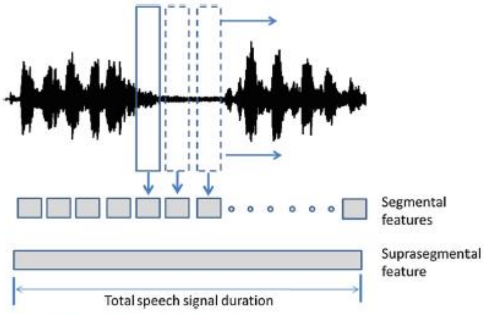


Fig. 2 Segmental and suprasegmental features in a speech signal

Automatic speech recognition technology has not yet reached the level of maturity required to perform consistently well in spontaneous speech. That's why linguistic features are not available to use as feature at present. In case, these features would be valuable, speaker dependent features can be used too and it makes implementation easier. From these, it is extracted that cross-cultural database are important and to be examined differently since bilingual people express themselves differently in both languages for the same expression.
However, until very recently, the researches usually made use of small, preselected, prototypical and often non-spontaneous emotional datasets and therefore comparability of results was inapplicable and the biggest problem is the results are not reproducible. -- New information: The INTERSPEECH 2009 Emotion Challenge was the first open public evaluation of speech-based emotion recognition systems with strict comparability, where all participants were using the same corpus.

AUTHORS: F. Weninger, F. Eyben, B. W. Schuller, M. Mortillaro,K. R. Scherer
TITLE: On the acoustics of emotion in audio: what speech, music, and sound have in common
SOURCE: Frontiers in Psychology
YEAR: 2013
NO OF QUOTES: 133
VOLUME: 4
ISSUE: --
PAGES: 1-12

AFFILIATION: Germany, Switzerland
KEYWORDS: audio signal processing; emotion recognition; feature selection; transfer learning; music perception; sound perception; speech perception
CATEGORY: Article
TYPE: Experimental
OBJECTIVE: A holistic computation model: starting from standard acoustic feature extraction schemes in the domains of speech,music, and sound analysis. Cross-domains arousal and valence dimensions.
SUMMARY: The ultimate goal is to recognize emotion automatically, in this study it has been focused on cross-domain generalization of features, which is novel. Firstly, to avoid speaker dependent features because of being able to benefit from AER, low level of acoustic descriptors were used (paralinguistic analysis of speech). Followingly, it is addressed the importance of acoustic descriptors for the automatic recognition of continuous arousal and valence in a “cross-domain” setting.4 databases used listed below: GEMEP: Enacted database created by actors including most known emotion speech VAM: Audio-visual recordings from a TV-show NTWICM: Songs compiled from 1983-2010 with different genres. ESD: Emotional sound database includes emotional connotations from FindSounds.com Dimensions of emotions: 1. Valence describes the intrinsic pleasantness or unpleasantness of a situation. 2. Arousal describes whether a stimulus puts a person into a state of increased or reduced activity. For all four of the databases, the individual listener annotations were averaged using the evaluator weighted estimator (EWE). As feature relevance, besides correlation coefficients (r) of features with the arousal or valence ratings,it is introduced the cross-domain correlation coefficient (CDCC) as criterion.In the present study, it is of particular interest to consider cross-domain evaluation, i.e., training on data from one domain (e.g.,enacted speech) and evaluating on another domain (e.g.,sound). The findings: In the result, cross-domain arousal and valence regression has been proven feasible, achieving significant correlations with the observer annotations. For VAM, it is observed that valence was more difficult to evaluate than arousal, while conversely, on ESD, raters agree more strongly on valence than arousal.

AUTHORS: F. Ren, C. Quan
TITLE: Linguistic-based emotion analysis and recognition for measuring consumer satisfaction: an application of affective computing

SOURCE: Information Technology and Management
YEAR: 2012
NO OF QUOTES: 44
VOLUME: 13
ISSUE: 4
PAGES: 321-332
AFFILIATION: Japan
KEYWORDS: affective computing; enterprise systems; linguistic feature; customer satisfaction
CATEGORY: Article
TYPE: Experimental Research
OBJECTIVE: A fine-grained emotion recognition system proposed to improve customer satisfaction

SUMMARY: This research focuses on emotion recognition application in customer satisfaction field. Albeit linguistic-based emotion recognition is built here, the work has been done using only words and comments(no speech signals). It presents a fine-grained emotion recognition system for measuring CS, that can recognize multiple emotions and make it possible to gain a rich feedback data from customers.
Firstly, in order to measure the customer satisfaction, it is looked if the review is + or -. Secondly, word by word, the data is examined and depending on the context, each word was measured in 8 different emotions. In this first step, the positive emotion, negative emotion, negation, conjunction and punctuation words were to be tested if they increase performance of recognition using with WEF(word emotion feature). The results show that it increases the performance. Also it is found that negative emotions are closely interrelated. When anxiety is detected, there is probably sorrow, anger and hate. Blended emotions: The system used to train and test the model is shown below. As machine learning algorithm, Naive Bayes, Voting feature intervals and Random forest were built. Blended emotions are covering 8 basic emotions.
Multinomial Naive Bayes estimation process produces a model which assigns a sentence Si to the emotion class C that has the highest probability.
In the voting feature intervals method, text is separately represented by a set of feature intervals on each feature dimension. Each feature participates in the classification by distributing real-valued votes among classes. The emotion class with the highest total vote is predicted to be the emotion class of a sentence.
RandomForest is an ensemble classifier consisting of multiple decision trees. Outputs of all trees are aggregated to produce one final prediction that is predicted by the majority of trees.
The evaluation results included recognizing (a) emotional and unemotional sentences and (b) the eight basic emotions for emotional sentences (single emotion matching). The results;

Table 2 Results of sentence emotion recognition using different machine learning algorithms

Machine learning algorithm	F-value	
	(a)	(b)
Multinomial Naive Bayes	90.2	77.3
Voting feature intervals method	77.0	57.0
RandomForest	87.7	75.4

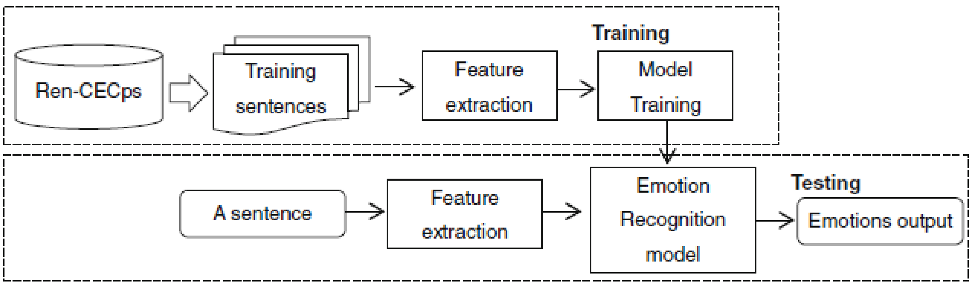
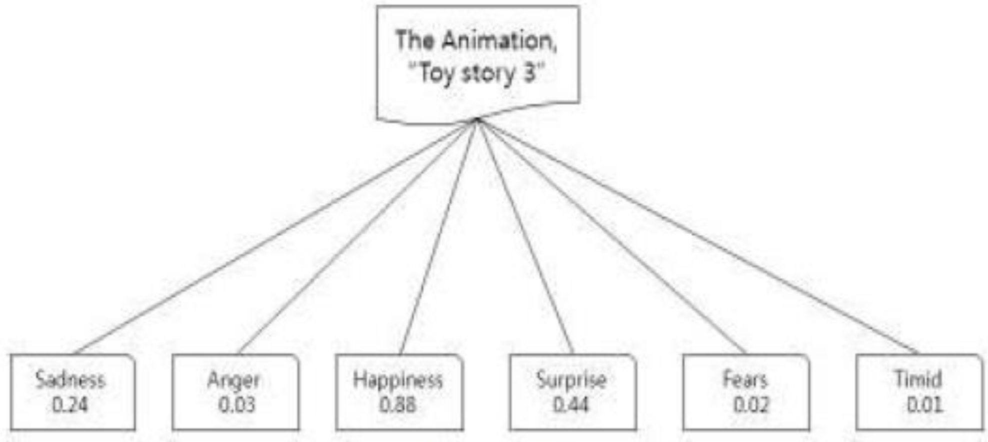


Fig. 4 Main steps of sentence emotion recognition

AUTHORS: H. Lee, J.Kwon
TITLE: Combining Context-Awareness with Wearable Computing for Emotion-based Contents Service
SOURCE: International Journal of Advanced Science and Technology
YEAR: 2010
NO OF QUOTES: 13
VOLUME: 22
ISSUE: --
PAGES: 13-25
AFFILIATION: South Korea
KEYWORDS: context awareness computing;, wearable computing; ubiquitous computing; emotion based content service
CATEGORY: Article
TYPE: Conceptual Paper
OBJECTIVE: To propose novel methodological algorithms that employs the convergence of wearable technique, human's emotional context and tagging technique of web in order to develop more realistic and robust emotion based contents services in ubiquitous computing environment.

SUMMARY: The paper explains the methods to provide the customer a customized service based on his/her emotion and need at that moment. The requirements behind this proposal is wearable computing system to be able to recognize the emotion, construction of emotion tagged repository, search of personalized contents.  **Fig 5. The example of emotion tagged content** The scenario can be applied to the music streaming service, mobile advertising contents service, media streaming service, etc. One is detailed below; • User B fails to get to sleep as her emotional status are much fears and surprised. So, user B wants to listen to music while she lies down on the bed and just turn on the music online streaming service application without inputting any queries. The wearable system attached to the application capture user B's current physical signals, define her current emotional status based on the captured physical signals and transfer her emotion contexts to the application. In this case, the application based on our approach automatically generates two type queries, one is category query, and 'music' and another is emotional queries, 'fears' and 'surprised' which are returned from the wearable system. Then, through searching process described in the above chapter, the application can provide comfortable music proper to insomnia not rock music or dance music.

AUTHORS: S. Gong, Y. Dai, J. Ji, J. Wang, H. Sun
TITLE: Emotion Analysis of Telephone Complaints from Customer Based on Affective Computing
SOURCE: Computational Intelligence and Neuroscience
YEAR: 2015
NO OF QUOTES: 6

VOLUME: 5
ISSUE: --
PAGES: 1-9
AFFILIATION: China
KEYWORDS: --
CATEGORY: Research Article
TYPE: Experimental
OBJECTIVE: The characteristics of telephone complaint speeches

SUMMARY: The paper designs the steps and explains the methodology to detect the customer feeling real-time from the conversations between service staff and the customers. Differently in this research, it is used customer identification information and right before classifying the emotion, this information is merged with the system. 3 emotions were monitored at the experiments: anger, calmness, discontent. detection. To identify the customer, cost sensitive technology is used.

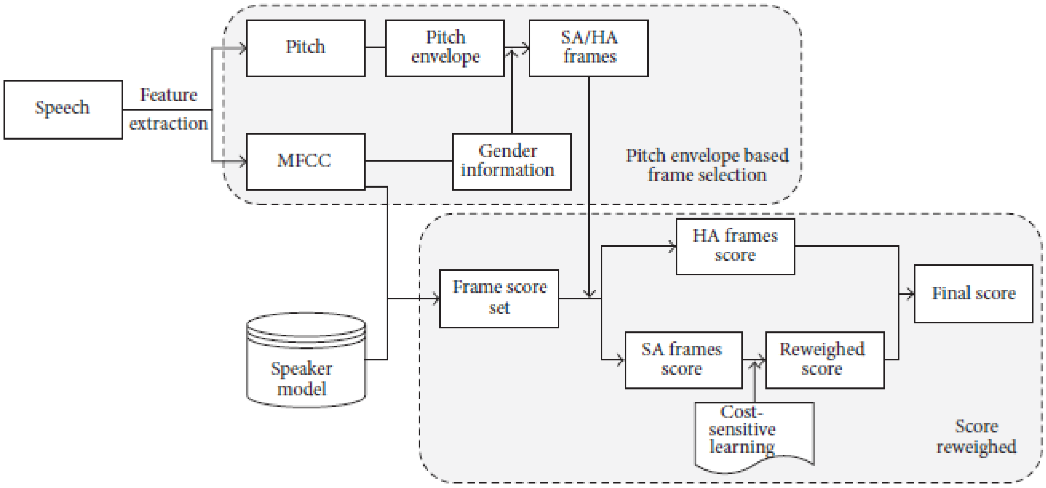


FIGURE 4: Speaker identification algorithm based on cost-sensitive learning technology.

TABLE 5: Average recognition rate of emotions.

Recognition methods	Recognition rate			
	Calmness	Discontent	Angry	Average
BPNN	81.22%	60.46%	80.92%	74.20%
SVM (12-order MFCC)	84.50%	61.40%	83.27%	76.39%
SVM (combined 12-order MFCC and short energy)	88.60%	61.83%	89.80%	80.08%

AUTHORS: N. Gupta, M. Gilbert, G. Di Fabbrizio
TITLE: Emotion Detection in Email Customer Care
SOURCE: Computational Intelligence
YEAR: 2010
NO OF QUOTES: 24
VOLUME: 29
ISSUE: 3
PAGES: 10-16
AFFILIATION: the United States

KEYWORDS: Baseline Systems; boosting; classification (of Information); customer retention; customer satisfaction; customer service; emotional email; salient features; text classification; text processing
CATEGORY: Conference Paper
TYPE: Experimental Research
OBJECTIVE: A method for emotional email identification.
SUMMARY: This paper describes a method for extracting salient features and identifying emotional emails in customer care. Salient features reflect customer frustration, dissatisfaction with the business, and threats to either leave, take legal action and/or report to authorities. Compared to a baseline system using word ngrams, the proposed approach with salient features resulted in a 20% absolute F measure improvement(0.52 to 0.72).

AUTHORS: F. Eyben, M. Unfried, G. Hagerer, B. Schuller
TITLE: Automatic Multi-lingual Arousal Detection From Voice Applied to Real Product Testing Applications
SOURCE:
YEAR: 2017
NO OF QUOTES: 4
VOLUME: --
ISSUE: --
PAGES: 5155-5159
AFFILIATION: the UK, Germany
KEYWORDS: emotion recognition; marketing research; arousal; speech; openSMILE

CATEGORY: Conference Paper
TYPE: Experimental Research
OBJECTIVE: To evaluate a state-of-the-art method which predicts emotional arousal from voice recordings of participants of real market-research studies
SUMMARY: In the paper a state-of-the-art method is evaluated which predicts emotional arousal from voice recordings of participants of real market-research studies The study focuses on arousal it's because it makes the consumer act (e.g. sharing the info). TV commercials and concepts of new products were used as data and annotation made manual by student.s Acoustic features are chosen according to being suitable with emotional arousal. Followingly, the new method is explained in detail in a technical way. The results give the comparison and show the differences between and within a large set of seven corpora collected in real marketing research studies across four different languages.

AUTHORS: S. Scheidt, Q. B. Chung
TITLE: 19. Making a case for speech analytics to improve customer service quality: Vision, implementation, and evaluation
SOURCE: International Journal of Information Management
YEAR: 2019
NO OF QUOTES: 2
VOLUME: 45
ISSUE: --
PAGES: 223–232

AFFILIATION: the United States
KEYWORDS: customer service quality; innovation; speech analytics; key performance; indicators; workforce management; customer experience
CATEGORY: Article
TYPE: Experimental
OBJECTIVE: Building a new customer service model with sophisticated digital technology to solve customer service issues
SUMMARY: The paper emphasises the importance of understanding the customer in terms of gaining competitive advantage over rivals. The experiment takes place in the call-centers of a brand in different states of the US. After the data collecting phase, data mining phase takes start and gives the new measures via speech analytics. This study helps customer to handle the qualitative evaluation of the conversations made by the agents. 3 approaches: phonetic, transcription, phase-matching. Each has different pros and cons. With the metrics chosen based on this particular call-center, the results give an improvement alongside suggesting further study areas. A metric, as an example, can be shown: “Appropriately greet the customer” Further study: “There is a need for research on additional variables such as customer inputs (e.g. education, culture, age, buying behavior, etc.) that may directly or indirectly influence the customer’s evaluation of quality performance and satisfaction”

AUTHORS: P. Loui, J.-P. Bachorik, H.-C. Li, G. Schlaug
TITLE: Effects of voice on emotional arousal
SOURCE: Frontiers in Psychology
YEAR: 2013
NO OF QUOTES: 14
VOLUME: 4

ISSUE: --

PAGES: 1-6

AFFILIATION: the United States

KEYWORDS: emotion; music; arousal; perception; gender; aging

CATEGORY: Article

TYPE: Experimental Research

OBJECTIVE: To investigate the effects of vocals in music on participants' perceived valence and arousal in songs.

SUMMARY: The study is about the measurement of arousal and valence coming from vocal music and only instrumental music using the differences of the two genders and different age groups.

The participants are asked to listed the chosen songs with vocal and without vocal and followingly it is examined which are increasing arousal and affecting valence. The age and gender differences are also taken into consideration since it is thought these demographic features can play role of perception of the music.

The results revealed that the vocal versions were more arousing overall, especially stronger for females than males. It is interpreted that the presence of the human voice, rather than recognizable words, led to the increase in arousal. Familiarity and age do not change significantly the results.

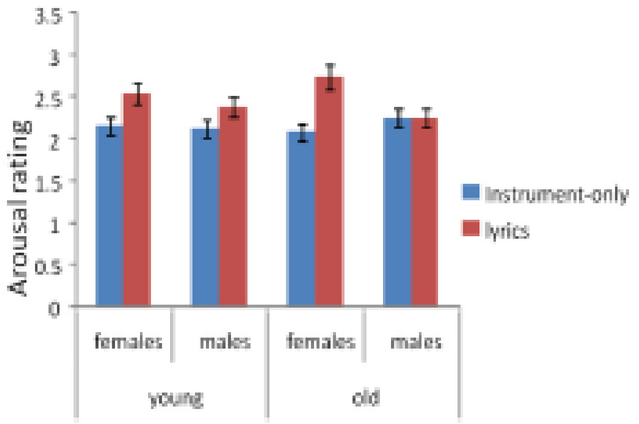
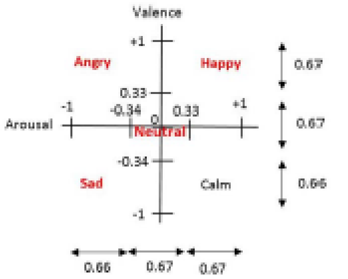


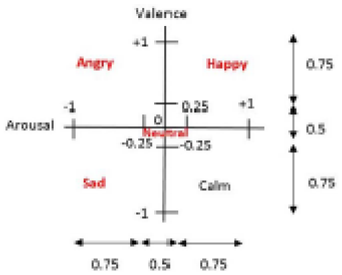
FIGURE 2 | Three-way interaction between vocals, sex, and age on arousal ratings.

AUTHORS: N. Kamaruddin, A.-W.-A. Rahman, A.-N.-R. Shah
TITLE: Measuring Customer Satisfaction through Speech using Valence-Arousal Approach
SOURCE: 6th International Conference on Information and Communication Technology for the Muslim World
YEAR: 2017
NO OF QUOTES: 1
VOLUME: --
ISSUE: --
PAGES: 298-303
AFFILIATION: Malaysia
KEYWORDS: speech emotion; customer satisfaction; affective space model; mel frequency cepstral coefficient; valence; arousal; adaptive neuro fuzzy Inference system
CATEGORY: Conference Paper
TYPE: Experimental Research
OBJECTIVE: In order to quantify the CSI, it is proposed the use of speech analysis based on valence and arousal of the customer, indicating their immediate emotion.

SUMMARY: The customer satisfaction index (CSI) is obtained using Affective Space Model; however, to do this, experiments were conducted by the researchers. They used two dimensions by naming the neutral phase as emotionless state of the speech. And they conducted the experiments by changing the size of this neutral state. During experimental setup phase, mfcc are extracted as features, dataset was chosen depending on the emotions selected and dataset is composed of videoclips, tv sitcom, etc. Among several options of features and classification methods, the best ones (according to the researchers) were chosen.
The results obtained are shown below;

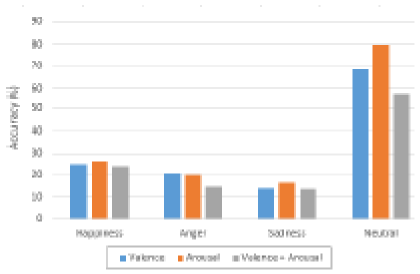


a. Neutral Region with 0.67 Threshold

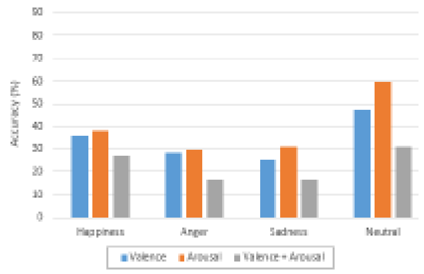


b. Neutral Region with 0.50 Threshold

Figure 4. Different Threshold values for Neutral Region.



a. Classification Performance for 0.67 Threshold Neutral Region



b. Classification Performance for 0.50 Threshold Neutral Region

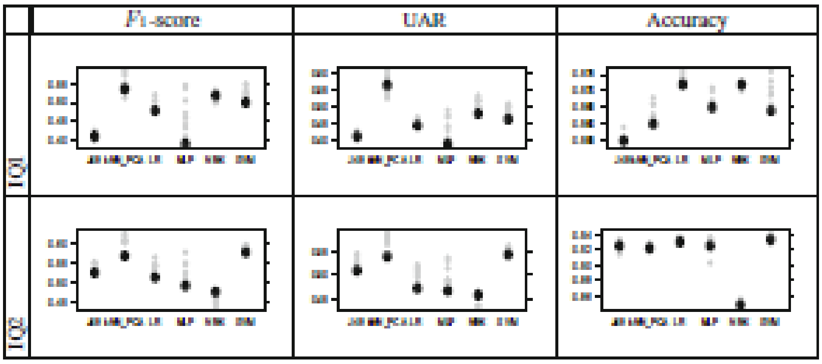
Figure 5. Different Classification Performance for Different Threshold Values of Neutral Region.

The difference coming from the different neutral sizes are due to the training datasets. For the first one, while accuracy is increasing in the emotional states, the accuracy of emotionless state is decreasing. It is because for the emotion states, there is more training data to detect the states.
Thus, it is possible to detect the customer satisfaction and dissatisfaction via the proposed method.
Further works are needed to tune the classifier to have the optimal set-up value.
Profiling neutral region for individual can help improve classification accuracy.
Exploration of the proposed method can be extended to analyze and profile driving behavior especially for sleepy driver where negative arousal and valence values can be used

AUTHORS: A. Spirina, W. Minker, M. Sidorov
TITLE: Could emotions be beneficial for interaction quality modelling in human-human conversations
SOURCE: Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics
YEAR: 2017
NO OF QUOTES: 2
VOLUME: --
ISSUE: --
PAGES: 447-455
AFFILIATION: Germany
KEYWORDS: human-human interaction; task-oriented dialogues; performances
CATEGORY: Conference Paper
TYPE: Experimental Research
OBJECTIVE: Emotions' role determination in Interaction Quality (like Customer Satisfaction) modeling for human-human conversations (HCC)

SUMMARY: To understand the problem happens in a dialogue, it is not necessary to detect the emotion. Emotion detection could be helpful when there is a task-oriented service. To understand the problem, it is seen that CS has been used widely. CS can have the emotion as a variable; but this paper questions whether if it is useful significantly or not, by conducting experiments. 2 datasets were used; one including emotions, the other is emotion-free.
As a result, the ANOVA demonstrates that including emotion data set does not contribute significantly to detect the IQ. The reason can be that the data is very unbalanced (the majority has the same ranking).

Table 3. The graphics of the obtained results for *IQ1* and *IQ2* in terms of F_1-score, UAR, and accuracy. The big black dots perform the results, which have been achieved on the dataset without emotion labels. Whereas the small grey dots mark the results, obtained with using information about emotions (eight emotion sets).



AUTHORS: Y. Park, S.-G. Gates
TITLE: Towards real-time measurement of customer satisfaction using automatically generated call transcripts
SOURCE: International Conference on Information and Knowledge Management, Proceedings
YEAR: 2009
NO OF QUOTES: 18
VOLUME: --
ISSUE: --

PAGES: 1387-1396

AFFILIATION: The United States

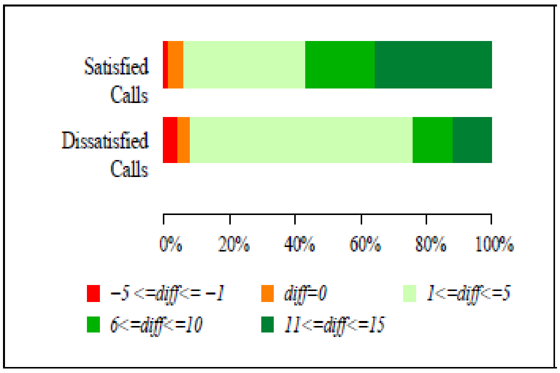
KEYWORDS: customer satisfaction, contact center calls, speech analytics, natural language processing, text mining, classification, machine learning

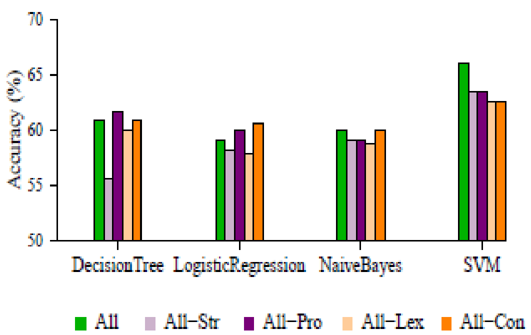
CATEGORY: Conference Proceedings

TYPE: Experimental Research

OBJECTIVE: A fully automated method for measuring customer satisfaction by analyzing automatically transcribed calls.

SUMMARY: The study examines the call center database of an automobile company and aims to measure the CS in near real-time to control the conversation and take action on time. The satisfaction scale is 2 and 5-point. The performance measurements are obtained with the accuracy, prediction and F-measures. The classifiers are human, machine and artificial classifiers. In the end, all these substitutes are compared in order to find the best or most suitable one. The features are of 4 groups: prosodic, lexical, contextual and structured features. Depending on the middle of calls and the end of calls, the features are selected. The point in this study to find out the best method to measure the customer satisfaction. The researchers extracted the features and classify the transcripts depending on different classifications, followingly compared the performances.
In terms of features; the leave-one-out method is chosen to see the differences between features. The result: All features contribute significantly to the CS measurement. (Using only positive and negative words are not enough to distinguish satisfied and dissatisfied customers.)





For classification methods; in general automatic classification methods outperform the human judgment. Shown below;

Methods	Classification Accuracy	*Satisfied* Calls			*Dissatisfied* Calls		
		Precision	Recall	F-measure	Precision	Recall	F-measure
DominantClass	75.96	75.96	100.0	86.34	0.00	n/a	n/a
CSRJudgment	86.54	89.16	93.67	91.36	76.19	64.00	69.57
Decision Tree	89.42	92.50	93.67	93.08	79.17	76.00	77.55
Logistic Regression	85.58	92.41	97.47	90.68	68.09	64.00	68.09
Naive Bayes	83.65	82.98	98.73	90.17	90.00	36.00	51.43
SVM	89.42	87.78	100.0	93.49	100.0	56.00	71.79

SVM is slightly a better option.
Comparing 2-point and 5-point CS, the 2-point has higher accuracy as expected.

AUTHORS: F. D’Errico, M. Paciella, B. De Carolis, A. Vattanid, G. Palestra, G. Anzivino

TITLE: Cognitive Emotions in E-Learning Processes and their Potential Relationship with Students’ Academic Adjustment

SOURCE: International Journal of Emotional Education

YEAR: 2018

NO OF QUOTES: 6

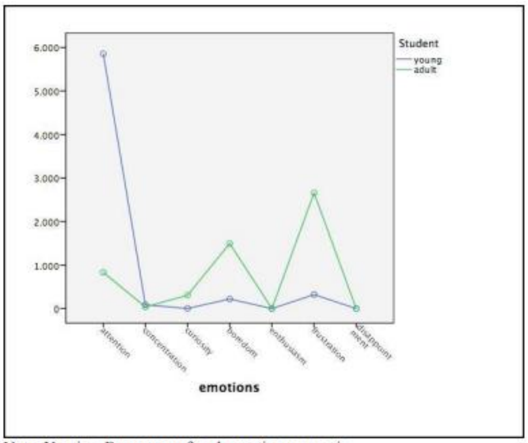
VOLUME: 10

ISSUE: 1

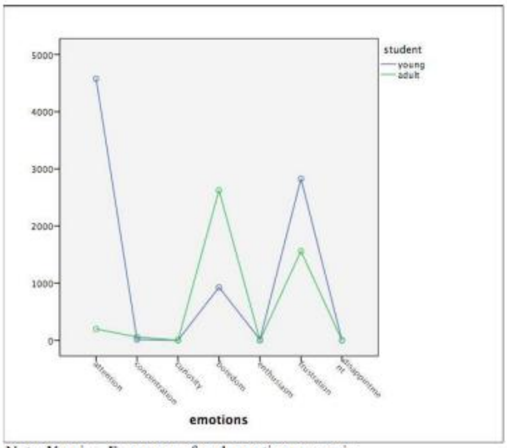
PAGES: 89-111

AFFILIATION: Malta

KEYWORDS: cognitive emotions; self-efficacy; academic adjustment; automatic detection of emotions; e-learning process
CATEGORY: Article
TYPE: Experimental Research
OBJECTIVE: Measuring emotion through state-of-the-art tools to improve e-learning
SUMMARY: The paper focuses on the relationship between cognitive emotions and outcomes of e-learning activities (e.g. achievement at the exams). Cognitive emotions lead people to approach the event in a biased manner. For example, when a student feels not confident about the topic, he/she approaches it with potential frustration emotion and this, at the end, changes the outcome of the event. However, this study creates the experiments in 2 ways: • 10 students watch an educational video • 10 students chat with a teacher during a session Emotions are chosen based on this special experiment. Instead of taking 6 basic emotions, the researcher chose the emotions as listed; "attention, interest, surprise, curiosity, concentration, enthusiasm, disappointment, boredom, confusion, annoyance, and frustration" The reason is that these emotions are of importance to understand better how e-learning affects students. Previous studies have used self-report measures and not considered the real-time emotion. In this study, the core is to detect the cognitive emotions in e-learning contexts in relation to personal beliefs, academic well-being and performance. The personal beliefs and academic well-being have been obtained applying questionnaires the participants. They were divided into 2 groups as young and old participants, high self-efficacy and low self-efficacy students. To get a reliable result from experiments, the researcher has used the proven methods alongside examining several different ways for some steps. And questionnaires have been interpreted using a 5-point likert-scale. To train the model, they have used another dataset since it is not possible to understand the emotions at the very first second when the participant engage e-learning tasks. The FER accuracy has been tested in 4 ways: (a) AVG Accuracy= No. of correctly recognized emotions of all types/No. of all the emotions (b) Precision= No. of correctly recognized emotions labeled as Q/ No. of all the emotions recognized as Q (c) Recall= No. of correctly recognized emotions labeled as Q/ No. of all the emotions labeled as Q (d) F-Measure=2*((Precision + Recall)/(Precision*Recall)) The Results:



Note: Y-axis = Frequency of each emotion on x-axis



Note: Y-axis = Frequency of each emotion on x-axis

Figure 3. Emotional Profiles in video-lectures*Age

Figure 4. Emotional Profiles in chat with tutor*Age

The results revealed that for old people and young people, emotions are different in both types of the experiment. Young people are less disturbed with the video, however in general they have better results than the old. On the other hand, old people prefer an engagement during the elearning activities.

Table III. Academic profiles across younger and older students with opposite levels of self-efficacy

Students			Exams			Well-being					Self-efficacy
	age		Number	Mean		SAT	PER	INT	GRA		e-Task
YS-HS	25		14	29		4.29	4.83	4.71	4.83		4.67
YS-LS	21		9	25		3.29	3.67	3.29	4		3.67
OS-HS	63		14	28		4.86	4	4.57	5		5
OS-LS	64		9	25		4.71	4.5	4.71	5		3.67

Note: SAT = satisfaction; PER= persistence; INT= interdependence; GRA = gratitude. YS-HS= younger students with high self-efficacy; YS-LS = younger students with low self-efficacy; OS-HS= older adult students with high self-efficacy; OS-LS= older adult students with low self-efficacy

Looking at the results in the individual-level, it can be interpreted that the achievements are independent from age relatively comparing to the self-efficacy of individuals. This means that people feeling more confident made better performance.
The overall result is that the relationship of cognitive emotions and performance is very significant while evaluating e-learning activities. How people approach the event can explain the behavior of the individual in the end.

AUTHORS: D. McDuff, R.-E. Kaliouby,J.-F. Cohn, R.-W. Picard

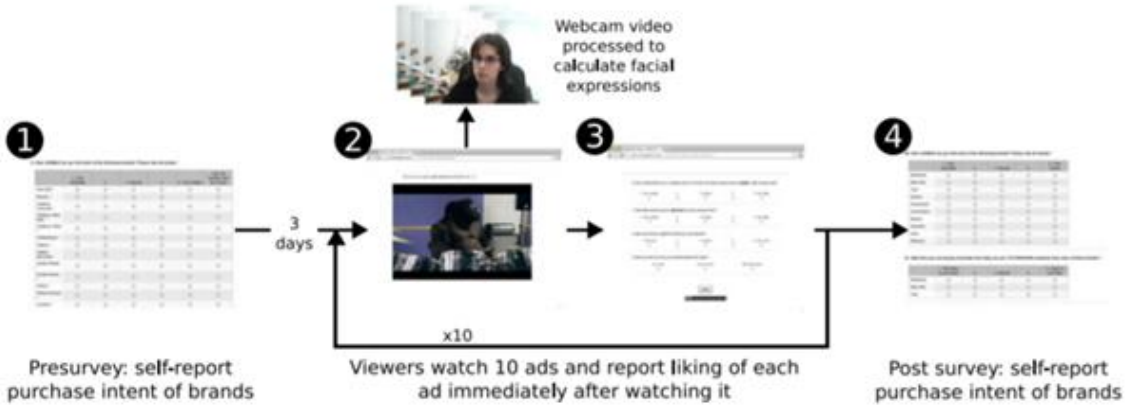
TITLE: Predicting Ad Liking and Purchase Intent: Large-Scale Analysis of Facial Responses to Ads

SOURCE: IEEE Transactions on Affective Computing
YEAR: 2015
NO OF QUOTES: 38
VOLUME: 6
ISSUE: 3
PAGES: 223 – 235
AFFILIATION: the United States
KEYWORDS: facial expressions; emotion; market research
CATEGORY: Article
TYPE: Experimental Research
OBJECTIVE: The relationship between facial responses and ad liking

SUMMARY: In this study, researchers tried to clarify the relationship between facial responses and ad liking and changed in purchase intent. Looking at the facial expressions, they pursued to model ad liking and purchase intention of the viewers in addition to find the relationship of these three metrics. An enormous dataset has been used for this study and the data have been collected from the real life, not a lab environment. Ultimately, the researchers had the propensity to identify the features of aggregate emotional responses that make an ad effective such as where to use brand name.

Previous studies claimed that psychological and self-reported measurements capture different information. Also, including emotion in an ad can keep the viewer paying more attention.

Experiments:
Data: 170 ads, 4 countries, 12230 facial responses, 3million+ frames. Product categories (Petcare, confectionary, food, other) Emotion categories (amusement, disgust, ...)
Survey: The participants filled 2 types of surveys which are presurvey that asks the intention of purchase before the ad and main survey which asks the PI after the ad. At least 70% of the participants were chosen according to the fact that they were using the product before.



Face Analysis: For this part of the study, the researcher has used the state of art methods like HOG for the feature extraction and SVM. The frames were examined looking at eyebrow raises, smiles, disgust and +/- valence expressions. For all have been used binary measurements. As an instance, existence of smile is 1 and nonexistence of smile is 0.
Judgments: The images from these datasets were labelled for the presence of an expression by human coders.
The result arises here is that looking at the face expressions only gave only 17% significant response from the viewers. although responses are sparse, different people respond to the ads differently and the ads elicited a range of expressions (from strong positive valence to negative valence). So, the researcher created the aggregate metrics and this way gave rich information about the effectiveness of the ads.

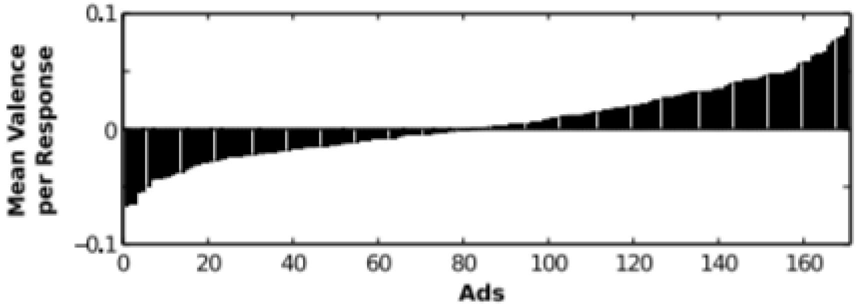


Fig. 10. Mean expression valence metrics for the 170 ads sorted by ascending mean valence.

Calculating aggregate metrics: - Contextual features: binary matrix depending on the product
-Facial metric features: using median, mean, max and min

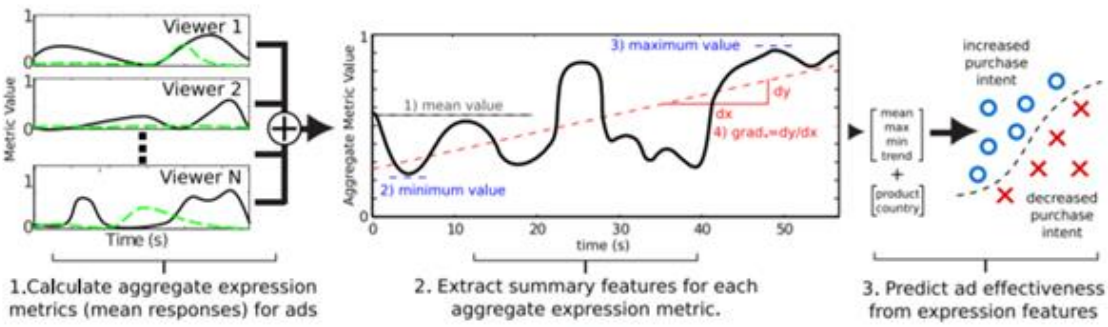


Fig. 11. 1) Aggregate metric tracks calculated from all viewers who watched the ad. 2) Features extracted from each of the aggregate metric tracks: a) mean value, b) minimum value, c) maximum value, d) the gradient of the linear trend. 3) Summary features extracted from the facial expression metrics used to predict ad effectiveness.

Computing labels: Liking score & purchase intent score: this score has been obtained from the self-reported responses of viewers to the questions
Training and test analysis: using leave one ad out method is chosen for the test and training part. SVM is chosen as a classification method.

The results:
- The emotion profiles of ads that generate high ad liking is a strong gradient and high peak in positive expressions (valence and smiles).
The prediction performance was lower for the PI model than for the liking model suggesting that the relationship between facial responses and changes in PI is more complex, as expected.
Brand appearances immediately prior to the peak positive emotion is a driver for increasing purchase intent

AUTHORS: P. Lewinski, M.-L. Fransen, E. S. H. Tan

TITLE: Predicting Advertising Effectiveness by Facial Expressions in Response to Amusing Persuasive Stimuli

SOURCE: Journal of Neuroscience, Psychology, and Economics
YEAR: 2014
NO OF QUOTES:64
VOLUME: 7
ISSUE: 1
PAGES: 1-14
AFFILIATION: the Netherlands
KEYWORDS: facial expressions of emotion; AAD; AB; FaceReader; amusement
CATEGORY: Article
TYPE: Experimental Study
OBJECTIVE: To establish the link between ad effectiveness and facial behavior toward ad stimuli

SUMMARY: In this study, researchers predict the ad effectiveness measuring facial emotional state and comparing its performance to the self-reports. FaceReader was chosen as a tool to detect the emotion in the participants' faces and it was used for similar reasons before as well.The stimuli was shown and the reactions of the viewers were recorded in the natural environment. Due to the topic of the research, only one emotional state was measured which is happiness(smile, amusement).
Advertising effectiveness is formulated as combination of ATTITUDE towards the ad and the ATTITUDE towards the brand. Each gives slightly different insight due to the fact that liking the ad does not necessarily mean the liking the brand. For this study, researchers employ facial emotion recognition tools because this sort of a concept is hard to capture via self-reports and plus, autonomic measures are already proven methods to convey better performance to gauge the emotion.
The main goal of the research is to differ the levels of amusement of the adverts and reach the ad effectiveness accordingly. There are 4 hypothesis were tested for the study, listed below:

H_1: Highly amusing video advertisements elicit more frequent and more intense facial expressions of happiness than medium and low amusing ones;

H_2: There is no difference between highly, medium and low amusing video advertisements in the frequency and intensity of facial expressions of all other basic emotions but happiness.

H_3: There is a positive correlation between frequency and intensity of facial expressions of happiness and self-reported measures of ad effectiveness (a) attitude toward the ad and (b) attitude toward the brand in the highly and medium but not in the low amusing video advertisements;

H_4: There is no correlation between frequency and intensity of facial expressions and all other basic emotions but happiness and self-reported measures of ad effectiveness (a) attitude toward the ad and (b) attitude toward the brand in any of the highly, medium or low amusing video advertisements.

The hypothesis basically compare the happiness with all the other emotions and the levels of happiness with the ad effectiveness.

Pretesting:
The stimuli was chosen by ad experts.The judgment was also made depending upon the length of the ad (M=30") The ads are not well-known but well-presented (service+product) 30 female and 30 male US citizen participants were recruited through Amazon MTurk with the aim of scoring the amusement level of ads. The participants made ranking (1 to 5, 5=very funny) In the end of process 16 ads were eliminated to 6 (high, medium and low level of amusement) and absolutely the participants were questioned about the familiarity of ad/brand and product before.

Result: No significant result recorded about the familiarity of ad or brand by participants or among two genders for the level of amusement.

Experiment: 51 men, 39 women with the same process of pretesting were recruited. The data collection was made through MTurk as well. Each participant watched 3 different types of ad.
FaceReader uses 3 layer neural network to analyze facial expressions of emotions.

1. Finds the face using Active Template Method
2. Creates a virtual, superimposed 3D Active Appearance Model of face (500 landmarks)
3. Scores for the intensity and probability of facial expressions for basic emotions

The neural network of system was trained using 10000 images judged by the human coders.
Experiment Results: Global mean average was used to perform all the calculations for the each video recordings. For each participant, it is used their facial expressions from lowest to highest level for that emotion and top 10% peak values for all expressions were calculated. Frequency of the facial expressions for that emotion was taken into account to calculate the reaction.The cut-off criterion did not influence the result, meaning that using top 10% peak values is safe.
Hypothesis testings:
H1: Friedman tests H2: Pairwise comparisons with Bonferroni correction
H3-H4: Spearman's rank-order correlations
To avoid subjectivity of participants (naturally having better temper,mood) they obtained sufficient amount of sample and used repeated measures. However, due to using product and service ads which can be less known or well known, it is subjected to face biases. Another limitation in this study is to measure only happiness and the other emotions could be evoked by stimuli as well.
Variables = stimuli characteristics(AAD, AB, emotions of the facial expressions, humor, brand, positive and negative cognitive responses) + (viewer characteristics (gender, sex, socioeconomic status) based on Eisen's review (2011).
Results: * FaceRead successfully distinguish amusive and no amusive ads looking at facial expressions.

- Cognitive responses and facereader results correlate. Thus, it is possible to measure ad effectiveness.
- When they smile, it was not noted as they like the ad or brand. Because they were primed after watching a nice video. Hence, no correlation between smiles and liking the ad.
- Poor quality of videos did not vaguely affect to measure emotion and analyze the faces.

AUTHORS: M. Horvat, D. Kukolja, D. Ivanec

TITLE: Comparing Affective Responses to Standardized Pictures and Videos: A study report

SOURCE: 38th International Convention on Information and Communication Technology, Electronics and Microelectronics

YEAR: 2015
NO OF QUOTES: 2
VOLUME --:
ISSUE: --
PAGES: 1394-1398
AFFILIATION: Croatia
KEYWORDS: arousal; emotion; happiness; human; personality test; psychometry; reference value; social dominance
CATEGORY: Conference Paper
TYPE: Experimental Research
OBJECTIVE: The comparison of still image and video-clip to detect the emotion
SUMMARY: Many stimuli are composed of still image and sound in the background; however, the video is stimulating more intensive emotions than this format. In the study, it is aimed to create video-clips as stimuli. Experiment: 10 (4 male, 6 female) participants were exposed to the images and images including sound and were told to fill a survey assessing the emotion type and intensity. Two dominant emotions were selected in the stimuli: happiness and fear. The stimuli was chosen from IADS and IAPS. The very first stimuli included 200 images and 100 samples depending on their intensity level of emotion norms. Subsequently, this number reduced to 20 images and 20 sounds stimuli with the eliminations done by psychology experts. Neutral stimuli had the characteristics low arousal and valence in picture and no dominant emotion in video. Each participant was exposed to 5 minute happiness and 5 minute fear emotion stimuli and followingly self assessment response. Meanwhile, their signals were recorded in the forms of skin conductance, electrocardiagram (ECG), respiration and skin temperature. All the experiment was held in the laboratory environment. Results: In arousal emotion dimension, according to self assessment videos are more effective than still images. In videos, it was more obvious to see higher pleasure in happiness sequences and less pleasure in fear sequences. In videos, arousal emotion dimension was more apparent than basic emotions in terms of showing the intensity of emotion.

However, having 10 participants in the experiments affected the results in terms of individuality. For this study, it was a drawback not to recruit more participants and get better database rather than IAPS and IADS.

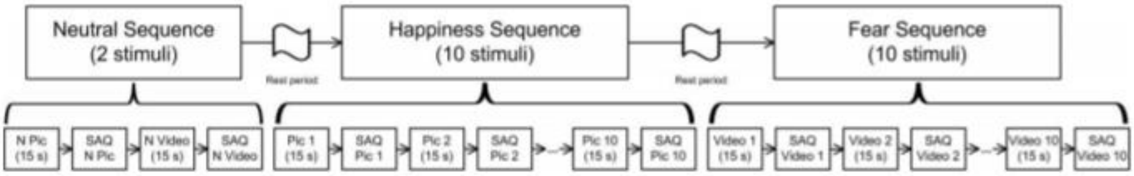


Fig. 2. The timeline of emotion stimulation paradigm. The order of dominant emotion (happiness/fear) and multimedia format (picture/video) were randomized.

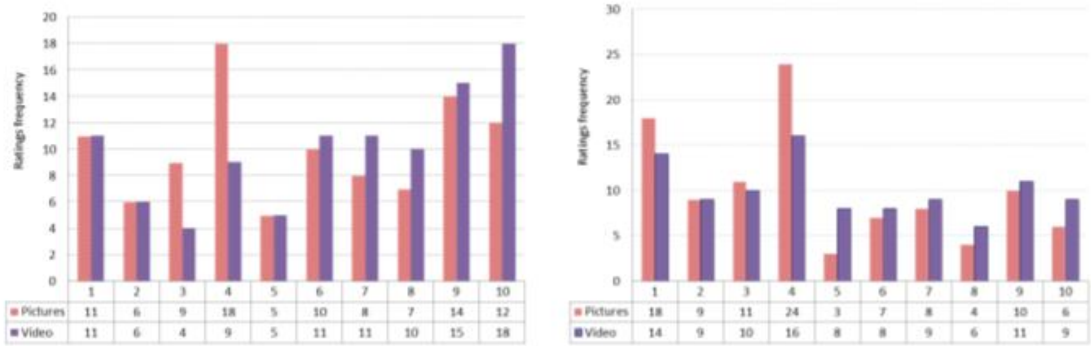


Fig. 4. Frequency distribution of N=10 participants' ratings after elicitation with IAPS and IADS video clips ("Video") and IAPS pictures ("Pictures"). Reported emotional dimensions arousal and pleasure (upper and middle rows), discrete emotions happiness and fear (bottom row).

AUTHORS: D. McDuff, M. Soleymani
TITLE: Large-Scale Affective Content Analysis: Combining Media Content Features and Facial Reactions
SOURCE: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)
YEAR: 2017
NO OF QUOTES:8
VOLUME: -
ISSUE: -
PAGES: 339-344

AFFILIATION: the United States, Switzerland

KEYWORDS: Videos; media; feature extraction; visualization; encoding; tagging; software

CATEGORY: Conference paper

TYPE: Experimental Research

OBJECTIVE: A novel multimodal of the combination of video characteristics and facial reactions to better detect affections

SUMMARY: Fusion model = Media Content (Visual+Audio+Sentiment Descriptors) + responses to media stimuli(facial action measurements)

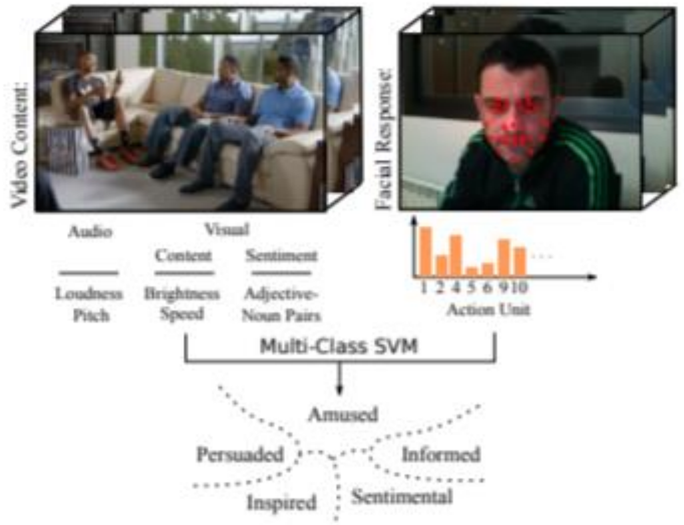


Fig. 1. Large-scale implicit affective content tagging using facial responses to media collected over the Internet. Audio and video features are extracted from the media content. Automated facial coding is used to quantify responses to the media content. Features are extracted and a classifier trained for discriminating between different types of affective content.

As functionalized above, the study focuses on detecting emotion in the video looking through both media content characteristics and facial responses to that media content. This combined model is supposed to give better results than either way. At the first step of the research, 2215 videos were selected labelled using Amazon Mechanical Turk. Afterwards, this number was reduced to 384 with the help of coders who agree on the videos within the classes of informative, persuasive, inspiring, sentimental, amusing.

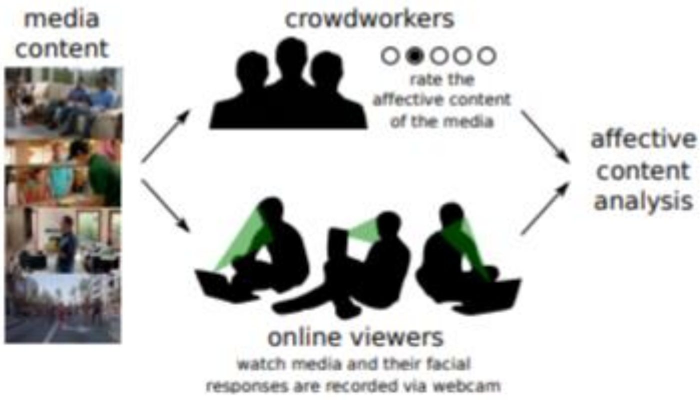


Fig. 2. For our affective content analysis, labels were collected from crowd-workers who rated the affective content of the video media. Independently viewers watched the media online and their facial responses were recorded and coded.

However, any type of ad, as stimuli, is supposed to be designed to be likeable and persuasive anyway. The experiment was applied on 10.000 participants which made it the most data-intensive study of its kind. In the literature, it is stated that there are normally two ways of understanding affect in the video:

Implicit tagging -> responses to media, facial expressions -> recommendation systems

Explicit tagging -> asking users to assess directly

In this study, both ways were to be used. Plus, FACS is chosen due to providing objective and comprehensive taxonomy.

In the video: audio= qualities (pitch, loudness, power in different frequency bands) and in the end 89 features were extracted and hand picked using openSMILE

visual= 51 simple low level visual descriptors (motion, contrast, entropy)

sentiment= adjective-noun pairs and images, then deep convolutional neural networks

STIMULI VISUAL FEATURES, THEIR DESCRIPTIONS AND DIMENSIONALITY.

Feature	Description	#
Visual features [24]	Entropy, exposure, balance, brightness, compression quality [25], contrast, sharpness, uniformity, image asymmetry (intensity), image asymmetry (histogram of gradients (HOG)), motion component (norm of difference between consecutive frames), color histogram (four bins for each color (RGB) channel), Contrast balance (Euclidean distance between the original image and the contrast-equalized image), video length, number of pixels	51
Sentiment descriptors [26]	Probabilities of adjective noun pairs related to sentiment	4342
Acoustic features	eGeMAPS[27] feature-set including pitch, loudness and Mel Frequency Cepstral Coefficients (MFCC)	89

Facial Expressions= automated software was used, 34 facial landmarks (eye, eyebrow, nose, mouth). To capture textural changes of face, HOG features were extracted from face. 18 facial actions were determined in the end.

Results: 63% accuracy using media content + facial action features to predict affect.

Using all these decriptors is significantly better in terms of accuracy than using only media descriptors or facial expressions. Nevertheless, obtaining better accuracy do not mean that this is a great classification because of no uni-modal facial response or universal expression.

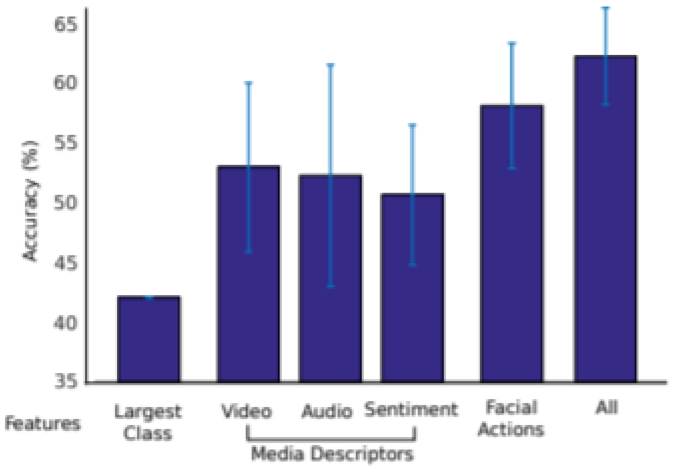


Fig. 4. Average accuracy across 10-fold testing. The error bars represent one standard deviation either side of the mean.

AUTHORS: D. McDuff, Student Member, R. El Kaliouby, R..W. Picard
TITLE: Crowdsourcing Facial Responses to Online Videos
SOURCE: IEEE Transactions on Affective Computing
YEAR: 2012
NO OF QUOTES: 52
VOLUME: 3
ISSUE: 4
PAGES: 456-468
AFFILIATION: the United States

KEYWORDS: crowdsourcing; facial expressions; nonverbal behavior; advertising; market research

CATEGORY: Article

TYPE: Experimental Research

OBJECTIVE: A novel framework to collect and analyze facial responses to online media content

SUMMARY: Through 3 commercials, a novel framework was created to collect and analyze the responses of participants. The paper compares traditional corpus with the novel one in terms of pose, lumination, position of participants. Intensity and dynamics of smiles towards the stimuli differ largely between the groups such as the ones who liked the ad and the ones who did not like the ad. The difference also occurs for another two groups like the ones who are familiar with the ad and the ones who are not. Unlike the aforementioned studies, in the research is observed the relationship between head movements and facial behaviors. In the non-laboratory environment, the environment has significantly greater influence on the viewer and causes more disturbances in the data.
In the study, researchers mostly focus on smiles depending on its effect to keep the viewer being attracted to the stimuli (avoiding “skip the ad” effect). It was also provided a survey of work on emotion measurement in advertising including evaluation of self-report and facial measurement techniques.

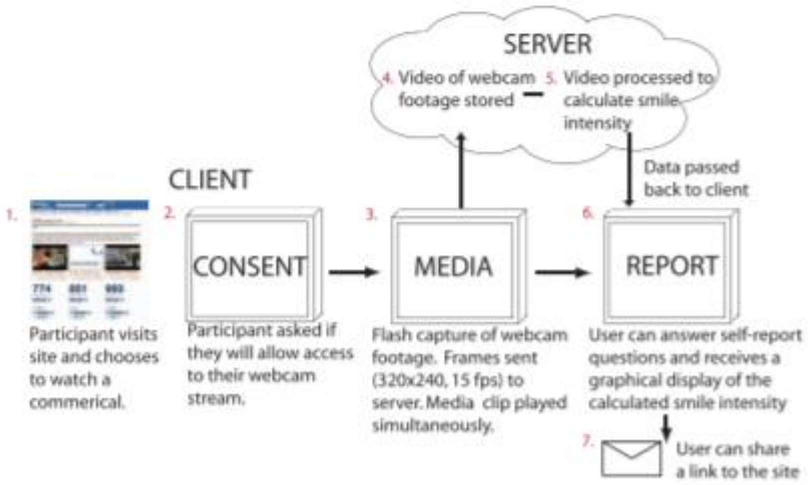


Fig. 2. Overview of what the user experience was like and Affectiva's (www.affectiva.com) web-based framework that was used to crowd-source the facial videos. From the viewer's perspective, all that is needed is a browser with Flash support and a webcam. The video from the webcam is streamed in real time to a server where automated facial expression analysis is performed, and the results are rendered back to the browser for display. All the video processing was done on the server side.

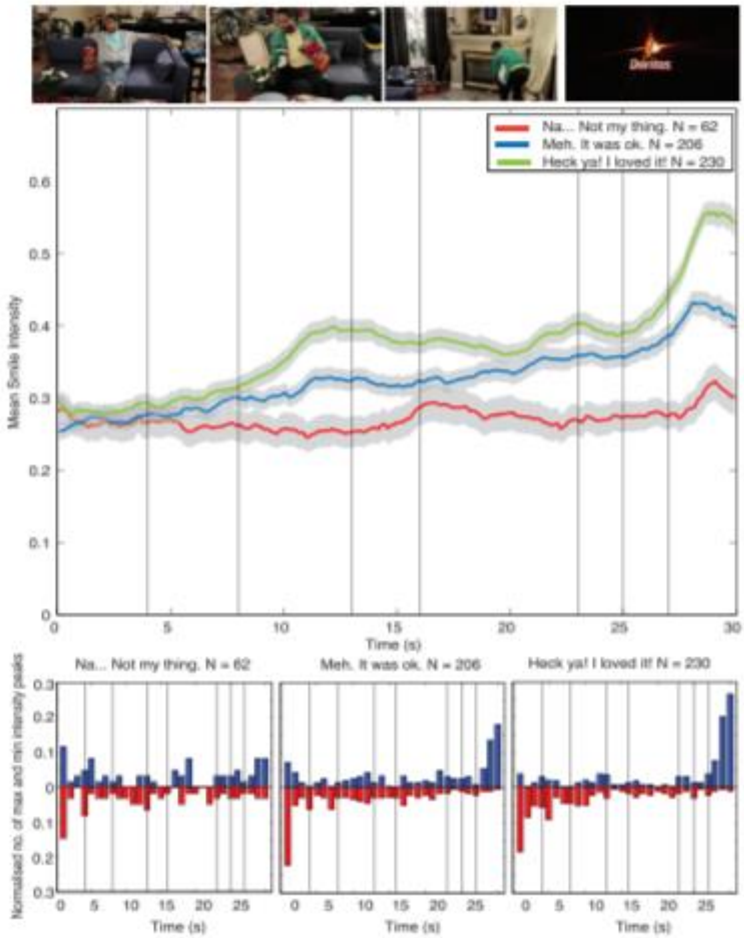


Fig. 19. There are significant differences in the smile responses between people that reported liking the ads more than others. Top: The mean smile intensity and standard error while watching the Doritos ad for the three self-report classes. Histograms of the maximum (blue) and minimum (red) smile intensity peak locations while watching the Doritos ad for the three self-report classes.

The most smiled ads were responded as most liked ads in the survey.

AUTHORS: B. de Gelder, J. Vroomen
TITLE: The perception of emotions by ear and by eye
SOURCE: Cognition and Emotion
YEAR: 2000

NO OF QUOTES: 358
VOLUME: 14
ISSUE: 3
PAGES: 289-311
AFFILIATION: Netherlands
KEYWORDS: affect; facial expression
CATEGORY: Article
TYPE: Experimental Research
OBJECTIVE: The differences of perception by ear (music) and eye (image) and their intertwisted effect
SUMMARY: The researchers conducted 3 experiments to explore the question of integration of emotion in voice and facial expressions in two ways as combined and separated. The first experiment asked participants to watch face with an accompanying voice and express the emotion inserted in the stimuli. In this experiment, there is a continuum of face pics changing the mood from very sad to very happy. Accompanying voices are sad, happy and neutral. At the end, the participant is affected by both. The second experiment asks them to ignore voice and follow the face to determine the emotion. However is seen that the voice has a significant effect on the decision. In the last experiment, the reverse case of second experiment was applied. All in all, it is realized that the participant is affected from voice and face significantly even though they were told to ignore voice or face.

AUTHORS: C. Maaoui, F. Abdat, A. Pruski
TITLE: Physio-visual data fusion for emotion recognition
SOURCE: IRBM
YEAR: 2014
NO OF QUOTES: 4
VOLUME: 35
ISSUE: 3
PAGES: 109-118
AFFILIATION: France
KEYWORDS: physiological features; facial expressions; bimodal; data fusion; decision-level; feature-level
CATEGORY: Article
TYPE: Experimental Research
OBJECTIVE: Using bimodal at feature level and decision level data fusion and comparing the results to the ones obtained for single-model emotion recognition

SUMMARY: The experiments of this study were applied upon the patients with anxiety disorders so as to gauge positive and negative emotions in valence dimension. The features were extracted from signals from facial expressions and physiological reactions. These features were used to being classified for emotion recognition as single modality; concurrently these features were transformed to be used in bimodal methods like vote, dynamic bayesian network and mutual information, PCA, concatenation. However, biosignals overperform facial expressions due to the inherent characteristics like not being easily controlled. Physiology features = heart beat, SC, ..
Visual features= facial expressions, gestures, postures, etc
Unlike from previous studies done in the similar area, this study provides a new means for ER and finds significant classification rates from bimodal data
As to measure visual features, the muscle movements beneath the skin were taken into account and dynamic and static points were determined in the face. Using distance in neutral and emotional face expressions were the way to extract features.
After extracting the all features needed as visual and physiological, the bimodal were established. Normally this can be applied at data level, feature level and decision level. In this study, last 2 methods were used.M utual Information= if the feature and class label give high information then that feature is selected relevant. Afterwards this feature set is transformed via PCA to reduce dimensionality.

The experiment: test phase was completed via 10 subjects (8male, 2 female) the stimuli was chosen from IAPS and was evaluated by psychologists. The real experiment stimuli is 60 images for each subject and for each subject 5 second was assigned to measure the + and - emotions. As classifiers, SVM was chosen. RBF classifies non-linear problems. Cross validation examines accuracy of each SVM with k-fold cross validation.

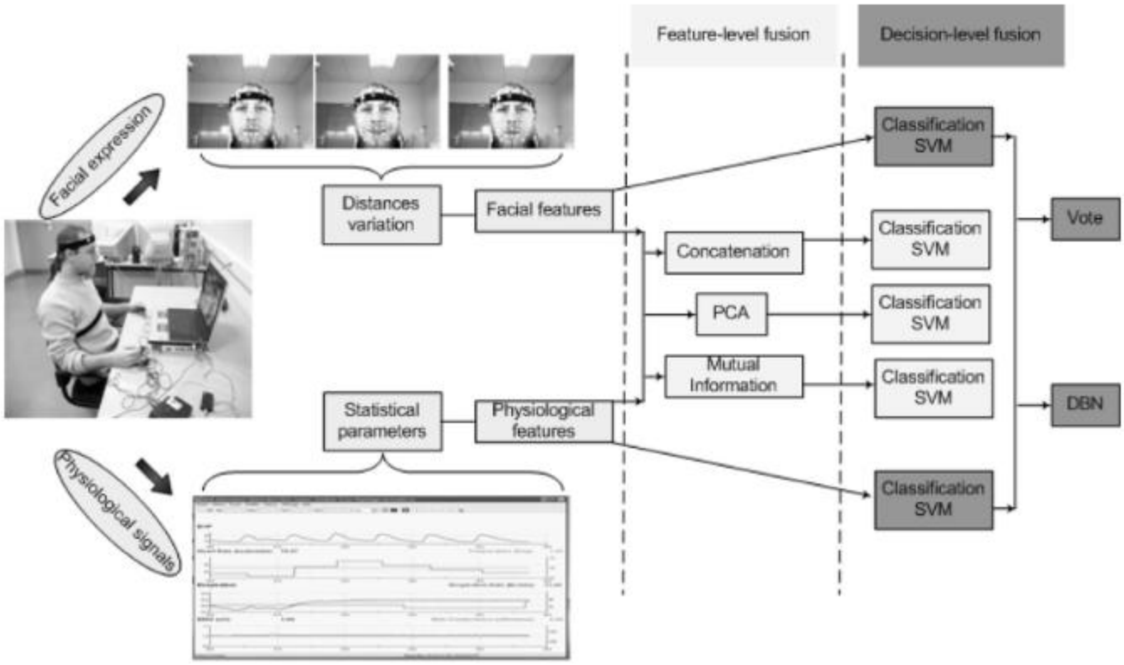


Fig. 3. Bimodal system outline with different levels of fusion.

Results: the results were acquired for 4-day study data and 1-day study data, also they were in two categories as person-dependent and person-independent in the latter one

test and training subjects were different. The reasons of obtaining dissimilar outcomes come from the fact that 4-day results can be varied with sugar, caffeine consumption, sleep and mood changes. To overcome this effect, using facial expressions are a better choice. However, owing to the aforementioned reasons, physiological features are better at some point too. Which is why it is suggested to use both feature types. Overall result for both comparisons, PCA outperforms other feature level fusion modalities. And feature level outperforms decision level fusion (in 1d and 4d, in person dependent and person independent). Moreover, self assessment performs better than IAPS annotation.

AUTHORS: Z. Zeng, M. Pantic, R.G. Roisman, T. S. Huang
TITLE: A Survey of Affect Recognition Methods: Audio, Visual and Spontaneous Expressions
SOURCE: PubMed
YEAR: 2009
NO OF QUOTES: 2009
VOLUME: 31
ISSUE: 1
PAGES: 39-58
AFFILIATION: the United States, the Netherlands

KEYWORDS: evaluation/methodology; human-centered computing; affective computing; introductory; survey

CATEGORY: Article

TYPE: A comprehensive review

OBJECTIVE: The emotion recognition methods review article

SUMMARY: Today's HCI designs are lack of the ability to convey implicit information and the current communication via HCI is perceived as cold and incompetent (keyboard, mouse). To be able to reach the rich information regarding affective computing, more human centered design is needed. There are the potential improvement in terms of performance in both commerce and research in case the HCI yields trustable results and can be easy to apply. For this reason, the AC field has been a great field to work on for researchers. Due to the difficulty to obtain robust data, it emerged the use of audiovisual modality and on top of that it provided better results in this field compared to single modalities (visual and audio).
The way to decide the emotion from a human perspective is described in three ways: discrete categories, dimensional description and appraisal-based approach. Each has some pros and cons.
About the relationship between affect, audio and visual signal, it can be said that psychologists and linguists have created a valuable bunch of information. Looking at the psychology department first, Facial Action Coding System (a manual system) describes 27 AUs and plenty of descriptors to detect the emotion in the face. From a linguistic point of view, people communicate through voice and words and these provide linguistic and paralinguistic information. All can be obtained through signals and match with the regarding emotion. However, in the literature, the results show that acted data is reliable to some extent using these data. Nonetheless, the natural data includes much dynamism coming from human behavior and it is quite challenging to overcome dynamism using audio and visual data. Therefore, audiovisual multimodality seems to be a need owing to yielding higher accuracy. Moreover, context provides useful information that can change the emotion completely (for example smile can represent happiness or greetings).
Regarding the database used in AC, since the AER is not well developed yet, the researchers tended to use acted data in order not to face the described input data. However, the acted data do not represent the real world data also for the facial expressions. Because in spontaneous data, people react differently (e.g., longer smiles). Also this leads the acted data failing to representing the real world data. On the other hand, due to technical necessities, the input data must resemble a portrayed one. Even though the portrayed data could solve this problem, it again fails to fulfill the requirement having the labeled data.
Looking at the studies working with discrete categories they are mostly the acted datasets whilst the natural datasets employ dimensional descriptors, positive vs negative states and application dependent affective states (the states not listed in basic six emotions, e.g., depression).
Visual Based Affect Recognition: Reading emotions from face can be done in two ways: face emotion recognition and facial muscle action. Facial action units are

perceived objective descriptors and with the limited number of AUs is possible to interpret tones of emotions. This way it is possible to create a fine grained AER system.
In the studies regarding visual based AC, the datasets are mostly from human human interaction, HCI and tv broadcasts. The features are generally categorized as geometric features, location of facial salient points, appearance features. While 2-D feature based methods require profile view face (less motion), 3-D face models provide the possibility of view independent facial expression. Few studies are found as to fuse facial expressions and body and head movements. Also, context play a crucial role to label the emotion at a final phase.
Vocal based AC bring forth the linguistic and paralinguistic features. Even though prosodic features yield very valuable results, they fail at spontaneous data. They cannot deal with the unexpected. Therefore, lexical features supply fruitful exploratory information in this sense. Nonetheless, today ASR in vocal cue is not very reliable anyway.
Audiovisual Affect Recognition is not so strong at using spontaneous data. Also it confronts the problem of human labeler inefficiency. While humans label the emotion, they concentrate either the face or the voice thus it is better to find the stimuli agreeing the same emotion and then make the further study. Fusion of modalities can be at feature level, decision level or modal level. It is suggested to do at earlier phases nevertheless the vast majority of studies were done in decision level fusion.
The challenges faced in this field that it is mostly investigated basic emotions and the acted data. This needs to opt to spontaneous data and non basic emotions. Regarding the database, it is hard to elicit certain emotions from the subjects in the lab however, social psychology provided some valuable strategies regarding the issue. Plus, while it was hard to label the dataset, using the FACS could solve the problem of subjectivity. Also, coding emotions manual is a troublesome work, this can be remedied by semisupervised active learning. About the vision based AC, a robust face detector that can detect the face in bad natural environment is needed to exploit this breakthrough. Moreover, the most important part is to find temporal structures and temporal correlation among various modalities. To detect the embarrassment, gaze, head movement, body pose, face and voice all needed at the same time.
To sum up, this survey focuses on the possibility of multimodal emotion detection using natural dataset.

AUTHORS: Jeffrey F. Cohn
TITLE: Foundations of Human Computing: Facial Expression and Emotion
SOURCE: Artifical Intelligence for Human Computing
YEAR: 2006

NO OF QUOTES: 111
VOLUME: --
ISSUE: --
PAGES: 233-238
AFFILIATION: the United States
KEYWORDS: emotion; facial expression; automatic facial image analysis; human-computer interaction; temporal dynamics.
CATEGORY: Conference Paper
TYPE: Survey
OBJECTIVE: Proposing a model for facial expression recognition based on the individual differences in order to reduce the complexity of emotion
SUMMARY: The researcher outlooks emotion recognition from a perspective that humans mostly interact in their daily with the intentions, actions and also subjective feeling, in essence they rarely interact with their feelings. For this reason, the focus is on the human-human and human-machine interaction alongside emotion recognition. Besides, emotions are not subjective; they are species-related patterns. The stability of showing emotions stabilize in the early development time and by the time of adulthood they are strongly stabilized. The ways to observe emotions are context, self-report, physiological indicators and expressive behavior. In this study expressive behavior is analyzed (facial emotion recognition).Moreover, facial expression correlates with self-assessment emotions. Annotation: facial expressions might be interpreted through two methods: message judgment and sign measurement. Message judgment is basically to detect the emotion of a smiling face as happy. However, in the sign measurement the smile is examined in more detailed like smile of happiness, smile of covering stress, etc. In the literature, there can be seen a major number of studies working with 6 basic emotions which are deemed to be recognized universally and proven in terms of validity of emotions. Thus, the message judgment employed in general these emotions in the studies. However, apart from these 6 basic emotions, the reason of “smile” should be engineered better to understand the intention of the action. In the sign measurement part, there were created many methods to manually label human emotion on the face but FACS outperforms the others in terms of differentiating same anatomically distinct movements. In the FACS, action units are observed additive and non additive. Additive AUs are like in speech emotion recognition. Each AU explains the emotion to some extent. In non-additive AUs, they provide complementary information. The joint production of two AUs could mean a whole different thing than their separate interpretations. Very importantly, the reliability of two methods (message judgment and

sign measurement) depend on occurence, temporal precision, intensity and aggregate. Timing which is temporal precision is of crucial importance. Because it consists of features like response latency and turn taking. For instance, people smile more intensely in the social groups than they do alone. These audience effects must be taken into account to make the work more reliable. Nonetheless, the manual measurement of timing is laborious. Intensity can be measured via the validation of dissimilar methods such as automatic facial analysis and human labelling. Aggregate signifies additive and non additive AUs.
Furthermore, to increase the reliability, also the individual differences which are temperament, personality, socialization, cultural background must be considered.