KEER2014, LINKÖPING | JUNE 11-13 2014
INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH

A Spontaneous Cross-Cultural Emotion Database: Latin-America vs. Japan

Maria Alejandra Quiros-Ramirez[1], Senya Polikovsky[2], Yoshinari Kameda[3], Takehisa Onisawa[4]

[1] *Graduate School of Systems and Information Engineering, U.of Tsukuba, Japan, alejandra@fhuman.esys.tsukuba.ac.jp*

[2] *Graduate School of Systems and Information Engineering, U.of Tsukuba, Japan, senya@image.esys.tsukuba.ac.jp*

[3] *Graduate School of Systems and Information Engineering, U.of Tsukuba, Japan, kameda@iit.tsukuba.ac.jp*

[4] *Graduate School of Systems and Information Engineering, U.of Tsukuba, Japan, onisawa@iit.tsukuba.ac.jp*

Abstract: In this paper, we present a new database to support the cross-cultural studies. Two cultural groups are selected: Latin America and Japan, to represent *western* and *oriental* cultures. Emotions are elicited through an experiment in which participants observe emotionally loaded stimuli and then rate their feelings in a *valence* (how positive or negative is the experienced emotion) and *arousal* (how intense is this emotion) scale. The interactions are recorded using audiovisual and thermal devices. This database features three innovative characteristics: *spontaneous emotion expressions*, *multiple synchronized sources of interaction*, *cross-cultural comparison support*. This set of characteristics is missing in the currently available emotion databases, making our database a unique open option for studying spontaneous expressiveness of emotions in a cross-cultural context.

Keywords: cultural specificity, universality, multimodal corpus, affect.

1. INTRODUCTION

Imagine you are observing a conversation between two individuals. They are talking when one of them suddenly frowns strongly; his body slowly tenses up… You can hear him saying some words in a rough tone, rushing to leave. The expression on the face of the person left behind makes you think this person did not understand why the first one had such a strong reaction and left. You seem to grasp a lot about their feelings just by observing two people in the distance, even without listening to any of the spoken words in detail.

Expression of emotions is a key component in the communication among people (Mehrabian, 1971). For decades now, emotions and their expression and understanding have been studied in

several fields and the evidence shows the important role they play in our daily lives (Gratch et al., 2009). Along with the study of emotions, the study on *development* of systems that can *automatically* recognize human emotions has increased as well in the last decade (Gunes et al., 2011). The widespread use of computers in daily life and the desire of making them easier to use, have encouraged the inclusion of human emotion recognition systems for human computer interaction (Beale & Peter, 2008).

The interest on the degree of universality of emotions arose actively more than half a century ago (Ekman & Friesen, 1971) and since then several hypothesis on the universality of emotions have been presented (Izard, 1994). 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.

Emotion databases are widely used in the studies of emotion expression. There are some databases currently available for this purpose (Gunes et al., 2011) but none of them can support the study of cross-culture variables. Our goal is to provide the scientific community with a robust database aimed to support such comparisons, thus, bringing the solution to the current lack of data to analyze the problem of universality or specificity of emotions.

Kashima (2000) defines culture as “a *relatively stable system of shared meanings, a repository of meaningful symbols, which provides structure to experience*”. The database introduced in this paper comprises two cultural groups: *Japanese* and *Latin-American*. These two cultures are selected to represent the western and oriental populations, following psychological studies carried out to compare cultural differences (Scherer et al., 2011). Both chosen cultural groups comply with the definition of culture. Our database presents spontaneous expressions of emotion: participants are not led to show a specific expression or enact any emotion. The database includes emotional labels of the emotions portrayed in each interaction as well as labels for facial features and head position in space. The database is open and distributable for the research community.

A review of our database background is presented in the following section. Section 3 explains the details of the data collection, methods and materials. In section 4 the actual database contents are explained and finally the conclusion in section 5.

2. BACKGROUND

Broadly accepted studies on the hypothesis of *universality* of emotions (Ekman & Keltner, 1997) suggest there is a group of *basic* and universal emotions that are equally expressed and recognized in all the different cultures around the globe. The initial methods used to support these studies are based usually in still photographs of subjects and comparison among them.

New break through studies in psychology (Jack et al., 2012) neurology (Chiao et al., 2008) (Adams et al., 2010) and affective computing (Quiros-Ramirez & Onisawa, 2013) (Kamaruddin et al., 2012) have presented new evidence to support the *specificity* of emotions: culture seems to play an important role on the expression and recognition of human emotions.

Even though there are a handful of open emotion databases, most of them do not consider cultural factors. The work of Caridakis et al. (2012) presents a gesture database with participants from Germany (21 subjects), Greece (11 subjects) and Italy (19 subjects) who go through a mood

immersion procedure by speaking emotionally loaded phrases aloud. The work of Makatchev et al. (2012) presents another cross-cultural corpus showing *receptionist encounters* with participants from English and Arabic speaking countries. Yet as they warn in their paper, the small amount of participants in the corpus makes it unsuitable for cross-cultural comparisons.

3. DATA COLLECTION METHODOLOGY

In this section we explain the details of execution and contents of our Latin-American/Japanese cross-cultural database.

3.1. 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.

3.2. Experimental materials

The data collection is carried in an empty dark room conditioned for the task. A group of cameras, display and microphone are set up to record the reactions of the participants during the experiment.

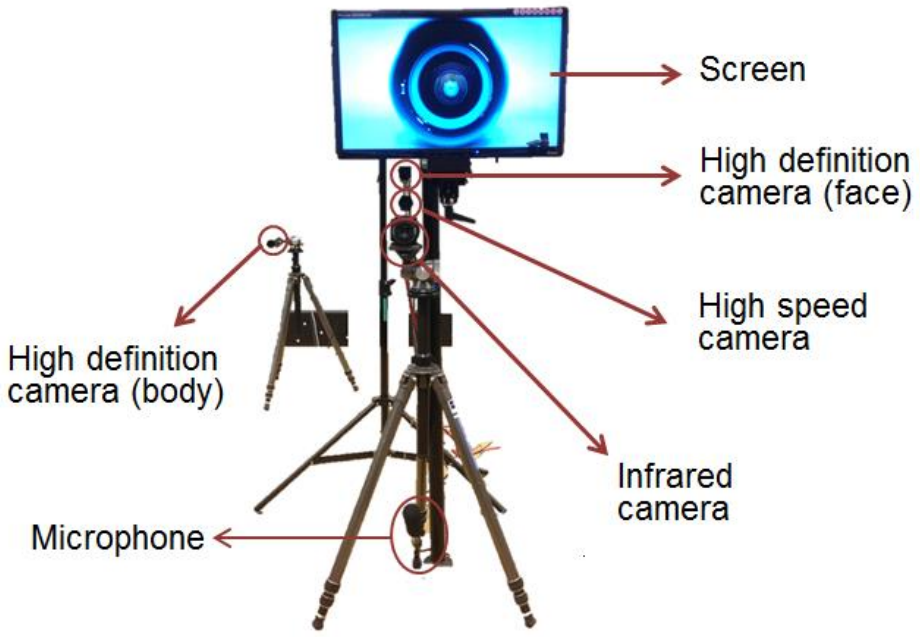

Figure 1: Set of devices utilized for the data collection

Figure 1 shows the array of devices used for the experiment. First, a screen (ProLite E2208HDS Widescreen) is placed 125cm over the floor. Below this screen is a set of cameras all focused to the participant's face: high definition camera (PointGrey Flea2CCD), high speed camera (PointGrey Grasshoper CCD) and infrared camera (FLIR SC300). Under this set of cameras there

is a directional microphone to record the subjects' utterances. In the back (left side of the picture), one more high definition camera is placed to record the participants' full body reactions. All the devices are synchronized to assure that the recorded data in the different channels can be accessed simultaneously in the same point in time for analysis (Quiros-Ramirez et al., 2012).

In front of the screen, at a distance of 130cm a chair is placed for the participant to sit and take the experiment. Three sources of light are used to improve and stabilize the room illumination. Two lights are placed aside the cameras and one above the display screen. The layout of the experimental room can be seen in figure 2.

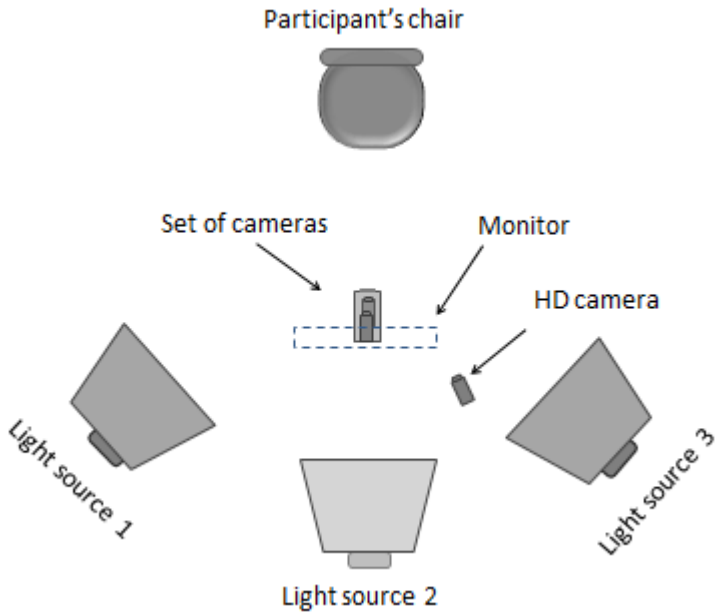

Figure 2: Layout of the experimental room

3.3. Data collection experiment

Previous research has shown the importance of natural or spontaneous emotions (Wilting et al., 2006) (Hoque & Picard, 2011). When emotions are enacted, they tend to be exaggerated or stereotyped making the resulting data not fit for emotion analysis or studies. Our goal in this database is to provide robust data for cross-cultural comparisons, therefore spontaneous emotion elicitation techniques are used to allow the participants to express freely even under laboratory conditions.

An emotion elicitation technique refers to experimental conditions that might induce the participant into an emotional state without any explicit guidance or request him or her to enact an emotion expression. The participant is not hinted about the final emotional goal.

3.3.1. Experimental Flow

The participant is guided to the experimental room and asked first to fill up two questionnaires about current mood and cultural traits. After this, the experimenter explains the experimental task to the participant. The experimenter informs to the participant that the interaction will be recorded and that the experiment can stop at any moment by request of the participant.

If the participant agrees to the conditions, he or she then signs a consent experimental form and is guided to the task area. First, the participant makes a short practice of the experiment to familiarize with the task. After answering any questions from the participant, the experimenter

leaves the room and the participant performs the task alone.

Once the experiment is finished, the experimenter returns to the room and the participant is asked to sign once again the consent form.

3.3.2. Experimental Stimuli

Emotionally loaded images (Dan-Glauser & Scherer, 2011) and videos (Hewig et al., 2005) are used as stimulus for the experiment. These are initially classified by valence content in positive, negative and neutral. Each participant is presented with 8 positive, 8 negative and 8 neutral images and one video of each category. Participant self-emotional rating could differ from the original label of the stimulus.

3.3.3. Experimental Task

The experimental task consists in rating images and videos that are presented for the participant in the screen. A total of 24 loaded images and 3 videos are then presented randomly. A grey screen is displayed between stimuli for 3 seconds to let the participant rest in between.

The participant is asked to observe each image or video (stimuli) and answer two questions: 1. his/her current emotion; 2. the intensity of this emotion. Both questions are rated in a 5 point scale (from -2 to 2). For the first question, the emotional *valence* is inquired: how positive or negative is the emotion he or she feels as a product of the image. In this question, the rating -2 represents a very negative feeling and 2 represents a very positive feeling. In the second question emotional *arousal* (intensity of the emotion he or she felt) is checked, -2 means a very weak feeling and 2 stands for a very strong feeling. The participant should answer to the questions aloud.

4. CROSS-CULTURAL EMOTION DATABASE

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. Thus, 27 sessions are obtained per participant. Using FaceTracker toolbox through *Open Frameworks* toolbox (Saragih, 2012) the participant's facial features as well as head position information are tracked and recorded. Table 1 explains the information contents of each session stored in the database.

Table 1: Contents per session in the database

Information	**Detail**
Observed stimulus	Information on the type and valence of observed stimulus
Emotional tag	Corresponds to the emotional valence and arousal self-report evaluated by the participant immediately after observing a stimulus
High definition videos	30FPS videos of the face and body of the participant
High speed video	180FPS video of the participant's face during the session
Infrared data	60FPS infrared data of the participant
Sound	Recorded utterances (includes the spoken evaluation)
Facial Feature Points Data	Frame by frame tracking of 68 facial points
Head Position	Frame by frame tracking of the head position in space

The database is stored in SQLite (Owens, 2006); 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 presents still frames of expressions captured by the high speed camera focused to the participant's face. The frames are grouped in three clusters: positive, negative and neutral emotions. 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.

Figure 3: Still frames from the high definition videos. Each block represents the emotion reported by the participant: negative, neutral and positive emotions respectively. All the participants in the top row are Japanese, and in the bottom row Latin-American.

Database characteristics and usability

The database introduced in this paper has a set of characteristic that makes it suitable for emotion analysis in different fields, such as psychology and affective computing. The following is an explanation of each characteristic and how it can support emotion expression studies.

Cross-cultural: Spontaneous emotional data from participants of two separate regions (Japanese and Latin American) is successfully recorded making this database the first open database of its kind. Up until now there is no available open and spontaneous data that can be used to analyze and compare differences and similarities between western and oriental population groups.

Spontaneous expressions of emotion: Given the results on previous studies (Wilting et al., 2006) (Hoque & Picard, 2011), the need for non-posed expressions is clear. The current database provides the community with a dataset of spontaneous expressions which is a robust tool to study the emotion phenomenon.

Multiple synchronized data channels: One important aspect of emotions is the timing and relation between different cues. The current database allows comparison between cues in time thanks to the 1ms order synchronization between data channels achieved through our data recording set-up

(Quiros-Ramirez et al., 2012). The synchronization issue is a key factor in the analysis of emotions, it is necessary to study the time correlations between different cues and movements when an emotion is being displayed in order to achieve a deeper understanding of the expressions of emotion (Scherer et al. 2011).

Open for research purposes: Our database is open and distributable online for research purposes. This database covers the current lack on open spontaneous and cross-cultural emotional data. By providing emotion and feature tags, the database is ready to be used in the analysis and training of emotion systems. Also, the structure of the database allows the extension of emotion and feature tags, allowing the users to create new labels to fit their own research needs.

We consider the construction of communication models using emotion recognition as one of the applications of the presented database because, as mentioned in the Introduction section, expression of emotions is a key role in the communication among people. Even though the current database is yet to be tested in the construction of automatic emotion recognition systems, its success is anticipated based on the results obtained previously employing the same data collection techniques for emotion recognition (Quiros-Ramirez & Onisawa, 2013).

5. CONCLUSION

The construction of an automatic emotion recognition system is a broad task with several problems in the middle. Several variables come to play while trying to understand the meaning of human behavior and expressions. Even though individuality comes to play in the expression of emotions, based on psychological studies culture seems to have a strong influence in our expression of emotions. In order to study and find the answers to this phenomenon it is necessary to have access to data fit for the problem. In this paper we introduce an open database to support the study of the cultural factor in emotions. The set of characteristics of our database is missing in the currently available emotion datasets, making our database a unique open option for studying spontaneous expressivity of emotions in a cross-cultural context.

REFERENCES

Adams Jr, R. B., Rule, N. O., Franklin Jr, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., ... & Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fMRI investigation. Journal of Cognitive Neuroscience, 22(1), 97-108.

Beale, R., Peter, C. (2008). The Role of Affect and Emotion in HCI. In Affect and Emotion in Human-Computer Interaction (pp. 1-11). Berlin: Springer.

Caridakis, G., Wagner, J., Raouzaiou, A., Lingenfelser, F., Karpouzis, K., & Andre, E. (2012) A cross-cultural, multimodal, affective corpus for gesture expressivity analysis. Journal on Multimodal User Interfaces, 1-14.

Chiao, J. Y., Iidaka, T., Gordon, H. L., Nogawa, J., Bar, M., Aminoff, E., ... & Ambady, N. (2008). Cultural specificity in amygdala response to fear faces. Journal of Cognitive Neuroscience, *20*(12), 2167-2174.

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

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. Journal of personality and social psychology, 17(2), 124.

Ekman, P., & Keltner, D. (1997). Universal facial expressions of emotion: an old controversy and new findings. Nonverbal communication: where nature meets culture, 27-46.

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

Kashima, Y. (2000), Conceptions of culture and person for psychology. Journal of Cross-cultural Psychology, 31:1, 14-32.

Gratch, J., Marsella, S., & Petta, P. (2009). Modeling the cognitive antecedents and consequences of emotion. Cognitive Systems Research, 10(1), 1-5.

Gunes, H., Schuller, B., Pantic, M., & Cowie, R. (2011). Emotion representation, analysis and synthesis in continuous space: A survey. In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on (pp. 827-834). IEEE.

Hewig, J., Hagemann, D., Seifert, J., Gollwitzer, M., Naumann, E., & Bartussek, D. (2005). A revised film set for the induction of basic emotions. Cognition and Emotion.

Hoque, M., & Picard, R. W. (2011). Acted vs. natural frustration and delight: Many people smile in natural frustration. In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on (pp. 354-359). IEEE.

Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research. Psychological Bulletin, 115(2), 288-299

Kamaruddin, N., Wahab, A., & Quek, C. (2012). Cultural dependency analysis for understanding speech emotion. Expert Systems with Applications, 39(5), 5115-5133.

Makatchev, M., Simmons, R., & Sakr, M. (2012). A Cross-cultural Corpus of Annotated Verbal and Nonverbal Behaviors in Receptionist Encounters. arXiv preprint arXiv:1203.2299.

Mehrabian, A. (1971). Silent Messages. Belmont, CA: Wadsworth.

Quiros-Ramirez, M. A., Polikovsky, S., Kameda, Y., & Onisawa, T. (2012). Towards developing robust multimodal databases for emotion analysis. In Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on (pp. 589-594). IEEE.

Quiros-Ramirez, M. A., & Onisawa, T. (2013) Considering cross-cultural context in the automatic recognition of emotions. International Journal of Machine Learning and Cybernetics, 1-9. DOI 10.1007/s13042-013-0192-2

Saragih, J. (2012). Non-Rigid Face Tracking. Mastering OpenCV with Practical Computer Vision Projects, 189-233.

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). 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.

Owens, M. (2006). The definitive guide to SQLite. Apress.

Wilting, J., Krahmer, E., & Swerts, M. (2006). Real vs. acted emotional speech. In *INTERSPEECH*.