Cross-Cultural Emotion Recognition Based on Facial Cues and its Application to Human Machine Interaction

Graduate School of Systems & Information Engineering
University of Tsukuba

November 2014

María Alejandra Quirós Ramírez

Thesis advisor: Takehisa Onisawa María Alejandra Quirós Ramírez

Cross-Cultural Emotion Recognition Based on Facial Cues and its Application to Human Machine Interaction

Abstract

Human beings possess an intrisic need for interaction and communication. People express their feelings indistinctively and even unconsciously towards other people but also towards innanimated objects. The digital era has created a gap between our emotional and communicative nature and the interaction with machines.

Although the field of affective computing has tried to close this gap by providing machines with understanding of the users' internal emotional state in order to provide a more human like interaction, currently it is missing one of the most important factors: the individual. Very few systems or models take in consideration the personal characteristics or context of the user, drifting away from the original intention of making a more ecological interaction.

In this thesis we study one of those individual factors: culture, and how it molds the expression of emotions and the interaction with machines that try to read those emotions. We have compared emotion recognition models trained with data of individuals with different cultural backgrounds to study further the question of universality or specificity of emotions. Then we have prepared a human machine interaction experiment that utilizes those emotion recognition models to understand the effect of culturally blind interactions.

The experimental results show that failing to consider the cultural background of the user could impact negatively the performance of the system and user satisfaction. It is not recommended to employ emotion recognition systems without considering who is going to use the system and what is the context of the individual and the interaction.

Contents

1 Introduction 1

1.1 Background 2

1.2 Aim of this thesis 3

1.3 Related research overview 5

1.4 Thesis overview 6

2 Emotions, expressions and culture 7

2.1 Introduction 7

2.2 Communication and human emotions 8

2.2.1 Theories of emotion 8

2.2.2 Human emotion and expressions 9

2.3 Cross-cultural aspect of human emotions 10

2.3.1 Definition of culture 10

2.3.2 Universality hypothesis 11

2.3.3 Recent findings on the cross-cultural factor of emotions 12

2.4 Affective computing and human machine interaction 13

2.4.1 The *whats* and *whys* of affective computing 13

2.4.2 The state of the art in automatic emotion recognition . . 13

2.4.3 Emotion recognition in human machine interaction . . 14

2.5 Discussion 14

3 Cross-cultural emotion recognition 16

3.1 Introduction 16

3.2 Experiment 1: Multicultural emotion recognition 18

3.2.1 Procedure 18

3.2.2 Participants 20

3.2.3 Data collection 22

3.2.4 Analysis and results 23

3.2.5 Summary 26

3.3 Experiment 2: Latin-American and Japanese emotion recognition 27

3.3.1 Procedure 28

3.3.2 Participants 28

3.3.3 Data collection 28

3.3.4 Analysis and results 29

3.4 Discussion 31

4 Providing emotional eyes to the machine **36**

4.1 Introduction 36

4.2 Methods 37

4.2.1 Eyeglass design system 37

4.2.2 Experimental design 39

4.2.3 Participants 44

4.2.4 Analysis and results 44

4.3 Discussion 49

5 Conclusion **50**

5.1 Cross-cultural context 50

5.2 Implications in human machine interaction systems 51

5.3 Future work 52

References **53**

List of Figures

1.2.1 Human Machine Interaction scenario 4

3.2.1 Examples of emotion elicitation stimulus 19
3.2.2 Experimental room set up 21
3.2.3 Experiment 1: Stillshots of expression examples 24
3.3.1 Experiment 2: Stillshots of spontaneous expressions (negative) . 32
3.3.2 Experiment 2: Stillshots of spontaneous expressions (neutral) . 33
3.3.3 Experiment 2: Stillshots of spontaneous expressions (positive) . 34

4.2.1 Eyeglass design system flow 37
4.2.2 Eyeglass design example 38
4.2.3 Best eyeglass selection screen 39
4.2.4 Eyeglass ranking screen 40
4.2.5 Facial feature tracking 42
4.2.6 Eyeglass system interaction flow 43
4.2.7 Example of a final eyeglasses set ranking 45
4.2.8 Experimental results: Scenario preference 46
4.2.9 Correct culture vs wrong culture emotion recognition 47
4.2.10Direct evaluation vs correct culture emotion recognition 48

List of Tables

3.2.1 Subjects Breakdown by Region 22
3.2.2 Recognition results for positive and negative valence and general accuracy of emotion recognition per culture 26
3.2.3 Accuracy rates per training/testing trial in the cross-cultural recognition paradigm 26
3.3.1 Intra-cultural emotion recognition precision 29
3.3.2 Cross-cultural emotion recognition precision 30
3.3.3 Cross-cultural emotion recognition precision detail (positive) . 30
3.3.4 Cross-cultural emotion recognition precision detail (negative) . 30

Para Rita y Sergio

Acknowledgments

It was the best of times. It was the worst of times.

I am deeply thankful to all the people who helped me out in this process, and for the circumstances that placed me where I am today. First of all I want to thank my advisor, Dr. Takehisa Onisawa, for his thoughtful guidance througout these years. The completion of this project would not have been possible without his support and insight since the very beginning.

I would also like to thank the members of my thesis committee: Dr. Tomonori Shirakawa, Dr. Takehito Utsuro, Dr. Yoshinari Kameda and Dr. Hiroaki Yano, for dedicating the time to reviewing and discussing my work.

Then, I would like to thank my dear colleague Dr. Senya Polikovsky, for all the years of collaboration and friendship and all his support at every step of the road. Following, I would like to thank the CCS department and Dr. Kameda por allowing me to use their facilities and equipment . My study would have not been possible without their help.

Thank you very much to my Tsukuba angels, Sandrine Fischer, Alejandra Vilaplana, Kavita Johnson, Lee-Ann Haslam, Satoshi Suzuki, Samie Carvalho and Megumi Oki; to my roaming angels, Gabriel Gonzalez Fuentes, Oscar Bolaños Carvajal, Carlos Alvarez Gonzalez and Hanzel Mata; and to my very special angel, Stephan Streuber. It has been a blessing to have them around and share my days with them. I do not think I would have been able to survive this project or even

Japan without their company and care :).

Also, I want to thank to Shogo Kimura for welcoming me to the Humanistic Systems Laboratory and helping me go around the Japanese bureaucracy and secret ways :) and to Dr. Hideo Joho for his external insight of my project and all the support in the final stages of this thesis process.

Very special thanks to the Nagakawa family for taking me in as one of their own and for all their support and encouragement since day one. Thank you for being my family here in Japan. I will never forget this.

Finally, to the most special people who have always been there for me, my parents and sisters, for their faith, love and patience.

.thx jah.

A wild longing for strong emotions and sensations seethes in me, a rage against this toneless, flat, normal and sterile life.

Herman Hesse, Steppenwolf

1

Introduction

Every individual's daily life requires a great deal of interaction and communication with other individuals. Our quality of life and even our survival strongly depend on our interpersonal skills. An individual's capacity to successfully express his or her needs in several areas may be the difference in every area of his or her life: from obtaining food and getting a partner to having fun, etc.

During communication, the interlocutor's internal emotional state is reflected along with the spoken contents, through different cues such as facial expressions, gestures, body language, voice intonation and so on. An individual's internal emotional state influences the communication and affects the person he or she communicates with. The capacity of an individual to read and consider this emotional content as part of the communication defines the degree of success in the interaction.

On the other hand, the last decades have brought a new technological era, full of gadgets and devices that we interact with constantly. Up to this day, users are expected to adapt to the machines and modify their behavior in order to match this kind of interaction. Yet, when we consider the sort of interaction and communication that humans are used to experience, the current paradigm of human machine interaction seems to be lacking the multimodal exchange of information and context inclusion required to create a rich interaction.

1.1 Background

Machines and computers are usually perceived as devices ruled by reason, logic and numbers. The concept of computers that are capable of empathy or to comprehend human emotions and their expressions seem unreal and, due to a lack of understanding, even unnecessary.

Since the decade of 1990, a new branch in Computer Science called *Affective Computing* [31] brought to the field innovative concepts about the importance of the consideration of emotions in the world of machines. The field brought the light into the key role that feelings play in the interaction between individuals. The goal of the field is to improve the interaction among human and machine by providing understanding of the user's internal emotional state.

The current human machine interaction paradigm places the machines in the center of the interaction, and expects the user to adapt to the machine in every situation. Affective computing proposes to bring the human to the center of the interaction, and provide the tools to create a more human-like interaction.

Currently, the field has advanced quickly and has spread awareness on the importance of human-like interaction in different areas of Human Machine interaction [5], yet it is still far from its final goal of placing the human as the center of the interaction.

Most of the research focused in understanding the emotional internal state of people tries to design and develop models that fit the population in an *universal* manner [39], skipping each user's individuality and context, which may account for peculiarities in the internal emotional state and its expressions. It has not been investigated so far what is the effect of different individual variables in the expression of emotion and what its their impact in the human machine interaction scenario.

1.2 Aim of this thesis

Emotions are key in human relations; they are intertwined with an individual's behavior, decision making and every sort of communication and interaction he or she engages in. The purpose of this thesis is to study the expression of emotions and its automatic recognition from a cultural perspective. So far among the majority of the related studies, the universality of emotions is taken for granted and is assumed as a fact. The cultural factor and context of the interaction in general has been dangerously overlooked in the attempts of including emotion recognition models to human machine interactions.

It is our goal to answer what is the influence of culture in the expression of emotions and what happens when we use culture blind emotion recognition systems in human machine interaction.

To find the answer to these questions, we have prepared two different types of experiments. First, culturally aware emotion recognition models are developed and tested to understand the effect of the cultural factor in the expression of emotions. After, these models are embedded in an interaction system. The original purpose of including an emotion recognition model in such system is to obtain deeper knowledge of the user's inner state to improve performance and user satisfaction [31]. Thus, this interaction system is used to explore the consequences of ignoring the cultural context in a human machine interaction scenario.

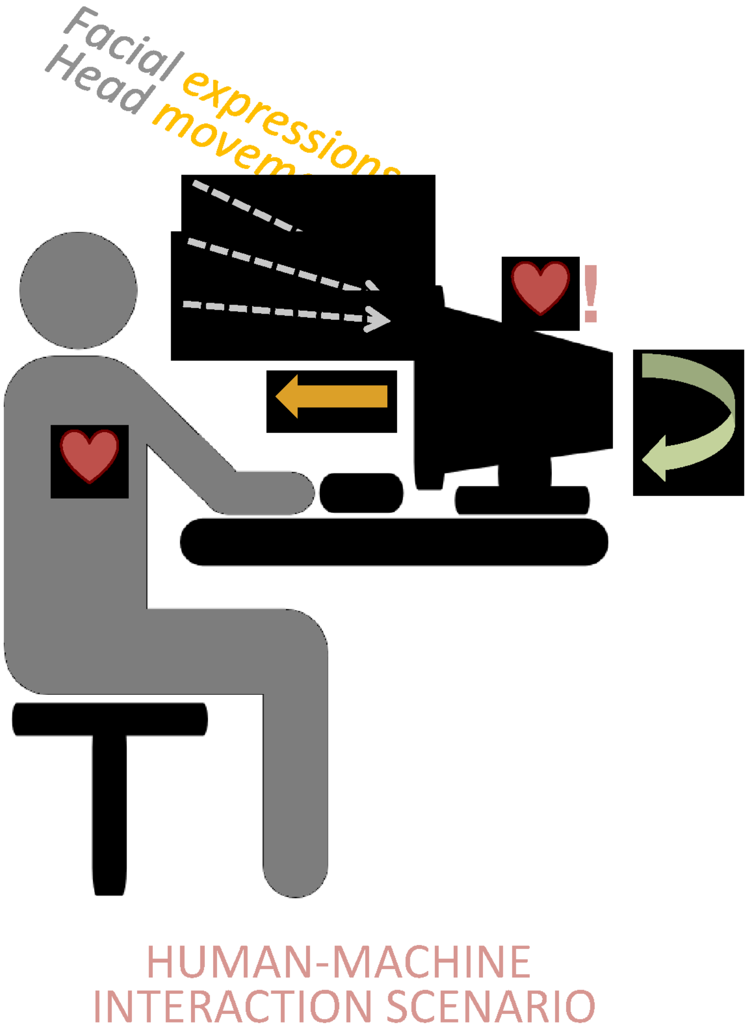


Figure 1.2.1: The figure summarizes the expected interaction paradigm between man and machine desired when considering the user's emotions. The user experiences constant internal emotional states while engaging in any sort of interaction. In this case, the intention is to obtain external expressions of the emotion, for example facial expressions and head movements, and feed them to the computer. This information will be processed using a previously trained emotion recognition model to allow the computer to understand the user's current emotional state. Using this hint, the computer take certain choices or modifications in its output to provide better support or better fit the user. Such interaction loop will repeat while the interaction among the user and the machine continues.

1.3 Related research overview

Up to this day, very few have been studied in the field of affective computing concerning the cultural context of the individuals.

Even though in the previous decades the complexity of emotion recognition systems has increased, culture related emotion recognition systems are still focused on one single cue, for example, body posture [25] or speech [24]. Very few information is found on facial expressions and gestures. Furthermore, these systems are focused on categorical classification of emotions, e.g., happy, sad, angry. Caridakis et al. [7] proposed a model to include some contextual information from the user, yet this study is also limited to the affective context only.

The data available to perform cross-cultural studies of emotion is limited as well. To our knowledge, the few examples of cultural dedicated corpuses correspond to the work of Caridakis et al. [8] on enacted multimodal gestures from three European population groups and the work of Makatchev et al. [27] with enacted interactions of native American English and Arabic speakers. The authors point out that the small number of participants (11 English speakers and 13 Arabic speakers) makes the corpus unsuitable for quantitative cross-cultural studies.

In addition, the issue on universality and specificity of emotions remains an open question. Scherer [39] describes expressions of emotion as a mix of psychobiological, sociocultural and epochal factors. In his study he presents evidence of the ongoing debate about universality versus specificity of emotions. His findings suggest that emotion encoding and decoding depend on the context of the interaction.

Another latent problem at the time of studying cross-cultural interactions is language. Haidt et al. [17] describe emotion words as *poor anchors* for cross-cultural comparisons. They point the need to look beyond the six most common emotions for this type of comparisons.

1.4 Thesis overview

The current thesis is organized as follows: Chapter 2 presents a deep introduction to the concepts of emotion and its relation to communication and culture. We explore in this chapter the field of affective computing and its role in human machine interaction. In chapter 3, the experiments carried out to study the effect of culture in automatic emotion recognition are introduced. Chapter 4 shows the effect of having culturally blind emotion recognition systems in human machine interaction settings. The final conclusions of this work are presented in Chapter 5, as well as some pointers and recommendations for future work.

We seldom realize, for example that our most private thoughts and emotions are not actually our own. For we think in terms of languages and images which we did not invent, but which were given to us by our society

Alan Watts

2

Emotions, expressions and culture

2.1 Introduction

Emotion as a word is used, lightly and constantly in daily life. There are several misconceptions when we refer to the emotions, for example, when a person is catalogued as *emotional* or when emotions are associated to a specific *gender* or *personality type*. But what is actually an emotion? Is there anybody exempt from the effect of emotions? And does every individual perceives and expresses emotions in the same manner?

A lot of the questions concerning emotions and how we experience them are still obscure and mysterious. Anyhow, it is now clear that emotions play a critical role in our activities, communication, decision making, etc. We are all emotional beings. It was just a matter of time before the necessity of affect inclusion in the top notch devices and gadgets appeared.

In this chapter some basic concepts of emotions are explained; the role they play in our daily life and communication. Also we go deeper in the question of individuality and culture as a variable that modifies our emotion expression and perception, and what is the current state of affairs of automatic emotion recognition in human machine interaction.

2.2 Communication and human emotions

Expressions of emotions are embedded in human-to-human communication [28]. Emotion is interpreted by the communicating parties using different hints such as facial expressions, body language and vocal pitch. The understanding of the exchanged emotional information is crucial in a successful communication [29]. For decades now, emotions, their expression and understanding have been studied in several fields and the evidence shows the important role they play in our daily lives [15]. Based on the available research, it seems to be impossible to separate emotions and their expressions from human interaction.

There are several propositions on the role of emotions in communication. It has been stated that emotions shape an individual's response to their social environment [15]. Furthermore, emotions in communication seem to influence the communicating partners reciprocally [2]. For example, the interpretation of the other's emotional expressions can increase or decrease the trust between communication parties [12] [18]. Thus, it is possible to say that the interpretation of the bodily expressions modifies the information being communicated and modulates the interaction between individuals. Unconsciously, an individual will try to induce the other's internal emotional state in order to give meaning to the information obtained.

2.2.1 Theories of emotion

There is evidence of the study of emotion that sets back to the 1800's with Darwin's observations [11] that hinted for the first time the universality of emotions. In this initial theory, Darwin suggests that both humans and animals expressed the same states of mind using the same physical movements. James [23] proposes that emotions are a result of physiological reactions to some specific event or external stimulus, and then the emotional reaction will depend on the interpretation of this event. This theory proposes that an individual would understand his or her internal mental state based on the perception of his or her own physical reactions.

Currently, emotion is studied from different focuses and specialties, yet there is still no generally accepted concept or definition [38]. The definition or model used to describe emotions determines as well how emotions are represented [28].

- *Discrete Emotions*. The theory of discrete emotions assumes that there is a set of basic or core emotions. Each of these basic emotions is supposed to have a unique set of responses and experiences. The number of emotions that compose the basic set varies depending on the theory [14], [21].

- *Dimensional Emotions*. On the other hand, the theory of dimensional emotions suggests that emotional states are organized not as basic cores of emotion, but in factors like valence (how positive or negative an emotion is) and arousal (how strong or weak such emotion is) [33].

- *Appraisal Theory*. This theory argues that an emotion is an episode of synchronized and related changes in the state of an organism's response to the evaluation of an stimulus. This stimulus or event may be external or internal [37], [26].

2.2.2 Human emotion and expressions

A sly smile, a frown or a sigh can tell us a lot about the internal feeling of the people surrounding us, either if we are in direct communication with them or not. There seems to be an intricate synchronization of different features and bodily movements when an emotion is expressed [41]. The set of correlations and timings that are naturally and spontaneously presented in each expression cannot be emulated consciously. On the other hand, it has been shown that stereotypical expressions of emotions do not necessarily match the portrayed expression [20]. A person that cries when feeling deep happiness is an example of this proposition.

2.3 Cross-cultural aspect of human emotions

Universality of emotions has been debated since Darwin [11]. In his work, Darwin presented correlations between facial expressions and emotions in different subjects. Ekman's work [13] in cross cultural studies has backed up the universality of emotions through studies carried out within different ethnic groups. Russell [34] presented strong evidence to disprove Ekman's theory of universality and recent evidence of differences in emotion perception [22] questions the universality of facial expressions.

It is important to point out that most of the work that has been done in order to assess universality of emotions, has been carried out through the study of facial cues. Furthermore, even though there are several theories of emotion available [28], these studies mainly focus on the discrete categorization of emotions, e.g. happy, sad, angry.

In emotion theory, the effect of cultural context is in sending and receiving subtle emotional cues is still an open question [39]. Several psychological studies suggest that culture plays a very important role in the mutual understanding of emotions.

2.3.1 Definition of culture

Culture is a broad concept as well as emotion. There are many definitions of the meaning of culture. As a general concept, culture can be defined as a *shared set of values and norms* [19] [40]. This concept allows for different levels of granularity, being able to group people from a same area, ethnicity, nation or region. Culture influences the individual's behavior and its interpretation unconsciously [32].

2.3.2 Universality hypothesis

Universality or specificity of emotions has been debated since the times of Darwin [11]. Universality of emotions suggests that emotions can be recognized regardless of the cultural backgrounds of the sender and receiver. This means that even if two people belong to different cultures, they would each be able to understand what emotion is being transmitted by the other based on visual and auditory cues. Ekman's [13] multicultural studies lead to the idea that there are six basic universal emotions.

On the other hand, specificity of emotions suggests that emotions are expressed and interpreted differently across cultures. Russell's work presents strong evidence disproving Ekman's theory of cultural universality [34]. Recent evidence in emotion perception questions the universality of facial emotions (Jack et al. 2009).

Although the question of cultural universality or specificity of emotion has been a hot topic for several decades, today it remains without a definite answer. Most of the work done to disentangle this question focuses on facial cues and utilizes discrete categorization of emotions.

The first issue for cross-cultural emotion studies is the lack of a common corpus that can be used for analysis, modeling, training and testing. There are very few open emotion databases [16] and none of these are constructed for the purpose of cross-cultural comparisons. Developing a cross-cultural corpus poses its own challenges: from basic and important points such as ways of gathering subjects from different cultures to complex points such as designing tasks simultaneously suitable for different cultures.

As mentioned in the previous section, most of the research done in this topic is based on single cue analysis. Scherer [39] describes expressions of emotion as a mix of psychobiological, sociocultural and epochal factors. His study presents evidence on the ongoing debate regarding cultural universality and specificity of emotions. His findings suggest that emotion encoding and decoding depend on the context of the interaction and suggests multimodality to more deeply study the question of cross-cultural emotions.

Working with multiple cultures might imply working with several languages as well. This is another unresolved issue at the time of the interaction and analysis. Haid and colleagues describe emotion words as poor anchors for cross-cultural comparisons [17]. Looking beyond the six most common emotions is suggested for comparing different cultures.

The interest on the degree of universality of emotions arose actively more than half a century ago [14] and since then several hypothesis on the universality of emotions have been presented [21]. The vast majority of works done on automatic recognition of emotions assume that emotions are universal and the cultural variable of each individual is ignored. Still, current works in psychology show that there is no agreement on the universality hypothesis. Further studies on the effect of individual cultural background are required in order to settle whether emotions are really universal or specific to each culture.

2.3.3 Recent findings on the cross-cultural factor of emotions

Scherer et al. describe in their state of the art review[39] the advances on the debate of both universality and specificity of emotions. In his paper, the surveyed works show that it is not possible to assume that emotions are universal, but there seems to be a lack of psychological evidence to support either position.

In previous years, the cultural question has been revisited. Failure to reach an emotion recognition agreement from individuals of India and US was found in [6], using a cross-cultural analysis similar to Ekman's research.

Through a visual perception analysis, the group of Jack et al. found strong evidence against the hypothesis of universality of emotions [22]. Their research shows that the westerners and easterners do not represent the hypothetical six basic emotions with the same facial movements. On top of this finding, the results of their work show that intensity of the emotions varies in eye dynamics among cultures. The results depict the strong influence culture has on shaping emotional behavior.

2.4 Affective computing and human machine interaction

2.4.1 The *whats* and *whys* of affective computing

One way to bridge the breach between human and computers is to provide tools that understand the internal mental state of the users [31]. Understanding the emotional state of a user allows the machine to modify its responses accordingly. This new paradigm introduces the possibility of changing the current character of interaction in which the user is typically expected to adapt to the computer instead of the opposite, ideal way. Ever since, the research on emotion recognition has advanced and evolved over time and yet, due to the complexity of the task, it remains an ongoing challenge [5].

The field of emotion recognition has been advancing quickly, yet the inclusion of the cultural aspect is still missing from currently available emotion recognition systems [16].

2.4.2 The state of the art in automatic emotion recognition

Due to the growth and advances in the field of *Affective Computing*, work on emotion databases has also evolved. Still, most of the existing databases are focused on the emotional content, neglecting technical aspects such as quality of the data and management of huge amounts of information. An important point that has not been explored deeply so far is the synchronization between multiple sensors for recording multimodal interaction, to find the relation between signals in time.

A prototypical emotion recognition system is developed by training a system with several subjects' emotional reactions. There are single or multiple cue emotion recognition systems. A single cue emotion recognition system is trained focusing in one of the physiological hints. For example, a single cue emotion recognition system may be based on facial expressions only. A multiple cue emotion recognition system mixes several hints, such as facial expression and voice. Besides audiovisual hints, other physiological signals such as brain waves or skin conductivity can also be used to train the emotion recognition system.

Although the multimodal (multiple cue) emotion recognition systems and the use of different emotion theories (e.g. categorical and dimensional [28]) continue to be popular among researchers, very few efforts have been made to include contextual information such as cultural context in interaction systems.

Overall, emotion recognition models do not consider the individual's context in any level. The recognition is usually performed in a straight forward manner, ignoring personal characteristics of the individuals. In daily life, however, several individual cues are a key in our understanding of another individual's expressions.

2.4.3 Emotion recognition in human machine interaction

The field of affective computing has advanced a lot in the previous decades [16]. Major efforts have been carried out in each of the steps required to build an emotion recognition system: data collection, modelling, analysis and interpretation. Multimodality and continuous affect are characteristics of the most recent systems. Even though researchers have explored different theories of emotion, for example, categorical and dimensional [28], most of the available systems assume universality of emotions. Little attention is put in the design of systems that consider cultural context, thus aiming to model a system that can decode emotions of any individual without considering his or her cultural background.

Several examples of the inclusion of emotions in human computer interaction have proven the importance of the consideration of internal emotional state [3], [1], [4].

2.5 Discussion

Even though the nature and mechanisms of emotion are still an open question, there are now several cues and hints that validate their importance in our behavior. Individuals constantly communicate their emotions as primary or secondary message during communication and interaction whether this takes place with other human beings or not. Given the main role emotions play in our interaction, it seems to be vital to include and consider them in the interactions that occur between human and machine.

The notion of affective computing is now widespread among several areas but there are still important key points missing in the development of the field, one of them being the individual factor. The theory of universality that has become pervasive in the general knowledge and even science has brought along several myths that are taken as facts, inducing the research of emotions in human machine interaction to skip the validation of the universality hypothesis from a point of view of emotion recognition. Then, there is a lack of understanding of the effect of the users' cultural background when considering the users' emotions during the interaction with the machine.

But the eyes are blind. One must look with the heart . . .

Antoine de Saint-Exupery, The Little Prince

3 Cross-cultural emotion recognition

3.1 Introduction

Communication is the basis of our daily interaction with other people and a complex process to transmit personal ideas to another individual. This process becomes even more challenging when the cultural background is different among interacting people. Emotions are basic components of the communication process as well. Emotional messages through non-verbal behavior support and modify our communication [2].

Automatic understanding and assessment of emotions could bring strong benefits to a wide variety of areas [9]. Since the beginnings of Affective Computing [31], the field has advanced quickly: with initial attempts of recognition of emotion from face-only pictures evolving to the current complex signal arrangements and multicue recognition systems.

Yet, not much focus has been given to cross-cultural recognition of emotions. It is still an open question whether or not emotions are universal [39]. Universality of emotions can represent a problem in trying to build a single detector that fits any individual despite cultural background.

In this chapter we introduce two experiments carried out to collect emotion data suitable to train models that are able to consider the individual's cultural background. In the first experiment, individuals from around the world were recruited and finally separated in three cultural groups: Europe, America and Asia. In the second experiment, in order to provide a more specific corpus, individuals from Latin America and Japan (as representatives of the Western and Eastern cultures) were selected to take part in the data collection.

Five general requirements are considered for the cross-cultural emotion recognition model construction:

- a) Cross-cultural corpus: In order to perform emotion recognition analysis considering cultural differences, it is necessary to design an emotion recognition experiment using an emotional corpus built with a focus on culture. Due to the lack of databases with the features required for this study, it is necessary to construct a corpus with emotional expressions and interactions to use as training and testing material for the experiment. This corpus needs to include people from different cultural backgrounds, and the interactions need to be natural. Conventional studies work with posed or acted interactions. Previous research have shown that when a person acts or poses an emotion, the result tends to differ from natural emotions in at least two points: the timing and synchronization between features and motions tend to be wrong, and the expressions exaggerated and, based on stereotypes of how the posed expression should look (Wilting et al. 2006, Hoque and Picard 2011). For our current purpose, we want to avoid these issues. Thus, we prepared an emotional corpus with subjects from different nationalities, interacting in situations that elicit emotional reactions in the participants. The expressions that appeared during such emotional interactions were recorded.

- b) Multimodality: The model requires analysis of several cues to further study the effect of culture on different audiovisual expressions. Facial expressions, head motions and body movements were considered as three different types of cues.

- c) Theories of emotion: Even though most emotion research revolves around the six basic emotions proposed by Ekman (1994), this study utilizes dimensional categorization of emotions (Russell 1980). Two dimensions are considered: valence, which means how positive or negative an emotion is; and arousal, which represents the intensity of this emotion.

- d) Language: The assessment of emotions forces subjects to assign linguistic symbols to their feelings. This is a point of bias in a cross-cultural context. The use of dimensional categorization of emotions diminishes the effect of linguistics. Besides assessment, stimuli that require deep understanding of language could bias the interaction as well. To avoid such bias, pictures are used as stimuli for the experimental interactions to record the emotion corpus.

- e) Cultural Comparisons: The final goal of the corpus construction and experiment is to compare the emotional expressions among different cultures. For this purpose, a model for each culture is prepared and then tested with data from the different cultures.

3.2 Experiment 1: Multicultural emotion recognition

3.2.1 Procedure

The recording devices used were two high speed cameras to capture facial and head information and two high definition cameras, one for the head and one for the body.

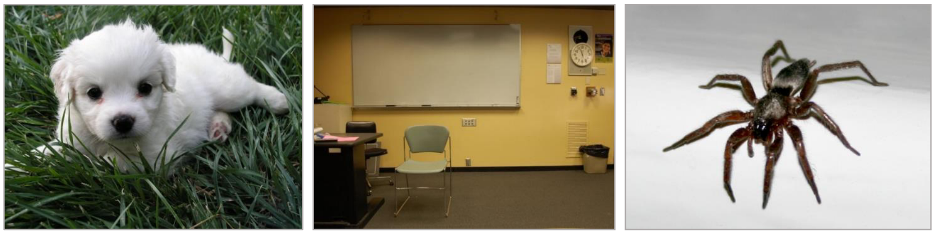


Figure 3.2.1: Images were presented to the participants during the data collection experiment to elicit some emotions in them. The images were obtained from the GAPED database. The images in the database were proved good emotion elicitators. In the figure on the left, there is an example of a positive image, in the middle a neutral one, and on the left a negative example. The images in the figure were obtained from the free stock image of *deviantart* as an illustrative example of the stimuli, since its is forbidden to disclose the original images of the database

A special room with no windows was used in order to control the experiment's illumination settings. Three different sources of lightning were used, two from the sides of the recording array and one on top of it. All the sensors were synchronized together in order to be able to retrieve the correct samples from each of the devices accurately in time to analyse multimodal cues.

Pictures were selected as a non-linguistic stimulus to obtain spontaneous emotional displays from the participants. Images from the GAPED database [10] were selected to elicit emotions from the participants. This database provides a value for each picture which corresponds to the emotional valence. This value was used to evaluate the picture as positive or negative, which represents the picture's *emotional ground truth*. The contents of the pictures range from pleasant images (for example cute animals or babies) to unpleasant images (like spiders and gross situations). Examples of the images can be seen in figure 3.2.1. It is important to remark that the information of the pictures is not used further in the study. The only purpose of the pictures is to create some feeling or reaction in the participants. After the participation of the subjects is recorded, there is no relation between the contents of the pictures and the emotional models we construct.

The emotional corpus is essential to developing correct emotion recognition models. The ecology of the corpus is important to reflect real behavior, not only for the analysis of cultural context, but also for future use in real life scenarios.

During the experiment, each participant was invited to enter the experimental room and sit in a chair placed one meter away from the monitor where the images were displayed. Participants were then instructed to watch the pictures that were automatically displayed on the screen. The experimental room design can be observed in figure 3.2.2

After observing a picture, the participant was asked to assess his or her own emotional state using a five point scale (from negative to positive, zero being neutral). Two high definition cameras synchronized to each other recorded the whole interaction. The first camera was focused on the face of the participant and the second on the full body. Each session was recorded continuously.

The pictures were presented in a random order for 5 seconds each, with a grey screen displayed for 3 seconds between pictures to let the participant rest. In total each participant observes 20 pictures: 8 positive, 8 negative and 4 neutral. To avoid language artifacts the study was based in dimensional description of emotions, using emotional valence for the experiments.

3.2.2 Participants

Thirty six naive people from different countries participated voluntarily in the experiment. Their ages range from 21 to 35, 14 of the participants were female and 19 male. A region breakdown of the participants can be observed in Table 3.2.1. Each of them had different educational backgrounds from undergraduate to postdoctoral fellows, from the University of Tsukuba and nearby research centers. English proficiency ranged from intermediate to native.

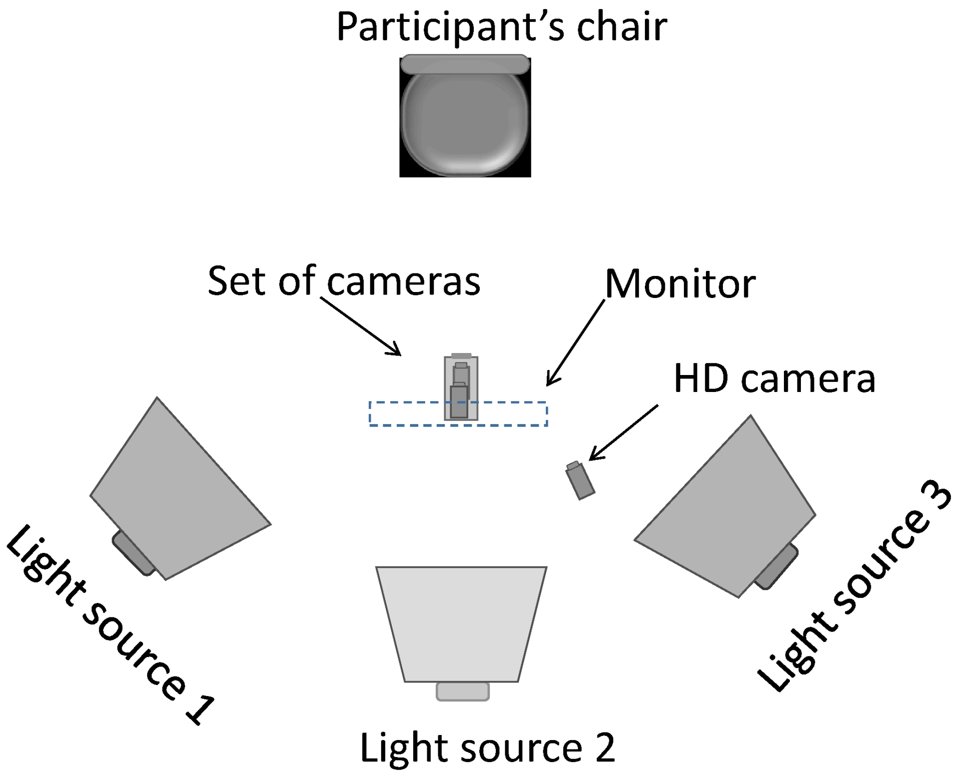


Figure 3.2.2: The data collection experiment was carried out in a close space. A chair was placed for the participant to sit, right in front of the arrange of cameras and the screen. A high speed camera was used to record the facial expressions and head moements and a high definition camera was employed to record the full body expressivity. In order to provide a better image recording quality, three sources of light (one in over the cameras and two to the sides of the participant) were used

Table 3.2.1: Subjects Breakdown by Region

Region	**Sub-Region**	#	**Total**
Africa	North	3	4
	Central	1	
America	Caribbean	2	8
	Central	3	
	South	3	
Asia	East	5	15
	Central	4	
	West	1	
	South-East	3	
	South	2	
Europe	West	2	7
	South	3	
	East	2	
Oceania	Australia	1	2
	Melanesia	1	

3.2.3 Data collection

Feature Annotation

Twenty-nine features were labeled for face, head motions and body movements. The features were chosen considering the frequency of movement among all participants. A feature is considered significant if it is observed more than 5 times in at least 2 independent participants from any cultural group.

Facial Features: Inner eyebrows up, outer eyebrow raiser, eyebrow lowerer, frown, eyelid tightener, eyelids towards each other, multiple blinks, smile, laugh, abnormal breathing, nose wrinkle, jaw drop, lip pressor, lip suck, lip corner puller, lip corner depressor, jaw sideways, swallow, chin raiser.

Head Features: move head, move head away, nod, say no, tilt head

Body Features: move finger up and down, move hands, touch or scratch with the hand, press hand, move leg.

Emotional Annotation

Emotional annotation refers to the emotion label assigned to an observed interaction. This label is considered the "real" emotion that the participant in the video segments is feeling. Labels are necessary to train a model and perform associations between interactions and elicited feelings.

There are several techniques to assign emotion labels to the segments. We chose to assign the participant's self-report of emotion as emotion label. Self-report of emotions is considered valid in the cases when subjects report "currently experienced" emotions[28]. For this experiment the participant's emotion was reported immediately after he or she observed the image, thus this labeling technique is appropriate for our investigation.

Since our interest in this study is to analyze the expression of emotions in different cultures, not understanding of emotions, we do not include the emotional judgment of third parties [39].

3.2.4 Analysis and results

The collected data was post processed in segments. Each segment consists in the time lapse between the instant when the stimulus is presented in the screen for the participant to observe until it is removed from the screen 5 seconds later. Examples of still shots of the videos recorded during the interaction can be observed in figure 3.2.3.

In order to study the effect of culture in the expressivity of emotion, participants were separated in three groups: America, Asia and Europe. Eight participants' data was feature-labeled, for a total of 160 segments. This 160 segments are represented by 67 negative pictures (47 rated as -2 and 20 rated as -1), 24 neutral pictures and 69 positive pictures (28 rated as 1 and 41 rated as 2).

There are several common classifiers used for emotion recognition tasks [16]. We chose Support Vector Machines (SVM) to train each model. An implementation of SVM from SVM-KM Toolbox [6] with Gaussian kernel was employed for

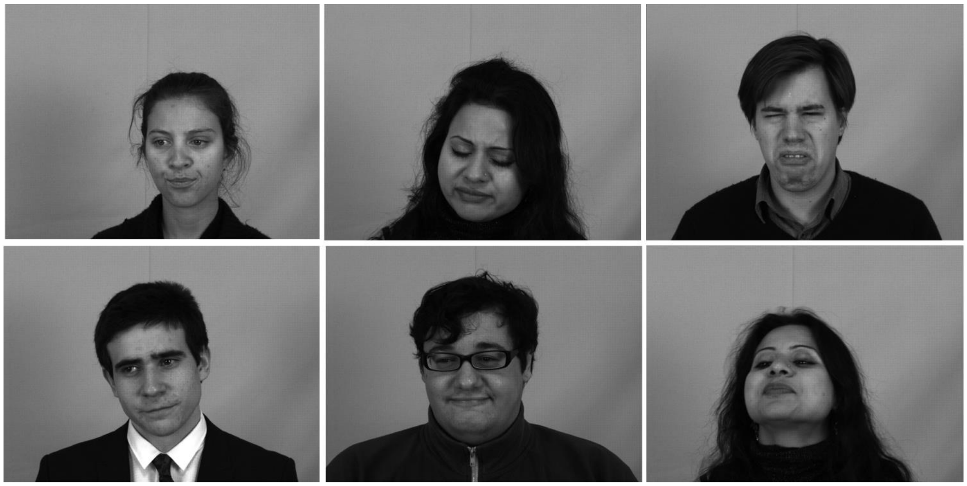


Figure 3.2.3: Participants from different cultural backgrounds joined the data collection experiment. In this figure, stillshots of spontaneous expressions shown by the participants during the experiment are presented. The top row shows participants that observed negative images. The bottom row shows participants that observed positive images. Participants nationalities: *(top row from left to right)* Costa Rica, India, *(bottom row)* Hungary, France, Brazil, India (Picture published with the permission of the participants)

training and testing.

The annotated data was coupled in vectors to train the emotion models, where the observed features are coupled to the reported emotion. In this case, we chose the emotion valence labels positive (+1, +2) and negative (-1, -2). Each training vector has the following shape:

$Eij = (fij1, fij2, \ldots, fij19, hij1, hij2, \ldots, hij5, b\ ij1, b\ ij2, \ldots, b\ ij5)$

where *i* represents the participant's ID, *j* the number of picture the participant observed, *E* is the reported emotion, *fijk* refers to each labeled facial feature (k = 1,2,...,19), *hijl* indicates the head motions (l = 1,2,...,5), *bijm* represents the body movements (m = 1,2,...,5). Vectors were chosen for training and testing the models based on participant *i*'s culture.

A leave one out cross-validation (LOOCV) procedure was selected in order to use all the vectors for training and testing each model, using each vector as an independent test exactly once. LOOCV consists of training a model with n-1 vectors and testing it with the remaining one, where n represents the total amount of vectors. The training is performed n times, testing a different vector each time. LOOCV has been chosen instead of data partitioning to avoid biasing the model towards specific participants to avoid individual expression bias. Two training strategies were employed: intra-cultural and cross-cultural.

Intra-Cultural Emotion Recognition refers to emotion recognition inside a single culture. That is, the model is trained and tested within the same culture. An Intra-Cultural Emotion Recognition experiment was performed for each of the three cultures in this study. It is necessary to examine the recognition results within a culture before proceeding to analyze cross-cultural scenarios.

Table 3.2.2 presents a summary of the recognition rates and recognition accuracy per culture in the intra-cultural emotion recognition paradigm. While participants from American and Asian cultures achieved a reasonable accuracy rate, participants from European cultures achieved a very low accuracy rate. In all three cultures it was easier for the model to recognize positive expressions of emotion than negative ones. Participant from Asian cultures achieved the best recognition

Table 3.2.2: Recognition results for positive and negative valence and general accuracy of emotion recognition per culture

Culture	Positive	Negative	Accuracy
American	0.64	0.59	0.62
Asian	0.73	0.63	0.68
European	0.46	0.40	0.43

Table 3.2.3: Accuracy rates per training/testing trial in the cross-cultural recognition paradigm

	Trained Culture		
Tested Culture	American	Asian	European
American		0.46	0.48
Asian	0.52		0.64
European	0.58	0.63	

accuracy among the three culture groups.

As for the Cross-Cultural Emotion Recognition, the models trained with data from one culture are then tested using data from a different culture. For example, an emotion model trained with American data is tested with Asian data. That is, the model is trained and tested within the same culture. This type of scenario represents the recognition attempt of individuals from one culture over the expressions of people from a different cultural background.

Table 3.2.3 presents a summary of the recognition rates and recognition accuracy per culture in the cross-cultural emotion recognition paradigm. A decrease in the recognition rates can be observed from the intra cultural rates except in the Asia-Europe scenario.

3.2.5 Summary

Emotion recognition of positive and negative emotions was obtained from our cross-cultural dataset. This finding suggest that it is possible to find agreement points between the expression of dimensional emotions between cultures.

The results indicate that the model is able to recognize the emotions for the American and Asian cultures with good accuracy. In the case of the European model, the recognition accuracy is low. Further analyzing this issue, variability inside the different cultures that are represented in our corpus by the label ”Europe”' was recognized as a possible reason for the low recognition accuracy of the European model. The corpus is based on subjects from Spain, France and Hungary. Although the countries belong to the same continent, their cultural backgrounds are quite different; behavioral expressivity seems to be different as well. From this issue, we can understand that it is necessary to define a more sensitive cultural filter. Continental grouping seems to be too broad to reflect the nuances of the different cultural populations.

Overall, the cross-cultural emotion recognition model had lower accuracy results compared with the intra-cultural models. Such decrease in the result indicates that a model trained to understand emotional expressions from a specific culture fails to recognize with the same accuracy emotional expressions from subjects of a different culture. This finding suggests cultural specificity of expression of emotions. Within a universal context, a model trained with subjects from a single culture should not suffer a recognition rate decrease when new subjects are tested, despite their cultural background.

It is possible to understand that expressions of participants in the European corpus had low similarity to those participants in the American corpus. On the other hand, there seems to be closer emotion expressivity between the European and Asian. Nevertheless, it is still necessary to refine both corpuses before reaching a conclusion regarding the relative similarity amongst the three cultural groups.

3.3 Experiment 2: Latin-American and Japanese emotion recognition

According to the results of experiment 1, we decided to make a more specific corpus in order to study further the cultural aspect of emotion expressions. One western culture (Latin-America) and one oriental culture (Japanese) were chosen for the comparisons. A similar experiment to the one described in the previous section was carried out.

3.3.1 Procedure

The experimental procedure for this second experiment follows the steps of the procedure in Experiment 1. The only difference lays in the number of stimuli presented to each participant. In this case 8 images of each valence group (negative, neutral, positive) were shown randomly to each participant.

3.3.2 Participants

Individuals from Japan and the Latin American countries are recruited to participate in the data collection experiment. In total 57 participants join the experiment: 30 Latin-American subjects (12 female – 18 male) and 27 Japanese subjects (10 female – 17 male) currently living in the city of Tsukuba, Japan. The average age of the participants is 29.8 years old (22 – 45 years old).

The recording sessions are carried in a period of two weeks. Each participant proceeds individually and voluntarily to the data collection session.

3.3.3 Data collection

Feature Annotation

The feature annotation on the collected data was performed automatically using the Open Frameworks' Face Tracking toolbox [36] to extract facial points and head position in three dimensions.

Emotional Annotation

Emotional annotation refers is in the same manner as the previous experiment, done by the participants' self-report of experienced emotion after observing each image.

Table 3.3.1: Intra-cultural emotion recognition precision

Culture	Positive	Negative
Latinamerican	60.9%	53.1%
Japanese	61.5%	57.7%
Mix	54.5%	45.5%

3.3.4 Analysis and results

After all the data is collected, the segmentation and database creation process are started. The first step to shape the database is to segment the recorded data in short session. Each observation and evaluation of a stimulus is considered a session in the database. The database is stored in SQLite, this makes it easy to browse through the data. Relevant non-private information about the participant is stored as well for example nationality, gender, wears glasses or not, etc.

Figures 3.3.1, 3.3.2 and 3.3.3 present still frames of expressions captured by the high speed camera focused to the participant's face. The figures portray three emotional clusters: positive, negative and neutral emotions, respectively. This classification corresponds to the emotion reported by the participants themselves and they might differ from the original emotional tag of the stimuli they saw; we consider the participant's self-report as the "real" emotional valence. The participants on the top row are Japanese and the participants from the bottom row are Latin-American. Even though the participants are not aware of the emotional goal of the experiment and they complete the stimuli rating task alone, expression changes can be observed in each emotional valence block.

In a similar manner as Experiment 1, two emotion recognition scenarios were studied: intra-cultural and cross-cultural. SVM were also used in this case, with the python toolbox scikit-learn [30]. Due to the amount of data k-fold cross validation was utilized.

Intra-cultural emotion recognition

Table 3.3.1 shows the precision of recognition per culture in positive and negative valence. It is important to note how mixing both cultures in a single set does

Table 3.3.2: Cross-cultural emotion recognition precision

	Positive		Negative	
Tested \Trained	Latinam.	Japanese	Latinam.	Japanese
Latinam.	60.9%	55%	53.1%	45%
Japanse	50%	61.5%	45%	57.7%

Table 3.3.3: Cross-cultural emotion recognition precision detail (positive)

Positive	+1		+2	
Tested \Trained	Latinam.	Japanese	Latinam.	Japanese
Latinam.	67.5%	60%	59.8%	47%
Japanse	64%	74%	38%	65.9%

not produce betterment in the recognition results. Given that the mix of cultures did not produce satisfactory results, we continue working with the two cultural groups only.

Cross-cultural emotion recognition

Table 3.3.2 presents the results of the cross-cultural test. The columns represent the culture used to train the models, while the rows represent the culture tested. For example, the original results of Japanese positive valence was 61.5%, but when the same data is tested using a recognition model trained with Latin-American data, the precision rate decreases to 50%. A performance decrease is observed in all the conditions.

Then, we proceed to analyze the degree of positive or negative valence using the same procedure. Tables 3.3.3 and 3.3.4 show the results of the test. In this case, it

Table 3.3.4: Cross-cultural emotion recognition precision detail (negative)

Negative	-1		-2	
Tested \Trained	Latinam.	Japanese	Latinam.	Japanese
Latinam.	61.9%	48%	61.8%	63%
Japanse	53%	62.9%	51%	61.1%

is easier for the model to recognize milder valence of emotion (+/- 1) in comparison with stronger valence of the emotion (+/-2), suggesting more variation in the expression of higher valence expressions.

An interesting case in our test in which the precision rate is similar to the original rate, happened when the Japanese data was tested using the Latin-American model. The original recognition precision rate was 61.1% and through the Latin-American model we obtain 63%. This finding suggests that there are similar expressions between Japanese and Latin American very negative emotions but there is more variation in the way Japanese people express them in comparison with the Latin American culture.

3.4 Discussion

Two different experiments for data collection and their respective emotion recognition models. Two different training and testing scenarios were presented: a) intra-cultural scenario, in which the model is trained with data of a culture and tested within the same culture. This scenario represents the individuals from a culture that perceive and recognize emotions of people within their same culture. b) cross-cultural scenario, in this case a model is trained with data of a culture and then tested with data from a different culture. This case represents the cases when an individual tries to recognize emotions from people outside of his or her own culture.

In the first experiment three geographical cultural filters were tested American, European and Asian. After the analyis, the results show hints of culture specificity yet due to the non specific characteristics of the original data it is not possible to draw strong conclusions from the results. In order to study further the hints obtained in this experimetn, a second experiment is carried out with more specialized data. Two cultural groups are chosen to represent the West and the East: Latin American and Japan.

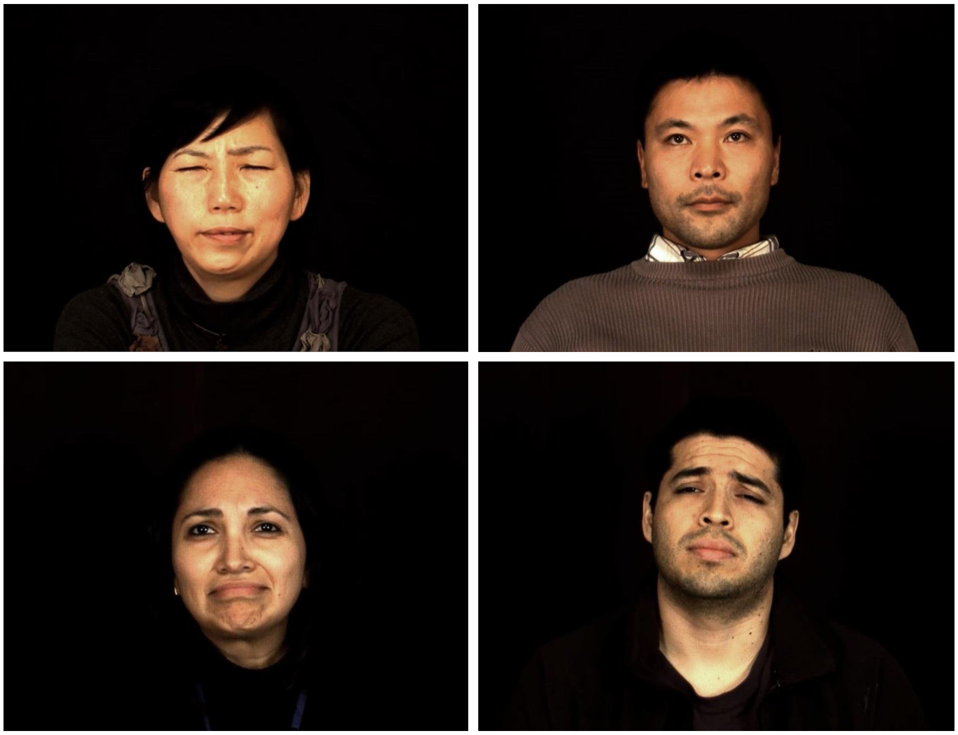


Figure 3.3.1: Japanese and Latin American participants joined the second data collection experiment. This experiment was carried in the same manner as experiment 1. The figure shows stillshots of spontaneous expressions of participants that observed **negative images**. The top row shows Japanese participants and the bottom one, Latin American.(Picture published with the permission of the participants)

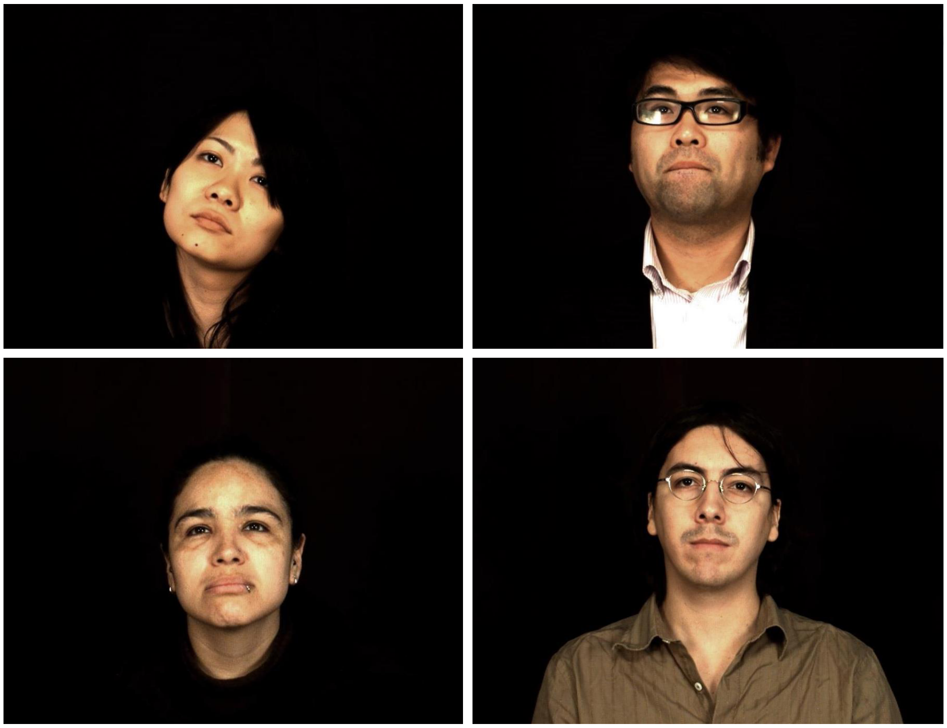


Figure 3.3.2: Japanese and Latin American participants joined the second data collection experiment. This experiment was carried in the same manner as experiment 1. The figure shows stillshots of spontaneous expressions of participants that observed **neutral images**. The top row shows Japanese participants and the bottom one, Latin American.(Picture published with the permission of the participants)

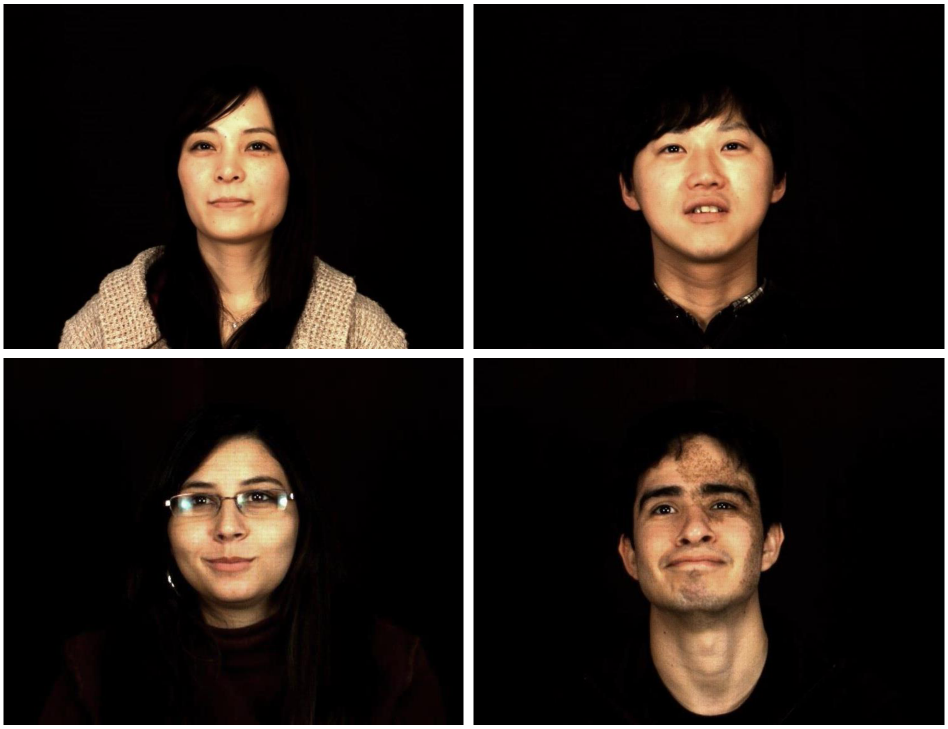


Figure 3.3.3: Japanese and Latin American participants joined the second data collection experiment. This experiment was carried in the same manner as experiment 1. The figure shows stillshots of spontaneous expressions of participants that observed **positive images**. The top row shows Japanese participants and the bottom one, Latin American.(Picture published with the permission of the participants)

The results of experiment two show a drop in the recognition rate when comparing the intra-cultural and cross-cultural experiments up to 20%. This drop suggests the correlation between the expression of positive and negative emotions with the person's cultural background, thus it is possible to say that there is a cultural specific factor when recognizing expressions of emotion.

Marvin: I've been talking to the main computer.
Arthur: And?
Marvin: It hates me.

Douglas Adams, Hitchhiker's Guide to the Galaxy

4

Providing emotional eyes to the machine

4.1 Introduction

In general, an emotion recognition model, is attached to interaction systems hoping to grasp more information about the users' internal state, in order to increase the system performance and the user's satisfaction. In this interaction experiment, our goal is to assess the consequences in performance produced by the disregard of users' cultural background when an emotion recognition model is included.

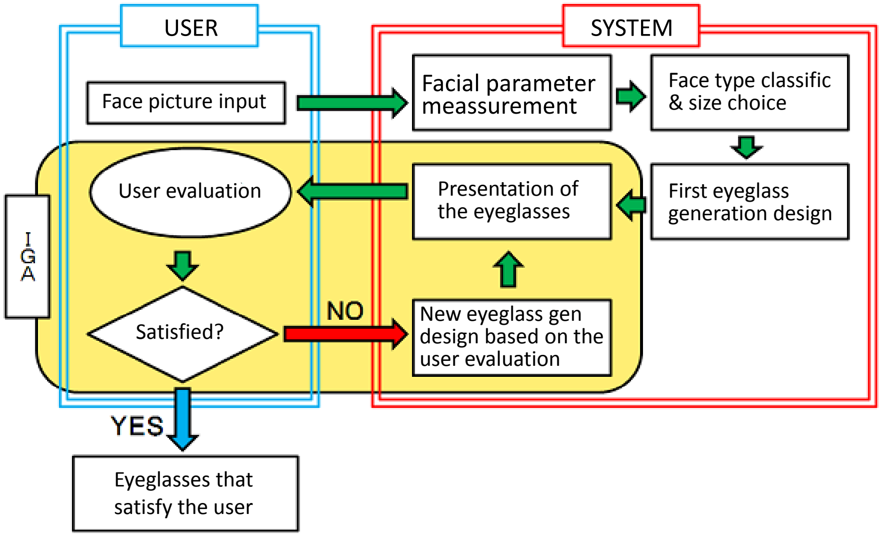

USER
SYSTEM
Face picture input
Facial parameter meassurement
Face type classific & size choice
User evaluation
Presentation of the eyeglasses
First eyeglass generation design
IGA
Satisfied?
NO
New eyeglass gen design based on the user evaluation
YES
Eyeglasses that satisfy the user

Figure 4.2.1: Eyeglass design system flow

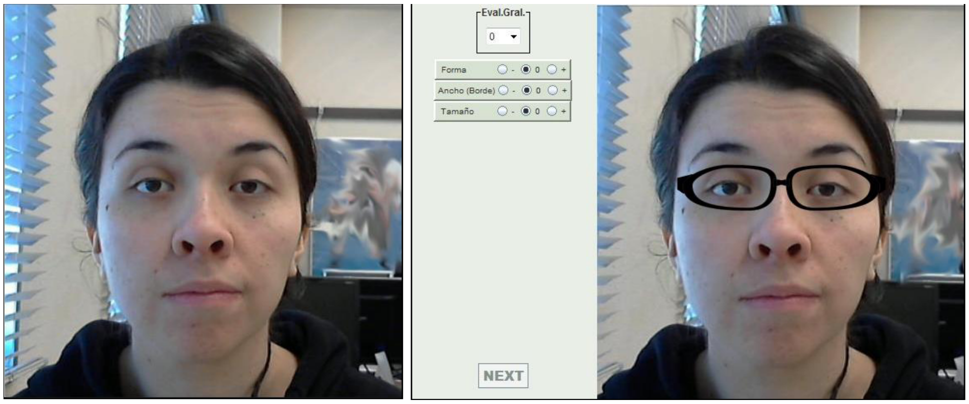
4.2 Methods

Having acknowledged the importance of culture in emotion recognition models, the next step of our study is to analyze the effect of the cultural dimension in a human machine interaction system scenario. For this experiment we chose an eyeglass design system [35].

This interactive system takes the user's subjective opinion of the produced eyeglasses to design in each iteration new eyeglasses that match better the user's preference or taste. We included our emotion recognition model in this interaction to obtain information on the user's internal state about each pair of eyeglasses presented to him or her, and to use this emotional information as an alternate input to the system to describe the user's preference to the product.

4.2.1 Eyeglass design system

The system receives as an initial input a picture of the user's face. Then, facial points are selected from the picture: three points from one eyebrow, one point per each iris and one point in the middle of this and the facial contour. Using this informa-



Eval.Gral.

0

Forma ○ - ● 0 ○ +

Ancho (Borde) ○ - ● 0 ○ +

Tamaño ○ - ● 0 ○ +

Figure 4.2.2: On the left the image taken at the beginning of the trial. On the right, a pair of eyeglasses created by the system with the evaluation screen. The user can evaluate the eyeglasses in general according to his or her satisfaction and partial evaluation of shape, frame thickness and size. The eyeglasses picture is presented to the user for 3 seconds before presenting the evaluation panel.

tion to feed an interactive genetic algorithm (IGA) based in eyeglass design rules; the system generates an initial group of glasses.

The user evaluates each pair of eyeglasses based on his or her opinion about them on a scale from -2 to 2, using an evaluation screen like the one that is presented in fig. 2 right. After, the user must evaluate three different characteristics of the glasses: shape, width of the border and size. These characteristics have to be evaluated as negative, neutral or positive. Finally, the user must choose the best glasses from each generation.

The general evaluation corresponds to the fitness evaluation for each individual (eyeglass). The partial evaluations are saved for mutation indices. Then, the individual chosen as best of the generation is chosen as elite individual, thus carried to the next generation. The user can continue getting and selecting eyeglasses until reaching satisfaction. The system flow can be observed in figure 4.2.1.

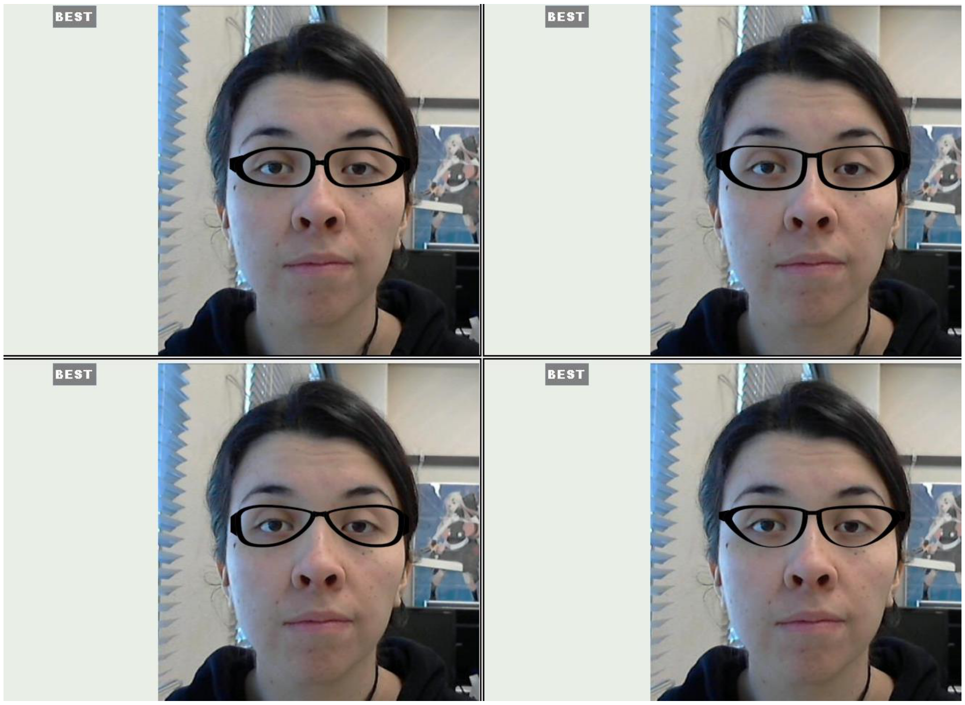


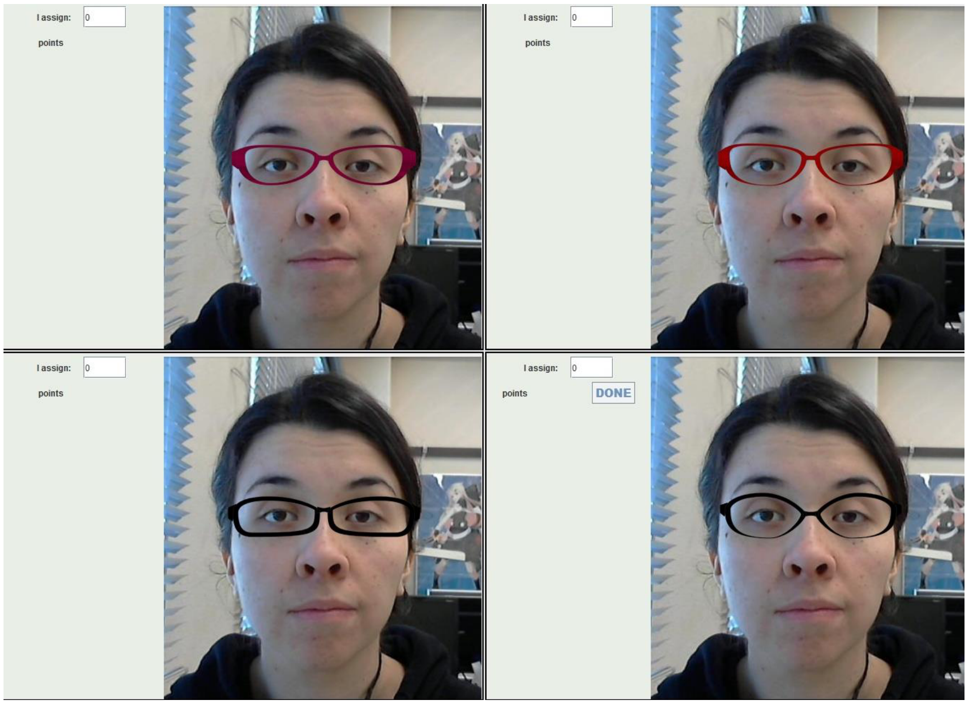
Figure 4.2.3: Best eyeglass selection screen

4.2.2 Experimental design

The eyeglass design system has been adapted for experimental purposes. From each generation, four eyeglasses are presented to the subject, one by one. First, the eyeglasses image is presented to the subject for 5 seconds, after this the evaluation panel is displayed. Figure 4.2.2 shows an example of an eyeglass designed for a subject. On the left, there is the picture that was inputted to the system at the beginning of the trial. On the right the subject's picture with eyeglasses designed by the system and the evaluation panel displayed.

After the user has evaluated the four eyeglasses, he/she must choose which one he/she considers to be the best. An example of this *best eyeglass* evaluation screen can be observed in fig. 4.2.3. The user can continue this process until reaching satisfaction, or after reaching the 10th generation.

In order to evaluate the system's performance, once the subject has chosen the



I assign: 0 points

I assign: 0 points

I assign: 0 points

I assign: 0 points

Figure 4.2.4: The final screen of the experiment presents the top best eyeglasses and two randomly generated eyeglasses. The user has 150 points to distribute among the 4 pairs of eyeglasses in order to create a preference rank.

glasses that satisfy him/her the most, the system presents a final screen with two of the best glasses generated by the system's evaluation process and two eyeglasses generated randomly. The user must rank the four eyeglasses according to preference. Fig. 4.2.4 shows an example of this screen. In the figure, the two eyeglasses in the top row correspond to the eyeglasses designed by the system and the bottom 2 eyeglasses correspond to random production. The user is naïve about the randomly produced eyeglasses.

For the human machine interaction experiment, we embedded the emotion recognitio models obtained in the Emotion Recognition trials of Chapter 3. The interaction flow for the experiment including emotion recognition is represented in figure 4.2.6. During the experimental trials, the emotion recognition models were fed with real time data from the users head and facial movements and expressions using a tracking toolbox based on the algorithm by Saragih et al [36]. Using the tracked data, feature vectors were created to be inputted in the models in order to obtain an internal emotion state estimation. Figure 4.2.5 shows on top the result of applying the tracking toolbox to an image. The toolbox provides information on 48 three dimensional facial points and head position, yet the 16 points presented in the bottom image were used to create the vectors in the same manner as explained in the previous chapter.

Three different experimental scenarios were set up: (1) Subject's correct culture emotion recognition model, (2) Subject's wrong culture emotion recognition model, (3) No emotion recognition (from now on we will refer to as "Direct Evaluation"). Each participant of the experiment evaluates each of the three scenarios. In the scenarios (1) and (2), when the direct evaluation is not used, the system considers the emotional reactions of the participants after observing each picture of the eyeglasses as "general evaluation of the eyeglasses". Thus, the system considers if the reaction was positive or negative and the degree of the emotion expression. For example, a very negative emotional expression at the time of observing the eyeglasses will represent a "-2" evaluation. In the scenario (3), as in the original system, the general opinion on the eyeglasses will be given directly through the interface by the user.

The input of the IGA fitness function will depend on the scenario. In scenario (1) and (2) the fitness value of each eyeglass is the valence output of the respective emotion recognition model. In the case of scenario (3), the fitness value corresponds to the direct evaluation of the user, through the general evaluation dropdown of the interface.

As a comparison evaluation, the best two top glasses of each scenario and a set of two random eyeglasses are printed and presented as sets to the participant. The final task of the participant is to rank the sets according to his/her preference. In this case also, the subject is unaware of the origin of each set of glasses; all the sets

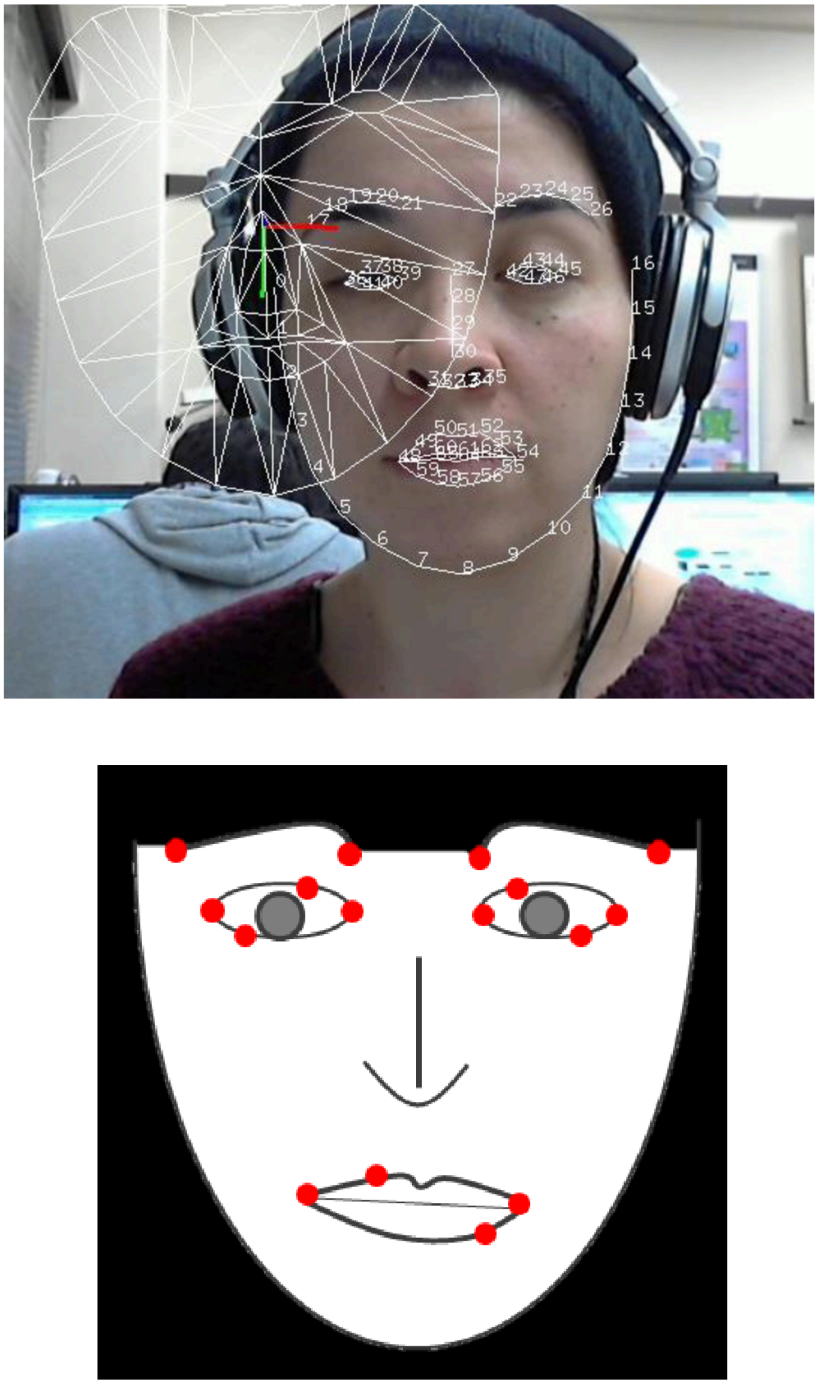


Figure 4.2.5: The user's facial expressions and movements were tracked from a 30FPS (frames per second) webcam. Facial points were tracked using the ofFaceTracker toolbox [36]. An example of the tracking can be seen on top of the image. Information of the head position is also extracted from the face points' information. Based on heuristic trials the 16 points shown at the bottom of the image were selected to create the vectors.

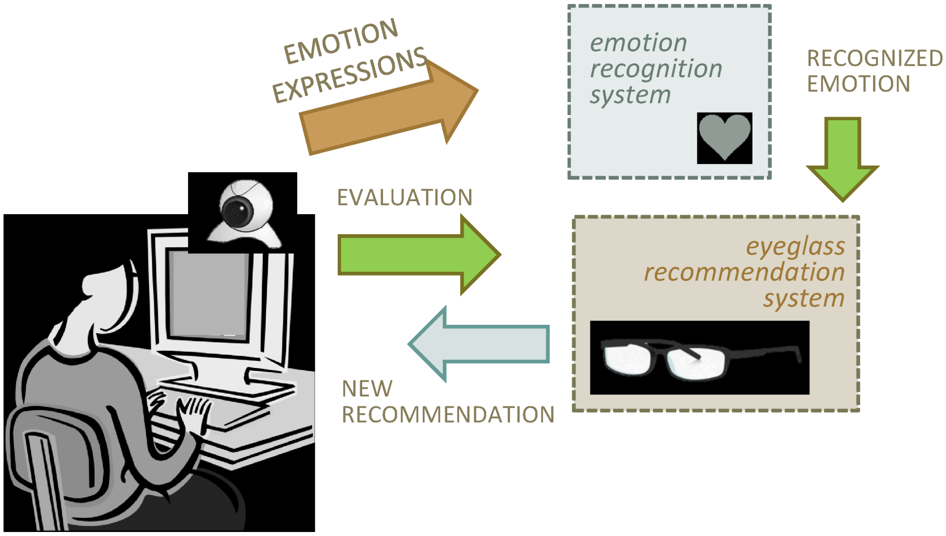


Figure 4.2.6: The figure shows the interaction flow between the user and the eyeglass system. The system recommends eyeglasses to the user, who evaluates them. The system considers the user's reactions and partial evaluation to provide new recommendations until the user finds a satisfying pair of eyeglasses.

are presented as results of the interaction system. An example of this evaluation stage can be seen in figure 4.2.7

4.2.3 Participants

Nineteen people (10 Latin-American and 9 Japanese) participated in the experiment. They all came voluntarily and signed a participation agreement after listening to the explanation of the tasks.

4.2.4 Analysis and results

The correctness of the system production was confirmed through the ranking among top eyeglasses and random eyeglasses in each scenario. This means that the system is indeed capable of designing eyeglasses that match the user's taste.

Fig. 4.2.8 shows the results of the comparison among all experimental scenarios. The eyeglasses produced by the correct emotion recognition system and the direct evaluation of the user were the most preferred. The random generated glasses were the least preferred.

Furthermore, fig. 4.2.8 presents the performance comparison of both emotion recognition models in the system. The label "Correct-Culture Emo" refers to the scenario in which the expressions of the user were analyzed using the emotion recognition model of his/her culture. For example, if the subject's culture is Japanese, it refers to the case in which the Japanese emotion recognition model was utilized. On the other hand the label "Wrong-Culture Emo" refers to the case in which the emotion recognition model does not correspond to the participant's culture. For example, if the subject's culture is Latin-American, it refers to the case in which the Japanese emotion recognition model was utilized.

The results show that the eyeglasses produced by the user's correct emotion recognition model are preferred in comparison with the eyeglasses produced by

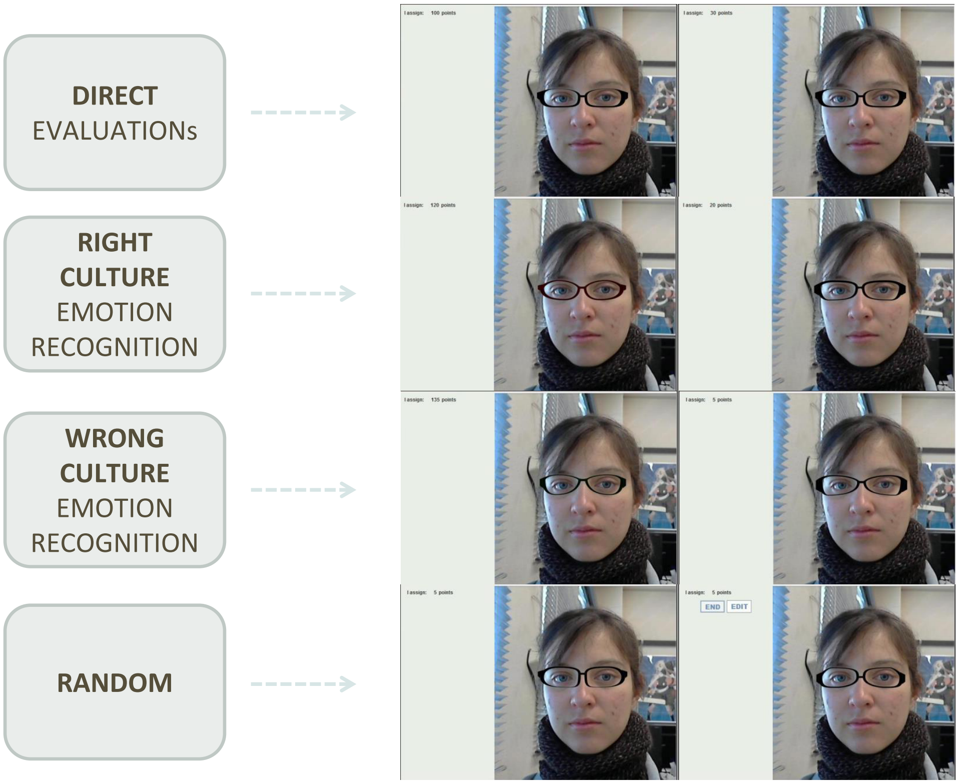


Figure 4.2.7: To evaluate the different experimental environments, the participant is asked to rank sets of top 2 ranked eyeglasses of each environment (selected through points by the participant) and a set of 2 randomly generated eyeglasses. In this example, the participant preferred the eyeglasses generated through direct evaluation, followed by the eyeglasses produced using the emotion recognition model of Latin American culture. The randomly generated eyeglasses were the least preferred. (Picture published with the permission of the participant)

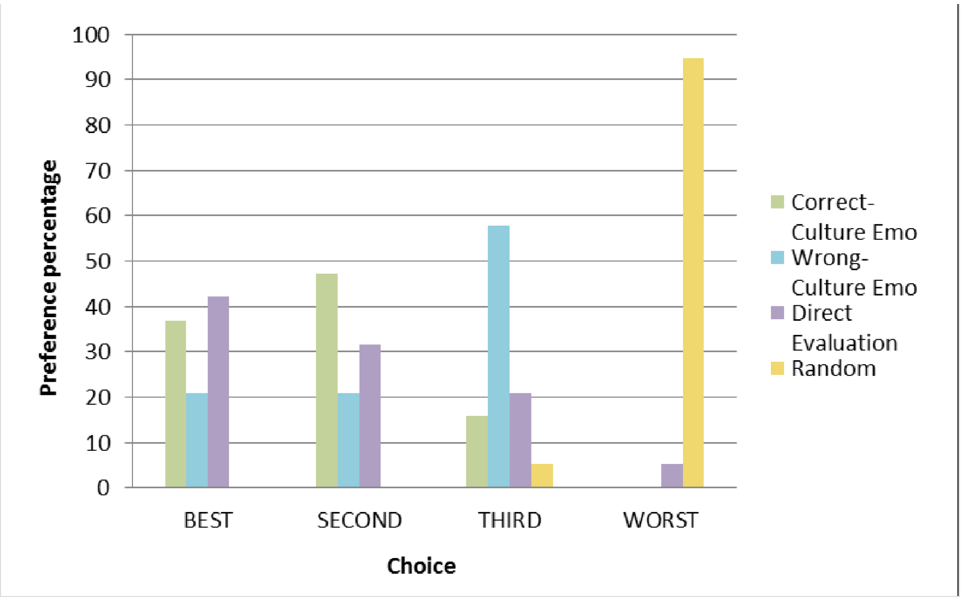


Figure 4.2.8: The graph shows the ranking results of the different experimental scenarios. It is considered a good result when the pair of eyeglasses was ranked first or second.

using the model that did not match the user's culture. This time we consider a "good" preference if the eyeglass is chosen as first or second in the final ranking performed by the user. Wilcoxon rank sum test shows that the results presented in fig. 4.2.8 have a significance of $p < 0.05$.

Finally, fig. 4.2.10 shows the performance comparison between the eyeglasses produced by the direct evaluations of the participants (through the interface general evaluation of each eyeglass) and the ones produced using the subject's reaction to each eyeglass considering the subject's culture.

The results do not show a significant difference between both groups, this finding suggest that the emotion recognition model considering the correct culture can perform as good as a direct evaluation from the users of the system. Within the three different scenarios of the experimental set up, the one that performed the worst was the scenario in which the emotion recognition system failed to match the participant's culture. This finding shows that whilst trying to obtain a better

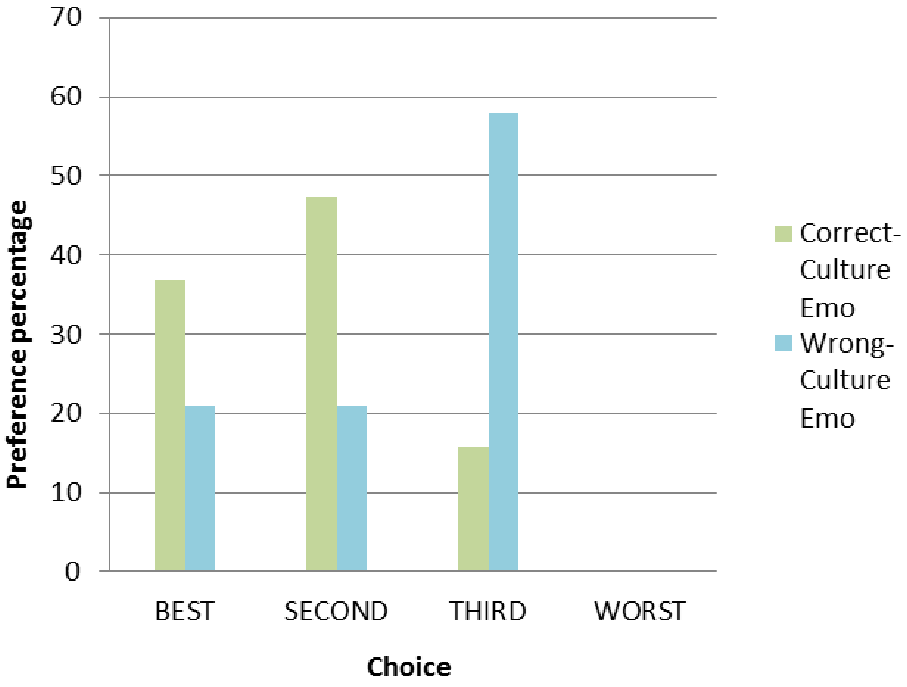


Figure 4.2.9: Comparison between the correct culture and wrong culture emotion recognition.

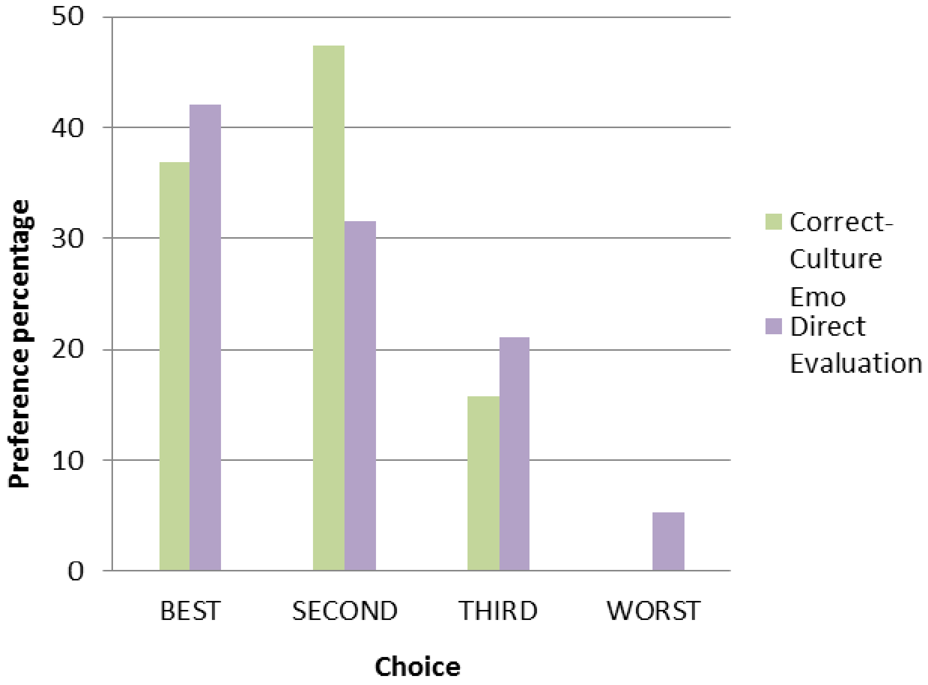


Figure 4.2.10: Comparison between direct evaluation and correct culture emotion recognition

performance by aiming to understanding the participant's emotions and expressions, the performance of the interaction system might be compromised if the cultural dimension is not considered.

4.3 Discussion

We have introduced an experiment to test the influence of culture for emotion recognition in human machine interaction scenarios.

The results show interesting findings: 1. the performance of the system and the user satisfaction dropped when the wrong culture's emotion recognition system was employed during the interaction. 2. the performance of the system and user satisfaction was as good in the cases where the user was asked directly compared with those that the system inferred from reading the user's emotional state. 3. Using the wrong culture emotion recognition system obtained worse results than using no emotion recognition system at all.

These results suggest that an emotion recognition system that assesses users' emotions considering the cultural background can be used as a reliable predictor of the user's satisfaction.

It is a far, far better thing that I do, than I have ever done;
it is a far, far better rest that I go to than I have ever known.

Charles Dickes, Tale of Two Cities

5 Conclusion

The interest of placing the user, as the center of human machine interaction has made clear the urge of understanding the internal state of people in such scenarios. Adding the user's emotional state as part of the input to a system is expected to benefit the interaction.

In this thesis, we studied the effect that culture has in the expression of emotions, and how this affects the interaction scenario between humans and machines.

5.1 Cross-cultural context

Aiming to study the expressivity differences by culture, three groups with data individuals from Europe, America and Asia were prepared to train and test emotion models. Following an intra culture and cross-culture schema the different models were trained and tested. The intra cultural results of the European group hinted an issue in the granularity of the cultural filter: having different cultures inside the group with different expressivity levels stopped the model from being able to recognize correctly the group's emotion expressions. Thus, we understood it was necessary to use a smaller filter of cultural grouping.

Using spontaneous data, several models were then trained grouping per cultures: a western culture (Latin-American) and an eastern culture (Japanese). After obtaining baseline results for recognition rates inside of the same culture, the trained models were switched and tested with data from the opposite culture. The decrease in the recognition rates shows that the cultural factor is important for emotion recognition models.

Based on this result, we consider it is important to have cultural aware emotion recognition models. Even though different cultures may have common expressiveness points – as presented, for example, in the case of Latin-American very negative expressions in comparison with those of Japanese individuals – in general, the model's performance is expected to decrease.

The obtained results suggest that there is no strong support to claim universality of emotions, since cultural background correlation has been found in the expressivity of emotions. Yet, it is important to point that there may be some common expression traits shared by different cultures.

5.2 Implications in human machine interaction systems

The experiment in Chapter 4 proves how the whole system performance suffers from using models that do not match or do not consider the user's culture. It is disadvantageous to ignore the cultural dimension. Thus, it is better to avoid including an emotional input when the cultures used to train the emotional model do not match the user's culture.

On the other hand, we confirmed with our experiment that using an emotion recognition model that matches the user's culture translates into good system performance.

There was no significant difference between the satisfaction of the user when the eyeglasses were based on the user's expressiveness or based on the direct feedback from the user. This finding suggests that an emotion recognition system that includes the cultural dimension is reliable and robust.

The results of the present study, as well as the research of several research groups from multiple fields sustain the importance of culture consideration when dealing with recognition of emotions.

Instead of trying to fit every single individual with a single global/universal emotion recognition system, we suggest that more individual dimensions and model purpose are considered in the system design stages.

The understanding of human beings in any level is multidimensional and requires a deep comprehension of the individual's background and context. We believe a dimension as important as culture should be strongly taken in consideration at the time of constructing affective models.

5.3 Future work

Even though we have provided the base that demonstrates the importance of individuality in emotion expression and human centered systems, there is still a long way to go before we can achieve an interaction that is really "human-like".

Currently, our emotion recognition system it is based only in the valence of the emotion (how positive or negative it is). It is necessary to include the intensity of the emotion and to find out when an emotion is closed to the *neutral* state.

We introduced culture as an individual variable; in the future it will be interesting to explore other variables such as age group, gender, etc.

Finally, we consider that including the interaction context at the time of the emotion recognition task and also during the interaction scenario is crucial to improve the communication between human and machine.

References

[1] A. Abbasi, T. Uno, M. Dailey, and N. Afzulpurkar. Towards knowledge-based affective interaction: situational interpretation of affect. *Affective Computing and Intelligent Interaction*, pages 452–463, 2007.

[2] A. Bartsch and S. Trewin. Emotional Communication - a Theoretical model. *Proceedings of the IGEL 2004*, 2004.

[3] R. Beale and C. Peter. The role of affect and emotion in HCI. *Affect and emotion in human-computer interaction*, pages 1–11, 2008.

[4] E. Blanchard, R. Mizoguchi, and S. Lajoie. Addressing the interplay of culture and affect in HCI: An ontological approach. *Human-Computer Interaction. ...* , pages 575–584, 2009.

[5] R. Calvo and S. D'Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *Affective Computing, IEEE Transactions on*, 1(1):18–37, 2010.

[6] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy. Svm and kernel methods matlab toolbox, 2005.

[7] G. Caridakis, K. Karpouzis, and S. Kollias. User and context adaptive neural networks for emotion recognition. *Neurocomputing*, 71(13-15):2553–2562, Aug. 2008.

[8] G. Caridakis, J. Wagner, a. Raouzaiou, F. Lingenfelser, K. Karpouzis, and E. Andre. A cross-cultural, multimodal, affective corpus for gesture expressivity analysis. *Journal on Multimodal User Interfaces*, 7(1-2):121–134, Oct. 2012.

[9] R. Cowie. Building the databases needed to understand rich, spontaneous human behaviour. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE, 2008.

[10] E. S. Dan-Glauser and K. R. Scherer. The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, 43(2):468–477, 2011.

[11] C. Darwin. *The expression of the emotions in man and animals*. John Murray, London, 1872. Freeman #1141.

[12] J. R. Dunn and M. E. Schweitzer. Feeling and believing: the influence of emotion on trust. *Journal of personality and social psychology*, 88(5):736–48, May 2005.

[13] P. Ekman. Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique., Mar. 1994.

[14] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124–9, Feb. 1971.

[15] J. Gratch, S. Marsella, and P. Petta. Modeling the cognitive antecedents and consequences of emotion. *Cognitive Systems Research*, 10(1967):1–5, 2009.

[16] H. Gunes, B. Schuller, and M. Pantic. Emotion representation, analysis and synthesis in continuous space: A survey. *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 827–834, 2011.

[17] J. Haidt. Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition &amp; Emotion*, 13(3), 1999.

[18] S. Hanibuchi and K. Ito. Feedback interface for visible congruent facial expression: Towards better face-to-face communication. In *Systems Man and Cybernetics*, pages 1004–1009, 2010.

[19] G. Hofstede. Culture and organizations. *International Studies of Management & Organization*, pages 15–41, 1980.

[20] M. E. Hoque and R. W. Picard. Acted vs . natural frustration and delight : Many people smile in natural frustration.

[21] C. E. Izard. Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm. *Perspectives on Psychological Science*, 2(3):260–280, Sept. 2007.

[22] R. E. Jack, C. Blais, C. Scheepers, P. G. Schyns, and R. Caldara. Cultural confusions show that facial expressions are not universal. *Current biology : CB*, 19(18):1543–8, Sept. 2009.

[23] W. James. Ii.—what is an emotion? *Mind*, (34):188–205, 1884.

[24] N. Kamaruddin, A. Wahab, and C. Quek. Cultural dependency analysis for understanding speech emotion. *Expert Systems with Applications*, 39(5):5115–5133, Apr. 2012.

[25] A. Kleinsmith, P. De Silva, and N. Bianchi-Berthouze. Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers*, 18(6):1371–1389, 2006.

[26] P. Kuppens. From Appraisal to Emotion. *Emotion Review*, 2(2):157–158, Mar. 2010.

[27] M. Makatchev, R. Simmons, and M. Sakr. A Cross-cultural Corpus of Annotated Verbal and Nonverbal Behaviors in Receptionist Encounters. page 7, Mar. 2012.

[28] I. B. Mauss and M. D. Robinson. Measures of emotion: A review. *Cognition & emotion*, 23(2):209–237, Feb. 2009.

[29] A. Mehrabian. Silent messages. 1971.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[31] R. W. Picard. *Affective Computing*. The MIT Press, Cambridge, Massachusetts, 1997.

[32] M. Rehm, Y. Nakano, T. Koda, and H. Winschiers-Theophilus. Culturally aware agent communication. In *Human-Computer Interaction: The Agency Perspective*, pages 411–436. Springer, 2012.

[33] J. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 1980.

[34] J. a. Russell. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin*, 115(1):102–41, Jan. 1994.

[35] T. Sakaguchi and T. Onisawa. Support system for glasses design matching user's face. *27th Fuzzy System Symposium*, 2011.

[36] J. M. Saragih, S. Lucey, and J. F. Cohn. Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision*, 91(2):200–215, 2011.

[37] K. R. Scherer. Appraisal theory. *Handbook of cognition and emotion*, pages 637–663, 1999.

[38] K. R. Scherer. What are emotions? And how can they be measured? *Social Science Information*, 44(4):695–729, Dec. 2005.

[39] K. R. Scherer, E. Clark-polner, M. Mortillaro, K. R. Scherer, E. Clark-polner, and M. Mortillaro. In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, 46(6):401–435, 2011.

[40] S. H. Schwartz and L. Sagiv. Identifying culture-specifics in the content and structure of values. *Journal of cross-cultural psychology*, 26(1):92–116, 1995.

[41] T. Vogt and E. André. Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. *Multimedia and Expo, 2005. ICME 2005. IEEE ...*, 2005.

Jah rules

論文目録

(1) Maria Alejandra Quiros-Ramirez and Takehisa Onisawa, “Considering cross-cultural context in the automatic recognition of emotions” International Journal of Machine Learning and Cybernetics.　doi: 10.1007/s13042-013-0192-2, 2013

(2) Maria Alejandra Quiros-Ramirez, Senya Polikovsly, Yoshinari Kameda, and Takehisa Onisawa, “Towards Developing Robust Multimodal Databases for Emotion Analysis” Joint 6th International Conference on Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS) 2012, pp. 589-594

(3) Maria Alejandra Quiros-Ramirez and Takehisa Onisawa, “Assessing Emotions in a Cross-Cultural Context” IEEE International Conference on Systems, Man and Cybernetics (SMC) 2012, pp. 2967-2972

(4) Maria Alejandra Quiros-Ramirez, Senya Polikovsky, Yoshinari Kameda and Takehisa Onisawa, “A Spontaneous Cross-Cultural Emotion Database: Latin-America vs. Japan” International Conference on Kansei Engineering & Emotion Research, KEER June 2014

(5) Maria Alejandra Quiros-Ramirez and Takehisa Onisawa, “Cultural dimension in emotion recognition for human machine interaction” IEEE International Conference on Systems, Man and Cybernetics (SMC), October 2014

(6) Senya Polikovsky, Maria Alejandra Quiros-Ramirez, Takehisa Onisawa, Yoshinari Kameda and Ohta Y, “A non-invasive multi-sensor capturing system for human physiological and behavioral responses analysis” Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction, 2012, pp. 63-70

(7) Senya Polikovsky, Maria Alejandra Quiros-Ramirez, Yoshinari Kameda, Judee Burgoon and Ohta Y. “ Benchmark Driven Framework for Development of Emotion Sensing Support Systems” Intelligence and Security Informatics Conference (EISIC), 2012 pp. 353-355

(8) Maria Alejandra Quiros-Ramirez, Yasuhiro Hatori and Ko Sakai　“Representation of Shape by Medial Axis in Cerebral Cortex”　映像情報メディア学会技術報告 Vol.34, No.11, 2010,　pp.1-4.