Cultural Dimension in Emotion Recognition for Human Machine Interaction

María Alejandra Quirós-Ramírez
Onisawa Lab., Graduate School of Systems and Information Engineering, University of Tsukuba
1-1-1 Tennodai, Tsukuba, 305-8573 Japan
alejandra@fhuman.esys.tsukuba.ac.jp

Takehisa Onisawa
Graduate School of Systems and Information Engineering, University of Tsukuba
1-1-1 Tennodai, Tsukuba, 305-8573Japan
onisawa@iit.tsukuba.ac.jp

Abstract—**Human emotion recognition is a multidimensional task. In this paper we study the effect of the cultural dimension in emotion recognition models and their use in human computer interaction. We prepared two experiments to analyze the consequences of disregarding culture in emotion recognition models. The results show that failing to consider the user's culture while applying emotion recognition techniques in interaction scenarios decreases the system's performance, making the emotional input meaningless and detrimental to the system and interaction.**

Keywords—cultural awareness; spontaneous expressions; decision making; affective computing

I. INTRODUCTION

The study of emotion recognition started more than two decades ago. The task of automatically recognizing emotions is complex and multidimensional, and it is still an ongoing challenge [3] [5].

Emotions and communication are inseparable [1]. The interaction between people strongly depends on the emotional understanding of the interacting parties. Based on this fact, it is necessary to include emotion understanding in human and machine interactions in order to improve mutual understanding and system efficacy.

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.

One basic and important dimension of an individual's context and identity is the cultural background. Culture molds and influences the cognitive and emotional experiences of individuals [8]. Thus, it is our interest to investigate the effect of the cultural background in emotion recognition models and its further effect in human machine interaction.

In this paper, we focus on the cultural aspect of the emotion recognition and we study further its effect, not only in emotion recognition models but also the consequences of ignoring culture in interactions between man and machine, where the user's emotions are being considered.

Using an existing emotion database [12], we build emotion recognition models using culture as a variable to study its effect on the recognition precision. Most of the available emotion recognition studies ignore the cultural factor. It is our interest to understand the consequences of disregarding this factor in an emotion recognition model.

To understand the effect of culture in human-machine interaction we embed the models in an interaction system. The original purpose of including an emotion recognition model in such system consists in obtaining deeper knowledge of the user's inner state in order to reach better performance and provide more satisfaction to the user. Thus, we will explore what happens to these goals when culture is ignored in the interaction.

The interaction system selected for the tests is an *eyeglass design system* [13] that considers the user's face type and his or her taste to design eyeglasses that are highly satisfying for the user. This system is based on the interaction between user and system. Its output is modified based on the user's feedback about eyeglass preference. We consider this system is a good candidate to test the emotion recognition models, since it would benefit from the extra knowledge of the user's internal state to provide a better eyeglass output for the user. Thus, the emotion recognition models are embedded in the eyeglass system for our human-machine interaction experiment.

In the next section we present a brief review of studies about culture and emotion in different fields. Sections III and IV describe our experiments and finally, in section V, we discuss our findings.

II. BACKGROUND

The field of affective computing originated by understanding the important role emotions play in human interaction, namely, decision making, perception, learning, etc. [10]. Ever since, the research on emotion recognition has advanced and evolved over time [5] [14] and yet, due to the complexity of the task, it remains an ongoing challenge. One of the goals of this research is to provide machines with the understanding of the emotional internal state of the user to improve the interaction.

Most of the research on emotion recognition is based on the assumption that the expressions of emotion are universal. This assumption has strong origins in the work of Paul Ekman and his findings on what he describes as a set of six basic universal expressions [4].

Scherer et al. describe in their state of the art review [14] 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 [7]. 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.

Based on these works, it is in our interest to analyze the effect of the cultural dimension in emotion recognition as well. The work in [11] shows initial hints of emotion specificity, while comparing among three geographical cultures: America, Asia and Europe.

It is our goal in the current paper to study further the effect of culture in the recognition of emotions. Then, based on the results, our second goal is to study the influence of culture in human support interaction systems.

III. CULTURALLY AWARE EMOTION RECOGNITION MODEL

The link between culture and the display or understanding of emotions has been demonstrated in several studies [2][7], including hints of its effect in the recognition of spontaneous emotions [11].

Using the data from Japanese and Latin-American subjects introduced in [12], we develop an emotion recognition system to analyze further the effect of culture consideration in automatic emotion recognition systems.

The emotional database contains spontaneous emotional expressions of subjects from Japan and Latin America and it includes both feature tagging of the face and emotional labels. Each entry in this database consists of 5 seconds of data in different formats with the reactions of a participant after observing a visual stimulus.

From the contents of the database, we have chosen to work with the *valence* dimension (how positive or negative is an emotion), using the emotional labels (self-report) included in the database. We have selected the positive and negative interactions and the degree of positive valence (+1 or +2) and negative valence (-1 or -2). Also, we have chosen to work with the facial points included in the database that were obtained from a high speed camera (30FPS) recording the subjects' reactions.

To create the emotion recognition models, we work with Support Vector Machines (SVM), RBF kernel, using the python toolbox *Scikit* [9]. After empirical trials using the selected emotional data, we have chosen to work with 1 seconds of video (30 frames) and a subset of the 68 facial points. A total of 16 points in three dimensions are selected to create the models. Fig. 1 shows the position of those 16 facial coordinates.

First, models were trained per culture to represent the recognition inside of the same culture. The precision results inside the same culture are obtained by presenting cases of the same culture to the model. Then, to understand the influence the cultural factor has in the recognition, those models are tested with data from the opposite culture.

There are two steps in our approach to recognize the participant's emotion. First, the expression is catalogued as positive or negative and then it is evaluated its degree of positivity or negativity.

A. In-culture emotion recognition

A Latin American and a Japanese emotion recognition models are trained using exclusive data from each culture. Also, for comparison purposes, all data was mixed together and used to train a third model. The data is segmented in 10 folds and through grid search procedures the optimal model for each set of data is selected and then tested through cross-validation.

Table 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 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.

B. Cross-cultural emotion recognition

The majority of the studies concerning recognition of emotions do not consider the cultural dimension. In this second part of the modeling, our interest is to understand the effect of disregarding user's culture when proceeding to recognize his or her emotion.

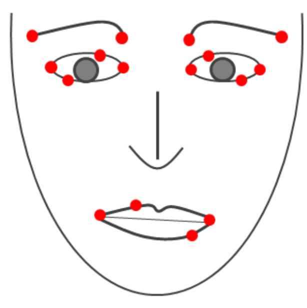

Fig. 1. Sixteen facial points chosen for the emotion recognition model. Each point corresponds to a 3D coordinate.

TABLE I. EMOTION RECOGNITION PRECISION

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

In order to do this, we classify the data from a cultural group using the model created with data from the opposite culture in the previous section, and thus, the Latin American model is tested with Japanese data and the opposite.

Table 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 and 4 show the results of the test. In this case, it 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.

The results of the cross-cultural test show that indeed there is a performance decrease when data from an unknown culture is introduced to the model. Thus, it is important to consider the cultural dimension of the individuals that are subject to the emotion recognition model.

IV. HUMAN MACHINE INTERACTION EXPERIMENT

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.

TABLE II. CROSS-CULTURE EMOTION RECOGNITION PRECISION

	Positive		Negative	
Test \ Train	Latinam.	Japanese	Latinam.	Japanese
Latinam.	60.9%	**55%**	53.1%	**45%**
Japanese	**50%**	61.5%	**45%**	57.7%

TABLE III. CROSS-CULTURE EMOTION RECOGNITION PRECISION INSIDE OF POSITIVE AND NEGATIVE VALENCE (POSITIVE)

Positive	+1		+2	
Test \ Train	Latinam.	Japanese	Latinam.	Japanese
Latinam.	67.5%	**60%**	59.8%	**47%**
Japanese	**64%**	74%	**38%**	65.9%

TABLE IV. CROSS-CULTURE EMOTION RECOGNITION PRECISION INSIDE OF POSITIVE AND NEGATIVE VALENCE (NEGATIVE)

Negative	-1		-2	
Test \ Train	Latinam.	Japanese	Latinam.	Japanese
Latinam.	61.9%	**48%**	*61.8%*	***63%***
Japanese	**53%**	62.9%	**51%**	61.1%

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 [13].

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.

A. Original eyeglass system layout and interaction

The system receives as an initial input a picture of the user's face. An example can be seen in fig. 2 left. 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 information 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.

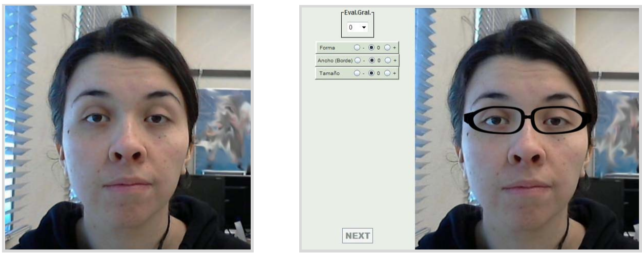

Fig. 2. Screenshots of the user eyeglass design system user interface. On the left an example of a picture inputted to the system. On the right the evaluation screen, with an example of a pair of glasses designed by the system based on the input picture and face feature points.

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.

B. Eyeglass system modifications and 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 (as shown in fig. 2, right).

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 evaluation screen can be observed in fig. 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 glasses that satisfy him/her the most, the system presents a final screen with two of the best glasses created by the system and two eyeglasses created randomly. The user must rank the four eyeglasses according to preference. Fig. 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.

In this experiment there are three different evaluation scenarios: (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.

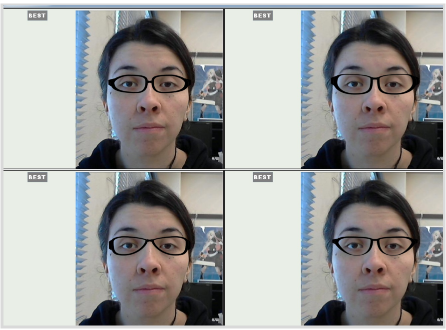

Fig. 3. *BEST* selection screen. After evaluation each of the four eyeglasses the user has to choose which he/she considers as the best pair of eyeglasses.

Fig. 4. *Ranking* selection screen. After reaching satisfaction, the user ranks four eyeglasses according to preference. In this example the glasses on the top row are products of the interaction, the bottom ones are randomly produced.

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 are presented as results of the interaction system.

C. Experimental results

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.

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. 5 shows the results of the comparison between each scenario. 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. 6 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 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. 6 have a significance of $p < 0.05$.

Finally, fig. 7 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 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.

V. Discussion

From the point of view of interaction between human and machine, adding the user's emotional state as part of the input to a system is expected to benefit the interaction. The pillars of affective computing are based on the ideal of creating machines and systems that can understand better its users to provide a richer and more human experience and to improve the system's performance in comparison with systems that cannot comprehend the user's emotional expressions.

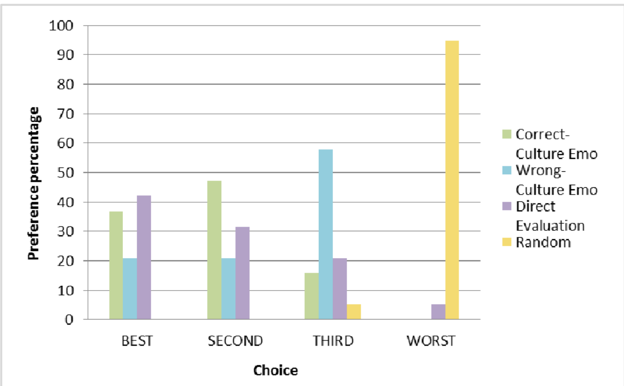

Fig. 5. General experimental results. The horizontal axis represents the ranking order performed by the participants of the 4 sets of glasses. The vertical axis represents percentually the preference of each type of scenario.

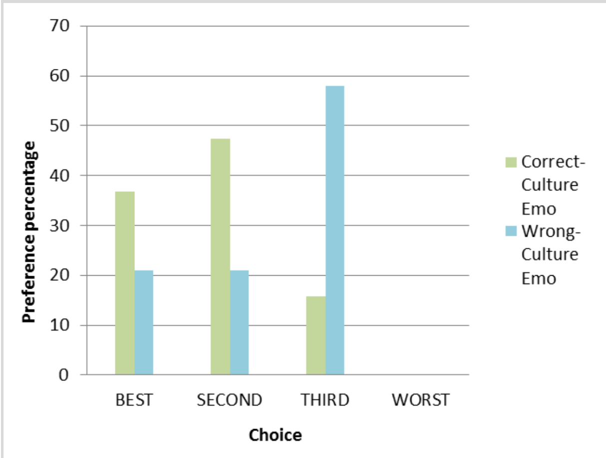

Fig. 6. Comparison of the experimental results per emotion recognition model. The eyeglasses produced using the emotional input recognized by the correct culture emotion recognition system are more preferred in comparison to the eyeglasses produced using the wrong cultural emotion recognition system.

In this paper we explored the cultural dimension of emotion recognition and human-machine interaction in order to understand what is the meaning and effects of considering culture in both paradigms.

A. Culture in emotion recognition models

Using spontaneous data, several models were 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 show 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.

B. Culture in interaction scenarions where emotion is considered

After observing the emotion recognition's rate decreasing as a result of ignoring the user's cultural background, our next step was to investigate the possible effects when using such models in interaction environments regardless of culture.

The experiment in section IV 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.

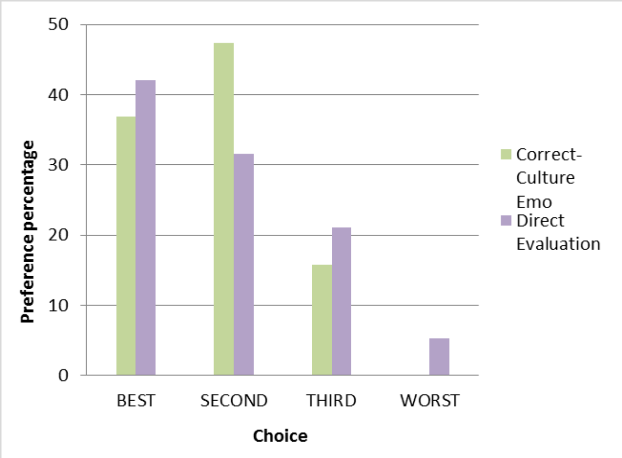

Fig. 7. Comparison between the eyeglasses produced in the direct evaluation scenario and the correct culture emotion recognition scenario. No significant difference was presented between both scenarios.

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.

C. *Implications for future work*

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.

ACKNOWLEDGMENT

The authors would like to thank Mr. Tomohiko Sakaguchi for kindly allowing us to use and adjust his eyeglass design system for our experiments.

REFERENCES

[1] P. Andersen, and LK. Guerrero. "Principles of communication and emotion in social interaction." Handbook of communication and emotion: Research, theory, applications, and contexts: 49-96. 1998

[2] J. Y. Chiao, T. Iidaka, H. L. Gordon, J. Nogawa, M. Bar, E. Aminoff, et al., "Cultural specificity in amygdala response to fear faces". Journal of Cognitive Neuroscience, 20(12), 2167-2174, 2008.

[3] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J.G. Taylor. "Emotion recognition in human-computer interaction." Signal Processing Magazine, IEEE 18(1), 32-80. 2001.

[4] P. Ekman "Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique". Psychol Bull 115:268–287, 1994

[5] H. Gunes, B. Schuller, "Categorical and dimensional affect analysis in continuous input: current trends and future directions". Imag Vis Comput 31(2):120–136, 2012.

[6] J. Haid, D. Keltner, "Culture and facial expression: openended methods find more expressions and a gradient recognition". Cogn Emot 13:225–266, 1999.

[7] R.E. Jack, O.G. Garrod, H. Yu, R. Caldara, and P. G. Schyns, "Facial expressions of emotion are not culturally universal". Proceedings of the National Academy of Sciences, 109(19), 7241-7244, 2012.

[8] H.R. Markus, and S. Kitayama. "Culture and the self: Implications for cognition, emotion, and motivation." Psychological review 98, no. 2: 224, 1991.

[9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, Grisel, O. et al., "Scikit-learn: Machine learning in Python." The Journal of Machine Learning Research, 12, 2825-2830. 2011.

[10] R. Picard, "Affective computing". MIT press, 2000.

[11] M.A. Quiros-Ramirez, and T. Onisawa, "Considering cross-cultural context in the automatic recognition of emotions." International Journal of Machine Learning and Cybernetics, 2013.

[12] M.A. Quiros-Ramirez, S. Polikovsky, Y. Kameda and T. Onisawa, "A Spontaneous Cross-Cultural Emotion Database: Latin-America vs. Japan." Kansei Engineering & Emotion Research International Conference, Sweden, 2014, *in press.*

[13] T. Sakaguchi, and T. Onisawa, "Support System for Glasses Design Matching User's Face". 27th Fuzzy System Symposium, 2011.

[14] K.R. Scherer, E. Clark-Polner, M. Mortillaro, "In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion." Int J Psychol 46:401–435, 2011.

[15] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. "A survey of affect recognition methods: Audio, visual, and spontaneous expressions." Pattern Analysis and Machine Intelligence, IEEE Transactions on 31(1), 39-58, 2009.