Bayesian Inference of Hidden Cognitive Performance and Arousal States in Presence of Music

Saman Khazaei , *Graduate Student Member, IEEE*, Md Rafiul Amin , *Member, IEEE*, Maryam Tahir , and Rose T. Faghih , *Senior Member, IEEE*

Abstract—Goal: **Poor arousal management may lead to reduced cognitive performance. Specifying a model and decoder to infer the cognitive arousal and performance contributes to arousal regulation via non-invasive actuators such as music.** *Methods:* **We employ a Bayesian filtering approach within an expectation-maximization framework to track the hidden states during the n-back task in the presence of calming and exciting music. We decode the arousal and performance states from the skin conductance and behavioral signals, respectively. We derive an arousal-performance model based on the Yerkes—Dodson law. We design a performance-based arousal decoder by considering the corresponding performance and skin conductance as the observation.** *Results:* **The quantified arousal and performance are presented. The existence of Yerkes—Dodson law can be interpreted from the arousal-performance relationship. Findings display higher matrices of performance within the exciting music.** *Conclusions:* **The performance-based arousal decoder has a better agreement with the Yerkes—Dodson law. Our study can be implemented in designing non-invasive closed-loop systems.**

Index Terms—**Affective computing, biomedical signal processing, estimation, state-space methods.**

Impact Statement—**There is no dynamic approach to decode the arousal from the arousal-performance link. Our algorithm decodes the arousal using skin conductance as well as cognitive performance.**

I. INTRODUCTION

THE word cognition refers to "the mental action of acquiring knowledge and understanding through thought and experience", which emphasizes the dynamics of learning as opposed to the participant's previous knowledge [3]. Human emotional status directly impacts the cognition [4]. Particularly, arousal, which refers to an intensity level of human emotions, can determine cognitive performance in performing a cognitive task [5]. Cognitive tasks are those that require a person to mentally process new information, retrieve that information from memory, and use it at a later time [6]. The term cognitive performance describes the overall performance of the cognitive functions over a cognitive task. Human cognitive functions are diverse and can be divided into two main branches, namely, basic functions and higher-level cognitive functions. Basic cognitive functions include attention, working memory (WM), and perception, while higher-level cognitive functions consist of speech and language, decision-making, and executive control [7]. In this research, we investigate the underlying arousal and performance state during the n-back task –which requires WM usage– in the presence of music.

WM is a basic cognitive function that enables the temporary storage and manipulation of information [8]. While working memory, by itself, is a basic cognitive function, it would serve as a core component of higher cognitive function, and several cognitive tasks involve working memory usage [9]. Here, the n-back task serves as a cognitive task of interest. The n-back task mainly executes the working memory. Additionally, there is evidence of executive control and attention involvement when performing the n-back task [10]. A single block of an n-back task includes sets of letters known as the stimulus. For each stimulus, a participant is supposed to realize whether the presented stimulus matches the nth previous one or not. The higher order of "n" would result in a higher cognitive load of WM [11]. The term cognitive load implies the occupied WM resources [12].

Human emotion has been modeled using different approaches. In one of the early and well-known models, emotion has been demonstrated by Ekman and Friesen using 6 distinct

categories: happiness, sadness, surprise, fear, anger, and disgust. However, there was a lack of continuity in that paradigm [13], [14]. A more advanced model has been developed by Russell such that emotion can be recognized continuously using two orthogonal axes—valence and arousal. The term arousal denotes the intensity level of emotion associated with the sympathetic nervous system, and valence has been related to the pleasantness and unpleasantness of the emotion [5], [14]. The electrodermal activity (EDA) can be considered as an informative index of arousal [15]: The human autonomic nervous system is composed of three main branches, namely, the sympathetic nervous system, the parasympathetic nervous system, and the enteric nervous system [16]. Sympathetic nerve fibers are responsible for the innervation of sweat glands [17]. The variations in sweat secretions can be measured from the skin. Therefore, the skin conductance signal (a measure of EDA) can be applied as a metric to monitor the arousal level [18].

The arousal and working memory association can depend on multiple factors, such as underlying neurotransmitter production, valence, working memory tasks, and personal characteristics [9]. Based on the observed association between norepinephrine – a type of neurotransmitter that is positively correlated with arousal – and performance in working memory tasks, it has been hypothesized that moderate levels of arousal can improve working memory, while extreme levels of arousal may impair working memory [19]. This hypothesis complies with the Yerkes—Dodson law. The Yerkes—Dodson law – known as an inverted-U law in psychology– explains that an extremely low arousal level can lead to a lack of attention while extremely high arousal may result in a distraction in which both cases would prevent reaching the optimal cognitive performance [20]. To support this hypothesis, one may exclusively investigate the n-back task as a working memory task of interest. Specifically, to perform the n-back task, the focus of attention serves as an essential factor. The focus of attention is mainly provided by attentional capacity, which is the limited capacity system. Optimal arousal can result in high attentional capacity, while excessively high and low arousal levels lead to reductions in attentional capacity, which follows the Yerkes—Dodson law [21], [22].

The inverted-U law offers us an opportunity to regulate arousal such that it boosts cognitive performance. External non-invasive interventions such as background music can be employed to influence the mood or arousal level of an individual [23]. Particularly, the type of music can be an effective factor in cognitive performance regulation and designing non-invasive arousal actuators. Previous studies reveal that rock music may reduce productivity in performing the n-back task, while no music or listening to country or jazz music can enhance the participant's performance in the course of the n-back task [24]. Given the non-invasiveness, accessibility, daily music listening time, and advances in music streaming platforms, the idea of using music to impact cognitive states seems to be far-reaching and worthy of consideration. Hence, we employ the collected behavioral data and skin conductance signal throughout the n-back tasks in the presence of two types of music selected by the participants [25]. The music component was used to mock the low and high arousing environment that can possibly affect the performance. Hence, the participants were asked to

provide music with calming and exciting content, which are mainly different in terms of arousal rather than valence [5].

The cognitive performance and arousal states are often presented as discrete measurements, such as discrete ratings provided by either subjects or observers. The discrete measurement prevents us from continuous tracking of the arousal and performance. Using the Bayesian filtering approach within an expectation-maximization framework, we decode the continuous performance and arousal state [14], [26].

The objectives of this study can be listed as decoding the arousal and performance in the presence of music, presenting performance indices within each music session as well as task difficulty, evaluating the arousal-performance link, and developing a performance-based arousal decoder accordingly. To obtain the sparse autonomic nervous system (ANS) activation from the skin conductance, we perform a signal deconvolution [27]. To estimate the arousal state from the recovered ANS activations, we use a marked point process (MPP) filtering. To estimate the latent cognitive performance state, we employ the sequence of correct/incorrect responses and the reaction time at each trial [28], [29]. In order to estimate performance-based arousal, we consider the combination of skin conductance data and cognitive performance as the effective observation. Particularly, we utilize the arousal-related events derived from ANS activation and the continuous performance to form the observation vector. Thereafter, we decode the arousal using the proposed Bayesian decoder.

II. MATERIALS AND METHODS

A. Dataset

The experimental data used in this research was collected under the approval of the Institutional Review Board at the University of Houston, Houston, Texas, USA. The experiment was originally conducted to investigate the viability of applying music as a neurofeedback mechanism in the course of n-back experiment [25]. Several behavioral data and physiological signals were recorded from 6 novice participants during the n-back task in the presence of two sessions of calming and exciting background music. The participants were asked to select the music with calming and exciting content. The music was applied to simulate the low and high arousing environment. Hence, the selected content of music was supposed to be different in terms of arousal rather than valence. According to Russel's emotion model, while calming music is pleasant and minimally arousing, music with an exciting subject is pleasant and highly arousing [5]. The calming background music was played in the first session, and the exciting background music was presented during the second session. To avoid fatigue and minimize learning impact on behavioral measurements, equal numbers of 1-back and 3-back task blocks were randomly distributed within two sessions of calming and exciting background music, and instructions were provided at the beginning of each trial. A total number of 32 task blocks were implemented (16 task blocks within each session). Each task block was initialized with the 5 seconds instruction period, followed by 22 trials, with 0.5 seconds for displaying the letter. In addition to 0.5 seconds display time, the participant had 1.5 seconds to deliver
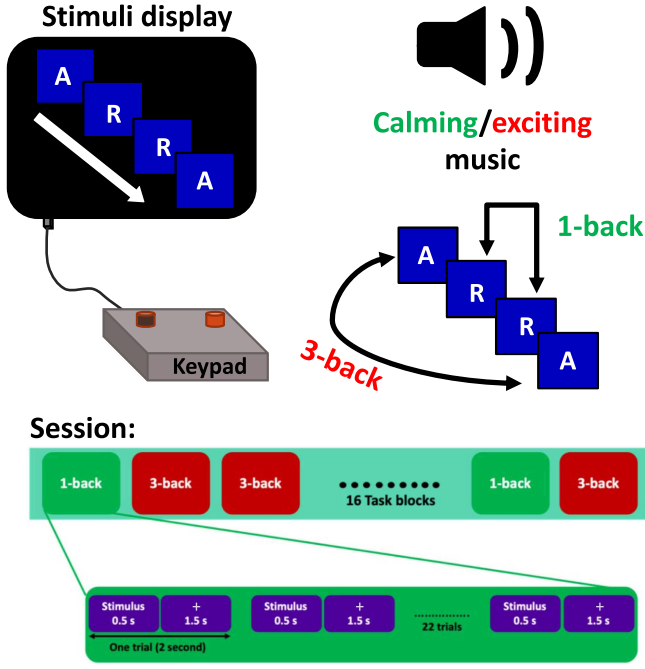
Fig. 1. Summary of the experimental setup (n-back task). The 1-back and 3-back task blocks were implemented within two sessions in the presence of calming and exciting background music.

the response and press a Chronos Keypad button to determine whether the presented letter was the same as the the nth previous letter (pressing the target button) or it was mismatched (pressing non-target button). Therefore, the total stimulus time at each task block was 49 seconds. In total, a participant performed 704 trials (i.e., 2 sessions × 16 task blocks × 22 trials). At the end of each task, a 10 seconds relaxation segment was contrived. After 8 blocks (halfway mark for each session), a 20 seconds relaxation section was implemented, and between the sessions, there was a 2-minute relaxation break. During the data collection process, the participants were asked to engage in the task and avoid unnecessary movements. The only required movement was related to pressing either the target button or the non-target button on the Chronos Keypad. The recorded behavioral data comprised of reaction time and sequence of correct/incorrect responses. A summary of the employed cognitive task is presented in Fig. 1.

The collected physiological data included functional near-infrared spectroscopy (fNIRS), electrocardiogram (ECG), respiration, skin surface temperature, electrodermal activity (EDA), photoplethysmography (PPG), electromyogram (EMG), and facial expression. In this study, we use the recorded behavioral data as well as EDA measurements. The electrodermal activity (EDA) recordings used in this study have been collected from have been collected from the MP160 BioPac system with a sampling frequency of 2 KHz. The system sensors were placed over the digitus quartus manus and digitus medius manus of the participant's non-dominant hand. The experiment is explained comprehensively in the supplementary information and in [25].

B. Inference of Brain Activation From Skin Conductance Measurements

In order to infer the neural impulse train from the raw skin conductance signal, an appropriate deconvolution method needs to be employed. Applying a coordinate descent approach we recover the sparse arousal events due to ANS activation [30], [31]. A detailed description of the approach can be found in the supplementary materials.

C. A Marked Point Process State-Space Model for Arousal

Similar to [14], we assume a random walk model for the hidden arousal state $\hat{x}_j$ such that

$$\hat{x}_j = \hat{x}_{j-1} + \epsilon_j, \tag{1}$$

where $\epsilon_j \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is the process noise and j stands for the time index. Following the marked point process filtering approach in [14], we consider Bernoulli distribution for the arousal events n_j, with probability mass function $a_j^{n_j}(1 - a_j)^{1-n_j}$ such that $P(n_j = 1) = a_j$. We relate $\hat{x}_j$ to a_j by applying a sigmoid transform similar to [26]. Thus,

$$a_j = \frac{1}{1 + e^{-(\hat{x}_j + \beta)}}, \tag{2}$$

where β is a constant that can be derived from $\beta \approx \log(\frac{a_0}{1-a_0})$ and a_0 is the average probability of observing an impulse during the experiment. Similar to [14], continuous-valued amplitude r_j of each neural impulse may be represented as

$$r_j = \hat{\gamma}_0 + \hat{\gamma}_1 \hat{x}_j + v_j, \tag{3}$$

where r_j is the amplitude of the observed impulse, $v_j \sim \mathcal{N}(0, \sigma_v^2)$ presents the sensor noise, $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are the unknown parameters to be determined. Consequently, the joint density function for the observed neural stimuli is

$$p(n_j \cap r_j | \hat{x}_j) = \begin{cases} 1 - a_j & \text{if } n_j = 0 \\ a_j \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{\frac{-(r_j - \hat{\gamma}_0 - \hat{\gamma}_1 \hat{x}_j)^2}{2\sigma_v^2}} & \text{if } n_j = 1 \end{cases}. \tag{4}$$

The derivation of a marked point process state-space decoder for arousal is described in the supplementary materials.

D. A State-Space Model for Performance

Inspired by the proposed state-space model in [28], we consider an autoregressive model for the cognitive performance state.

$$z_k = \rho z_{k-1} + w_k, \tag{5}$$

where z_k is a hidden performance state, $w_k \sim \mathcal{N}(0, \sigma_w^2)$ stands for the process noise, ρ is the unknown coefficient, and k is the trial number during the experiment.

Similar to [28], we form the observation model by specifying one binary observation (correct/incorrect response at k^{th} trial) and one continuous observation (reaction time of the corresponding response). The Bernoulli probability model is assumed for the binary responses with the probability mass function of

TABLE I
AROUSAL-PERFORMANCE REGRESSION MODEL RESULTS

p-value			
Participant	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$
1	3.14e-05	0.0197	0.0048
2	4.13e-09	2.04e-25	5.71e-06
3	1.20e-07	0.10985	0.0001
4	5.89e-45	0.0594	8.34e-25
5	3.53e-18	0.0110	4.13e-08
6	2.91e-18	0.0455	1.7e-10

$p_k^{m_k}(1-p_k)^{1-m_k}$. Applying sigmoid transform we may express the p_k in terms of z_k such that

$$p_k = \frac{1}{1 + e^{-(z_k+\mu)}}. \quad (6)$$

The constant term μ can be determined from $\mu \approx \log(\frac{p_0}{1-p_0})$ where p_0 is the average probability of having a correct response.

The reaction time τ_k can be related to the performance state using:

$$l_k = \log(\tau_k) = \alpha_0 + \alpha_1 z_k + \delta_k, \quad (7)$$

where $\delta_k \sim \mathcal{N}(0, \sigma_\delta^2)$, and l_k is the log of reaction time at each trial.

The performance state decoder's equations can be found in the supplementary materials.

E. Performance as a Function of Arousal

By utilizing both decoded arousal ($\hat{x}_j$) and performance states (z_k), we define an arousal-performance function inspired by the inverted-U law [1], [20]:

$$Y_k = \hat{\lambda}_1 X_k^2 + \hat{\lambda}_2 X_k + \hat{\lambda}_3 + e_k, \quad (8)$$

where Y_k presents the standard score of the performance state at each n-back trial and X_k stands for the standard score of the average arousal—derived from marked point process filter—at each n-back trial. Thus, the observed data points consist of (X,Y); e_k is assumed to follow a white noise structure, $e_k \sim \mathcal{N}(0, \sigma_e^2)$ and, $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\lambda}_3$ are the unknown parameters that can be determined by robust fitting with bisquare weighting. It is not advised to use the ordinary least-squares method since the data points here consist of different layers from multiple different trials. Instead, a robust fitting with bisquare weights can be employed using a MATLAB function (*fitlm*).

F. A Performance-Based Arousal State-Space Model

According to Table I, the p-values for the linear term $\hat{\lambda}_2$ are high for most of the participants which display that the statistical significance of $\hat{\lambda}_2$ is considerably low compared to $\hat{\lambda}_1$ and $\hat{\lambda}_3$. Based on the inverted quadratic relationship between arousal and performance and by ignoring the linear term $\hat{\lambda}_2$, we may modify the arousal state observation model based on the performance. Considering the arousal level at each trial, the state follows the

previous random walk model such that

$$\tilde{x}_k = \tilde{x}_{k-1} + u_k, \quad (9)$$

where $\tilde{x}_k$ stands for the performance-based arousal at k^{th} trial and $u_k \sim \mathcal{N}(0, \sigma_u^2)$ is the process noise. Similar to the proposed marked point process approach, we specify a Bernoulli distribution for $\tilde{n}_k$ at each trial where $\tilde{n}_k$ stands for the arousal events at each trials (average of neural impulses over each trial). Note that each trial takes 2 seconds and we might have more than one impulse at each trial; however, it does not affect our arousal events' vector $\tilde{N} = \{\tilde{n}_1, \tilde{n}_2, \ldots, \tilde{n}_k\}$ since the hidden state is defined based on trials. Hence, $\tilde{n}_k$ only takes 0 or 1 to indicate the arousal events in particular period. We relate the state $\tilde{x}_k$ to the probability of arousal events occurrence ϕ at trial k as before

$$\phi_k = \frac{1}{1 + e^{-(\tilde{x}_k+q_0)}}, \quad (10)$$

where q_0 is a constant that can be derived from $q_0 \approx \log(\frac{\phi_0}{1-\phi_0})$ and ϕ_0 is the average probability of having the arousal event.

Additionally, the observation model contains the continuous value of performance state z_k and, continuous-valued arousal event's amplitude $\tilde{r}_k$. Hence,

$$z_k = \tilde{\lambda}_1 \tilde{x}_k^2 + \tilde{\lambda}_0 + \psi_k, \quad (11)$$

where $\tilde{\lambda}_1$ and $\tilde{\lambda}_0$ are the unknown parameters, and

$$\tilde{r}_k = \tilde{\gamma}_0 + \tilde{\gamma}_1 \tilde{x}_k + \zeta_k, \quad (12)$$

where, $\tilde{r}_k$ stands for the average value of neural impulses at each trial. Also, $\tilde{\gamma}_0$ and, $\tilde{\gamma}_1$ are the unknown parameters. Similarly, $\psi_k \sim \mathcal{N}(0, \sigma_\psi^2)$ and $\zeta_k \sim \mathcal{N}(0, \sigma_\zeta^2)$ are assumed to be Gaussian. The joint density function for the occurred arousal event is

$$p(\tilde{n}_k \cap \tilde{r}_k | \tilde{x}_k) = \begin{cases} 1 - \phi_k & \text{if } \tilde{n}_k = 0 \\ \phi_k \frac{1}{\sqrt{2\pi\sigma_\zeta^2}} e^{\frac{-(\tilde{r}_k - \tilde{\gamma}_0 - \tilde{\gamma}_1 \tilde{x}_k)^2}{2\sigma_\zeta^2}} & \text{if } \tilde{n}_k = 1 \end{cases}. \quad (13)$$

G. A Performance-Based Arousal Decoder

We derive a decoder based on the performance-based arousal model. The unknown parameters vector would be $\tilde{\theta} = \{\sigma_u^2, \tilde{\lambda}_1, \tilde{\lambda}_0, \sigma_\psi^2, \tilde{\gamma}_0, \tilde{\gamma}_1, \sigma_\zeta^2\}$, and we apply the EM algorithm to jointly estimate $\tilde{x}_k$ and $\tilde{\theta}$.

1) E-Step: The E-step consists two subsections namely, forward filtering and backward smoothing. Based upon the observation $\tilde{R}^K = \{(\tilde{n}_1, \tilde{r}_1, z_1), \ldots, (\tilde{n}_k, \tilde{r}_k, z_k)\}$ up to time K, E-step equations can be formulated.
Predict:

$$\tilde{x}_{k|k-1} = \tilde{x}_{k-1|k-1}, \quad (14)$$

$$\tilde{\sigma}_{k|k-1}^2 = \tilde{\sigma}_{k-1|k-1}^2 + \tilde{\sigma}_u^2, \quad (15)$$

Update:

if $\tilde{n}_k = 0$

$$\tilde{x}_{k|k} = \tilde{\sigma}^2_{k|k-1} \left[(\tilde{n}_k - \phi_{k|k}) \right. \tag{16}$$

$$\left. + \frac{(2\tilde{\lambda}_1 \tilde{x}_{k|k})(z_k - \tilde{\lambda}_0 - \tilde{\lambda}_1 \tilde{x}^2_{k|k})}{\sigma^2_\psi} \right] + \tilde{x}_{k|k-1},$$

$$\tilde{\sigma}^2_{k|k} = \left[\phi_{k|k}(1 - \phi_{k|k}) + \frac{1}{\tilde{\sigma}^2_{k|k-1}} \right.$$

$$\left. - \frac{1}{\sigma^2_\psi} \left[2\tilde{\lambda}_1(z_k - \tilde{\lambda}_0 - \tilde{\lambda}_1 \tilde{x}^2_{k|k}) - (2\tilde{\lambda}_1 \tilde{x}_{k|k})^2 \right] \right]^{-1}. \tag{17}$$

if $\tilde{n}_k = 1$

$$\tilde{x}_{k|k} = \tilde{\sigma}^2_{k|k-1} \left[(\tilde{n}_k - \phi_{k|k}) + \frac{\tilde{\gamma}_1(\tilde{r}_k - \tilde{\gamma}_0 - \tilde{\gamma}_1 \tilde{x}_{k|k})}{\sigma^2_\zeta} \right. \tag{18}$$

$$\left. + \frac{(2\tilde{\lambda}_1 \tilde{x}_{k|k})(z_k - \tilde{\lambda}_0 - \tilde{\lambda}_1 \tilde{x}^2_{k|k})}{\sigma^2_\psi} \right] + \tilde{x}_{k|k-1},$$

$$\tilde{\sigma}^2_{k|k} = \left[\phi_{k|k}(1 - \phi_{k|k}) + \frac{1}{\tilde{\sigma}^2_{k|k-1}} + \frac{\tilde{\gamma}^2_1}{\sigma^2_\zeta} \right.$$

$$\left. - \frac{1}{\sigma^2_\psi} \left[2\tilde{\lambda}_1(z_k - \tilde{\lambda}_0 - \tilde{\lambda}_1 \tilde{x}^2_{k|k}) - (2\tilde{\lambda}_1 \tilde{x}_{k|k})^2 \right] \right]^{-1}. \tag{19}$$

In order to solve for $\tilde{x}_{k|k}$, we have utilized a MATLAB function called *fzero* which solves for the roots of non-linear equations.

In order to acquire the smoothed state $\tilde{x}_{k|K}$ and smoothed variance $\tilde{\sigma}^2_{k|K}$, we reverse the direction:

$$\tilde{A}_k = \frac{\tilde{\sigma}^2_{k|K}}{\tilde{\sigma}^2_{k+1|k}}, \tag{20}$$

$$\tilde{x}_{k|K} = \tilde{x}_{k|k} + \tilde{A}_k(\tilde{x}_{k+1|K} - \tilde{x}_{k+1|k}), \tag{21}$$

$$\tilde{\sigma}^2_{k|K} = \tilde{\sigma}^2_{k|k} + \tilde{A}^2_k(\tilde{\sigma}^2_{k+1|K} - \tilde{\sigma}^2_{k+1|k}). \tag{22}$$

By utilizing the proposed approach in [14] and [26], we derive the expected values of $\tilde{x}^2_k$, and $\tilde{x}_k \tilde{x}_{k-1}$ as

$$\mathbb{E}[\tilde{x}^2_k] = \tilde{x}^2_{k|K} + \tilde{\sigma}^2_{k|K}, \tag{23}$$

$$\mathbb{E}[\tilde{x}_{k+1} \tilde{x}_k] = \tilde{x}_{k+1|K} \tilde{x}_{k|K} + \tilde{A}_k \tilde{\sigma}^2_{k+1|K}. \tag{24}$$

Since we consider a quadratic function to relate the arousal to performance, we first derive an expression for $\mathbb{E}[\tilde{x}^4_k]$.

$$\mathbb{E}[\tilde{x}^4_k] = \mathbb{E}[(\tilde{x}^2_k)^2]$$

$$= \left(\mathbb{E}[\tilde{x}^2_k]\right)^2 + var(\tilde{x}^2_k). \tag{25}$$

For approximating the last term ($var(\tilde{x}^2_k)$) in (25), similar to [32], we employ the second order Taylor series such that

$$var(\tilde{x}^2_k) \approx 4 \left(\mathbb{E}[(\tilde{x}_k)]\right)^2 \tilde{\sigma}^2_{k|k} + 2 \left(\tilde{\sigma}^2_{k|k}\right)^2. \tag{26}$$

Therefore,

$$\mathbb{E}[\tilde{x}^4_k] \approx \left(\mathbb{E}[\tilde{x}^2_k]\right)^2 + 4 \left(\mathbb{E}[(\tilde{x}_k)]\right)^2 \tilde{\sigma}^2_{k|k} + 2 \left(\tilde{\sigma}^2_{k|k}\right)^2. \tag{27}$$

2) M-Step: We denote the location of the trials where arousal events occurred at $\tilde{K} = \{k | \tilde{n}_k = 1\}$. Based on the E-step results, we can form a log-likelihood function Q_3, and find the unknown parameters $\tilde{\theta} = \{\sigma^2_u, \tilde{\lambda}_1, \tilde{\lambda}_0, \sigma^2_\psi, \tilde{\gamma}_0, \tilde{\gamma}_1, \sigma^2_\zeta\}$ such that $\tilde{\theta}$ maximizes Q_3.

$$Q_3 = \sum_{k=1}^{K} \mathbb{E}[\tilde{n}_k(q_0 + \tilde{x}_k) - \log(1 + e^{q_0 + \tilde{x}_k})]$$

$$+ \frac{-\tilde{K}}{2} \log(2\pi\sigma^2_\zeta) - \sum_{k \in \tilde{K}} \frac{\mathbb{E}\left[(\tilde{r}_k - \tilde{\gamma}_0 - \tilde{\gamma}_1 \tilde{x}_k)^2\right]}{2\sigma^2_\zeta}$$

$$+ \frac{-K}{2} \log(2\pi\sigma^2_\psi) - \sum_{k=1}^{K} \frac{\mathbb{E}\left[(z_k - \tilde{\lambda}_0 - \tilde{\lambda}_1 \tilde{x}^2_k)^2\right]}{2\sigma^2_\psi}$$

$$+ \frac{-K}{2} \log(2\pi\sigma^2_u) - \sum_{k=1}^{K} \frac{\mathbb{E}\left[(\tilde{x}_k - \tilde{x}_k)^2\right]}{2\sigma^2_u}. \tag{28}$$

The algorithm iterates between the E-step and the M-step until the convergence.

III. RESULTS

The collected behavioral data at each session and type of n-back task can be found in Fig. 2. Mainly, the number of correct responses and the average reaction times for all participants can be found in Fig. 2. The blue bars correspond to a calming session, and the red bars correspond to an exciting session. The dark intense bars indicate the 3-back data, and the brighter ones stand for the 1-back task.

Fig. 3 presents the distribution of the performance state while different types of music were presented.

The estimated arousal state based on the MPP type observation for all participants are available in the supplementary information. Furthermore, we depict the distribution of average MPP-based arousal state within trials with respect to each tasks difficulty (supplementary information). We find the point-biserial correlation coefficients between the task difficulty and arousal by considering the task difficulty as dichotomous variable and the average arousal state within the trials as the continuous variable. In turn, the point-biserial correlation coefficients for participants 1 to 6 are 0.0603, -0.0268, 0.0166, 0.0150, -0.0206, and 0.0405, respectively.

In Fig. 4, we investigate the link between the estimated arousal from the MPP filter and the estimated performance. The blue and red points are associated with the 1-back and 3-back tasks,
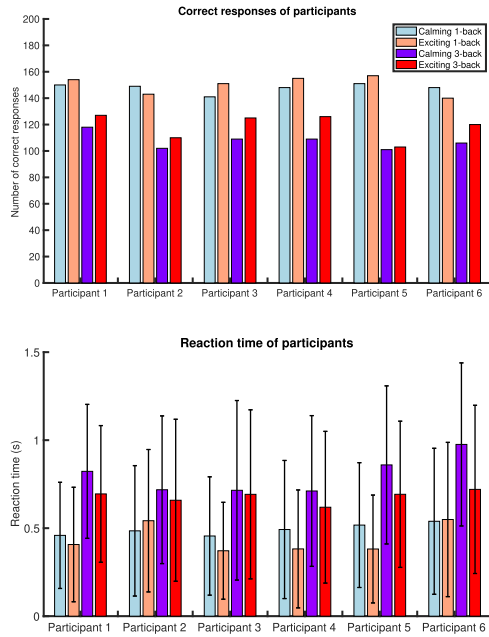
Fig. 2. Number of correct responses and reaction time with respect to the type music and task for all participants. The top sub-panel presents the number of correct responses with respect to each session and n-back task block. The bottom sub-panel displays the average reaction times (the bars), and the error bars show the data within the first and third quartiles with respect to each session and n-back task block. The blue and red colors present the calming and exciting sessions, respectively. The darker colors with more intensity stand for the 3-back task blocks and the brighter ones present the 1-back task blocks.
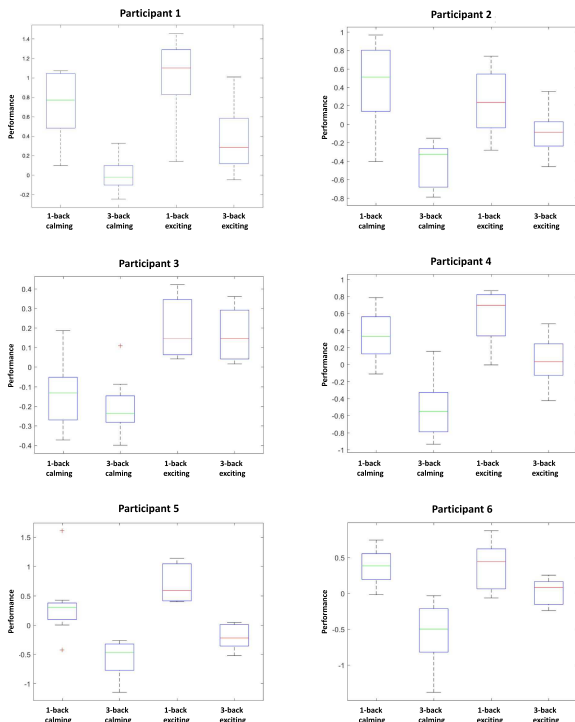

Fig. 3. Distribution of performance state within different tasks and types of music. Each sub-figure shows the box plot of the performance state.
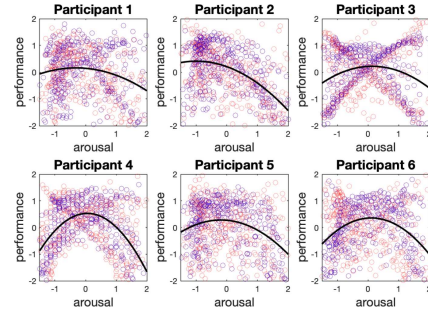

Fig. 4. Arousal-performance diagram within the whole experiment. The x-axis represents the standard score of the estimated arousal derived from the marked point process estimator, and the y-axis stands for the standard score of the performance states. The red and blue data points show the observed pairs of arousal and performance withing the 3-back trials and 1-back trials, respectively. The black curve presents the fitted model.

respectively. The data points can form an inverted-U shape for all participants with different coefficients. The p-values of the model parameters are provided in Table I. We can observe that the p-values for $\hat{\lambda}_1$ are statistically significant for all participants, while the coefficient $\hat{\lambda}_2$ is not statistically significant.

We represent the performance-based arousal estimator results for participants 1 and 6 in Fig. 5. The first two subplots at each column present the observed performance and the average of arousal events at each trial, which together comprise the decoder observation. The third and fourth subplots show the estimated arousal state and the probability of observing arousal events, respectively. To study the personalized trajectories of arousal, we formulate a high arousal index (HAI) to generalize the estimated arousal level of participants ($\tilde{x}_k$) [14]. The HAI can be calculated from $p(\tilde{x}_k > x_{\text{threshold}})$, where the threshold has been set to the median of the state values. It indicates the probability that a binary event occurs more than just by chance over the experiment. The performance-based arousal state estimation results for other participants can be found in the supplementary materials.

In Fig. 6, we compare the arousal-performance link derived from the performance-based arousal state and the estimated arousal from the MPP decoder.

According to the proposed models, we simulate sets of internal states (Fig. 7). The subplots of Fig. 7 present simulated arousal events with their amplitudes and reconstructed $\tilde{r}_k$, simulated performance state and reconstructed one, simulated arousal state (ground truth) and estimated one, simulated probability of observing an arousal event and estimated one, the quantile-quantile (QQ) plot of arousal state residual error, and the arousal-performance link, in turns. The R-squared value for the estimated arousal state is 0.8541.

IV. DISCUSSION

In most of the affective studies, the absence of ground truth resulted in implicit evaluation and validation of estimated arousal. An instance of such implicit evaluation would be using experiment information such as presented emotional stimuli to evaluate
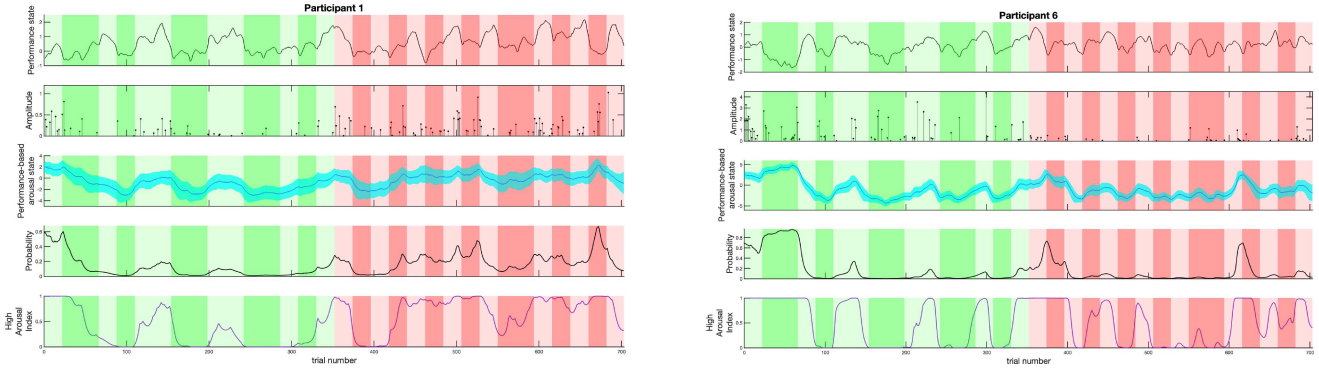
Fig. 5. Arousal state estimation from performance-based decoder for two participants. The sub-panels of the figure at each column depict: The performance state signal (z_k); The average of the deconvolved neural impulses during trials –arousal events– ($\tilde{r}_k$); The estimated state ($\tilde{x}_k$) and its 95% confidence limits; The probability of impulse occurrence (ϕ_k); The high arousal index. The background colors in each sub-panel depict: The 1-back task during the calming session (light green); The 3-back task during the calming session (dark green); The 1-back task during the exciting session (light red); The 3-back task during the exciting session (dark red).
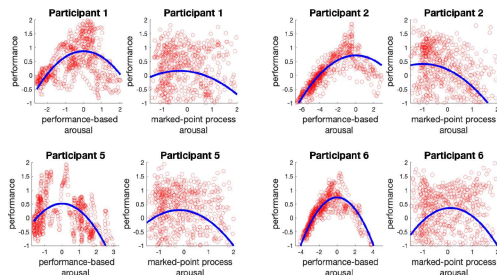

Fig. 6. Arousal-performance diagram comparison. In each pair of arousal-performance diagrams, the left sub-panel displays the arousal-performance data points constructed from the performance-based arousal estimator; The right subplot demonstrates the standard score of data points derived from the marked point process arousal estimator and performance state decoder. The blue curves show the fitted model for the corresponding data points (red).

the decoder's outcome [33]. In this study, the experiment was designed with two types of task blocks (i.e., 1-back and 3-back tasks) and two different background music. We evaluate and discuss the decoded arousal and performance state with respect to each session and each type of task. Particularly, since the high cognitive load can reduce cognitive performance, we would expect to see lower performance levels in 1-back trials compared to 3-back trials in the decoded performance [34]. Furthermore, due to the exciting content of the selected music in the second session, the participant may experience a higher level of arousal within the second session compared to the first session, and the decoded arousal state may be evaluated accordingly [25], [35], [36].

The collected behavioral data that is used as an observation to decode the performance is presented in Fig. 2; the decoded performance is demonstrated in Fig. 3. Apart from Participant 2, the higher number of correct responses, as well as the faster reaction time, can be seen within the second session. These higher performance metrics in the second session are aligned with the decoded performance state. Also, the low performance during the 3-back and high performance during the 1-back task blocks are noticeable in both the decoded performance

and the recorded behavioral data. Overall, there is a decent agreement between the estimated performance state and the observed performance metrics. Considering the variation of performance metrics within the sessions, we may notice that the performance variation from calming to exciting in 3-back task blocks is higher than the 1-back ones. One possible explanation for the observed behavior is that participants already perform well enough within the 1-back tasks to the point that no considerable difference can be made by changing the condition [37]. This is aligned with the findings in [38], where participants performed near ceiling level at 1-back regardless of the applied training.

In general, our findings present higher performance matrices associated with the exciting session for five out of six participants. One may argue that the improved performance is an indication of the arousal establishment within the desired range using music. However, other factors such as, the learning effect, the nature of the task, and the participant's baseline can be involved which hinder us from drawing any definite conclusion. Particularly, it should be noted that the exciting session was implemented as the second session; it is possible that the participant outperformed in the second session due to learning the task [39]. Hence, while it might be viable to impact the performance via personalized music, further studies are needed to make any solid judgment on the impact of music on performance. Specifically, including a control group in this context can provide a better insight into the impact of music as well as the presence of learning.

It is worth highlighting participant 2 as the only participant that presents a lower performance within the second session. This may indicate that the exciting component of music does not improve this person's performance. Perhaps the exciting music makes the person excessively aroused while the person's ideal arousal level is located within the lower arousal range. Another interpretation from the observed trend would be the absence of plausible learning. Looking into the arousal-performance link for participant 2, Table I and Fig. 4 reveal a relatively strong linear component in the arousal-performance relationship. Commonly, the linear arousal-performance link can be expected
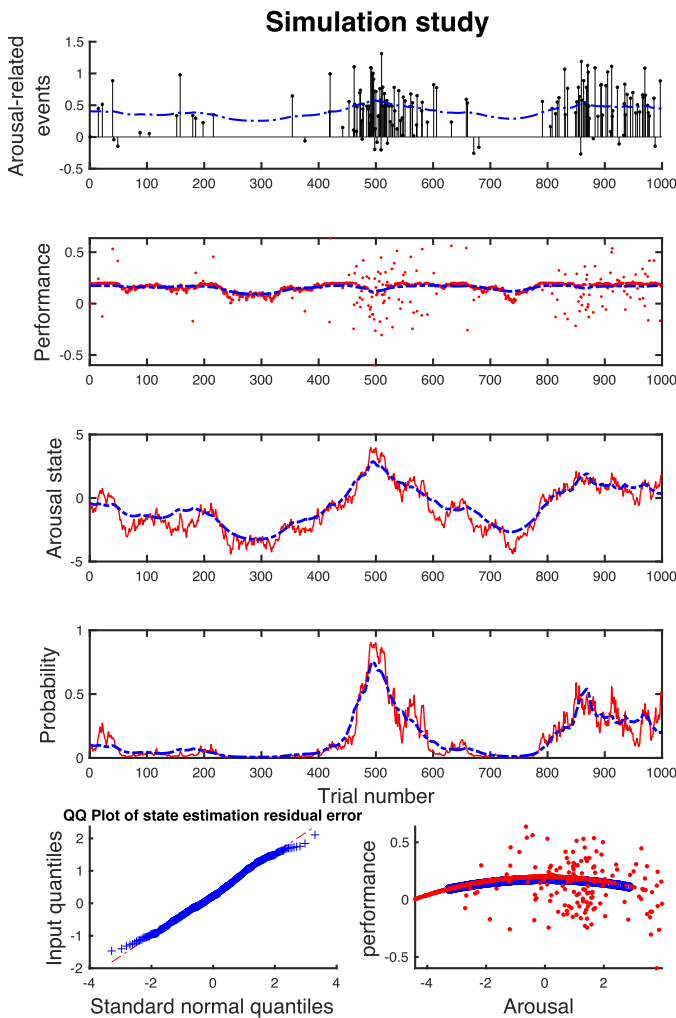
Fig. 7. Performance-based arousal state estimation on simulated data. The sub-panels of the figure depict, in turns: Simulated arousal events with their amplitudes (black impulses) and reconstructed $\tilde{r}_k$ (blue); Simulated performance state (red dots) and reconstructed one (blue); Simulated arousal state as ground truth (red) and estimated state (blue); Simulated probability of observing an arousal event (red) and estimated one (blue); The quantile-quantile (QQ) plot of arousal state residual error; The arousal-performance link such that the simulated data points are shown by red dots and the estimated ones are presented by blue dots.

when a person does not experience high enough arousal [40]. However, for participant 2, we have a different scenario in which the presence of high performance in the low arousal range and the absence of high performance in the high arousal range region is seen. A linear model can describe the arousal-performance link in such cases where either the rise or decay of the performance is presented solely. Nevertheless, if the rise or decay occurs exponentially, the exponential model also can be a good fit.

In this research, to avoid the potential reduced statistical power [41], we consider all the data collected during the experiment to perform the regression analysis and identify the presence of Yerkes-Dodson law. One may interpret the presence of Yerkes-Dodson law from the presented analysis. One crucial point that needs to be addressed is the extent to which the task difficulty and distribution impact the underlying arousal and, subsequently, the observed arousal-performance link. The point-biserial correlation coefficients do not reveal a significant association between the n-back task difficulty and the arousal, which agrees with the findings in [42]. However, we should keep in mind that in this research, the arousal state is derived from the skin conductance signal, and employing other physiological signals as an arousal index may produce a different outcome, in which task difficulty plays a confounding role [42], [43].

The participants were asked to provide calming and exciting music with no quantification or rating of the elicited emotion. In particular, the applied music was meant to have a person-specific impact rather than equal emotion elicitation across the whole sample size. Hence, we should be cautious in generalizing the findings. Instead, we may investigate the results with an individualized viewpoint. Thus, we use the participant's baseline to present person-specific metrics of arousal and performance (i.e., HAI and HPI), which can be found in the supplementary information.

While the applied personalized music mimics the personalized closed-loop architecture, it can induce the impact of familiarity on arousal and performance. One possible way to reduce the effect of music familiarity and preserve the personalized nature of the music intervention is to employ new generative deep-learning models and produce new music based on the participant's preferences person [44].

Using the performance-based arousal decoder, we can obtain the arousal level corresponding to the performance of each trial (Fig. 5). The performance-based arousal decoder can benefit from having a performance as one of the observations. Specifically, for participant 6, in spite of the few observed ANS activations during the exciting session, the higher baseline of performance in the exciting session would prevent an excessive drop of arousal. As it can be seen in Fig. 6, the arousal-performance links derived from the performance-based model (Fig. 5) tend to maintain the inverted quadratic shape and follow the inverted-U law. Additionally, since we are using both behavioral data and skin conductance signal for the performance-based arousal decoder, the results would be less affected by the possible artifacts from the skin conductance signal recording, solely.

The simulation study illustrates a decent performance of the proposed decoder. Specifically, the R-squared value and QQ plot display an agreement between the decoded state trajectory and the ground truth. While the Bayesian state-space approach is a powerful estimation tool, there are cases that might suffer from the overfitting issue [33]. This simulation study (first two subplots of Fig. 7) demonstrates that the estimates do not overfit to either the provided MPP observation or the continuous observation.

V. CONCLUSION

In the studied n-back experiment, two types of music were presented to investigate the potential of developing a safe neurofeedback via music. Using a state-space modeling approach, we

decode cognitive arousal and performance states of participants. To obtain a better insight into the arousal and performance relationship, we evaluate the plausible Yerkes—Dodson law via regression analysis. The existence of Yerkes—Dodson law would be one possible interpretation from the observed results while the small sample size and a lack of a control group would hinder us from establishing a definitive conclusion.

Our study would shed light on the primary idea of enhancing cognitive performance and shifting one's arousal using music. It might be feasible to impact the arousal and performance via music [45]. However, it should be highlighted that several factors, such as the learning effect, the nature of the task, the participant's baseline, and the type of applied music, can impact the outcome. Hence, a more comprehensive experiment with a larger sample size, control group, shuffled cognitive tasks, and various types of music would be helpful for having a settled resolution on the music's effect on arousal and performance.

We design a performance-based arousal decoder that estimates the arousal level of individuals based on their performance. This type of decoder conforms to the Yerkes—Dodson law. The ultimate goal of this performance-based arousal decoder is to be implemented within safe closed-loop systems, and the proposed decoder can be further investigated in different behavioral experiments [46], [47], [48]. In the future, we aim to test decoders in different experiments and quantify the arousal and performance in various environments. Also, given the developed performance-based arousal decoder, informative signals such as pupil size can be used in parallel with the skin conductance signal to decode the hidden arousal and evaluate the performance of the decoder [49].

Supplementary Materials: Additional figures, details of the methods, and further discussion are provided in supplementary materials.

Conflict of Interest Statement: RTF and MRA are co-inventors of a patent application filed by the University of Houston reltaed to this research.

Author Contributions: Conceptualization: RTF; Formal analysis: SK; Funding acquisition: RTF; Investigation: SK, MRA, MT, and RTF; Methodology: SK, MRA, and RTF; Validation: SK; Visualization: SK, MRA, and MT; Writing — original draft: SK; Writing — review and editing: SK, MRA, MT, and RTF.

REFERENCES

[1] S. Khazaei, M. R. Amin, and R. T. Faghih, "Decoding a neurofeedback-modulated cognitive arousal state to investigateperformance regulation by the Yerkes-Dodson law," in *Proc. IEEE 43rd Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2021, pp. 6551–6557.

[2] M. R. Amin, M. Tahir, and R. T. Faghih, "A state-space investigation of impact of music on cognitive performance during a working memory experiment," in *Proc. IEEE 43rd Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2021, pp. 757–762.

[3] N. J. Cooke, "Varieties of knowledge elicitation techniques," *Int. J. Hum.-Comput. Stud.*, vol. 41, no. 6, pp. 801–849, 1994.

[4] T. Brosch, K. Scherer, D. Grandjean, and D. Sander, "The impact of emotion on perception, attention, memory, and decision-making," *Swiss Med. Weekly*, vol. 143, no. 1920, 2013, Art. no. w13786.

[5] J. A. Russell, "A circumplex model of affect," *J. Pers. Social Psychol.*, vol. 39, no. 6, 1980, Art. no. 1161.

[6] N. M. Seel, *Encyclopedia of the Sciences of Learning*. Berlin, Germany: Springer, 2011.

[7] E. L. Glisky, "Changes in cognitive function in human aging," in *Brain Aging*. Boca Raton, FL, USA: CRC Press, 2007, pp. 3–20.

[8] A. Baddeley, "Working memory," *Science*, vol. 255, no. 5044, pp. 556–559, 1992.

[9] T.-Y. Hou and W.-P. Cai, "What emotion dimensions can affect working memory performance in healthy adults? A review," *World J. Clin. Cases*, vol. 10, no. 2, pp. 401–411, 2022.

[10] M. Mather, K. J. Mitchell, C. L. Raye, D. L. Novak, E. J. Greene, and M. K. Johnson, "Emotional arousal can impair feature binding in working memory," *J. Cogn. Neurosci.*, vol. 18, no. 4, pp. 614–625, Apr. 2006.

[11] X. Li, Z. Ouyang, and Y.-J. Luo, "The effect of cognitive load on interaction pattern of emotion and working memory: An ERP study," in *Proc. IEEE 9th Int. Conf. Cogn. Informat.*, 2010, pp. 61–67.

[12] J. Jordan, J. Wagner, D. E. Manthey, M. Wolff, S. Santen, and S. J. Cico, "Optimizing lectures from a cognitive load perspective," *AEM Educ. Training*, vol. 4, no. 3, pp. 306–312, 2020.

[13] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Pers. Social Psychol.*, vol. 17, no. 2, 1971, Art. no. 124.

[14] D. S. Wickramasuriya and R. T. Faghih, "A marked point process filtering approach for tracking sympathetic arousal from skin conductance," *IEEE Access*, vol. 8, pp. 68499–68513, 2020.

[15] H. F. Azgomi and R. T. Faghih, "Enhancement of closed-loop cognitive stress regulation using supervised control architectures," *IEEE Open J. Eng. Med. Biol.*, vol. 3, pp. 7–17, 2022.

[16] A. S. Ravindran, S. Nakagome, D. S. Wickramasuriya, J. L. Contreras-Vidal, and R. T. Faghih, "Emotion recognition by point process characterization of heartbeat dynamics," in *Proc. IEEE Healthcare Innovations Point Care Technol.*, 2019, pp. 13–16.

[17] R. Amin and R. T. Faghih, "Physiological characterization of electrodermal activity enables scalable near real-time autonomic nervous system activation inference," *PLoS Comput. Biol.*, vol. 18, no. 7, 2022, Art. no. e1010275.

[18] H. D. Critchley, R. N. Melmed, E. Featherstone, C. J. Mathias, and R. J. Dolan, "Volitional control of autonomic arousal: A functional magnetic resonance study," *Neuroimage*, vol. 16, no. 4, pp. 909–919, 2002.

[19] F. G. Ashby, V. V. Valentin, and A. Turken, "The effects of positive affect and arousal on working memory and executive attention," *Emotional Cogn.*, vol. 44, pp. 245–288, 2002.

[20] R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation," *J. Comp. Neurol. Psychol.*, vol. 18, pp. 459–482, 1908.

[21] N. Unsworth and M. K. Robison, "The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study," *J. Exp. Psychol.: Learn., Memory, Cogn.*, vol. 43, no. 12, 2017, Art. no. 1962.

[22] D. Kahneman, *Attention and Effort*, vol. 1063. Englewood Cliffs, NJ, USA: Prentice-Hall, 1973.

[23] J. A. Lehmann and T. Seufert, "The influence of background music on learning in the light of different theoretical perspectives and the role of working memory capacity," *Front. Psychol.*, vol. 8, 2017, Art. no. 297754.

[24] M. Nithya.S, "Analysis of music effect on N-back task performance," *Int. J. Latest Eng. Res. Appl.*, vol. 1, no. 3, pp. 38–42, Jun. 2016.

[25] S. Parshi, "A functional-near infrared spectroscopy investigation of mental workload," Ph.D. dissertation, University Houston, Houston, TX, USA, 2020.

[26] A. C. Smith et al., "Dynamic analysis of learning in behavioral experiments," *J. Neurosci.*, vol. 24, no. 2, pp. 447–461, 2004.

[27] M. R. Amin and R. T. Faghih, "Identification of sympathetic nervous system activation from skin conductance: A sparse decomposition approach with physiological priors," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 5, pp. 1726–1736, May 2021.

[28] M. J. Prerau et al., "Characterizing learning by simultaneous analysis of continuous and binary measures of performance," *J. Neuriophysiol.*, vol. 102, no. 5, pp. 3060–3072, 2009.

[29] M. J. Prerau, A. C. Smith, U. T. Eden, M. Yanike, W. A. Suzuki, and E. N. Brown, "A mixed filter algorithm for cognitive state estimation from simultaneously recorded continuous and binary measures of performance," *Biol. Cybern.*, vol. 99, no. 1, pp. 1–14, 2008.

[30] D. S. Wickramasuriya, M. Amin, and R. T. Faghih, "Skin conductance as a viable alternative for closing the deep brain stimulation loop in neuropsychiatric disorders," *Front. Neurosci.*, vol. 13, 2019, Art. no. 470066.

[31] R. T. Faghih, M. A. Dahleh, G. K. Adler, E. B. Klerman, and E. N. Brown, "Quantifying pituitary-adrenal dynamics and deconvolution of concurrent cortisol and adrenocorticotropic hormone data by compressed sensing," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 10, pp. 2379–2388, Oct. 2015.

[32] G. Hendeby and F. Gustafsson, "On nonlinear transformations of Gaussian distributions," Linköpings Universitet, Linköpings, Sweden, Tech. Rep. SE-581, 2007.

[33] D. S. Wickramasuriya and R. T. Faghih, "A mixed filter algorithm for sympathetic arousal tracking from skin conductance and heart rate measurements in pavlovian fear conditioning," *PLoS One*, vol. 15, no. 4, 2020, Art. no. e0231659.

[34] H. J. Kim, F. Fallahtafti, J. Yentes, D. Venema, and J. B. Boron, "High cognitive load situations decrease both gait and cognitive performance," *Innov. Aging*, vol. 4, pp. 299–299, 2020.

[35] A. Marti-Marca, T. Nguyen, and J. A. Grahn, "Keep calm and pump up the jams: How musical mood and arousal affect visual attention," *Music Sci.*, vol. 3, 2020, Art. no. 2059204320922737.

[36] G. Husain, W. F. Thompson, and E. G. Schellenberg, "Effects of musical tempo and mode on arousal, mood, and spatial abilities," *Music Percep.*, vol. 20, no. 2, pp. 151–171, 2002.

[37] S. M. Jaeggi, M. Buschkuehl, W. J. Perrig, and B. Meier, "The concurrent validity of the n-back task as a working memory measure," *Memory*, vol. 18, no. 4, pp. 394–412, 2010.

[38] M. Buschkuehl, L. Hernandez-Garcia, S. M. Jaeggi, J. A. Bernard, and J. Jonides, "Neural effects of short-term training on working memory," *Cogn., Affect., Behav. Neurosci.*, vol. 14, pp. 147–160, 2014.

[39] S. M. Jaeggi, M. Buschkuehl, J. Jonides, and W. J. Perrig, "Improving fluid intelligence with training on working memory," *Proc. Nat. Acad. Sci.*, vol. 105, no. 19, pp. 6829–6833, 2008.

[40] J. A. Bargh and J. L. Cohen, "Mediating factors in the arousal-performance relationship," *Motivation Emotion*, vol. 2, pp. 243–257, 1978.

[41] E. Dumas-Mallet, K. S. Button, T. Boraud, F. Gonon, and M. R. Munafò, "Low statistical power in biomedical science: A review of three human research domains," *Roy. Soc. Open Sci.*, vol. 4, no. 2, 2017, Art. no. 160254.

[42] Wiediartini, U. Ciptomulyono, and R. S. Dewi, "Evaluation of physiological responses to mental workload in n-back and arithmetic tasks," *Ergonomics*, pp. 1–13, 2023, doi: 10.1080/00140139.2023.2284677.

[43] M. K. Yeung, T. L. Lee, Y. M. Han, and A. S. Chan, "Prefrontal activation and pupil dilation during n-back task performance: A combined fnirs and pupillometry study," *Neuropsychologia*, vol. 159, 2021, Art. no. 107954.

[44] H. F. Azgomi, L. R. F. Branco, M. R. Amin, S. Khazaei, and R. T. Faghih, "Regulation of brain cognitive states through auditory, gustatory, and olfactory stimulation with wearable monitoring," *Sci. Rep.*, vol. 13, no. 1, 2023, Art. no. 12399.

[45] J. Faller, J. Cummings, S. Saproo, and P. Sajda, "Regulation of arousal via online neurofeedback improves human performance in a demanding sensory-motor task," *Proc. Nat. Acad. Sci.*, vol. 116, no. 13, pp. 6482–6490, 2019.

[46] H. F. Azgomi, D. S. Wickramasuriya, and R. T. Faghih, "State-space modeling and fuzzy feedback control of cognitive stress," in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2019, pp. 6327–6330.

[47] H. F. Azgomi, I. Cajigas, and R. T. Faghih, "Closed-loop cognitive stress regulation using fuzzy control in wearable-machine interface architectures," *IEEE Access*, vol. 9, pp. 106202–106219, 2021.

[48] L. R. Branco, A. Ehteshami, H. F. Azgomi, and R. T. Faghih, "Closed-loop tracking and regulation of emotional valence state from facial electromyogram measurements," *Front. Comput. Neurosci.*, vol. 16, 2022, Art. no. 747735.

[49] D. Hulsey, K. Zumwalt, L. Mazzucato, D. A. McCormick, and S. Jaramillo, "Decision-making dynamics are predicted by arousal and uninstructed movements," *Cell Rep.*, vol. 43, no. 2, 2024.