Analysing Physiology of Interpersonal Conflicts Using a Wrist Device*

Junoš Lukan✉1 [0000−0002−5325−8385], Martin Gjoreski[1], Heidi Mauersberger[2], Annekatrin Hoppe[2], Ursula Hess[2], and Mitja Luštrek[1]

[1] Jožef Stefan Institute, Jamova cesta 39, Ljubljana, Slovenia
{junos.lukan,martin.gjoreski,mitja.lustrek}@ijs.si
[2] Department of Psychology, Humbold-Universität zu Berlin, Unter den Linden 6, Berlin, Germany
{heidi.mauersberger,annekatrin.hoppe,ursula.hess}@hu-berlin.de

Abstract. We present a study in which 59 participants logged their interpersonal conflicts while wearing an Empatica E4 wristband. They marked the beginnings and endings of the conflicts, as well as their intensity. In this paper, the dataset is described and a preliminary analysis is performed. We describe data segmentation and feature calculation process. Next, the interrelationships between the features and labels are explored. A logistic regression model for conflict recognition was built and significant features were selected. Finally, we constructed a machine learning model and proposed how to improve it.

Keywords: Interpersonal conflicts · Context · Real life · Wrist device · Machine learning

1 Introduction

It is well known that understanding the user's context is essential for ambient intelligence. Most research is dealing with physical context, while psychological context remains a challenge. In this work we tackle perceived interpersonal conflicts in the workplace, an important source of stress [13] whose management could contribute to workers' satisfaction and productivity.

The psychophysiology of interpersonal conflicts has already been studied [e.g. 5, 9, 14]. The physiological data in these studies, however, is limited to short recordings and gathered in a laboratory setting. There have also not been, to our knowledge, any attempts to develop machine learning models of physiological responses during interpersonal conflicts.

In this study, participants tracked their interpersonal interactions with a focus on conflicts. They were asked to log their conflicts in the workplace for several

* Data collection for this study was supported by a grant to Heidi Mauersberger from the structured graduate program "Self-Regulation Dynamics Across Adulthood and Old Age: Potentials and Limits". Junoš Lukan's expenses were covered by Slovenian Research Agency (ARRS, project ref. N2-0081).

days and fill out a questionnaire after each such interaction. At the same time, they were wearing Empatica E4 wristbands to track their physiological data.

For the first time, to our knowledge, a large dataset of physiological data was thus collected in a field study and labelled with times and durations of interpersonal interactions. The dataset could turn out to be a valuable resource for building physiology-based models of interpersonal conflicts.

2 Data Collection and Analysis

Physiological Data. A total of 59 participants, 35 women and 24 men, with a mean age of 32.5 years ($SD = 11$ years) participated in this study. The participants, full-time employees, wore an Empatica E4 wristband throughout their work day for several days, resulting in 43.7 h of data per participant on average.

The following physiological signals were collected: a) three-axis acceleration, b) blood volume pulse (BVP) from a photoplethysmograph, c) electrodermal activity (EDA), d) heart rate and interbeat intervals as calculated from BVP on device, and e) skin temperature.

Description of Conflicts. The participants were asked to keep track of their interpersonal interactions in the workplace, paying particular attention to disagreements. They were instructed to note the time and duration as soon as possible and no more than 15 minutes after the conflict took place.

The participants reported 379 conflicts. They were further classified into task (304) or relationship conflicts (58) or both (17 conflicts). This referred to answers to two questions:

- Did you experience disagreements with your interaction partner regarding content or the implementation of the work being done?
- Did you experience personal attacks during the interaction?

The participants also answered other questions pertaining to the interaction, such as assessing its intensity and positive and negative affect during the interaction.

In a couple of days of data collection, each participant logged 6.7 conflicts on average (see Fig. 1). The mean duration of all conflicts was 10.9 min, however, almost half of them were shorter than 5 min. The distribution of the conflicts' durations is shown in Fig. 2.

Data Segmentation. To analyse the physiological data, we divided it in shorter segments on which features were calculated. We defined three periods related to each conflict: pre-conflict period, which acts as a baseline, a gap, which accounts for transitory changes before the interaction, and the conflict itself. The length of the pre-conflict period was chosen as 25 min and the length of the gap was 5 min.

The pre-conflict and conflict periods were split into two-minute segments, so that they could be compared without the effect of period duration. This length was chosen, because that is the lower bound for the frequency-domain methods used to calculate heart rate variability (HRV, see [7]).

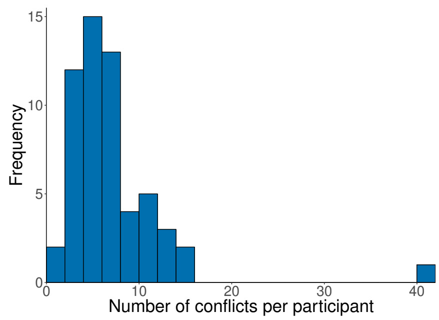

Fig. 1. The distribution of participants by the number of conflicts each of them logged. A couple of participants logged a high number of (short) conflicts.

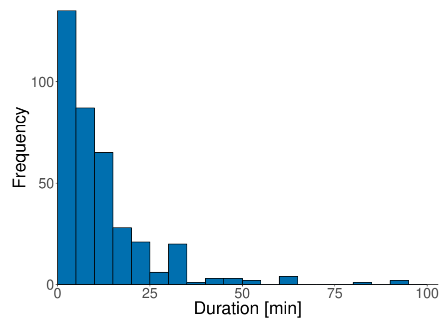

Fig. 2. The distribution of conflicts by their reported durations. Almost half of the conflicts fall within the first bin of width 5 min.

Feature Calculation. Features that can be calculated from Empatica signals are multiple and varied (for a good overview of commonly used ones see Alberdi et al. [1], and for heart rate variability, specifically, see Malik et al. [7]). Table 1 lists the features used in this work.

Table 1. Features calculated from each signal.

Signal	Features
Skin temperature	mean, median, standard deviation
Heart rate	mean, median, standard deviation
Interbeat intervals	mean, median, SDNN, SDSD, RMSSD, pNN20, pNN50, $\sqrt{SDNN}$, $\sqrt{\lvert 2SDSD^2 - 0.5SDNN^2\rvert}$, VLF, LF, HF, LF normalised, HR normalised, LF/HF
Electrodermal activity	mean, median, standard deviation, first and third quartile, interquartile range, mean derivative, number of SCRs, rate of SCRs, sum of amplitudes of SCRs, mean amplitude of SCRs, mean phasic component, mean tonic comp., mean derivative of tonic comp., integral of tonic comp. in different regions
Physical activity	mean, standard deviation, mode

For definitions of commonly used abbreviations in heart rate variability, see Malik et al. [7].

The electrodermal activity signal was first filtered using a fourth-order lowpass filter with a cut-off frequency of 1 Hz. Next, using `peakutils` Python library [8], the signal was separated into a tonic and phasic component, where the former is a slowly changing baseline and the latter characterises fast changing skin conductance responses (SCRs).

A custom algorithm was used to identify peaks in blood volume pulse and calculate interbeat intervals [12]. These were transformed using a Lomb-Scargle periodogram [6, 11] from `SciPy` [4] and typical heart rate variability features [7] were calculated.

To determine the physical activity from the acceleration data, another custom algorithm was used [3], which classifies activities into either lying, sitting, walking or standing, and running or cycling. Besides using the mode of these categories as a feature (see Table 1), they were also assigned activity intensity 1, 2, 4, and 5, respectively, so that the mean and standard deviation could be calculated.

Finally, the features were normalized. This was done within person, i.e. for each person, their mean was subtracted and the result divided by the standard deviation.

3 Results

After segmentation of the data and calculation of features, performed as described in Section 2, relationships between features and labels were explored. In Fig. 3, a correlation matrix of representative features is shown. Some correlations follow from the interrelatedness of definitions of features, such as the one between the rate of skin conductance responses (SCRratePeaks) and the sum of their amplitudes (SCRpeaksAmpSum) and some of these were intentionally excluded from the matrix in Fig. 3. Others illustrate that different features contain different information: for example, there is no correlation between the standard deviation of heart rate (HRsd) and another measure of heart rate variability, normalized power in the lower part of the spectrum (LFnorm). There are also some weak correlations between unrelated physiological parameters, such as skin temperature and galvanic skin response, specifically their standard deviations.

There are weak correlations between standardized features and conflict intensity and duration. For example, heart rate variability is somewhat lower in longer and more intensive (relationship) conflicts, which is in line with results from some existing literature (see, for example, Alberdi et al. [1]). To capture these relationships, a logistic regression model was fitted to classify conflicts and pre-conflict periods. Only conflicts longer than 2 min were analysed.

A hierarchical logistic regression model was built step-wise. Predictors from each physiological domain were included using the “forward” method, i.e., only the predictors which significantly decreased the Akaike information criterion (AIC) were retained in the model. This procedure yielded the following features: median skin temperature; normalized power in high-frequency spectrum of heart rate variability; interquartile range of electrodermal activity, and mean of skin conductance responses. No feature describing physical activity decreased AIC significantly.

Finally, a model containing the selected features was built, in which all predictors were statistically significant at $p < 0.1$ level. The interquartile range of electrodermal activity had the highest standardized coefficient $\beta = 0.096$ with $z = 2.94$ and $p = 0.003$.

An attempt to construct a machine learning predictive model was also made. Implementation of logistic regression in `scikit-learn` [10] was used. Specifically, a leave-one-subject-out (LOSO) cross-validation procedure was used [see e.g. 2], where one subject was taken as a test set, while all the others were put in

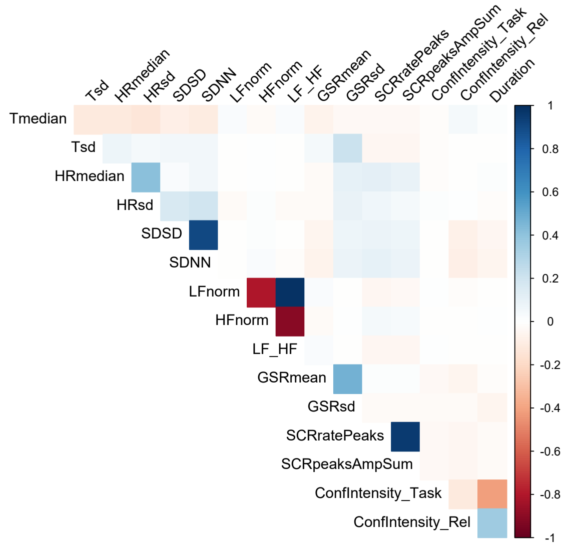

Fig. 3. Correlation matrix between selected standardized features and intensity of two types of conflicts and their durations.

the training set. The average accuracy obtained in this way was 65.8 % when distinguishing the pre-conflict period from the labelled conflicts themselves. Because the pre-conflict periods were longer than the average conflict, they made up the majority of instances, which was 63.0 %.

4 Conclusion and Outlook

In this paper, we presented a large real-life physiological dataset labelled with interpersonal conflicts. To our knowledge, this is the first example of such data being tackled by ambient-intelligence methods. Detecting conflicts in physiological data appears to be challenging and requires further work.

Several improvements to the machine learning model have already been tested. The dataset was limited to task conflicts only or conflicts with higher intensity were considered. Another way of reducing the dataset is to select only some participants. A model with subjects with more than 60 min of conflicts was tested. Different lengths of pre-conflict periods were tried out. Various types of machine learning models were also tested for predictive power, such as random forest, support vector machines etc.

None of these modifications improved the results significantly. A likely reason are the labels of the conflicts themselves. A possible workaround might be to search for the beginning of conflicts by considering significant changes of features.

In conclusion, an interesting dataset was collected with more possible pathways of analysis. The dataset is available on request.

Bibliography

[1] Alberdi, A., Aztiria, A., Basarab, A.: Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. Journal of Biomedical Informatics **59**, 49–75 (2016). https://doi.org/10.1016/j.jbi.2015.11.007

[2] Arlot, S., Celisse, A.: A survey of cross-validation procedures for model selection. Statistics Surveys **4**, 40–79 (2010). https://doi.org/10.1214/09-SS054

[3] Gjoreski, M., Gjoreski, H., Luštrek, M., Gams, M.: How accurately can your wrist device recognize daily activities and detect falls? Sensors **16**(6) (2016). https://doi.org/10.3390/s16060800

[4] Jones, E., Oliphant, T., Peterson, P., et al.: SciPy: Open source scientific tools for Python (2001–2018), http://www.scipy.org/

[5] Lavoie, K.L., Miller, S.B., Conway, M., Fleet, R.P.: Anger, negative emotions, and cardiovascular reactivity during interpersonal conflict in women. Journal of Psychosomatic Research **51**(3), 503–512 (2001). https://doi.org/10.1016/S0022-3999(01)00217-3

[6] Lomb, N.R.: Least-squares frequency analysis of unequally spaced data. Astrophysics and Space Science **39**, 447–462 (Feb 1976). https://doi.org/10.1007/BF00648343

[7] Malik, M., Bigger, J.T., Camm, A.J., Kleiger, R.E., Malliani, A., Moss, A.J., Schwartz, P.J.: Heart rate variability. European Heart Journal **17**(3), 354–381 (1996)

[8] Negri, L.H., Vestri, C.: lucashn/peakutils: v1.1.0 (Sep 2017). https://doi.org/10.5281/zenodo.887917

[9] Neumann, S.A., Maier, K.J., Brown, J.P., Giggey, P.P., Cooper, D.C., Synowski, S.J., Goble, L.A., Suarez, E.C., Waldstein, S.R.: Cardiovascular and psychological reactivity and recovery from harassment in a biracial sample of high and low hostile men and women. International Journal of Behavioral Medicine **18**(1), 52–64 (Mar 2011). https://doi.org/10.1007/s12529-010-9110-0

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

[11] Scargle, J.D.: Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. Astrophysical Journal **263**, 835–853 (Dec 1982). https://doi.org/10.1086/160554

[12] Slapničar, G., Luštrek, M., Marinko, M.: Continuous blood pressure estimation from PPG signal. Informatica **42**(1), 33–42 (2018), http://www.informatica.si/index.php/informatica/article/view/2229

[13] Stephenson, E., King, D., DeLongis, A.: Coping process. In: Fink, G. (ed.) Stress: Concepts, Cognition, Emotion, and Behavior, chap. 44, pp. 359–364. Academic Press, San Diego (2016). https://doi.org/10.1016/B978-0-12-800951-2.00045-5

[14] Tada, M., Uchida, H., Maeda, T., Konishi, M., Umeda, S., Terasawa, Y., Nakajima, S., Mimura, M., Miyazaki, T., Takahashi, T.: Fear conditioning induced by interpersonal conflicts in healthy individuals. PLOS ONE **10**(5), 1–14 (05 2015). https://doi.org/10.1371/journal.pone.0125729