Continuous Affect State Annotation Using a Joystick-Based User Interface: Exploratory Data Analysis

K. Sharma[1], C. Castellini[1] and Egon L. van den Broek[2]

[1]Robotics and Mechatronics Center, German Aerospace Center, Oberpfaffenhofen, Germany
[2]Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
karan.sharma@dlr.de and vandenbroek@acm.org

Abstract

The DLR (German Aerospace Center) aims to assess user's affective state in motion simulators. To facilitate this goal, a joystick-based user interface was used to gather reports on user's emotions. This user interface allowed continuous annotations, while video clips were watched. In parallel, several physiological parameters (e.g., electrodermal activity, heart and respiration rate) were acquired to record affective responses. An exploratory data analysis of the users' ratings (incl. several visualizations) that unveils several interesting data patterns is presented.

Introduction

The work presented here is a continuation of the research summarized in our submission to *Measuring Behavior 2014* [1]. Hence, a short re-introduction to this preceding work is presented in this section. The main focus of the aforementioned submission was to introduce the Data Acquisition (DAQ) and the video-playback system that we have developed for undertaking tests involving 'affect elicitation from videos' conditions. The two main aspects of this system are as follows:

1. **The Annotation User Interface (UI)**: features the annotation interface embedded in the video playback screen. More details on the underlying concepts and design of the UI can be found in [1].

2. **The Data Acquisition System:** simultaneously acquires the participants' physiological parameters (e.g., electrodermal activity, heart and respiration rate) [1,2,4,7-10,14,18] and self-reported affect state through the joystick. For the same, the participants were instructed to position the joystick on a 2-D plane such that the x and y position of the joystick indicates the experienced valence and arousal levels, respectively (cf. [15]).

Traditionally, experimental studies investigating participants' affective response predominantly tend to use Likert scale based self-reporting techniques, wherein the participants report their affect states on questionnaires post-stimuli [3,4,10]. Some advantages of our joystick based method over these techniques are:

1. A video is a dynamic stimulus [2,4,11,20], therefore continuous self-reporting by the participants allows us to link their responses to the events in the video and also to investigate how their affect state evolved during the course of the video, thereby providing more insight into the data [5,9,12,17].

2. A joystick based annotation system is intuitive to use (cf. [19]) and allows the participants to report their affect state at the same time as it is elicited [17].

3. As the participants report their perceived valence and arousal levels instead of discrete emotional states, the UI can also account for mixed emotional experiences [6,13-17].

A preliminary data analysis of the annotation data (i.e., the joystick ratings) recorded during the experiment was presented in an earlier submission [1]. In this current work, we extend the previous preliminary analysis by undertaking an exploratory data analysis (EDA) of the same annotation data. To perform the EDA, the data is first pre-processed and thereafter summarized. Then, exploratory graphs of this summarized data are presented.

A.J. Spink *et al.* (Eds.): Measuring Behavior 2016, ISBN 978-1-873769-59-1
Dublin, Ireland, 25-27 May 2016

These graphs provide insights into the underlying main characteristics of the data and are helpful in determining the appropriate methods to be used for model fitting and hypothesis testing.

Data Pre-processing

Before EDA can be performed on the annotation data, the data needs to be pre-processed and labelled [3,13]. The pre-processing step involves:

1. Using the *ffprobe* tool from the FFmpeg multimedia framework to determine the exact duration of the video sequences (unique for every participant) as well as the individual videos.

2. Using the duration information extracted in the last step, video-labels are added to the annotation data that contains joystick timing and co-ordinate position data. The corresponding video-label information needs to be added to this data for further data analysis. The resulting labelled joystick data for a single subject is shown in Figure 1.

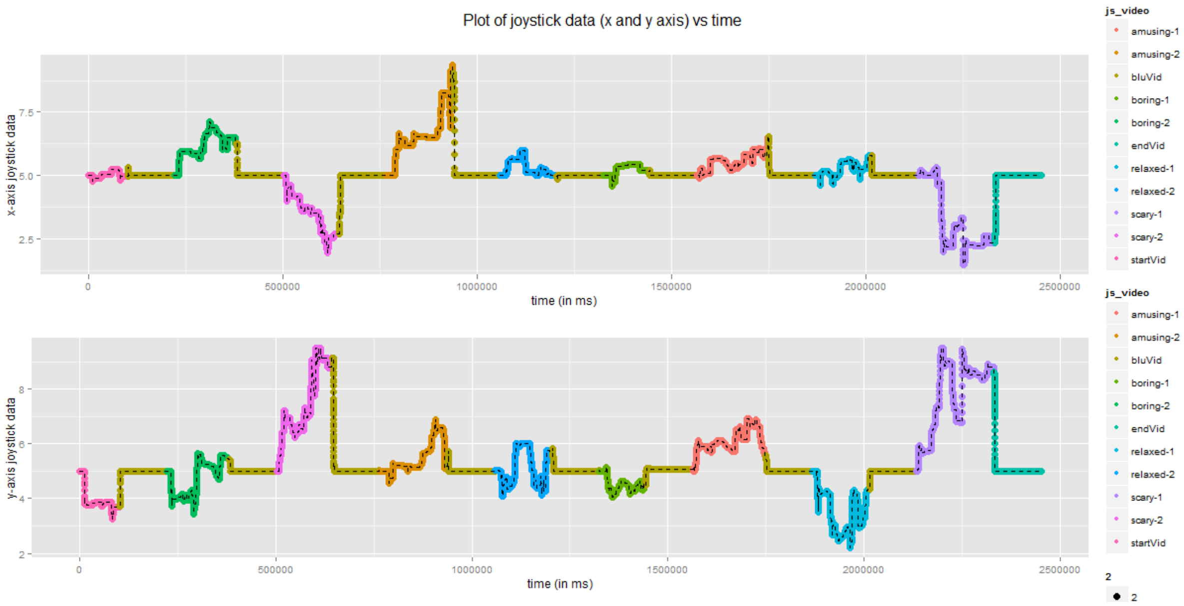

Figure 1. Plot of the joystick data-points (connected by a dashed black line) for one subject.

Data Analysis Levels

It can be seen in Figure 1, that the labelled joystick data is comprised of x and y position time series. This data can be analysed at two levels [14,17,20]:

1. **Mean ratings per video**: the *mean* valence and arousal ratings (x and y-axis joystick data, respectively) by a participant for every video are computed by calculating the arithmetic mean of the x and y positions across the complete time series (as seen in Figure 1). As there are 30 participants and 8 videos of interest in the study; this computation results in 30 mean ratings (i.e. 30 values of $\bar{x}$ and $\bar{y}$ each) for each video, where each participant provides mean ratings for 8 videos.

2. **Time series data:** continuous self-reporting through the joystick results in valence and arousal time series. By analysing these time series, a participant's response can be precisely co-related to the events in the video, and the evolution of user affect states during a video can be better analysed.

A detailed EDA of the time-series data is beyond the scope of this publication and hence will not be presented in the following sections. In the remainder of this publication, an EDA of the *mean ratings per video* data is presented.

Exploratory Data Analysis (EDA)

Figure 2 contains multiple scatter plots of the mean ratings for each of the 8 videos. In these plots, each black dot represents a mean valence and arousal rating ($\bar{x}$ and $\bar{y}$, respectively) from one subject for that video. Therefore, all the 8 scatter plots contain 30 data points each.

The scatter plots show the spread of the mean ratings for different types of videos. The blue line in these plots represents a simple linear regression line that has been fit to the given data. The shaded grey areas around the blue lines depict the 95% confidence region of the regression fit. The blue regression lines in these scatter plots show the relationship between valence and arousal ratings for different videos.

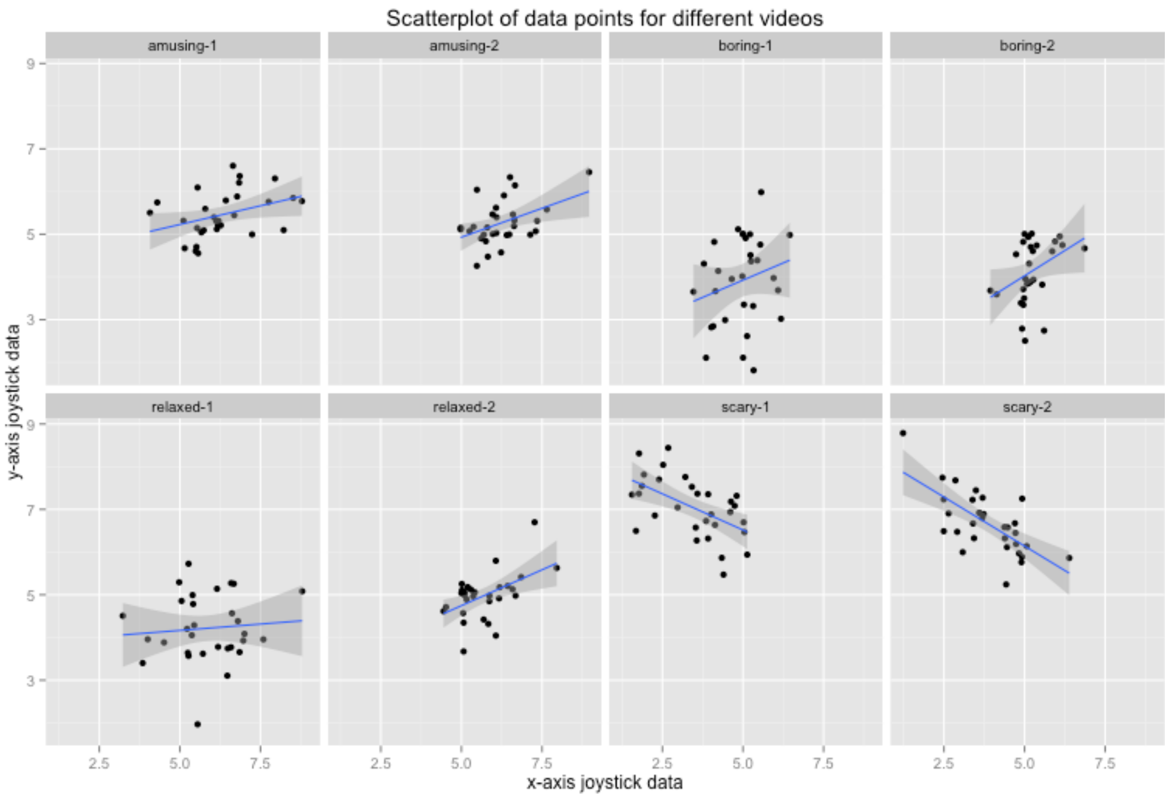

Figure 2. Scatter plot with regression lines (in blue) for different types of videos.

The box plots for the mean ratings are shown in Figure 3. This figure contains two box plots: the left and the right sub-figures show the box plots for the mean valence ($\bar{x}$) and arousal ($\bar{y}$) ratings, respectively.

In these box plots, solid black dots signify outliers; solid coloured boxes signify the Interquartile Range (*IQR*) of the data; thin black lines/whiskers signify the range Q1 - 1.5 *IQR* -- Q3 + 1.5 *IQR*; horizontal lines in the box signify the median of the mean values; and the star points signify the mean of mean values (cf. [14,18]).

Box plots are useful in visualising the spread (variance) of data in different categories (videos in our case) and facilitate a visual comparison of the data distributions across different categories.

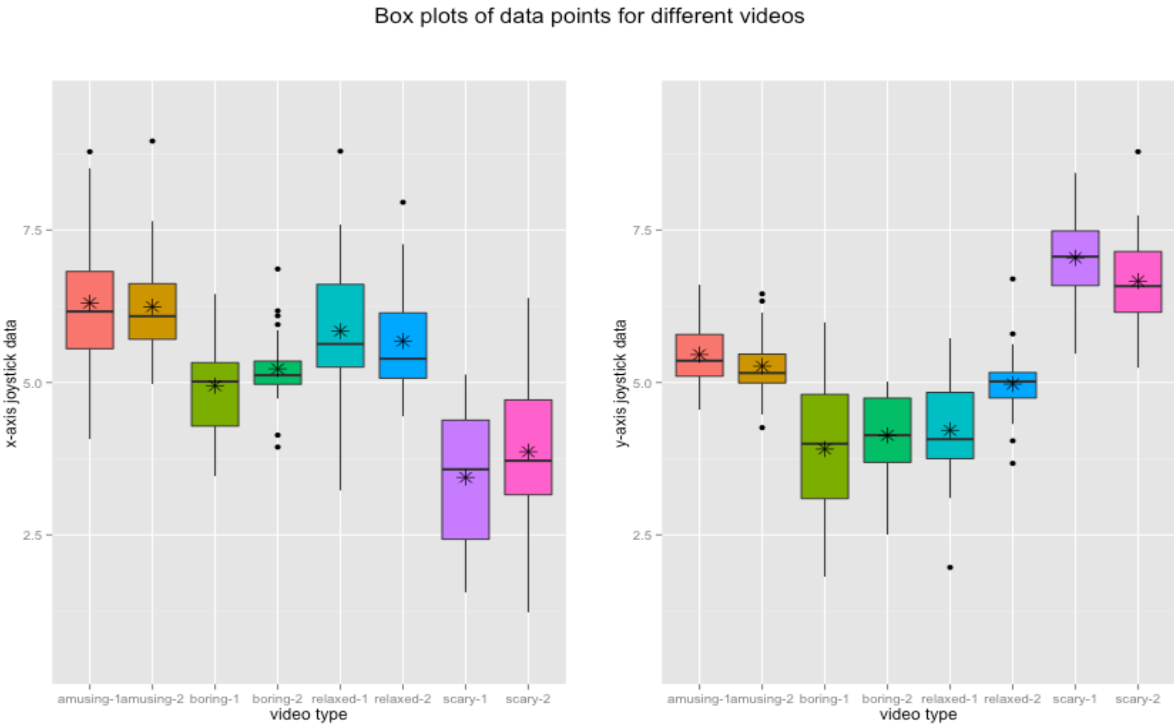

Figure 3. Box plots for the mean valence (x-axis) and arousal (y-axis) ratings.

Results

The scatter plots in Figure 3 show the location and spread of the mean ratings across different videos. For example, the mean ratings for scary-1 and 2 videos generally tend to be in the second quadrant; whereas, the ratings for amusing-1 and 2 videos tend to be in the first quadrant. The relationship between the mean ratings also differs across videos: for boring-2 video the variation of data in the *x*-axis (valence ratings) is less than the variation in the y-axis (arousal ratings); whereas, the variation of data in both axes is generally the same for the scary-1 video.

The distribution of the ratings data is better illustrated by the box plots. The data is said to be normally distributed if among others, the median is equal to the arithmetic mean. For example, for the scary-1 video, the mean and the median of the arousal (*y*-axis joystick data) ratings overlap with each other and both of them are approximately at the centre of the *IQR*. Based on these observations, we can state that arousal ratings for scary-1 video should be approximately normally distributed. Similarly, boxes that are larger in size than others have a larger *IQR*, which in turn implies that the distribution of the data is more spread out [14].

The skewness of a distribution can also be determined using the box plots (cf. [14,17,20]). Generally, if the median is not in the centre of the *IQR* box, the data is either negatively or positively skewed. For example, the median of the valence (*x*-axis joystick data) ratings for boring-1 video is not at the centre of the *IQR*, but rather shifted towards the top of the *IQR* box. This implies that the distribution is skewed to the left (negative skew).

The outliers (black dots in the box plots) in the data skew the mean in their direction; therefore, for distributions with large number of outliers the mean is skewed in one direction and hence is different from the median of the data e.g. arousal (*y*-axis joystick data) ratings for the amusing-2 video.

Conclusions

Given the multivariate nature of the presented dataset, a suitable model for hypothesis testing would be the Repeated Measures (RM) Multivariate Analysis of Variance (MANOVA) model. The EDA presented in this submission is an important initial step in determining if a chosen model is appropriate for the given dataset. For example, based on whether the data is normally distributed or not, either a parametric or a non-parametric analysis approach is chosen for the hypothesis tests.

To precisely determine which model is appropriate, thorough assumption testing must be undertaken. Nevertheless, the results of an EDA are essential building blocks that lead to comprehensive assumption testing. For example, one of the main assumptions for parametric MANOVA is that the correlation between any two dependent variables is the same in all groups of data. Using the scatter plots generated during the EDA, we can hypothesize that for the given dataset, this assumption might not be always fulfilled. Similarly, another assumption regarding the normality of data can be addressed through the box plots. The box plots presented here provide an initial indication regarding the normality or non-normality of the data. This initial indication can be then affirmed using specific tests.

Based on the EDA methods presented in this submission, an initial indication regarding the characteristics of the data was drawn. The EDA can be extended by including tests that check for specific assumptions related to models for hypothesis testing. These hypothesis tests will allow us to draw important insights and inferences about the joystick-based annotation system.

References

1. Antony, J., Sharma, K., van den Broek, E. L., Castellini, C., and Borst, C. (2014). Continuous affect state annotation using a joystick-based user interface. *Proceedings of Measuring Behavior 2014: 9th International Conference on Methods and Techniques in Behavioral Research* (Wageningen, The Netherlands), 268–271.
2. Bellmann, T., Heindl, J., Hellener, M., Kuchar, R., Sharma, K., Hirzinger, G. (2011). The DLR Robot Motion Simulator Part I: Design and Setup. In *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA 2011)* (Shanghai, China, May 9-13), p. 4694 –4701.
3. Bradley, M.M. and Lang, P.J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25(1), 49-59.
4. Janssen, J.H., Tacken, P., De Vries, G.-J., Van den Broek, E.L., Westerink, J.H.D.M., Haselager, P., IJsselsteijn, W.A. (2013). Machines outperform lay persons in recognizing emotions elicited by autobiographical recollection. *Human-Computer Interaction* 28(6), 479-517.
5. Janssen, J.H. Van den Broek, E.L., Westerink, J.H.D.M. (2009). Personalized affective music player. In J.F. Cohn, A. Nijholt, M. Pantic (Eds.), *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction (ACII2009)*, 472-477. Los Alamitos, CA, USA: IEEE Computer Society.
6. Lang, P.J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist* 50(5), 372-385.
7. Levenson, R.W. (1988). Emotion and the autonomic nervous system: a prospectus for research on autonomic specificity. In H.L. Wagner (Ed.), *Social psychophysiology and emotion: Theory and clinical applications, Chapter 2*, 17-42. Chichester, West Sussex, Great Britain: John Wiley & Sons Ltd.
8. Liu, C., Rani, P., Sarkar, N. (2006). Human-robot interaction using affective cues. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)* (Hatfield, UK, September 6-8), p. 285–290.
9. Nagel, F., Kopiez, R., Grewe, O., Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods* 39(2), 283-290.
10. Picard, R.W. (1995). *Affective computing*. Technical Report 321, M.I.T Media Laboratory Perceptual Computing Section, Cambridge, MA, USA.
11. Rottenberg, J., Ray, R.D., Gross, J.J. (2007). Emotion elicitation using films. In J.A. Coan, J.J.B. Allen

(Eds.), *Handbook of emotion elicitation and assessment, Chapter 1*, 9-28. New York, NY, USA: Oxford University Press, Inc.

12. Ruef, A.M. and Levenson, R.W. (2007). Continuous measurement of emotion: The affect rating dial. In J.A. Coan, J.J.B. Allen (Eds.), *Handbook of emotion elicitation and assessment, Chapter 17*, p. 286-297. New York, NY, USA: Oxford University Press, Inc.
13. Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review* 110(1), 145-172.
14. Van den Broek, E.L. (2011). *Affective Signal Processing (ASP): Unravelling the mystery of emotions.* CTIT Ph.D.-thesis series No. 11-204, Human Media Interaction (HMI), Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands.
15. Van den Broek, E.L. (2013). Ubiquitous emotion-aware computing. *Personal and Ubiquitous Computing* 17(1), 53-67.
16. Van den Broek, E.L., Lisý, V., Janssen, J.H., Westerink, J.H.D.M. , Schut, M.H., Tuinenbreijer, K. (2010). Affective Man-Machine Interface: Unveiling human emotions through biosignals. In A. Fred, J. Filipe, H. Gamboa (Eds.), *Biomedical Engineering Systems and Technologies,* p. 21-47. Berlin/Heidelberg, Germany: Springer-Verlag.
17. Van den Broek, E.L., Westerink, J.H.D.M. (2009). Considerations for emotion-aware consumer products. *Applied Ergonomics* 40(6), 1055-1064.
18. Van Drunen, A., Van den Broek, E.L., Spink, A.J., Heffelaar, T. (2009). Exploring workload and attention measurements with uLog mouse data. *Behavior Research Methods 41(3)*, 868-875.
19. Vuurpijl, L., Schomaker, L., Van den Broek, E.L. (2002). Vind(x): Using the user through cooperative annotation. In A.D. Williams (Ed.), *Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02)*, 221-226. Los Alamitos, CA, USA: IEEE Computer Society.
20. Westerink, J.H.D.M., Van den Broek, E.L., Schut, M.H., Van Herk, J., Tuinenbreijer, K. (2008). Computing emotion awareness through galvanic skin response and facial electromyography. In J.H.D.M. Westerink, M. Ouwerkerk, T. Overbeek, W.F. Pasveer, B. de Ruyter (Eds.), *Probing Experience: From academic research to commercial propositions*, p. 149-162. Dordrecht, The Netherlands: Springer Science Business Media B.V.