Emotion Recognition by Point Process Characterization of Heartbeat Dynamics

Akshay Sujatha Ravindran, *Student Member, IEEE*, Sho Nakagome, *Student Member, IEEE*, Dilranjan S. Wickramasuriya, *Student Member, IEEE*, Jose L. Contreras-Vidal, *Fellow, IEEE* and Rose T. Faghih, *Member, IEEE*

Abstract— **Recognizing human emotion from heartbeat information alone is a challenging but ongoing research area. Here, we utilize a point process model to characterize heartbeats. We extract features from the model and train an ensemble learner to classify these features into high/low valence, arousal, dominance and liking classes based on subject ratings. On average, we achieved over 60% classification accuracy which is comparable to other methods that use a combination of multiple types of physiological signals as opposed to only one type of physiological signal used here. Informative features were identified for the affective states and statistical testing was performed to check for significant differences. From the results we found that the ratio of low to high frequency band power, the mean and the 10$^{\text{th}}$ RR-interval percentile were the most significant features for distinguishing between low and high levels of valence and dominance. These findings enable the possibility of augmenting electrocardiogram or photoplethysmogram monitoring wearable devices with automated human emotion recognition capabilities for mental health applications.**

I. INTRODUCTION

Automated human emotion recognition has been an interesting and ongoing research area involving multidisciplinary expertise. It has been reported that over two million US citizens have been diagnosed with bipolar disorder [1]. Despite the rise in prevalence, current practices used for assessing emotions are mainly conducted by means of basic questionnaires or are solely based on physician experience. Some of the commonly used emotional spaces are: the discrete emotion model proposed by Ekman [2] with six universal emotions (happiness, surprise, anger, disgust, sadness and fear), the two dimensional valance-arousal model by Russell [3] which categorizes emotion according to scales of valance and arousal, and the PAD (pleasure, arousal, and dominance) model [4] which describes human emotions in terms of pleasure, arousal and dominance axes.

In general, people tend to express their emotions through the tone of their voice, gestures, posture and facial expressions [5]. The usage of gesture, facial expression and speech-based emotion detection techniques are susceptible to social masking as they can be easily modulated/suppressed by the subjects themselves [6]. This led to the popularity of emotion recognition techniques using physiological signals within the last decade, as they originate from Autonomic Nervous System (ANS) activity and hence cannot be triggered by volitional control [7].

Experimental evidence has demonstrated that the analysis of Heart Rate Variability (HRV) in both the time and frequency domains can provide insight into changes associated with emotion processing [8], [9]. The ANS is composed of both the sympathetic and parasympathetic branches, both of which are innervated to the heart in the sinoatrial node, which is in charge of heart's neuromodulation in response to sympathetic and vagal activities [10]. An increase in sympathetic neural activity is associated with an increase in heart rate while a relative increase in parasympathetic activity has the opposite effect [9]. The parasympathetic influences are typically vagal, and manifest over the entire HRV spectrum of the heart while the sympathetic influences "roll off" at 0.15 Hz [8].

Recent advances in wearable devices have galvanized the widespread use of consumer products capable of measuring physiological signals. Such device are expected to play a supportive role in health care, emotion recognition, and health management applications. However, the development of reliable health monitoring systems using commercially available wearable devices are still in an early stage of development [11]. While improvements in hardware such as better sensors can increase reliability, software improvements including better algorithms to estimate behavior would also be essential.

In this study, we aim to implement a point process model for heartbeats with parameters optimally chosen using Maximum Likelihood Estimation (MLE) and the Bayesian Information Criterion (BIC) [12]. The model is used to extract features for characterizing different emotional states. Photoplethysmography (PPG) signals are used here for deriving heart rate. HRV-based features are finally classified for emotion recognition.

II. METHODS

A. Data

The Database for Emotion Analysis using Physiological Signals (DEAP) is an open source data set [13] containing multimodal physiological signals – electroencephalography (EEG), electromyography (EMG), skin conductance, respiration, PPG and body temperature. The data was recorded

A. S. Ravindran, S. Nakagome, D. S. Wickramasuriya, J. L. Contreras-Vidal and R. T. Faghih are with the Department of Electrical and Computer Engineering at the University of Houston, Houston, TX 77004 USA. (e-mail:{asujatharavindran, snakagome, dswickramasuriya, jlcontreras-vidal, rtfaghih}@uh.edu). This work was supported in part by NSF grant 1755780 – CRII: CPS: Wearable-Machine Interface Architectures. Correspondence should be addressed to senior author Rose T. Faghih.

from 32 healthy subjects (16 male; 16 female) while they watched a series of music videos meant to elicit different emotions [13]. In this study, we make use of the data from the first 22 subjects to characterize emotions solely using heart rate modelled as a point process. The data from the rest of the subjects were collected from a different location and were not used in this study.

Subjects were instructed to provide feedback on each music video they viewed and rate them separately on valence, arousal, dominance, liking and familiarity scales ranging from 1-9. Due to the potential for subjective bias and differences added due to the use of Self Assessment Scores instead of standardized scores, we classified the trials to belong to high/low categories of arousal, valence, dominance and liking by imposing an inclusion criterion. The self assessment scores for each individual were first sorted and the 10 scores in the middle were removed to reduce the effects of inexperience in rating and subjective bias. This also ensured that we had an equal number of trials in the high/low affective states. The scores on either side of this criterion were then labeled as high or low accordingly.

B. Preprocessing

PPG signals from the DEAP dataset were first high pass filtered at 2 Hz to remove drift. Peak detection was used to detect the valleys of the signal and the differences between successive peaks were calculated as RR-intervals. Missed detections and false peaks were identified and manually corrected before constructing the RR-intervals.

C. Point Process Modeling

We used the point process model in [14] for characterizing heartbeat dynamics. A point process provides a mathematical means of modeling physical activities such as heartbeats that have homogeneous or inhomogeneous Poisson arrival times. Assume K successive R-peak occurrences at times u_k during the observation interval $(0,T]$ such that $0 < u_1 < u_2 < \cdots < u_K \leq T$. We can define the RR-intervals as $w_k = u_k - u_{k-1}$ and the history term as $H_k = \{u_k, w_k, w_{k-1}, \ldots, w_{k-p+1}\}$ where p is the order for the model.

At time $t > u_k$, the inter-arrival time for the next R-peak can be modeled using a History Dependent Inverse Gaussian (HDIG) density function:

$$f(t|H_k,\theta) = \left[\frac{\theta_{p+1}}{2\pi(t-u_k)^3}\right]^{\frac{1}{2}} \times \exp\left\{-\frac{1}{2}\frac{\theta_{p+1}[t-u_k-\mu(H_k,\theta)]^2}{\mu(H_k,\theta)^2(t-u_k)}\right\}, \quad (1)$$

where $\mu(H_k,\theta) = \theta_0 + \sum_{j=1}^{p}\theta_j w_{k-j+1} > 0$ and $\theta = (\theta_0, \theta_1, \ldots, \theta_{p+1})$.

Considering the dynamics of the parasympathetic and the sympathetic inputs to the sinoatrial node as continuous, we consider the parameter θ to be time-varying and use the local maximum likelihood procedure described in [14] to track the instantaneous RR-intervals. The parameters are updated at each discrete time bin $j\Delta$ for $j = 1, \ldots, J$, where $\Delta = T/J$. Here, we use $\Delta = 10$ ms.

D. Model Selection

BIC is one of the commonly used criterion for model selection. We used MLE to choose the optimal order p for each subject and the θ_j estimates at each discrete time instant. MLE and BIC were run iteratively to determine the model order that gave the lowest BIC values.

$$BIC = -2\log(likelihood) + K \times \log(N) \quad (2)$$

where K is the number of parameters estimated and N is the sample size.

E. Features

We extracted the following features from the RR-intervals to characterize the HRV variations associated with different emotions in each of the trials. These features were selected as they have shown sensitivity towards emotion recognition in prior studies.

The following were computed from the HDIG model:

Frequency domain features

1) HP (High Frequency Power): Power in 0.15 to 0.4 Hz.
2) LP (Low Frequency Power): Power in 0.04 to 0.15 Hz.
3) VLP (Very Low Frequency Power): Power $<$ 0.04 Hz.
4) LP/HP (Power ratio): Ratio of the LP to HP
5) Total Power: The total power over the trial.

Time domain features

1) Mean RR-Interval
2) Variance of the RR-interval
3) Max RR-interval: The maximum RR-interval
4) 98 Percentile of the RR-interval
5) Min RR-interval
6) 10th Percentile of RR-interval

The following were computed from the original RR-intervals and not from the HDIG model:

7) SDSD: Standard deviation of the successive difference between RR-intervals.
8) RMSSD: Root mean square of the successive difference between RR-intervals.
9) pNN50: Ratio of the number of successive RR-intervals which differ by more than 50 ms by the total number of RR-intervals.

Complexity feature

1) Sample Entropy: Complexity measure used to indicate the regularity of the signal [15].

All the features were z-scored for each individual for the remainder of the analysis.

F. Statistical Testing

Initially the mean values of individual features for each subgroup (low valence, high valence, low arousal, high arousal, low dominance, high dominance, low liking, high liking) was computer for each subject. Then, the normality of the distribution of these mean values from all the subjects was assessed using one-sample Kolmogorov-Smirnov test. Since most of them did not satisfy the normality assumption, a non-parametric paired statistical test was performed using the Wilcoxon signed rank test, on all features. The mean feature value during high emotional state trials were paired with the corresponding values during the low emotional trials. A p-value < 0.05 was considered to be statistically significant.

G. Classification and Feature Learning

We implemented ensemble learning using the *fitcensemble* function in Matlab 2018b (MathWorks, Inc., Natick, Massachusetts, United States), to classify the HRV features based on the emotion scores (each for valence, arousal, dominance, liking). Bootstrap aggregation that bags tree-stump based weak learners were used for learning. Multiple bootstrapped replicas are selected randomly from the data with repetition and these are used to grow decision trees. The average response of the prediction from all the trees gives the final prediction. Using such tree-based bagging methods also allows for understanding the importance of the features used in the model according to the change in risk/impurity associated with the split on each feature. The *Predictor Importance* function in Matlab was used to compute this estimate for each feature.

We used both subject-specific and subject-independent models. The latter model was trained on the pooled data from all the subjects whereas the former was trained for each subject individually, to fine tune for each subject. Five-fold cross validation was performed to report the accuracy.

III. Results

A. Goodness of Fit

We estimated the optimal order (9.63 ± 1.84 across subjects) of the inverse Gaussian model for each individual subject by systematically increasing the model order from 2 to 15 and selecting the one that yielded the least value for BIC. This model order was then kept fixed for each subject, independent of the trial.

B. Statistical Testing

The features which exhibited statistically significantly differences between the groups at a significance level of 0.05 were identified and are shown Fig. 1. For arousal and liking, no feature showed a significant difference between the high vs. low classes. Low frequency band power, the ratio of band powers and the total power were statistically significantly different in the frequency domain for valence whereas no frequency domain feature seemed to be significantly different for high vs. low dominance. The mean and 10^{th} RR-interval percentile were the significant time domain features for both valence and dominance classification. The dominance condition also showed significant difference for the 90^{th} RR-interval percentile.

C. Classification and Feature Learning

The 5-fold cross validated accuracy for the ensemble learners is summarized in Table I. The model was able to discriminate between high and low valence and dominance better than for arousal and liking when data from all the subjects were combined. However, for the subject-specific models, predicting based on arousal rating yielded the highest accuracy.

TABLE I

Cross Validation accuracy of classifiers: All values are in %

Classifier	Valence	Dominance	Arousal	Liking
Subject independent	57.57	55.45	51.96	52.72
Subject specific (mean)	63.2	64.1	66.5	63.8
Subject specific (max)	70.2	70.6	76.7	71.1

Fig. 1 shows an example of the most informative features for valence and dominance. The arousal and liking groups are not considered as we did not find any feature that showed a statistically significant difference across the high/low conditions. For valence, band power ratio and the lowest 10^{th} RR-interval percentile are the most relevant features. For dominance, the mean RR-interval and the 90^{th} RR-interval percentile are the two most relevant features.

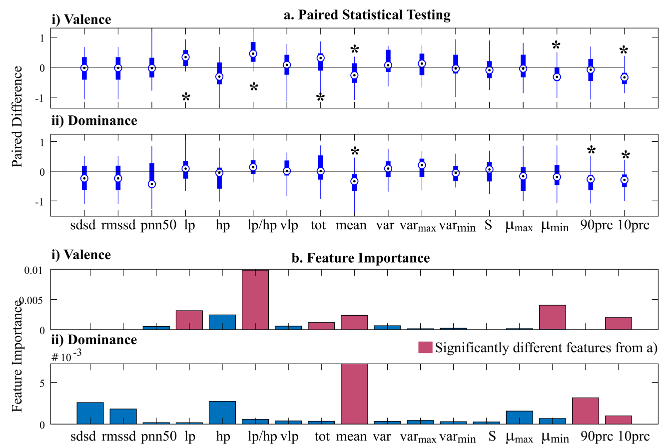

Fig. 1. Feature learning. a)distribution of the paired differences in features between high and low emotional states; *statistical significance at $p < 0.05$. b) estimated feature importance scores (subject independent classifier)

D. Continuous Monitoring of Important Features over Trials

Based on the statistical and the feature importance analysis, the ratio of band power, mean and extreme values of the RR-intervals during the trial were the most relevant features to discriminate between levels of valence and dominance. We then observed the change in these features for one of the subjects who yielded more than 65% accuracy on both valence and dominance classification. We can see that the from Fig.2, ratio is lower in the low valence/dominance condition compared to the higher rating conditions in general. Similarly the mean HR is lower in the high valence/dominance compared to lower rated trials. For instance, the mean RR-intervals increase with a reduction in dominance rating for the first 3 trials in dominance.

IV. Discussion

In the original DEAP paper, Koelstra *et al.* [13] reported an overall F1 score of 60.8% for valence, 53.3 % for arousal and 53.8% for liking. However, they used all the peripheral

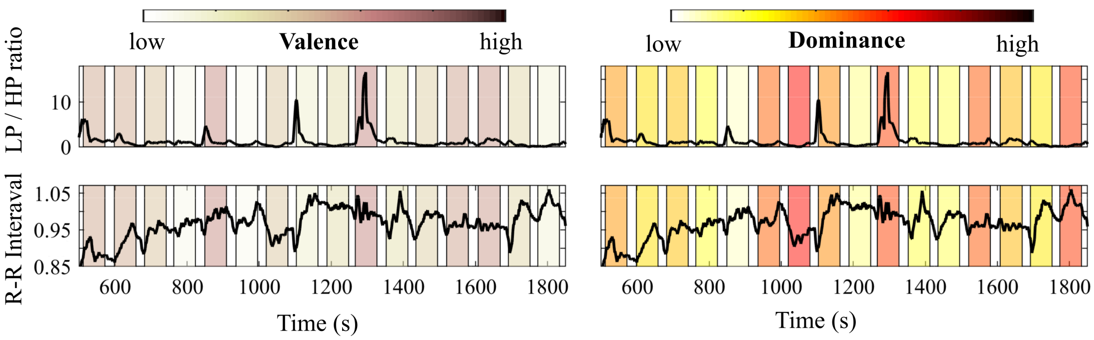

Fig. 2. Continuous monitoring of band power ratio and RR-intervals during the first half of the experiment for subject 18; the colored blocks represents different trials and the intensity corresponds to the emotion ratings.

signals (skin conductance, PPG, EMG, temperature and respiration) for the classifier. In this study, we were able to distinguish between different levels of emotion solely using heartbeat dynamics. We also identified that the ratio of band power is a key feature in differentiating between levels of valence. This is in agreement with prior studies that showed a similar relationship [16], [8]. Similarly, mean RR-interval and the 10^{th} RR-interval percentile is found to be lower in the high valence/dominance class suggesting that heart rate generally higher in such conditions. All the statistically significantly different features were identified as relevant features by the ensemble learner as well. At the same time, none of the features which gave the least or no importance were statistically different either. Similar to [13], we had a higher accuracy in classifying valence. We feel the model accuracy might have suffered due to the use of subjective ratings in the study, which might make it hard to compare across subjects and might be prone to subjective bias and inexperience.

In conclusion, we were able to distinguish between high and low levels of valence and dominance using only heartbeat dynamics. An HDIG model optimized using MLE and BIC for each subject was used to characterize RR-intervals. We identified relevant features for identifying levels of emotion, which is in agreement with prior literature as well. Performing non-parametric paired statistical tests, we found that these features were indeed statistically different. Future work would involve incorporating EEG features for improving classification accuracy. Examining physiological signal changes in additional scenarios (e.g. art, dance and drama) that evoke different emotions would be yet another direction of research.

References

[1] R. C. Kessler, K. A. McGonagle, S. Zhao, C. B. Nelson, M. Hughes, S. Eshleman, H.-U. Wittchen, and K. S. Kendler, "Lifetime and 12-month prevalence of dsm-iii-r psychiatric disorders in the united states: results from the national comorbidity survey," *Archives of general psychiatry*, vol. 51, no. 1, pp. 8–19, 1994.

[2] P. Ekman, W. V. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti *et al.*, "Universals and cultural differences in the judgments of facial expressions of emotion." *Journal of personality and social psychology*, vol. 53, no. 4, p. 712, 1987.

[3] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[4] A. Mehrabian, "Framework for a comprehensive description and measurement of emotional states." *Genetic, social, and general psychology monographs*, 1995.

[5] T. Bänziger, D. Grandjean, and K. R. Scherer, "Emotion recognition from expressions in face, voice, and body: the multimodal emotion recognition test (mert)." *Emotion*, vol. 9, no. 5, p. 691, 2009.

[6] P. Rani, C. Liu, N. Sarkar, and E. Vanman, "An empirical study of machine learning techniques for affect recognition in human–robot interaction," *Pattern Analysis and Applications*, vol. 9, no. 1, pp. 58–69, 2006.

[7] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: a review," in *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*. IEEE, 2011, pp. 410–415.

[8] M. T. Valderas, J. Bolea, P. Laguna, M. Vallverdú, and R. Bailón, "Human emotion recognition using heart rate variability analysis with spectral bands based on respiration," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE, 2015, pp. 6134–6137.

[9] R. D. Lane, K. McRae, E. M. Reiman, K. Chen, G. L. Ahern, and J. F. Thayer, "Neural correlates of heart rate variability during emotion," *Neuroimage*, vol. 44, no. 1, pp. 213–222, 2009.

[10] K. Sunagawa, T. Kawada, and T. Nakahara, "Dynamic nonlinear vago-sympathetic interaction in regulating heart rate," *Heart and vessels*, vol. 13, no. 4, pp. 157–174, 1998.

[11] M. Cardinale and M. C. Varley, "Wearable training-monitoring technology: applications, challenges, and opportunities," *International journal of sports physiology and performance*, vol. 12, no. Suppl 2, pp. S2–55, 2017.

[12] E. Wit, E. v. d. Heuvel, and J.-W. Romeijn, "all models are wrong...: an introduction to model uncertainty," *Statistica Neerlandica*, vol. 66, no. 3, pp. 217–236, 2012.

[13] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[14] R. Barbieri, E. C. Matten, A. A. Alabi, and E. N. Brown, "A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 288, no. 1, pp. H424–H435, 2005.

[15] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[16] H. Ferdinando, L. Ye, T. Seppänen, and E. Alasaarela, "Emotion recognition by heart rate variability," Ph.D. dissertation, Petra Christian University, 2014.