Regression of Multiple Conversation Aspects using Dyadic Physiological Measurements

Iman Chatterjee
Deptartment of Electrical and Computer Engineering
University of Cincinnati
Cincinnati, USA
chattein@mail.uc.edu

Maja Goršič
College of Engineering & Applied Science
University of Wisconsin Milwaukee
Milwaukee, USA
gorsic@uwm.edu

Robert A. H. Kaya
Department of Psychology
University of Wyoming
Laramie, USA
rkaya@uwyo.edu

Collyn J. Erion
Department of Psychology
University of Wyoming
Laramie, USA
cerion@uwyo.edu

Joshua D. Clapp
Department of Psychology
University of Wyoming
Laramie, USA
jclapp@uwyo.edu

Vesna D. Novak
Department of Electrical and Computer Engineering
University of Cincinnati
Cincinnati, USA
novakdn@ucmail.uc.edu

***Abstract*—Dyadic physiological responses are correlated with the quality of interpersonal processes – for example, the degree of "connectedness" in education and mental health counseling. Pattern recognition algorithms could be applied to such dyadic responses to identify the states of specific dyads, but such pattern recognition has primarily focused on classification. This paper instead uses regression algorithms to estimate three conversation aspects (valence, arousal, balance) from heart rate, skin conductance, respiration, and skin temperature. Data were collected from 35 dyads who engaged in 20 minutes of conversation, divided into 10 two-minute intervals. Each interval was rated with regard to conversation valence, arousal, and balance by an observer. When regression algorithms (support vector machines and Gaussian process regression) were trained on other data from the same dyad, they were able to estimate valence, arousal and balance with lower errors than a simple baseline estimator. However, when algorithms were trained on data from other dyads, errors were not lower than those of the baseline estimator. Overall, results indicate that, as long as training data from the same dyad are available, autonomic nervous system responses can be combined with regression algorithms to estimate multiple dyadic conversation aspects with some accuracy. This has applications in education and mental health counseling, though fundamental issues remain to be addressed before the technology is used in practice.**

Keywords—affective computing, autonomic nervous system responses, conversation, physiological signals, psychophysiology, regression

I. INTRODUCTION

In the fields of affective computing and applied psychophysiology, physiological responses are frequently combined with pattern recognition algorithms to automatically recognize human psychological states [1]. This can include, for example, workload levels in air traffic controllers [2], engagement levels in patients undergoing rehabilitation [3], drowsiness/inattention levels in drivers [4], and enjoyment levels in computer game players [5]. Machines can then take action to bring users into a more desirable psychological state by, e.g., changing the amount of automated assistance or adapting the difficulty of an exercise [1].

While most research in this area has focused on individuals, studies have also used physiological responses to identify psychological states of dyads (pairs) who compete, cooperate, or simply communicate with each other. Dyadic physiological responses are known to correlate with the intensity of competition [6], quality of cooperation [7] and the degree of "connectedness" in processes like education [8] and mental health counseling [9]. Pattern recognition algorithms could thus also be applied to such dyadic physiological responses to identify the state of a specific dyad. This would have many potential applications. For example, visual feedback about the interaction could be presented to users, allowing them to modify their own behavior in order to improve the interaction outcome [10], which may lead to better outcomes of technology-assisted education and mental health counseling.

Pattern recognition algorithms to identify the psychological state of a specific dyad are predominantly based on supervised machine learning – either classification or regression. Most commonly, classification algorithms are used to classify psychological states into one of two classes (e.g., engaged or unengaged dyads) based on a single physiological response type (e.g., only electrocardiography or only electroencephalography) [11]–[16]. Less commonly, some studies perform classification into 3 or 4 classes based on a single physiological response type [17], [18] or multiple physiological response types [19], [20]. To our knowledge, regression algorithms have been used by only two dyadic physiology studies [21], [22], and classification is more popular than regression in similar studies involving individuals rather than dyads as well [1]. Nonetheless,

This work was supported by the National Institute of Mental Health of the National Institutes of Health, grant number R03MH128633, and by the National Science Foundation, grant number 2151464.

regression does have several advantages over classification, such as the potential to provide more granular information about psychological states, and thus represents a promising alternative.

The goal of the current study is to combine multiple autonomic nervous system responses with supervised regression algorithms in order to automatically extract multiple psychological aspects of a dyadic conversation based on third-person (observer) ratings of the conversation. The current study is based on our previous dyadic regression study [22], but with several expansions and modifications. Specifically, the previous study used physiological responses and regression to estimate a single conversation variable (engagement) obtained using an ad-hoc self-report questionnaire completed by participant dyads. Conversely, this study estimates multiple conversation variables as coded by external observers. The observers were trained to produce consistent ratings of participant behavior, thus avoiding some of the weaknesses inherent in psychophysiological studies that are limited to self-report data [23].

II. Materials and Methods

A. Participants

Thirty-five dyads were recruited for the study from students and staff of the University of Cincinnati. All dyads knew each other prior to the session, and recruitment materials specifically requested that participants volunteer in self-selected pairs. Participants were 21.1 ± 3.9 years old (mean ± standard deviation). There were 11 dyads with two women, 9 dyads with two men, 13 dyads with one woman and one man, 2 dyads with one woman and one nonbinary participant, and one dyad with two nonbinary participants. Each participant signed an informed consent form.

B. Hardware

Participants' physiological responses were recorded at 600 Hz using two g.USBamp biosignal amplifiers (g.tec Medical Engineering GmbH, Austria) and their add-on sensors. Four disposable wired electrodes were placed on the trunk to record the electrocardiogram (ECG) using a placement suggested by the manufacturer. The g.GSRsensor2 was used to record skin conductance via two dry electrodes placed on distal phalanges of the index and middle finger of the nondominant hand. The g.Sensor Respiration Airflow was used to record respiration rate via a thermistor-based sensor placed under the nose and above the mouth. Finally, the g.Temp was used to record peripheral skin temperature via a sensor on the distal phalanx of the fifth finger of the nondominant hand. This sensor setup was very similar to that in our previous work [19], [22]. It is worth noting that, with this sensor setup, the respiration signal is likely influenced by participants' speech, and we will investigate alternatives in future work.

Additionally, audio and video of the participants were recorded throughout the session using a pair of consumer-grade webcams (one pointed at each participant) and a Yeti X microphone (Blue Microphones, USA) placed between the participants. This sensor setup was also similar to that in our previous work [22]. The sensors can be seen in Fig. 1, which illustrates the overall study setup.

Fig. 1. The study setup. Participants sit facing each other with physiological sensors attached and connected to signal amplifiers (bottom of photo). Webcams and a microphone between the participants are used to collect audio and video of the conversation. Participants also filled out self-report questionnaires, which were not analyzed for this paper.

C. Study Protocol

The study was approved by the University of Cincinnati Institutional Review Board, protocol 2021-1107. Each dyad took part in a single session. Participants sat at a table approximately 1.5 m apart facing each other (Fig. 1). They self-applied physiological sensors following the experimenter's guidance. Each dyad was randomly assigned an initial conversation tone: positive, neutral, or negative. Dyads in the positive group were asked to self-select an initial topic that both participants liked and agreed on; conversely, dyads in the negative group were asked to select a topic that they disagreed on. In both groups, these turned out to be mostly "non-serious" topics such as food, music and sports.

After the assignment and topic selection, participants first sat facing each other silently for 2 minutes with eyes open to obtain baseline physiological recordings. They then talked to each other for 20 minutes, divided into ten 2-minute intervals. They began the first interval by talking about their self-selected topic (in the case of positive/negative groups) or about career goals (for the neutral group). However, they could switch topics at any time. After each 2-minute interval, the experimenter raised a hand to pause the conversation, during which participants filled out a brief self-report questionnaire (not analyzed in current study). The conversation then continued. At the end of the session, participants removed the sensors and received $15 Amazon gift cards as compensation.

To introduce more variability into conversations (which were a priori expected to be mostly neutral-to-positive [22]), one of the two participants was also provided with secret prompts to modify the conversation after the fourth and seventh 2-minute interval. That participant was told about the prompts at the start of the session (by being asked to briefly leave the room by the researcher) while the other participant was not told about them until the end of the session. The prompts were "During the next 2-minute interval, show absolutely no emotion" and "During the next 2-minute interval, point out every possible flaw with whatever the other person says". The two prompts were given in random order (half the dyads receiving the no-emotion prompt first) and were shown on the paper self-report questionnaire between 2-minute intervals, though only to one of the two participants.

D. *Analysis of Audio and Video Data*

Each dyad's audio/video recordings were divided into the ten 2-minute conversation intervals. Segments not belonging to these intervals were not analyzed further. Each 2-min interval was reviewed by independent researchers who were not involved in primary data collection or the analysis of physiological responses. Behavioral codes were developed to correspond to items from a modified version of the Self-Assessment Manikin (SAM) [24]. The original SAM is a 3-item questionnaire assessing valence (positive vs. negative mood), arousal (degree of energy/activation), and dominance, with ratings made on a 9-point graphical scale [24]. For the modified SAM, the first two items were retained while the third was replaced with ratings of the balance of the interaction (i.e., the relative contribution of each participant to the conversation). Behavioral anchors for observer ratings of valence, arousal, and balance at the dyad level were developed using video from 5 randomly selected participant pairs. Intervals from an additional three randomly selected dyads (30 unique intervals) were used to verify the consistency of ratings across coders before finalizing scores for the remaining 27 pairs. Nine of the remaining cases were selected at random to be coded by both reviewers as a formal assessment of interrater reliability. Estimates of consistency were excellent for valence (ICC = .87), arousal (ICC = .81), and balance (ICC = .91) codes.

E. *Physiological Feature Extraction*

Physiological recordings were divided into the 2-minute baseline interval and ten 2-minute conversation intervals. All recordings were analog and digital bandpass filtered with the same filters as in our previous work [22]. Multiple physiological features were extracted from each interval. These were divided into individual features (extracted from a single participant's signal) and synchrony features (extracted from the same signal type of both participants – e.g., from both participants' ECG signals).

Individual ECG features consisted of each participant's mean heart rate, minimum heart rate, maximum heart rate, standard deviation of interbeat intervals, root mean square value of consecutive differences between interbeat intervals, percentage of consecutive interbeat intervals with a difference greater than 50 ms (pNN50), power in low-frequency band, power in high-frequency band, and the ratio of the two powers (LF/HF ratio). These are standard time-domain and frequency-domain measures of heart rate variability [25]. Individual skin conductance features consisted of mean skin conductance, final skin conductance, the difference between initial and final skin conductance values, number of skin conductance responses, mean skin conductance amplitude, and standard deviation of skin conductance response amplitudes. Skin conductance responses were detected using code from our previous work [20]. Individual respiration features consisted of mean respiration rate and standard deviation of respiration rate, again calculated using code from our previous work [20]. Finally, individual skin temperature features consisted of mean temperature, final temperature, and the difference between initial and final temperature values.

Synchrony features were calculated from instantaneous heart rate and respiration rate signals (computed as a function of time from raw electrocardiogram and nose respiration signals [20]) as well as from bandpass-filtered skin conductance and temperature signals. The same features were calculated for all 4 signals (ECG, respiration, skin conductance, skin temperature). They consisted of dynamic time warping distance (as introduced by Muszynski et al. [16]), nonlinear interdependence (also introduced by Muszynski et al. [16]), coherence (same algorithm used in our previous work [20]) and cross-correlation (same algorithm used in our previous work [20]).

F. *Regression*

Physiological features consisted of individual and synchrony features, obtained from 35 dyads with 10 data points (intervals) per dyad. Audio/video features consisted of observer ratings of valence, arousal, and balance, again obtained from 35 dyads with 10 data points (intervals) per dyad. The goal of the study was to create regression algorithms to infer valence, arousal, and balance ratings (outputs) from physiological features (input). Two regression methods were used for this: support vector machines (SVM - implemented using *fitrsvm* function in MATLAB 2022b) and Gaussian process regression (GPR - using *fitgrp* function in MATLAB 2022b). Additionally, two feature selection methods were tested prior to regression: bidirectional stepwise feature selection (using *stepwisefit* function in MATLAB 2022b) and recursive feature elimination with correlation bias reduction (RFE), implemented using open-source MATLAB code [26].

As all data were collected in advance, crossvalidation was used to train and test the feature selection and regression methods. In crossvalidation, the algorithms are trained on a subset of data (selecting regression coefficients and selection thresholds, choosing "best" selection/regression method) and then tested on the remaining data. Specifically, two different types of crossvalidation were used: leave-interval-out crossvalidation and leave-dyad-out crossvalidation. In leave-interval-out crossvalidation, the algorithms were trained on 9 data points of a dyad and then tested on the remaining data point. This was repeated 10 times per dyad, with each data point serving as the "test" point once, and performed for each dyad. Conversely, in leave-dyad-out crossvalidation, the algorithms were trained on all data from 34 dyads and then tested on all 10 data points from the remaining dyad. This was repeated 35 times, with each dyad serving as the "test" dyad once. Thus, leave-interval-out crossvalidation uses training data from the same dyad, but only has 9 training data points; on the other hand, leave-dyad-out crossvalidation has no training data from the same dyad, but has 340 training data points.

Errors in regression of valence/arousal/balance were calculated as the difference between the output of the regression algorithm and the corresponding valence/arousal/balance rating actually made by the observer. In both leave-interval-out and leave-dyad-out crossvalidation, a dyad's regression performance was calculated as the root-mean-square (RMS) error over the 10 data points of that dyad: each of the 10 error values was squared, the mean of the 10 squared values was calculated, and the root of the mean was used as the dyad's RMS error. The same outcome metric was used in our previous work [22].

As RMS errors in regression may be difficult to contextualize, a "baseline" estimator was used as well. This baseline estimator did not rely on physiological data. Instead, in leave-interval-out crossvalidation, the baseline estimator calculated the valence/arousal/balance of the test data point as the median valence/arousal/balance of the 9 training data points from a dyad. In leave-dyad-out crossvalidation, it calculated the valence/arousal/balance of the 10 "test" data points as the median valence/arousal/balance of the 340 training data points. A similar approach was used in our previous work [22].

III. Results

In leave-interval-out crossvalidation, the best results were obtained using RFE and SVMs for all three outcome variables. In leave-dyad-out crossvalidation, the best results were obtained using stepwise feature selection and SVMs for valence, but using stepwise feature selection and GPR for arousal and balance. Table 1 shows means and standard deviations of RMS errors across all 35 dyads in both crossvalidation types.

In leave-interval-out crossvalidation, paired t-tests showed that RMS errors with SVMs were significantly lower than with the baseline estimator for valence, arousal, and balance ($p < 0.001$ for all three comparisons). In leave-dyad-out crossvalidation, however, paired t-tests found no significant differences (lowest $p = 0.07$ for balance).

TABLE I. Root-mean-square errors when estimating three aspects of the conversation: valence, arousal, and balance. Errors were calculated in leave-interval-out and leave-dyad-out crossvalidation. Both crossvalidation types focused on regression algorithms for estimation, but a simple baseline estimator was also used. All values are means ± standard deviations. As valence, arousal, and balance are measured on 9-point scales, the minimum possible error is 0 and the maximum possible error is 8.

	Root-mean-square errors		
	Valence	***Arousal***	***Balance***
Leave-interval-out baseline estimator	1.10 ± 0.40	1.07 ± 0.30	2.07 ± 0.60
Leave-interval-out regression algorithm	0.69 ± 0.26	0.74 ± 0.25	1.34 ± 0.44
Leave-dyad-out baseline estimator	1.16 ± 0.39	1.25 ± 0.45	1.99 ± 0.50
Leave-dyad-out regression algorithm	1.15 ± 0.37	1.15 ± 0.41	1.91 ± 0.42

IV. Discussion

In leave-interval-out crossvalidation, SVM-based regression achieved lower RMS errors than the simple baseline estimator. This indicates that, given training data from the same participants, physiological measurements can be used to estimate multiple conversation aspects (valence, arousal, balance) with some degree of precision. These results are actually better than the results of leave-interval-out crossvalidation in our previous regression study [22], which did not find a difference in RMS errors between the baseline estimator and any regression estimator in leave-interval-out crossvalidation. The improvement relative to our previous study is likely due to methodological improvements such as the use of a more standard rating procedure – the SAM used in the current work is a very well-validated questionnaire compared to the ad-hoc questionnaire used in our previous work [22]. It is particularly noteworthy that positive results were obtained even though the training dataset was very small (9 data points), and even better results could potentially be obtained using more training data from the same dyad.

However, in leave-dyad-out crossvalidation, no regression algorithm achieved significantly lower RMS errors than the baseline estimator. This indicates that, without training data from the same participants, it is not possible to accurately estimate conversation aspects in this study design using these physiological signals and regression methods. While some improvements could likely be made by, e.g., introducing new physiological features or new regression algorithms, we believe that significant improvement of leave-dyad-out regression results would require either conceptually different algorithms or a different study protocol. For example, instead of training regression algorithms on data from all other dyads, the algorithms could be trained only on data from dyads that are similar to the "test" dyad with regard to age, personality traits or other characteristics. Alternatively, instead of observer ratings, self-report ratings could be used to obtain better insights into internal psychological processes that may not be visible to observers, though self-report ratings may have their own reliability issues [23].

Overall, results indicate that, as long as training data from the same dyad are available, autonomic nervous system responses can be combined with regression algorithms to automatically estimate multiple dyadic conversation aspects with some accuracy. Thus, while such approaches may not be useful in situations where participants meet only once, they could potentially be useful in situations where people interact with each other over longer time periods – for example, in education [8] and mental health counseling [9], where teachers/students and therapists/clients establish longer-term relationships. In such situations, data from a first session could potentially be used to train the algorithms. The trained algorithms could then be used to analyze the quality of interpersonal interaction in further sessions – for example, to identify moments of efficient vs. inefficient communication. Alternatively, the algorithms could be used to provide real-time feedback about the conversation as it occurs, allowing participants to modify their own behavior if they realize the conversation is going poorly [10].

However, several additional issues would need to be addressed before the regression algorithms could be used in practical situations. For example, dyadic physiological responses are different in populations such as chronically depressed people [27], which may limit their usability in mental health counseling. Furthermore, it is unclear whether the obtained accuracies are practically useful, as an RMS error of approximately 0.7 on a 9-point scale may be no better than what the participants themselves can glean from the interaction. Other technologies such as facial expression analysis and eye tracking may also be able to achieve better performance, and similar issues of practical usefulness have been raised in other areas of affective computing [28]. Finally, due to intrasubject variability, it is unclear whether the trained regression algorithms would be stable on a day-to-day basis, and multiple sessions on multiple days would be needed to evaluate their robustness. We will continue exploring these topics in our future work.

V. Conclusion

In leave-interval-out crossvalidation, regression algorithms applied to physiological data were able to estimate conversation valence, arousal and balance with an accuracy better than that of a simple baseline estimator. However, in leave-dyad-out crossvalidation, regression algorithms did not outperform the baseline estimator. This indicates that autonomic nervous system responses can be combined with regression algorithms to automatically estimate multiple dyadic conversation aspects with some accuracy, but only if training data are available from the same dyad. In the long term, such analysis of physiological responses could potentially be used to automatically analyze the quality of interpersonal interactions. It could even be used to provide real-time feedback about the interaction as it occurs, allowing participants to modify their own behavior if they realize the conversation is going poorly. However, several questions would need to be answered before the technology is used in practice, such as whether physiological responses provide any information that cannot be obtained more easily through other means.

References

[1] S. D'Mello, A. Kappas, and J. Gratch, "The affective computing approach to affect measurement," *Emot. Rev.*, vol. 10, no. 2, pp. 174–183, 2018.

[2] G. F. Wilson and C. A. Russell, "Operator functional state classification using multiple psychophysiological features in an air traffic control task," *Hum. Factors*, vol. 45, pp. 381–389, 2003.

[3] C. Rodriguez-Guerrero, K. Knaepen, J. C. Fraile-Marinero, J. Perez-Turiel, V. Gonzalez-de-Garibay, and D. Lefeber, "Improving challenge/skill ratio in a multimodal interface by simultaneously adapting game difficulty and haptic assistance through psychophysiological and performance feedback," *Front. Neurosci.*, vol. 11, p. 242, 2017.

[4] A. Darzi, S. Gaweesh, M. Ahmed, and D. Novak, "Identifying the causes of drivers' negative states using driver characteristics, vehicle kinematics and physiological measurements," *Front. Neurosci.*, vol. 12, p. 568, 2018.

[5] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," *IEEE Trans. Syst. Man Cybern. – Part A Syst. Humans*, vol. 41, no. 6, pp. 1052–1063, 2011.

[6] T. Liu, G. Saito, C. Lin, and H. Saito, "Inter-brain network underlying turn-based cooperation and competition: A hyperscanning study using near-infrared spectroscopy," *Sci. Rep.*, vol. 7, p. 8684, 2017.

[7] Y. Hu, Y. Pan, X. Shi, Q. Cai, X. Li, and X. Cheng, "Inter-brain synchrony and cooperation context in interactive decision making," *Biol. Psychol.*, vol. 133, pp. 54–62, 2018.

[8] L. Zheng *et al.*, "Affiliative bonding between teachers and students through interpersonal synchronisation in brain activity," *Soc. Cogn. Affect. Neurosci.*, vol. 15, no. 1, pp. 97–109, 2020.

[9] W. Tschacher and D. Meier, "Physiological synchrony in psychotherapy sessions," *Psychother. Res.*, vol. 30, no. 5, pp. 558–573, 2020.

[10] M. Salminen *et al.*, "Evoking physiological synchrony and empathy using social VR with biofeedback," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 746–755, 2022.

[11] I. Konvalinka, M. Bauer, C. Stahlhut, L. K. Hansen, A. Roepstorff, and C. D. Frith, "Frontal alpha oscillations distinguish leaders from followers: Multivariate decoding of mutually interacting brains," *Neuroimage*, vol. 94, pp. 79–88, 2014.

[12] Y. Pan, S. Dikker, P. Goldstein, Y. Zhu, C. Yang, and Y. Hu, "Instructor-learner brain coupling discriminates between instructional approaches and predicts learning," *Neuroimage*, vol. 211, p. 116657, 2020.

[13] J. Hernandez, I. Riobo, A. Rozga, G. D. Abowd, and R. W. Picard, "Using electrodermal activity to recognize ease of engagement in children during social interactions," in *UbiComp '14: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014.

[14] L. Zhu, F. Lotte, G. Cui, J. Li, C. Zhou, and A. Cichocki, "Neural mechanisms of social emotion perception: An EEG hyper-scanning study," in *2018 International Conference on Cyberworlds*, 2018.

[15] A. M. Brouwer, I. V. Stuldreher, and N. Thammasan, "Shared attention reflected in EEG, electrodermal activity and heart rate," in *CEUR Workshop Proceedings, 2019 Socio-Affective Technologies: An Interdisciplinary Approach, SAT 2019*, 2019, pp. 27–31.

[16] M. Muszynski, T. Kostoulas, P. Lombardo, T. Pun, and G. Chanel, "Aesthetic highlight detection in movies based on synchronization of spectators' reactions," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 14, no. 3, pp. 1–23, 2018.

[17] K. J. Verdiere, F. Dehais, and R. N. Roy, "Spectral EEG-based classification for operator dyads' workload and cooperation level estimation," in *2019 IEEE International Conference on Systems, Man and Cybernetics*, 2019.

[18] C. Simar *et al.*, "Hyperscanning EEG and classification based on Riemannian geometry for festive and violent mental state discrimination," *Front. Neurosci.*, vol. 14, p. 588357, 2020.

[19] I. Chatterjee, M. Goršič, M. S. Hossain, J. D. Clapp, and Novak, "Automated classification of dyadic conversation scenarios using autonomic nervous system responses," *IEEE Trans. Affect. Comput.*, vol. 14, pp. 3388-3395, 2023.

[20] A. Darzi and D. Novak, "Automated affect classification and task difficulty adaptation in a competitive scenario based on physiological linkage: an exploratory study," *Int. J. Hum. Comput. Stud.*, vol. 153, p. 102673, 2021.

[21] Y. Ding, X. Hu, Z. Xia, Y. J. Liu, and D. Zhang, "Inter-brain EEG feature extraction and analysis for continuous implicit emotion tagging during video watching," *IEEE Trans. Affect. Comput.*, vol. 12, no. 1, pp. 92–102, 2021.

[22] I. Chatterjee, M. Goršič, J. D. Clapp, and D. Novak, "Automatic estimation of interpersonal engagement during naturalistic conversation using dyadic physiological measurements," *Front. Neurosci.*, vol. 15, p. 757381, 2021.

[23] A. Schwerdtfeger, "Predicting autonomic reactivity to public speaking: don't get fixed on self-report data!," *Int. J. Psychophysiol.*, vol. 52, pp. 217–224, 2004.

[24] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *J. Behav. Ther. Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[25] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, no. 3, pp. 354–381, 1996.

[26] K. Yan, "Feature selection with SVM-RFE," *MATLAB Central File Exchange*, 2015. https://www.mathworks.com/matlabcentral/fileexchange/50701-feature-selection-with-svm-rfe (accessed Aug. 15, 2023).

[27] M. L. Woody, C. Feurer, E. E. Sosoo, P. D. Hastings, and B. E. Gibb, "Synchrony of physiological activity during mother-child interaction: moderation by maternal history of major depressive disorder," *J. Child Psychol. Psychiatry*, vol. 57, no. 7, pp. 843–850, 2016.

[28] V. D. Novak, D. Hass, M. S. Hossain, A. F. Sowers, and J. D. Clapp, "Effects of adaptation accuracy and magnitude in affect-aware difficulty adaptation for the Multi-Attribute Task Battery," *Int. J. Hum. Comput. Stud.*, vol. 183, p. 103180, 2024.