AN ATTRIBUTE-INVARIANT VARIATIONAL LEARNING FOR EMOTION RECOGNITION USING PHYSIOLOGY

Hao-Chun Yang, Chi-Chun Lee

Department of Electrical Engineering, National Tsing Hua University, Taiwan
MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan

ABSTRACT

Studies have shown that people with different personalities would result in a different physiological reaction when encountering emotional stimulus. In this work, we propose an attribute-invariance loss embedded variational autoencoder (AI-VAE) to learn the personality-invariant physiological signal representation. The AI-VAE includes an additional loss aiming to perturb features from different personality polarity to obtain emotion discriminative representation. We evaluate our framework on a large emotion corpus of physiological data. Our method achieves a state of the art unweighted accuracy of 68.8% and 67.0% in a binary classification of arousal and valence, which improves over the baseline vanilla VAE by 5.5% and 6.5%. Further analysis reveals that several EEG features are statistically relevant between different personalities types across emotional states, and ECG features are also specifically correlated to personality dimension of "Creativeness", underscoring the importance of personality in modulating psychophysiological processes.

***Index Terms*—** personality, physiological representation, emotion recognition, variational learning, psychophysiology

1. INTRODUCTION

Affective response is a psychophysiological process triggered by conscious and/or unconscious stimuli and is often manifested through humans observable behaviors [1]. Automatic emotion recognition has largely been developed by modeling human's observable behaviors such as speech, facial expression, and linguistic content. Recently, the advancement of miniaturized physiological sensors and mobile computing technologies has enabled continuous monitoring of human internal physiological signals such as electroencephalography (EEG), electrocardiography (ECG), and electrodermal activity (EDA). This has drawn increasing interest for researchers to model the interrelationship between the measured physiological signals and the psycho-physiological process. For example, research has shown that features derived from EEG and ECG are correlated highly with symptoms of depression [2]; the neuro-perceptual response measured by functional Magnetic Resonance Imaging (fMRI) can be used to automatic decoding the emotion stimuli [3].

Affective responses manifested through physiology are known to be modulated by a person's personality. Specifically, the personality-affect relationship has been extensively studied in the Eysenck's personality model [4]. Eysenck states that the personality trait of extraversion is correlated with cortical arousal, i.e., extraverts require higher external stimuli to reaches the same physiological status than introverts. Eysenck also suggests that neurotics are more sensitive to external stressors and are likely to appraise their environment to be more stressful. Other research also indicates that people with high agreeableness have higher emotional self-regulation [5]. Winter and Kuiper extensively examine the relationship between personality and emotion and propose a self-schema model to theorize the underlying mechanism [6]. Komulainen et al's study shows that personality traits lead to different reactions of daily emotion process, e.g., neurotic people have a more negative affect, and high conscientiousness people have lower reactivity to negative affect [7].

Although personalities play a major role in emotional responses, few works have systematically considered their relationship while developing automated emotion recognition system using physiological data. The most related work is the use of hypergraph-based model recently proposed by Zhao et al. [8], which utilizes network graph to constrain the higher order interactions among personal attributes, emotional responses, and affective labels. However, the framework is restricted to be applied in a speaker-dependent scenario only. In this work, we develop a mechanism in learning representation with a personality-invariance loss to mitigate the issue of individual idiosyncrasy to achieve a more robust subject independent emotion recognition from physiology.

Specifically, we propose an attribute-invariance loss embedded variational autoencoder (AI-VAE) to learn physiological signal representation for the task of inferring a subjects emotion status. Our framework is evaluated on the publicly-available large-scale AMIGOS dataset [9], which includes recordings of EEG, ECG, and EDA. The framework achieves an unweighted accuracy of 68.8% and 67.0% for arousal and valence recognition, which is 5.5% and 6.5% relative improvement over the standard VAE method. We further provide analysis of the measured internal physiological responses as a function on the personality-affect relationship.

978-1-5386-4658-8/18/$31.00 ©2019 IEEE

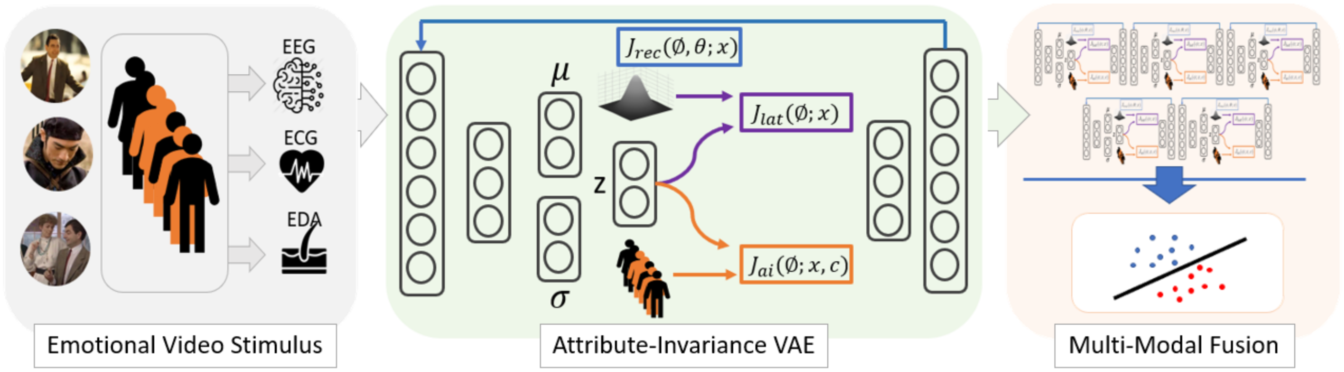

Fig. 1. Our proposed Attribute-Invariance loss embedded VAE *(AI-VAE)* to recognize emotion controlling for personality polarities. The VAEs are pretrained under five different personality dimensions *(Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism)*, and the final result is aggregated by late fusion of each SVM classifier.

2. RESEARCH METHODOLOGY

2.1. AMIGOS Dataset

This study uses AMIGOS Dataset [9] for developing and evaluation. The dataset is composed of 40 subjects with each watching 16 short emotional videos (duration<250s) and 4 long videos (duration>14min). The video stimuli are carefully chosen from two other databases designed to study physiological response to emotional videos. These videos are reannotated to select a subset of those having the most emotional content. EEG, ECG and EDA and frontal video (RGB) are recorded simultaneously. For each subject, the Big-Five personality traits are measured with an online form using the Big-Five marker scale questionnaire [10]. For each trial, the individuals internal affective annotations (self-assessment) are performed before and after each trial, and external annotations (observational-assessment) are conducted by 3 annotators indicating the participant's arousal and valence scores. A total of 800 recordings are in the dataset, but only 766 records are used due to the ill quality issues. We carry out our experiments as a binary classification problem using the mean of external observers' annotations. The two classes (high and low arousal and valence) are divided by reference to the median of each affective dimensions computed over the entire corpus.

2.2. Computational Framework

We will elaborate our proposed computational framework in this section. Figure1 depicts our overall framework, including low-level feature descriptors, variational autoencoder, and the proposed attribute-invariance multi-modal fusion network.

2.2.1. *Low-Level Physiological Descriptors*

For EEG, a bandpass filter from 4-45Hz is applied. ECG and EDA are both filtered by a low-pass filter with 60Hz cut-off frequency. Then, several standard LLDs are extracted using [11]. The detailed features are listed in Table 1. Besides, a standard z-normalization is performed on each feature dimension for each subject to mitigate the issue of individual difference.

Table 1. A summary of physiological llds. "F*" states for 15 statistical functions.[1] EEG functions are calculated for each channel then concatenated as a single feature vector.

Modality	Low-Level Descriptors
EEG(378)	Hjorth, Kurtosis, Skewness, First_diff_mean, First_diff_max, SecDiffMean, SecDiffMax, Slope_mean, Slope_var, Wavelets, MaxPwelch,Entropy, AutoRegressiveParameters
ECG(50)	number_of_artifacts, RMSSD, meanNN, sdNN, cvNN, CVSD, medianNN, madNN, mcvNN, pNN50, pNN20, Triang, Shannon_h, ULF, VLF, LF, HF, VHF, Total_Power, LFn,HFn, LF/HF, LF/P, HF/P, DFA_1, DFA_2,Shannon, Sample_Entropy, Correlation_Dimension, Entropy_Multiscale_AUC, Entropy_SVD, Entropy_Spectral_VLF, Entropy_Spectral_LF, Entropy_Spectral_HF, Fisher_Info, FD_Petrosian, FD_Higushi, Average_Signal_Quality, F* Cardiac_Cycles_Signal_Quality
EDA(60)	F*SCR_Onsets, F*SCR_Peaks_Amplitudes, F*EDA_Phasic, F*EDA_Tonic_Component

2.2.2. *Maximum-Mean Discrepancy VAE (MMD-VAE)*

Variational autoencoder has shown its great performance in learning informative latent representation in an unsupervised manner for affective multimodal data. However, recently researcher has observed that the learned latent information may not be informative due to the Kullback-Leibler Divergence (KLD) criteria used while optimizing the evidence lower bound (ELBO). To learn a more representative latent representation of our physiological data, we use the MMD-VAE that has been proven to obtain better modeling power on both quantitative and qualitative metric [12]. Given an input x, the standard VAE contains an encoder and decoder parameterized by ϕ and θ respectively. Then, we learn the network weights by maximization of variational evidence lower bound:

$$logp_\theta(x) \leq -(J_{lat}(x) + J_{rec}(x)) \tag{1}$$

$$J_{lat}(\phi;x) = D_{KL}(q_\phi(z)||p_\theta(z)) \tag{2}$$

$$J_{rec}(\phi,\theta;x) = -E_{q_\phi(z)}[logp_\theta(x|z)] \tag{3}$$

[1] max, min, mean, median, std, skewness, kurtosis, min position, max position, 25_percentile, 75_percentile, 75_percentile-25_percentile, 1_percentile, 99_percentile, 99_percentile-1_percentile

Table 2. A summary of prediction results. VAE: intrinsic VAE. A-VAE: VAE with domain adversarial discriminator loss. AI-VAE: proposed framework which is attribute-invariance loss embedded VAE. LF: Late fusion among 5 personalities.

UAR	SVM	VAE	A-VAE						C-VAE						AI-VAE					
			Agr	Con	Cre	Emo	Ext	LF	Agr	Con	Cre	Emo	Ext	LF	Agr	Con	Cre	Emo	Ext	LF
Low	0.614	0.680	0.638	0.641	0.650	0.638	0.632	0.638	0.656	0.656	0.613	0.619	0.613	0.674	0.683	0.665	0.656	0.659	0.674	0.740
High	0.589	0.586	0.523	0.517	0.495	0.523	0.523	0.523	0.619	0.607	0.656	0.656	0.659	0.637	0.631	0.634	0.637	0.634	0.616	0.637
Arousal	0.601	0.633	0.58	0.578	0.572	0.58	0.577	0.554	**0.637**	0.631	0.634	**0.637**	0.636	**0.655**	**0.657**	0.649	0.646	0.645	0.651	**0.688**
Low	0.575	0.743	0.680	0.620	0.623	0.620	0.599	0.620	0.671	0.677	0.656	0.668	0.659	0.751	0.668	0.671	0.668	0.686	0.677	0.743
High	0.613	0.468	0.498	0.571	0.559	0.571	0.556	0.571	0.562	0.580	0.562	0.565	0.568	0.514	0.604	0.604	0.580	0.604	0.592	0.598
Valence	0.594	0.605	0.589	0.595	0.59	0.595	0.577	0.595	0.616	**0.628**	0.609	0.616	0.613	**0.633**	0.636	0.637	0.624	**0.645**	0.634	**0.670**

where z is our latent code vector, and $P_\theta(z)$ is the variational gaussian normal prior of z. Here, we replace the KLD with Maximum-Mean Discrepancy (MMD):

$$D_{MMD}(q||p) = E_{p(z),p(z')}[k(z,z')] - 2E_{q(z),p(z')}[k(z,z')] + E_{q(z),q(z')}[k(z,z')] \tag{4}$$

where $k(z, z')$ is an positive definite kernel as $e^{-\|z-z'\|_2}$ and the l2 norm term is empirically divided by the dimension of z. $D_{MMD} = 0$ if and only if $p = q$.

2.2.3. *Attribute-Invariance VAE*

The objective of this work is to utilize a persons personality profile to control the variability in learning a more discriminative physiological representation to affect recognition. Hence, in this research, we propose a novel framework which embeds an attribute invariance loss into the original MMD-VAE formulation. Given a batch of data x and their personality attribute c, we first binarize the data pairs according to their personality attribute score into sub-batches (x_h, c_h) and (x_l, c_l), which in our experiments is defined as being higher or lower than the database's median value. Then we calculate the attribute invariance loss by:

$$J_{ai}(\phi; x, c) = D(p_\phi(z_h|x_h; c_h)||p_\phi(z_l|x_l; c_l)) \tag{5}$$

which aims to minimize the distance between two groups of distribution. The loss can be implemented by any divergence families, such as KLD and Hellinger distance, or Wasserstein distance potentially leading to a variety of results. In this research, we choose the MMD as our attribute invariance criteria due to its ease of implementation and robust modeling capacity. Note that the loss also includes hyperparameters alpha and lambda to balance different losses contribution. Hence, the final update criteria are summarized as follow:

$$\begin{aligned} \phi &\xleftarrow{\text{update}} -\nabla_\phi(\alpha * J_{lat} + J_{rec} + \lambda * J_{ai}) \\ \theta &\xleftarrow{\text{update}} -\nabla_\theta(J_{rec}) \end{aligned} \tag{6}$$

3. EXPERIMENTAL SETUP AND RESULT

3.1. Experimental Setup

The exact architecture of our personality-invariance loss embedded network includes: five attribute-invariance VAE (AI-VAE) proposed in section 2, each corresponding to the five personality dimensions. Each AI-VAE is composed of $[488 - 244 - 100 - 244 - 488]$ with standard fully-connected layers using Leaky ReLU as activation function. An early stop technique is performed and hyperparameters alpha and lambda are grid searched with the range of $[100, 200, 400]$ and $[500, 1000, 2000]$ respectively. Then we use linear support vector machine [13] as the classifier by inputting latent feature from the AI-VAE encoder, and a late fusion is applied to combine recognition results from different personality-dependent AI-VAE to obtain final emotion recognition accuracy. We carry out our experiments using a subject independent 10-fold cross-validation. The final evaluation metric used is the unweighted average recall (UAR).

3.1.1. *Comparison Models*

We first conduct our experiments utilizing linear SVM and vanilla VAE without considering personal attributes. Then there are several other algorithms developed with an aim to tackle similar attribute-invariance embedding problems:

- **Conditional Variational Autoencoder (C-VAE)**: C-VAE can be used to learn a hidden vector which maximizes the conditional probability under an encoder-decoder structure which adds a further constraint on the latent representation by giving it a conditional probability. The learned attribute-constrained encoded vector has been shown to be effective on audio and video classification tasks [14, 15], but could suffer performance issue due to its characteristic of weak constraint.

- **Adversarial Network Training (A-VAE)**: A-VAE can learn a discriminative and attribute-invariance representation by jointly optimizing an attribute-aware discriminator. The network structure is often set up in a way to leverage data from different domains [16] and design the discriminator to force the latent code to be domain-invariance. However, the discriminator could suffer from convergence problem if the attribute is only vaguely correlated to the data.

We implement the above two attribute embedding networks A-VAE (VAE with Adversarial training on personality) and C-VAE (Conditional VAE) as a comparison to our proposed model. The C-VAE is constructed the same as [14], and A-VAE includes an additional discriminator with layers of node number $[100, 10, 1]$ that is added to the vanilla VAE. For fairness, we keep the parameters exactly the same for the rest of the layers and nodes number used in the autoencoder.

Table 3. Comparison of mean F1-scores(mean F1-score for both classes). Note that SVM* states for the method in [9].

F1	SVM*	SVM	VAE	A-VAE	C-VAE	AI-VAE
Arousal	0.564	0.596	0.614	0.554	0.648	0.644
Valence	0.560	0.601	0.543	0.585	0.583	0.671

3.2. Personality Invariance Emotion Recognition Results

Table 2 summarizes our emotion recognition results. Our proposed personality invariance physiological latent encoding reaches the best UAR of 68.8% and 67.0% in the binary classification of arousal and valence, which is a relative gain of 5.5% and 6.5% over the vanilla VAE. Several notable observations can be summarized. Firstly, C-VAE demonstrates a slightly better discriminative power under personality constraints over vanilla VAE, while A-VAE fails to learn a meaningful latent embedding overall. We hypothesize that A-VAE has limited capability in regulating the learning trajectory of feature representation when the discriminator losses functionality and ends up converging in a non-meaningful minima. Secondly, we can observe from C-VAE and AI-VAE that the regulation of Agreeableness both improve the recognition for arousal. As for valence, the proposed AI-VAE achieves the highest result by controlling for the unwanted physiological signal variability with personality attribute of Emotional Stability. At last, although arousal recognition achieves higher UAR in the experiment, the framework benefits much for the valence dimension. We can also view the same pattern comparing the results in Table 3, which indicates that the valence-related physiological response has a closer relationship with personality dimensions, and the elimination of the potential variance improves the recognition results.

3.3. Statistical Analyses

In this section, to understand whether personalities do act as a latent control on affective physiological signals, we perform a standard two-sided Students t-tests on the extracted LLDs. We first cluster the data according to their emotion status, then we conduct the t-test to examine whether each physiological LLD would show a difference between high versus low for each personality dimension given the participant is in the same emotional state. Table 3 summarizes the most statistically significant features (t-value>4 and p-value<0.01), and several observations can be made from the comparison. Firstly, the Hjorth and ARMPB feature are consistently selected across different emotional states. Hjorth parameters have been carefully studied especially on alcohol-induced personality deterioration [17], while ARMPB is a key indicator of schizotypal personality [18], both give us supporting evidence that personality indeed induces a hidden variability on specific EEG representations. Secondly, we can see that different physiological modalities respond differently to each personality attribute. For example, EDA signal has a little effect among all dimensions, while ECGs features mostly correlate with "Creativeness". It is interesting to conclude from our analysis that there truly exists interrelationships between emotions and personalities as manifested in the measured physiological responses, and through our proposed method, which eliminates such a personality-induced latent factor, we obtain a "cleaner" emotional physiological signal. Our approach is beneficial for advancing the personal attribute-aware emotion recognition framework.

Table 4. A summary of representative features which are significant on the double-sided t-Test on personality polarities. ”F*” states for statistical functional. ARMPB: Autoregressive model parameters using Burg method. CCSQ: Cardiac cycles signal quality.

	High Arousal			**Low Arousal**		
	EDA	ECG	EEG	EDA	ECG	EEG
Agr	F*EDA_Tonic	HF/P	hjorth, ARMPB	F*EDA_Tonic, F*EDA_Phasic		hjorth, ARMPB
Con		CVSD, RMSSD, pNN50, ASQ, F*CCSQ	ARMPB			wavelet
Cre	F*EDA_Tonic	ASQ CVSD, F*CCSQ, RMSSD, LF/HF	hjorth, ARMPB	F*CCSQ,	ASQ	wavelet, hjorth, ARMPB
Emo		HF, CVSD	wavelet, ARMPB	F*EDA_Tonic		
Ext		Total_Power, LFn, LF/HF, HF/P	hjorth, ARMPB			hjorth, ARMPB
	High Valence			**Low Valence**		
	EDA	ECG	EEG	EDA	ECG	EEG
Agr	F*EDA_Tonic	HF/P	hjorth, ARMPB	F*EDA_Tonic, F*EDA_Phasic		hjorth, ARMPB
Con			wavelet		ASQ, F*CCSQ	
Cre		ASQ, CVSD, F*CCSQ,	hjorth, wavelet, maxPwelch, ARMPB		ASQ, CVSD, F*CCSQ, RMSSD	hjorth
Emo				F*EDA_Tonic		
Ext			ARMPB, hjorth	F*EDA_Phasic	pNN50, mcvNN	hjorth, wavelet, ARMPB

4. CONCLUSION

In this work, we present a novel framework of attribute-invariance loss embedded VAE to recognize the emotion given personality profiles. We apply our method to an emotional video stimulus physiological signal dataset and compare it with other attribute-aware architectures. The experiments show our methodology improves the recognition results by 5.5% and 6.5% on arousal and valence classification over standard VAE. Our analysis of the extracted physiological LLDs further reveals that "Hjorth" and "ARMPB" from EEG are key factors in bringing insight on how personality affects physiological emotion reaction, and "Creativeness" has a more prominent effect on the cardiovascular measurement. To our best knowledge, this is one of the first work in handling the attribute learning problem by the elimination of personal differences on physiological emotion recognition. There are multiple future directions. An immediate one would be to include expressive modality, such as facial expressions, to continuously explore attribute-affect relationship from both explicit and implicit behavior information. Second, other personal attributes like gender and age or even data-driven meta clustering should be added in. All of these could be utilized to enhance both the accuracy and robustness of the model that can be integrated for a variety of human behavior modeling tasks [19, 20].

5. REFERENCES

[1] John T Cacioppo, Louis G Tassinary, and Gary Berntson, *Handbook of psychophysiology*, Cambridge University Press, 2007.

[2] Julian Koenig, Andrew H Kemp, Theodore P Beauchaine, Julian F Thayer, and Michael Kaess, "Depression and resting state heart rate variability in children and adolescentsa systematic review and meta-analysis," *Clinical psychology review*, vol. 46, pp. 136–150, 2016.

[3] Ya-Tse Wu, Hsuan-Yu Chen, Yu-Hsien Liao, Li-Wei Kuo, and Chi-Chun Lee, "Modeling perceivers neural-responses using lobe-dependent convolutional neural network to improve speech emotion recognition," *Proc. Interspeech 2017*, pp. 3261–3265, 2017.

[4] Hans J Eysenck, *Dimensions of personality*, vol. 5, Transaction Publishers, 1950.

[5] Renée M Tobin, William G Graziano, Eric J Vanman, and Louis G Tassinary, "Personality, emotional experience, and efforts to control emotions.," *Journal of personality and social psychology*, vol. 79, no. 4, pp. 656, 2000.

[6] Kathy A Winter and Nicholas A Kuiper, "Individual differences in the experience of emotions," *Clinical psychology review*, vol. 17, no. 7, pp. 791–821, 1997.

[7] Emma Komulainen, Katarina Meskanen, Jari Lipsanen, Jari Marko Lahti, Pekka Jylhä, Tarja Melartin, Marieke Wichers, Erkki Isometsä, and Jesper Ekelund, "The effect of personality on daily life emotional processes," *PLoS One*, vol. 9, no. 10, pp. e110907, 2014.

[8] Sicheng Zhao, Guiguang Ding, Jungong Han, and Yue Gao, "Personality-aware personalized emotion recognition from physiological signals.," in *IJCAI*, 2018, pp. 1660–1667.

[9] Juan Abdon Miranda-Correa, Mojtaba Khomami Abadi, Nicu Sebe, and Ioannis Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," *arXiv preprint arXiv:1702.02510*, 2017.

[10] Murray R Barrick and Michael K Mount, "The big five personality dimensions and job performance: a meta-analysis," *Personnel psychology*, vol. 44, no. 1, pp. 1–26, 1991.

[11] Makowski. D, "Neurokit: A python toolbox for statistics and neurophysiological signal processing (eeg, eda, ecg, emg...)," Day, 01 November 2016, Paris, France.

[12] Shengjia Zhao, Jiaming Song, and Stefano Ermon, "Infovae: Information maximizing variational autoencoders," *arXiv preprint arXiv:1706.02262*, 2017.

[13] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin, "Liblinear: A library for large linear classification," *Journal of machine learning research*, vol. 9, no. Aug, pp. 1871–1874, 2008.

[14] Jeng-Lin Li, Yi-Ming Weng, Chip-Jin Ng, and Chi-Chun Lee, "Learning conditional acoustic latent representation with gender and age attributes for automatic pain level recognition," *Proc. Interspeech 2018*, pp. 3438–3442, 2018.

[15] Panna Felsen, Patrick Lucey, and Sujoy Ganguly, "Where will they go? predicting fine-grained adversarial multi-agent motion using conditional variational autoencoders," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[16] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell, "Adversarial discriminative domain adaptation," in *Computer Vision and Pattern Recognition (CVPR)*, 2017, vol. 1, p. 4.

[17] Wolfgang Spehr and Gerhard Stemmler, "Determinants of chlormethiazole-induced eeg effects in chronic alcoholics," *Neuropsychobiology*, vol. 12, no. 4, pp. 265–269, 1984.

[18] Gleb V Tcheslavski, "Effects of tobacco smoking and schizotypal personality on spectral contents of spontaneous eeg," *International Journal of Psychophysiology*, vol. 70, no. 1, pp. 88–93, 2008.

[19] Shrikanth Narayanan and Panayiotis G Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1203–1233, 2013.

[20] Daniel Bone, Chi-Chun Lee, Theodora Chaspari, James Gibson, and Shrikanth Narayanan, "Signal processing and machine learning for mental health research and clinical applications [perspectives]," *IEEE Signal Processing Magazine*, vol. 34, no. 5, pp. 196–195, 2017.