Decoding a Music-Modulated Cognitive Arousal State using Electrodermal Activity and Functional Near-infrared Spectroscopy Measurements

Anan Yaghmour \textit{Student Member, IEEE}, Md. Rafiul Amin \textit{Student Member, IEEE} and Rose T. Faghih \textit{Member, IEEE}

Abstract—Biofeedback systems sense different physiological activities and help with gaining self-awareness. Understanding music’s impact on the arousal state is of great importance for biofeedback stress management systems. In this study, we investigate a cognitive-stress-related arousal state modulated by different types of music. During our experiments, each subject was presented with neurological stimuli that elicit a cognitive-stress-related arousal response in a working memory experiment. Moreover, this cognitive-stress-related arousal was modulated by calming and vexing music played in the background. Electrodermal activity and functional near-infrared spectroscopy (fNIRS) measurements both contain information related to cognitive arousal and were collected in our study. By considering various fNIRS features, we selected three features based on variance, root mean square, and local fNIRS peaks as the most informative fNIRS observations in terms of cognitive arousal. The rate of neural impulse occurring under electrodermal activity (EDA) was taken as a binary observation. To retain a low computational complexity for our decoder and select the best fNIRS-based observations, two features were chosen as fNIRS-based observations at a time. A decoder based on one binary and two continuous observations was utilized to estimate the hidden cognitive-stress-related arousal state. This was done by using a Bayesian filtering approach within an expectation-maximization framework. Our results indicate that the decoded cognitive arousal modulated by vexing music was higher than calming music. Among the three fNIRS observations selected, a combination of observations based on root mean square and local fNIRS peaks resulted in the best decoded states for our experimental settings. This study serves as a proof of concept for utilizing fNIRS and EDA measurements to develop a low-dimensional decoder for tracking cognitive-stress-related arousal levels.

I. INTRODUCTION

Human emotions are convoluted structures processed and regulated by the central nervous system (CNS). Stress is of great importance to maintain a healthy and productive life. One manifestation of stress emotion is the fight or flight response that serves as a survival mechanism of a human subject in threatening situations. However, an abnormal level of stress could lead to serious mental disorders. For instance, anxiety disorders, which can be defined as a heightened level of activation in the sympathetic nervous system accompanied with intense fear and worry, are considered to be the epidemic of the 21 century due to their high prevalence [1]. Moreover, anxiety is responsible for detrimental consequences on every aspect of a patient’s life, for example, loss of jobs and relationships due to the adverse symptoms related to anxiety, like difficulty concentrating and social avoidance [2].

Stress as a psychological process cannot be directly measured, instead, it manifests itself through changes in the patterns of physiological signals such as increased blood pressure, heart rate, and sweating [1]. This idea lies in the heart of the automatic emotion recognition field of research, in which physiological signals are processed and analyzed to infer the emotional state of a participant [3]. Biofeedback systems are an example of the automatic emotion recognition principle. Biofeedback systems sense different physiological activities in real time and continuously send related information back to the user. Such systems are of great importance for users to gain self-awareness of their internal states and to learn how they can maintain their physiological activities in their healthy levels [4]. Furthermore, such systems can be utilized in remote health monitoring, smart environments, entertainment, and human-computer interface [5].

According to [4], stress-related biofeedback systems rely mainly on heart rate variability (HRV), heart rate (HR), respiratory (RSP), and electrodermal activity (EDA) measurements to estimate stress levels. However, stress-related biofeedback systems have not utilized functional near-infrared spectroscopy (fNIRS) combined with EDA measurements for cognitive-stress estimation. fNIRS is a relatively new neuroimaging technique that uses near-infrared light to capture the changes in oxygenated/deoxygenated hemoglobin levels in the outer cortex of the brain [6]. fNIRS has been utilized by Bigliassi et al. to explore the relation between music and prefrontal cortex activation [7]. Also, several mental workload studies have deployed fNIRS measurements in their analysis [8], [9], [10]. Skin conductance (SC) is a measure of EDA. Due to variations in the activity of the sweat glands as a function of neurological stimuli generated by the sympathetic nervous system (SNS), SC encodes important information about the arousal state in the form of fluctuations in the SC response (SCR) [11]. SC has been used to detect stress in real-world driving tasks [12]. Additionally, different machine learning algorithms have been utilized in the domain of arousal estimation using SC measurements, such as recurrent neural networks, regression trees, etc [5].

In this study, we take one further step by estimating the arousal state from fNIRS and EDA measurements for the...
sake of investigating how it will be affected by different types of music. We do so by exposing a participant to neurological stimuli in a working memory experiment that will elicit a cognitive-stress-related arousal response while playing a different type of music. The working memory experiment used in our investigation is called the n-back task. In this task, participants are subjected to a series of visual stimuli, then they’re required to determine whether it matches the stimulus from \( n \) trials before. Understanding music’s impact on the arousal state is of great importance for biofeedback stress management systems. Such systems could utilize music in closing the loop to control the participant’s stress level. Controlling stress levels through music can be utilized practically in numerous applications, one of which is designing a more optimized educational environment using appropriate music to maintain the student’s optimal cognitive performance through maintaining an optimal level of arousal.

To decode cognitive arousal using fNIRS and EDA measurements, we utilize a state-space model that describes the relationship between the observations and the internal arousal state similar to the approaches in [1], [5], [13]– [19]. By considering various fNIRS features, we select three features based on variance, root mean square, and local fNIRS peaks as the most informative fNIRS observations in terms of cognitive arousal. The neural impulse occurrence underlying EDA is taken as a binary observation. To retain a low computational complexity for our decoder and select the best fNIRS-based observations, two features are chosen as fNIRS-based observations at a time. Hence, we design the decoder observations to be one binary observation extracted from SC and two continuous observations extracted from fNIRS measurements. Then, the hidden arousal state is estimated using a Bayesian filtering approach within an expectation-maximization (EM) framework. Finally, the best combination of fNIRS observations for decoding cognitive arousal is selected. This study serves as a proof of concept for utilizing fNIRS and EDA measurements to develop a low-dimensional decoder for tracking cognitive-stress-related arousal levels.

II. METHODS

A. Music n-back Experiment

We have designed our version of the n-back experiment that consists of two main sessions separated by a rest period of 2 minutes. In the first session, calming music was played in the background and a vexing one for the second session. Calming, and vexing music, respectively, have been defined on an individual basis, where each participant has been asked to specify one type of music that makes him feel more relaxed (calming effect), and another type of music that makes him feel more activated (vexing effect). Each session consists of 16 trials of 1-back and 3-back tasks. The first eight trials of each session are followed by a 20 second rest period. A trial is selected to be a 1-back or 3-back task with 0.5 probability. A single trial can be composed of 5-second instructions followed by 22 of the 2-second blocks. The first part of each 2-second block is a stimulus that lasts for 0.5 seconds followed by a 1.5 seconds of fixation cross. Therefore, in total, each trial is approximately 49 seconds in length. A detailed description of the experiment is provided in [10].

B. Data Collection

EDA and fNIRS measurements have been collected from five participants: three females and two males with ages ranging from 22-24 years old. We used NIRSport 2 noninvasive sensors to measure fNIRS signals. The system uses two wavelengths of 760 and 850 nm. To measure hemodynamic activity in the prefrontal cortex and the occipital areas of the brain, we asked each participant to wear a head cap equipped with 16 sources and 14 detectors forming 44 channels in total. One EDA channel has been captured using the Biopac data acquisition system with a sampling frequency of 2 KHz.

C. Preprocessing

1) EDA Measurements: We have followed the preprocessing framework described in [20]. This framework can be described in three main steps: 1) manual discontinuity removal, 2) the application of a finite impulse response (FIR) low pass filter of order 64, with the cut-off frequency of 0.1 Hz, 3) decomposing SC into the phasic and tonic components using cvxEDA [21], where the phasic component is a high-frequency part of the SC and the tonic component is the low-frequency part.

2) fNIRS Measurements: All preprocessing steps were accomplished using Nirslab software obtained from www.nitrc.org. Those steps include:

- Truncating parts of the signal that do not correspond to the experiment time frame.
- Bandpass filtering with a low cut-off freq of 0.01 Hz and a high cut-off freq of 0.2 Hz.
- Converting raw light intensity signals to oxygenated hemoglobin (HbO), deoxygenated hemoglobin (HbR), and total hemoglobin (HbT) concentrations, respectively.

D. Observations Extraction

For the observation extraction step, one binary observation is taken from SC measurements and three continuous observations from fNIRS measurements. We should note, however, that we only use one binary and two continuous observations at a time for hidden arousal state estimation.

1) fNIRS-Related Continuous Observations: Since HbO responses in the prefrontal cortex are related to cognitive arousal, we only analyzed the HbO channels recorded from the prefrontal cortex.

We can divide the trails into four tasks: 1) Task 1: 1-back task with calming music. 2) Task 2: 3-back task with calming music. 3) Task 3: 1-back task with vexing music. 4) Task 4: 3-back task with vexing music. Based on our experiment setup, the arousal intensity levels are expected to be the lowest during Task 1 and highest during Task 4. Hence, we chose only the prefrontal channels that demonstrate distinct changes between the two extreme intensity levels of the arousal state - Task 1 and Task 4. We did so by
calculating $\ell_2$-norm for each prefrontal channel across the two extreme levels. Next, we subtracted $\ell_2$-norm of Task 1 from $\ell_2$-norm of Task 4, and we chose only the channels that have a corresponding value of greater or equal to 0.0001. Fig. 1 shows the evaluation of $\ell_2$-norm for each prefrontal channel across the two extreme intensity levels for one of the participants.

We have investigated various features based on fNIRS, such as a moving average filter, a moving variance filter, a moving skewness filter, a moving kurtosis filter, and a filter that finds the number of peaks in a moving window to select the continuous observations. Finally, we selected three fNIRS-based observations that show a significant variation between the calming music session and the vexing music session. For the first continuous observation, we fed the average of the selected channels to a moving variance filter with a window length of 44 seconds, followed by a moving average filter with window length equals 900 seconds to form the first continuous fNIRS observation signal. Panel (b) in Fig. 2 shows an example of the first observation. For the second continuous observation, we applied a moving root mean square (RMS) filter with a window length of 900 seconds. Panel (c) in Fig. 2 shows an example of the second observation. For the third continuous observation, we applied a moving local peak finder filter with a window length of 44 seconds followed by a moving average filter with a window length of 37.22 minutes applied over each window peak sample mean. Panel (d) in Fig. 2 shows an example of the third observation.

![Fig. 1: Different Channels’ $\ell_2$-norm Corresponding to Task 1 and Task 4.](image)

The blue-dashed line and the red-solid curves show the $\ell_2$-norms corresponding to Task 1 and 4 for different channels, respectively.

2) SC Related Observation: Since SCRs can be modeled as the sweat glands’ response to an SNS-generated burst of neuroelectric stimuli (a train of neuroelectric pulses), we extract the neuroelectric stimuli by the deconvolution algorithm proposed by [18]. SCRs’ occurrence rate (equivalent to the occurrence rate of the neural impulses underlying the SCRs), and amplitude are of the most common skin conductance-related measures of hidden arousal state used in the literature [5]. The deconvolution algorithm provides an estimate for neuroelectric stimuli and the parameters of the model governing sweat diffusion and evaporation. The estimation process relies on a two-step coordinate descent strategy [20] that incorporates the coupled differential equation formulation for the sweat diffusion and evaporation model described in [22]. Panel (a) in Fig. 2 shows an example of the binary observation.

E. State-Space Model

We utilize a point process state space (PPSS) framework to analyze and model the observations. We model the arousal state $x_k$ by using the following first-order auto-regressive model:

$$x_k = \rho x_{k-1} + \epsilon_k$$

(1)

where $\epsilon_k \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is noise process, and $\rho$ is a memory factor. Also, $\sigma_\epsilon^2$ and $\rho$ are to be estimated from the observations. Let $m_k$ be a binary version of the estimated neuroelectric stimuli, i.e., $m_k = 1$ when there is a neural stimulus at time $k$, and $m_k = 0$ otherwise. Since the occurrence of a neuroelectric stimulus follows a Bernoulli distribution with probability $p_k$, we related $x_k$ to $p_k$ as described in [5]

$$p_k = \frac{1}{1 + e^{-\beta + \omega_k}}$$

(2)

Let $p_0$ be the average probability that $m_k = 1$, and $x_0 = 0$. Then a participant-specific baseline parameter $\beta$ can be found by substituting $p_0$ and $x_0$ in (2) and solve for $\beta$.

For the continuous-valued observations $s_k$ and $r_k$, we relate them to $x_k$ using linear models [5]: $s_k$ and $r_k$ are related to $x_k$ as follows

$$s_k = \delta_0 + \delta_1 x_k + w_k$$

(3)

$$r_k = \gamma_0 + \gamma_1 x_k + v_k$$

(4)

where $w_k \sim \mathcal{N}(0, \sigma_w^2)$, $v_k \sim \mathcal{N}(0, \sigma_v^2)$ are the modelling error for $s_k$ and $r_k$, respectively, and $\delta_0, \delta_1, \gamma_0, \gamma_1$ are to be determined. To avoid overfitting to one of the continuous observations, we assume $w_k$ and $v_k$ are equal each other in our model.

Let the observations $M^\mathcal{K} = \{m_1, m_2, \ldots, m_K\}$, $R^\mathcal{K} = \{r_1, r_2, \ldots, r_K\}$, and $S^\mathcal{K} = \{s_1, s_2, \ldots, s_K\}$. Also, let $\mathcal{Y}^\mathcal{K} = \{M^\mathcal{K}, R^\mathcal{K}, S^\mathcal{K}\}$, $\theta = \{\rho, \delta_0, \delta_1, \gamma_0, \gamma_1, \sigma_\epsilon^2, \sigma_w^2, \sigma_v^2\}$. Then our goal is to estimate $\mathcal{X}^\mathcal{K} = \{x_1, x_2, \ldots, x_K\}$. This can be achieved through the application of Bayesian filtering within an expectation-maximization framework. In the E-step we aim to estimate $\mathcal{X}^\mathcal{K}$ conditioned on $\mathcal{Y}^\mathcal{K}$ and $\hat{\theta}$, and in the M-step we solve for $\theta$ that maximizes the log-likelihood of the complete data. The algorithm keeps alternating between the two steps until convergence.

F. E-Step

In this step, $\mathcal{X}^\mathcal{K}$ can be estimated first by applying a forward (causal) filter, and then followed by a backward (non-causal) filter to obtain smoother estimates. To obtain Kalman-like filter equations, we make a Gaussian approximation to the posterior distribution $p(x_k|y^k)$, following the exact same
procedure described in [5]. For $k = 2 : K$ we obtain the following

$$x_{k|k-1} = \rho x_{k-1}$$

$$\sigma_k^2 = \rho^2 \sigma_{k-1}^2 + \sigma_\epsilon^2$$

$$C_k = \frac{\sigma_k^2}{\sigma_\epsilon^2 + \sigma_{k-1}^2(\gamma_1^2 \sigma_m^2 + \delta_1^2 \sigma_m^2)}$$

$$x_{k|k} = x_{k|k-1} + C_k \left( m_k - p_{k|k} \right)$$

$$+ \gamma_1 \sigma_m^2 (r_k - \gamma_0 - \gamma_1 x_{k|k-1})$$

$$+ \delta_1 \sigma_m^2 (s_k - \delta_0 - \delta_1 x_{k|k-1})$$

$$\sigma_{k|k}^2 = \left[ \frac{1}{\sigma_{k|k-1}} + p_{k|k} (1 - p_{k|k}) + \frac{\gamma_1^2}{\sigma_m^2} + \frac{\delta_1^2}{\sigma_m^2} \right]^{-1}$$

Now by substituting $x_{k|k}$ in (2), we obtain

$$p_{k|k} = \frac{1}{1 + e^{-\beta + x_{k|k}}}$$

Substituting equation (10) in equation (8), results in $x_{k|k}$ appearing on both sides of the equation. Thus, Newton-Raphson numerical method should be applied to solve for $x_{k|k}$ [5].

For the backward filter, we incorporate all observations up to time $K$ to estimate the mean and the variance of the arousal state at time $k$, $x_{k|K}$, $\sigma_k^2$ respectively.

$$a_k \rho^2 \sigma_k^2$$

$$x_{k|K} = x_{k|k} + a_k (x_{k+1|K} - x_{k+1|k})$$

$$\sigma_k^2 = \sigma_{k|k}^2 + a_k (\sigma_{k+1|K}^2 - \sigma_{k+1|k}^2)$$

G. M-Step

In the M-step, we estimate $\theta$ by finding the one that maximizes the log-likelihood of the complete data. We use the same equations described in [5] for implementing the M-step. However, we change the maximization step as a constrained one to avoid any potential overfitting to one of the two continuous observations. We consider $\sigma_\nu = \sigma_w$ as the constraint on the maximization problem.

III. RESULTS AND DISCUSSION

We have estimated the hidden arousal state and the model parameters by using the EDA binary observation and every possible combinations of two fNIRS continuous observations. Fig. 3 shows the estimated hidden arousal state for each participant for every possible combinations of two continuous observations. From visual inspection, we see that there is an increase in the arousal state during the vexing music session regardless of the selected observations’ combination.

We observe from the results, that all three combinations of observations have captured the increase in the hidden arousal state.
Fig. 3: Estimated Arousal States for Five Participants. Each subfigure represents the results for a certain participant in three sub panels: (1) The hidden arousal state estimated using binary and continuous observations 1 & 2; (2) The hidden arousal state estimated using binary and continuous observations 1 & 3; (3) The hidden arousal state estimated using binary and continuous observations 2 & 3. The colored rectangles represent 1-back task with a calming music, 3-back task with a calming music, 1-back task with a vexing music, 3-back task with a vexing music, and resting period.

state during the vexing music session. This can be interpreted because usually the vexing music is composed of intense auditory stimulation with a higher level tempo that increases brain activity. In other words, vexing music helps the user to obtain a higher level of emotional arousal that reduces any potential boredom. Based on visual investigation, we see that the variation in the state estimates is very small for a given type of music. In other words, the variation in arousal from one n-back task to another n-back is not visible. This is because the blood hemodynamic response from fNIRS measurements has a lower frequency bandwidth and it does not capture the fast transitions in the level of brain activity. Therefore, while fNIRS measurements can capture the slow variation of the changes that occur over a long period of time, they do not capture fast fluctuations.

Even though the results of the three combinations look similar, the hidden arousal state estimated using observation 1 and observation 3 shows a delay in transitioning from a low arousal state to a higher state when moving from the calming music session to the vexing one. Furthermore, as illustrated in Fig. 3, the arousal state estimates in the other two combinations (i.e., observation 1-2 and observation 2-
3) look very similar. It is evident from the results in these five participants that these combinations share information and state estimates from one combination can be derived from another one. Among all combinations, the state estimate based on observations 2 and 3 is the best one for estimating the hidden arousal state.

IV. CONCLUSION

In this study, we utilize a state-space approach to combine physiological recordings (i.e., EDA) and brain hemodynamic recording (i.e., fNIRS) information to obtain one continuous arousal state. We also propose to obtain fNIRS observations that can differentiate between tasks with different arousal levels. Our estimation results also show that different types of music can influence the arousal level, where vexing music has driven the arousal state to higher levels compared to calming music.

In this study, we have shown that EDA measurements along with the fNIRS measurements can carry useful information about the hidden states of the brain, which can be applied in a wide range of applications. One potential application is in designing a wearable machine interface (WMI) that utilizes music to control stress in a non-invasive way [11].

By combining different physiological measurements from different locations in the body, we proposed a framework for a more reliable hidden arousal state estimation compared to a single physiological measurement. As this study serves as a proof of concept of the feasibility of combining EDA and fNIRS measurements for hidden arousal state estimation, in future, we will perform further experiments and utilize the current framework to investigate the benefit of using both physiological recordings such as EDA and direct brain recording over a single type of measurement.

ACKNOWLEDGMENT

The authors thank Dilranjan S. Wickramasuriya and Srinidhi Parshi for their help during data collection and dataset preparation. Anan Yaghmour would like to also thank Dilranjan S. Wickramasuriya for useful discussions related to filter design.

REFERENCES