CONTEXT-BASED SIGNAL DESCRIPTORS OF HEART-RATE VARIABILITY FOR ANXIETY ASSESSMENT

Moitreya Chatterjee, Giota Stratou, Stefan Scherer and Louis-Philippe Morency
Institute for Creative Technologies, University of Southern California, Playa Vista, CA

ABSTRACT

In this paper, we investigate the role of multiple context-based heart-rate variability descriptors for evaluating a person's psychological health, specifically anxiety disorders. The descriptors are extracted from visually sensed heart-rate signals obtained during the course of a semi-structured interview with a virtual human and can potentially integrate question context as well. The proposed descriptors are motivated by prior related work and are constructed based on histogram-based approaches, time and frequency domain analysis of heart-rate variability. In order to contextualize our descriptors, we use information about the polarity and intimacy levels of the questions asked. Our experiments reveal that the descriptors, both with and without context, perform far better than chance in predicting anxiety. Further on, we perform at-a-par with the state-of-the-art in predicting anxiety and other psychological disorders when we integrate the question context information into the descriptors.

***Index Terms*—** Heart-Rate Variability, Descriptors, Anxiety, Classification

1. INTRODUCTION

The heart-rate (measured in beats per minute) is a major physiological parameter. Its variability is largely involuntary [1] and has been used for detecting several clinical disorders. For example, a method is suggested for using heart-rate variability to detect perinatal hypoxia in [2]. Research also suggests that it is indicative of the stimuli a person is subjected to [3], [4].

Bradley et al. reports, that the response of the heart-rate variability is consistent with the motivational hypothesis [3], which suggests a relative deceleration in the heart-rate of individuals when subjected to stimuli that induce strongly negative affect (e.g. highly polar and intimate interview questioning). They observe that humans exhibit such a behavior in order to gain control of the challenge the scenario presents. Hjortskov et al. states that even common stressful activities such as, mental arithmetic, could result in significant heart-rate variations [4]. These are two major works studying the effect of stimuli on heart-rate.

Prior research work in the field of psychology, sociology and medicine has further on, looked into the strong association between heart-rate variability and

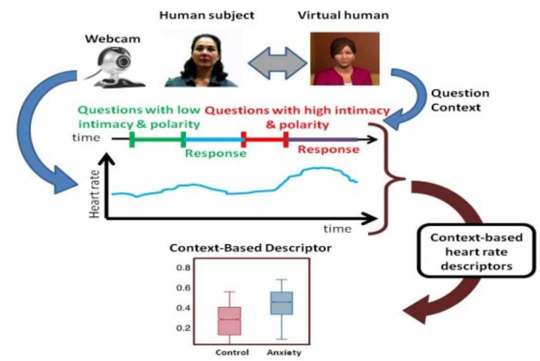

Fig. 1. A broad overview of our proposed method which shows a representation of the interview setting and how the descriptors are computed.

psychological disorders such as anxiety. Gorman et al. suggest an unopposed stimulation of the heart and reduced heart-rate variability in patients with mood and anxiety disorders [5]. They attribute these findings to the reduced parasympathetic innervations of the heart for individuals with such psychological disorders. While the line of research in [5,6] is promising, it lacks automatic descriptors and predictive models of anxiety disorder.

Moreover most of the prior work related to heart-rate variability pattern analysis, used contact-based techniques for measuring the heart-rate of the individuals. For a face-to-face conversation this could severely bias the non-verbal response of the participant, as the attached sensors could make them more self-conscious. This impedes rapport building and further interaction between the interviewer (here a virtual human) and the participant. This implies a need for non-contact based measurement techniques of heart-rate. Recently, an innovative approach for addressing this issue was proposed in [7]. The proposed algorithm, automatically analyzes the change of skin color around the face region (from the video obtained using a simple webcam) and applies blind source-separation strategies to infer heart-rate, with high accuracy. We therefore, adopt this strategy for heart-rate estimation.

In this paper we propose a computational approach to automatically predict psychological anxiety based on visually inferred heart-rate measurements. We investigate multiple heart-rate variability descriptors that can integrate affective information from the question context (viz. the degree of intimacy and polarity of the questions) along with the discriminative information contained in the heart-rate variability. We then employ a standard probabilistic machine learning based approach to evaluate the descriptors by categorizing a population of healthy individuals from those suffering from anxiety disorders. Figure 1 provides a broad overview of our approach. This proposed research is complementary to acoustic and visual descriptors for psychological anxiety [8], [9].

1.1. Research Goals

A look at the aforementioned prior findings, demands that we address the following major research questions.

1) Can heart-rate variability descriptors, computed from heart-rate signals measured using non-contact based techniques, be predictive of a person's psychological condition (especially anxiety disorders)?
2) What is the impact of contextualizing our descriptors using information about the nature of the questions asked (in terms of polarity and intimacy)?

A proper methodology to address these research goals could go a long way in helping clinicians and healthcare providers perform a quick and proper detection and administer appropriate treatment to those in need.

In the following section we describe the dataset that we used. Section 3 describes the Heart Rate Variability descriptors and how to integrate context information with it. Section 4 discusses the experimental methodology and Section 5 presents the results of our experiments. In Section 6 we discuss our findings. Finally in Section 7, we conclude putting forth some approaches for the future.

2. DATASET

In this section we introduce the dataset that we used for our experiment. The corpus used in our experiment is an extension of the Wizard-of-Oz Corpus introduced in [9] and is called the Virtual Human Distress Assessment Interview Corpus for Anxiety Analysis. This new extension to the corpus is mainly intended for analyzing the participant's mental health conditions in terms of anxiety disorders.

2.1. Procedure

Participants for the experiment were recruited via Craigslist and their interactions with the virtual humans were recorded with a webcam (30 fps), kinect sensor and a microphone. In total 48 interactions, lasting from 5 to 15 minutes were recorded. The participants were all above 18 years of age. 30 of the 48 participants were males and the rest were females. The interactions were recorded in a Wizard-of-Oz setup, similar to the one outlined in [9]. For a clinical diagnosis of the individuals we used the State/Trait Anxiety Inventory as suggested in [8]. A threshold level of 44 (mean + 1S.D.) was selected for categorizing the participants as ones suffering from anxiety versus a healthy populace. 33 of the 48 showed anxiety disorders.

2.2. Polarity and Intimacy Levels of Questions

The questions asked by the Virtual Human varied in their polarity (negative, neutral or positive) and intimacy levels (more intimate or less intimate). The questions affective polarity was evaluated on a Likert Scale from -2 (strongly negative) to +2 (strongly positive) while the intimacy levels were rated from 1 (not intimate) to 3 (very intimate), by two independent coders. The coder agreement as measured by Krippendorff's α, was 0.86 which indicates a high degree of inter-coder agreement. All the interviews were triphasic in pattern (i.e. rapport building-intimate and strongly polar questioning-cool off phase).

We categorize the questions asked by the Virtual Human into two broad categories, which we call *sympathetic* and *non-sympathetic* in order to use them as context. The sympathetic questions are those which had an average polarity rating of more than 1.5 or less than -1.5 (either strongly positive or strongly negative) and had an average intimacy rating of more than 2.5 (very intimate). All other questions were considered as non-sympathetic.

3. CONTEXT-BASED SIGNAL DESCRIPTORS OF HEART-RATE VARIABILITY

In order to aptly capture the discriminatory information contained in the heart-rate variability (HRV) signal obtained using non-contact vision based technique [7], we look into 8 distinct feature descriptors. For building all of these descriptors we compute a feature vector for the heart-rate variability for each question-response.

We categorize the 8 feature descriptors based on their types into a few subgroups.

A. Histogram Based Descriptors: A very standard approach that is adopted for computing signal descriptors is to adopt a histogram-based representation [10]. The following 3 descriptors draw inspiration from such techniques.

i. **Top Time/Bottom Time:** A significant influence of affect on heart-rate variability is reported in [3], [4]. Inspired by such work, we first compute the range of variation of the heart-rate during a question-response. Following this we construct a 2d feature, the first containing a count of the number of seconds the heart-rate is in the top quartile of the range, while the count of the bottom quartile is kept in the second bin.

ii. **Rise Time/Fall Time:** This feature is also represented using a 2-bin histogram with the first holding the count of the number of seconds the heart-rate is rising while the other one is for the falling during a ques.-response.

iii. **Mean Deviation Feature:** It is essential to capture the variation of the signal from the global mean. So we develop a 2-bin histogram descriptor, with each keeping a count for how many seconds the signal deviates above or below the mean.

B. Time Domain Descriptors: The time-domain representation of the HRV contains significant discriminatory information as well. We therefore look into the following descriptors.

i. **Range Feature:** This is a single quantity describing the range of HRV during a question-response, computed by taking the difference between the maximum and minimum values.

ii. **Number of Peaks and Peak Time Lag:** The former is a count of the number of peaks during a question-response while the later is the average time lag between two consecutive peaks.

iii. **Standard Deviation:** For this we compute the standard deviation of the HRV during a ques.-response.

C. Frequency Domain Descriptors: Some prior works suggest looking into the frequency domain response of HRV [2], [11], [12]. We therefore build a feature, drawing inspiration as such.

i. **FFT Coefficient:** We use the amplitude of the coefficient corresponding to the lowest frequency obtained by computing the FFT of the heart-rate signal, as a feature. We consider the entire question-response as a window for this task. This is because for ailing people we observe HRV reduction [5], [6]. We expect this to be manifested in the frequency domain.

D. Response Time: We also use the response time of the participants as a feature. It is a measure of the duration of the response for every question.

In order to integrate question context information we categorize these features, obtained during every question-response of an individual, into two groups (ones corresponding to sympathetic and non-sympathetic questions). For each of the two groups we compute the average value for each feature of HRV over all the question-responses in that group. This gives us a set of 24-dimensional descriptors (2x12=24) corresponding to every individual.

4. EXPERIMENTS

We validate our findings through the following experiments and also present our results.

4.1. Validation of Vision-based Heart Rate Estimation

As a first step we validate our implementation of the heart-rate signal estimation technique [7] with the ground truth. A high accuracy is reported using the technique in [7], where Viola- Jones is used for face-detection, [13]. We however used the confidence values obtained from OKAO [14] to ignore the outliers in the detection. Whenever a zero confidence is reported, we used the detection for the previous frame as an estimate. This improves the performance of the algorithm (Section 5.1). Here we note that, even though a window size of 30s is used but an increment of only 1s ensures that major variations in the signal is not lost.

Descriptor	With Context		Without Context
	Sym	Non-Sym	
Bottom Time	0.376	**0.008**	0.086
Top Time	0.443	0.350	0.376
Rise Time	0.151	0.100	0.069
Fall Time	0.311	0.075	0.122
Mean Dev. Bin 1	0.239	0.371	0.266
Mean Dev. Bin 2	**0.016**	0.448	0.236
Range Feature	0.349	**0.039**	0.255
Nos. of Peaks	0.393	0.092	0.202
Peak Duration	0.417	0.100	0.234
Standard Dev.	0.238	**0.043**	0.307
Response Time	0.258	**0.050**	0.144
FFT Coefficient	0.449	0.051	0.241

Table 1. The p-values for the various descriptors. The ones in bold have a p-value <= 0.05 (5 features).

For accuracy measurement the average RMSE scores between the ground truth (measured using a Biopac) and the estimated heart-rate signal for 6 separate individuals and also their correlation measure was recorded.

Moreover to study the extent of impact of head-pose variation of the participant on the computed heart-rate in our dataset, we measured the correlation between the change of head-orientation angle obtained using [15] and the recorded heart-rate.

4.2. Feature and Context Analysis

In order to test the discriminative power of our descriptors (with or without context) Student's T-test were performed.

For evaluating the effectiveness of the descriptors independent of the question context, we first compute the descriptors for the heart-rate variability (HRV) during each question-response without categorizing it into the two groups. We then just take an average over all the question-responses for each feature. This gives us a 12-dimensional descriptor for each participant. We then perform a t-test for each of the 12 descriptors to test their significance.

In order to construct the context-based descriptor, we adopt the approach suggested in Section 3 to obtain a 24d-descriptor for each participant and a t-test is performed for each descriptor.

4.3. Anxiety Prediction

In order to classify individuals, we divide the data into an independent training and a test set (for both context-based and otherwise). We also balanced the test-set equally between the two classes of anxiety and control by keeping equal number of participants in both classes.

We perform feature selection for the classification task. The optimal number of features was automatically validated by performing a greedy backward selection using a leave-one-candidate-out validation. We used Logistic Regression, Naïve Bayes [16] and a Bayesian Network (BayesNet) models for this task.

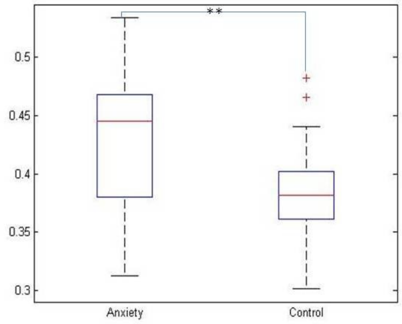

Fig. 2. Exemplary visualization of the context-based bottom time feature for anxiety and control population.

For the BayesNet structure learning we perform a Simulated Annealing Search in the space of candidate networks. For inference, we estimate the conditional probabilities using maximum likelihood, assuming an underlying Gaussian distribution for each node [17].

5. RESULTS

5.1. Ground Truth Validation of Heart-Rate

The RMSE scores between the ground truth and the obtained heart-rate for 6 individuals was 2.18bpm (better than 2.29bpm, [7]) and the correlation score was above 0.9 on the average. The Pearson's correlation between the head orientation angle and the signal was low at about 0.01.

5.2. Role of Context

The statistical significance of all the features used was determined. Table 1 lists the p-values of the features with both types of contexts and the p-values for features obtained without using context. As our experiments reveal, introducing contexts results in a significant improvement. It is noteworthy that the non-sympathetic context was more discriminatory. The p-values of the features are also comparable to the proposed anxiety indicators in [8]. Figure 2 puts forth an exemplary instance.

5.3. Anxiety Prediction

After performing feature selection, all the features had a p-value<=0.10 for context-based descriptors and <=0.30 for context-independent descriptors and only 10 features were used in both cases. Results obtained for accuracy measurement for the anxiety prediction tasks may be found in Table 2. A majority vote classifier on the test data would be 50% accurate. The BayesNet based inference presents a clear improvement over other models for classification and performs much higher than chance. Our performance is also comparable to the best reported accuracies in [8] for PTSD and depression detection, which were 72.09% and 75.00% respectively using a SVM.

6. DISCUSSION

A look at Table 1 clearly reveals the usefulness of using context-based feature descriptors. In fact, many of our context-based descriptors are statistically more significant than those reported in [8] (only 3 were below 0.05) for anxiety assessment.

	Type of Classifier	**Accuracy**	**F1 Pos/F1 Neg**
With Context	Logistic Regression	63.33%	0.560/0.685
	Naïve Bayes	66.67%	0.705/0.615
	Bayesian Network	**73.33%**	**0.750/0.714**
Without Context	Logistic Regression	60.00%	0.571/0.625
	Naïve Bayes	60.00%	0.538/0.647
	Bayesian Network	**63.33%**	**0.645/0.620**
	Baseline	50.00%	-

Table 2. The classification performances with and without contexts for various classifiers.

As far as our original research goals are concerned, we address all of them and observe the following:

R1: We find discriminative features contained in the heart-rate variability signal for discerning anxiety, as was motivated in the related prior works. This is shown in our experiments as well, where we observe that even without the knowledge of context, we perform noticeably better than chance. Moreover we also note the usefulness of capturing the conditional dependence amongst the various features using a BayesNet, as manifested by a higher accuracy.

R2: We see that knowledge of the question context, in terms of its polarity and intimacy levels adds significantly useful information for the classification task.

The computational complexity for all our automatically computed features is linear or log-linear.

7. CONCLUSION & FUTURE WORK

Our experiments reveal the importance of using heart-rate variability as a physiological parameter for measuring anxiety. Additionally, we also observe the benefit of utilizing the context information. Moreover, the fact that all our features are automatically constructed from video and using question context, allows its usability in decision-support tools for clinicians. Our immediate road ahead, is to integrate multimodal features to improve the recognition.

8. ACKNOWLEDGEMENT

The effort described here is supported by DARPA under contract W911NF-04-D-0005 and the U.S. Army. Any opinion, content or information presented does not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

9. REFERENCES

[1] Brener, J., and Hothersall, D. 1966. Heart Rate Control Under Conditions of Augmented Sensory Feedback. *Psychophysiology* (1966), 23-28.

[2] Dong, S., Boashash, B., Azemi, G., Lingwood, B.E., and Colditz, P.B. Detection of Perinatal Hypoxia Using Time-Frequency Analysis of Heart-Rate Variability Signals. *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013*.

[3] Bradley, M.M., Codispoti M., Cuthbert B.N., and Lang P.J. 2001. Emotion and Motivation I: Defensive and Apetitive Reactions in Picture Processing. *Emotion* (2001) 276-298.

[4] Hjortskov, N., Rissen, D., Blangsted, A.K.., Fallentin, N., Lundberg, U., and Sogaard, K. 2004. The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology* (2004). 92: 84-89.

[5] Gorman, J. M., and Sloan, R.P. 2000. Heart rate variability in depressive and anxiety disorders. *AM heart J*, 2000; 140:S77-83.

[6] Nashoni, E., Aizenberg, D., Sigler, M., Zalsman, G., Strasberg, B., Imbhar, S., Adler, E., and Weizman, A. 2004. Heart Rate Variability in Patients With Major Depression. *Psychosomatics* (2004); 45:129-134.

[7] Poh, M.-Z., McDuff, D.J., and Picard, R.W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express* (2010).

[8] Scherer, S., Stratou, G., Boberg, J., Gratch, J., Rizzo, A., and Morency, L.-P. Automatic Behavior Descriptors for Psychological Disorder Analysis. *10th IEEE Conference on Automatic Face and Gesture Recognition (FG 2013),* Shanghai, China.

[9] Scherer, S., Stratou, G., Gratch, J., and Morency, L.-P. Investigating Voice Quality as a Speaker-Independent Indicator of Depression and PTSD. *14th Annual Conference of the International Speech Association (InterSpeech 2013)* (Lyon, France, August 25 - 29, 2013).

[10] Gallardo-Antolin, A., and Montero, J.M. Histogram Equalization-Based Features for Speech, Music, and Song Discrimination. *IEEE Signal Processing Letters, Volume 1, No. 7, July 2010*.

[11] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[12] Laar, J.,V., Porath, M.,M., Peters, C.,H.,L., and Oei, S.,G. "Spectral analysis of fetal heart rate variability for fetal surveillance: review of the literature," *Acta Obstetricia et Gynecologica*, vol. 87, no. 3, pp. 300–306, 2008.

[13] Viola, P., and Jones, M. Rapid object detection using a boosted cascade of simple features. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2001 (CVPR 2001)* Kauai, HI, USA.

[14] OKAO Vision: *URL www.omron.com*.

[15] Morency, L.-P., Whitehill, J., and Movellan, J. Monocular Head Pose Estimation Using Generalized Adaptive View-based Appearance Model. *Image and Vision Computing* (2009) DOI: 10.1016/j.imavis.2009.08.004.

[16] Bishop, C. M. Pattern Recognition and Machine Learning, Springer (2006).

[17] Bouckaert, R., R. Bayesian Network Classifiers in Weka for Version 3-5-7. *URL www.cs.waikato.ac.nz/~remco/weka.bn.pdf*.