A New Look at Breathing for Affective Studies

Nanfei Sun and Ioannis Pavlidis, *IEEE Fellow*

Abstract—In affective computing, breathing has seen lighter use than the heart and EDA channels. Several reasons have contributed to this, including difficulties in disambiguating affective from speech effects and perceived lack of generalizability. Here we report a framework that addresses these issues. The cornerstone of the framework is a comprehensive set of physiologically informed features, comprised of three groups: breathing depth, respiratory time quotient (RTQ), and breathing speed features. The breathing depth features capture either mental arousal or fear effects. The RTQ features capture speech production. The breathing speed features capture arousal effects due to emotional influences. The said framework appears to have broad applicability. In the naturalistic *Office Tasks 2019* dataset with speaking sessions, the said features used either in regression or random forest models led to robust classification of arousal ($\overline{\text{AUC}}$ in [0.75, 0.96]) stemming from three different conditions: a) mental-emotional stressor effected through a time-pressured knowledge task; b) pure mental stressor effected through a long knowledge task; c) mental-social stressor effected through a public speech task. In the stylized *CASE* dataset with silent sessions, the same features and algorithms led to solid classification of arousal ($\overline{\text{AUC}}$ in [0.71, 0.85]) stemming from scary vs. non-scary movie clips.

Index Terms—Affective computing, breathing, breathing features, tidal volume, respiratory time quotient, breathing rate, arousal, fight or flight, mental stressor, social stressor, public speech, multinomial logistic regression, random forest, knowledge work, fear

1 Introduction

In its quest to determine arousal and emotional states, affective computing uses an array of audio, video, and physiological channels [1]. For the latter, the emphasis is on peripheral physiology, which includes electrodermal activity (EDA), heart function, and breathing function. Among these three peripheral channels, the one that has received the least research attention is the breathing channel. For instance, among five well-known multimodal datasets - DEAP [2], SEED [3], DSdRD [4], AMIGOS [5], and WESAD [6] - only DEAP and WESAD feature a breathing channel. It is not entirely clear why this is the case. Irrespective of the reasons for the research community's benign neglect, breathing has some unique advantages with respect to the other two channels, but also presents distinct challenges. A key advantage of breathing lies in its natural features, like depth and speed, which, we posit, can deliver nuanced arousal and emotion classifications. A key challenge of breathing lies in the fact that is the channel most affected by volitional actions, like speaking [7].

Here we put forward an integrated methodology for selecting and analyzing breathing features in affective computing studies. This methodology fills a gap in the literature, by addressing the issue of speech influences on breathing and providing insights into the nature of one's aroused state. We show that our feature set can differentiate among arousals stemming from time pressure, sustained mental load, and public speaking - three scenarios largely associated with work in the knowledge economy [7]. As such, our method stands to greatly benefit knowledge work studies. Moreover, we demonstrate that our feature set can detect arousals stemming from fear responses, in the context of movie watching [8]. To facilitate broad use of the presented physiological framework, we make publicly available under one repository, the relevant code and curated datasets[1].

- *N. Sun is with the Department of Management Information Systems, University of Houston-Clear Lake, Houston, TX, 77058.*
- *I. Pavlidis is with the Affective and Data Computing Laboratory, University of Houston, Houston TX 77004. E-mail: ipavlidis@uh.edu*

Manuscript received on Nov. 16, 2022; 1st revision on December 31, 2023; 2nd revision on June 1, 2024.

1.1 Relevant Work

The literature relevant to the present research is divided into two branches - psychophysiology and affective computing. These two branches ought to be closely linked, informing one another; in this direction, there is still much room for improvement. Our work could be viewed as an effort to bring a psychophysiological perspective in the affective computing analysis of breath. Indeed, to appreciate our approach, one should first understand the nature of the breathing function. Breathing is controlled by two physiological systems, catering to the metabolic and behavioral needs of individuals, respectively [9]. The metabolic breathing control system is in charge of the body's basal metabolic needs and is located in the brainstem. Its mission is to regulate arterial blood gas through autonomic activation of the respiratory muscles. Thanks to the metabolic breathing control system, humans can breathe even when they are asleep or unconscious. In contradistinction, the behavioral breathing control system activates the respiratory muscles for needs unrelated to basal metabolism. Behavioral breathing control is effected through either actions yielding explicit breathing control or states that precipitate involuntary breathing influences. Examples of the former include speech, singing, and yoga, while examples of the latter include emotions and mental activity.

Dramatic manifestations of explicit breathing control include breathing as fast as possible and breath holding for as long as possible. In adults, fast breathing can reach up to 40 times resting ventilation, while breath hold can last 70 times the duration of a normal breath [9]. These are impressive

1. https://github.com/UH-CPL/AffectiveBreathing

demonstrations of the power of behavioral breathing control. This power, however, is checked by metabolic control, as the ultimate limit of the breath hold suggests. Park et al. have recently shed light in another aspect of the conflict between the metabolic and behavioral control of breathing. They showed that subjects initiate voluntary actions more frequently during expiration, while such breathing-reaction coupling is absent during externally triggered actions [10]. This has implications for speech production; it is also a stunning demonstration of the incursion of metabolic breathing control into the sphere of free will. Volitional breathing control was also found to be effective against stress and to hold promise for mental health improvement [11].

The competition between the metabolic and behavioral breathing control systems can create significant analytical problems, which are exemplified in speech. Subjects sometimes hyperventilate while they speak to accommodate parallel metabolic and speech production demands [9]. Hyperventilation - which is characterized by deep and/or rapid breaths - is not only associated with speech but also with stress [12]. Hence, the metabolic-behavioral competition confounds breathing signal characteristics, making challenging the detection of arousal in speaking subjects. It leaves open the following question: How much of the observed hyperventilation is due to speech production and how much is due to emotions? This is particularly problematic in the affective analysis of public speakers, who are often in a state of anxiety. For this reason, researchers who focused on physiological methods to study public speaking, resorted to the EDA and heart channels [7]. For example, Kusserow et al. proposed talk assistants to regulate stressful hyperarousal and improve the public speech experience [13]. Their talk assistants detect stress through EDA and heart sensing, providing personalized feedback conveyed through the speaker's on-body system or podium displays. As an alternative to the physiological measurement approach, some researchers tried to detect breathing rates from speech recordings, using cepstrogram matrix as a feature for classifying breath vs. non-breath speech frames [14]. Other investigators used psychometric methods to study anxiety in public speaking. For example, Kelsen found that extraversion, neuroticism, conscientiousness, and openness explained 10 to 23% of the variance for public speaking anxiety [15].

For nonspeaking subjects, breathing analysis becomes easier, as a major confounding factor is out. Reports in the psychophysiological literature indicate that anxious subjects in a nonspeaking state exhibit increases in the speed, amplitude, and irregularity of their breathing [16], that is, they hyperventilate. In affective computing, much of the breathing analysis for nonspeaking subjects has been carried out in the context of the two dimensional arousal-valence model of human emotions [17]. This line of inquiry arrived at results that are in agreement with the psychophysiological literature, demonstrating that breathing patterns encapsulate rich emotional information. For instance, researchers found that depth and speed of human breathing varies with emotions; hyperventilation, in the form of deep and rapid breathing, indicates excitement that may be caused by anxiety, happiness or anger [17]. Moreover, affective computing investigators used machine learning (ML) methods, such as sparse auto-encoder and logistic regression [18], to classify emotions from breathing data. A physiological dataset that is often used to benchmark emotion classification is DEAP [2]. A deep learning framework operating on DEAP's breathing signals achieved arousal and valence classification accuracy of 85.89% and 83.72%, respectively [18]. However, these results are reported in the absence of cross-validation, without additional important metrics (e.g., F1), and without open code, thus leaving a number of unanswered questions.

Instead of using breathing as a stand-alone channel, some researchers combined breathing with other physiological variables to feed ML algorithms for the determination of arousal and valence levels. Wu et al. fed a support vector machine with skin conductance, breathing, electrocardiogram (ECG), and electroencephalogram (EEG) responses to classify arousal into three levels: low, optimal, high [19]. Valenza et al. used non-linearly extracted features from skin conductance, breathing, and ECG to classify valence and arousal at multiple levels [20]. A strong motivation behind the efforts to determine arousal through physiological variables is the desire to provide biofeedback to subjects. Biofeedback aims to help subjects develop self-awareness of stressful situations and improve their coping styles [21].

In addition to the influence of emotions on breathing, mental activity has also been shown to significantly affect breathing patterns. In the psychophysiological literature, Asmussen reported that an activity as simple as opening one's eyes to read, increased the subjects' breathing by an average of 16% [22]. Shea found that with respect to baseline, mental arithmetic induces a more rapid, shallow breathing pattern, accompanied by an increase in oxygen consumption [9]. Carroll et al. studied how this pattern regresses, showing that increases in mental task difficulty were associated with increases in breathing rate and subjects' self-reports of active engagement and arousal [23]. In a broader context, other researchers also corroborated that the more involved the subjects are in a task, the more their breathing rate increases [24].

In agreement with the aforementioned positive psychophysiological reports, some affective computing investigators reported that mental load during computer-based tasks can be assessed though physiological measures [25]. Some other affective computing researchers, however, published results that cast doubt on breathing as a mental activity predictor. McDuff et al. used heart rate, breathing rate, and heart rate variability (HRV) to differentiate mental stress between two computer-based tasks: ball control and card sorting. They found breathing rates during these two tasks to be significantly different than rest periods, indicating the stressful nature of mental work. Nevertheless, they did not detect any significant differences in the mean heart and breathing rates between the two tasks, concluding that heart and breathing rates alone are not discriminating enough to identify different types of mental stress [26].

1.2 Contributions

In summary, breathing has been largely used in affective computing research to track emotions or mental activity of nonspeaking subjects. In emotion classification, results were promising and in agreement with psychophysiological

investigations [2], [17], [18]. In mental activity classification, results were mixed [25], [26]. For speaking subjects, there is little reporting due to the challenging nature of the problem [7]. The novel framework of breathing analysis that we introduce in this paper makes two major contributions: First, it can handle both speaking and nonspeaking conditions; second, it addresses the problem of mental activity classification. The said framework draws its power from the detangling of the metabolic, the volitional behavioral (speech), and the involuntary behavioral (emotions and mental loading) components of breathing. Viewing things from an application perspective, the present paper contributes a potent analytical tool for naturalistic affective computing studies in knowledge work environments, where speaking and non-speaking states freely intermix.

Our method manages to account for several contributing sources of breathing patterns by using a comprehensive array of physiologically informed features. The set of features includes breathing depth, phase shift, and speed measures. We show that in sitting subjects: a) Breathing depth features help to account for mental loading as well as sighing associated with fear stimuli. b) Phase shift features, revealing of inspiration-expiration patterns, help to account for the presence/absence of speech. c) Breathing speed features help to account for various emotional effects. Breathing depth features have not been widely used in the affective computing literature, which partly explains the mixed record of prior research in differentiating mental stressors. Breathing phase shift features have not been used at all in the affective computing literature, which explains the difficulty of prior work in dealing with speech. By contrast, breathing speed features have been widely used in the affective computing literature, which partly explains the relative success of prior breathing analysis in nonspeaking emotional contexts.

To further enhance the discriminating power of our feature set, we factor out the basal metabolic contribution in each breathing feature, thus reducing inter-individual variability. To achieve this, our method uses an estimate of people's baseline, where only their metabolic breathing control system is at work. Dividing any observed breathing level by the corresponding baseline level, yields a normalized ratio that tracks the pure behavioral contributions to breathing. To estimate the breathing baseline of subjects, they must be put for a few minutes into a relaxed condition, characterized by minimal sensory activation, motor action, and mental activity [24]. Presently, an adequate baseline procedure is often lacking in affective computing studies, with serious downstream implications for the reliability of breathing analysis [7].

As we are looking to establish the value added of several features, it is important our analytical model to be explainable. Accordingly, we optimize and test the proposed feature set through a regression modeling process. The model brings to the fore the discriminating power of distinct feature groups against the mental stressors in *Office Tasks 2019* - a naturalistic knowledge work dataset [7]. As far as we can tell, this is the first time that such a variety of mental stressors is successfully classified through breathing, even in the presence of speech.

To investigate the broader value of our feature set, we also test it on a second open dataset, associated with controlled viewing of emotion-eliciting movie clips [8]. This dataset has little to do with naturalistic knowledge work in the presence/absence of speech, for which our feature set was primarily designed. Nevertheless, our feature set's respectable performance in this second test case, demonstrates the flexibility of our framework. Finally, to examine the relative performance of our framework in different ML algorithms, we carry out classification not only with regression but also with random forest in both datasets.

In the remainder of the paper, we organize the description of our research as follows: First, we describe the datasets we used to test our framework, explaining the rationale for their selection (section 2). Next, we provide details of our signal processing and feature extraction methods (section 3). Then, we present the results of our exploratory and model-based analysis (section 4), before closing with a balanced discussion (section 5).

2 DATASETS

2.1 Office Tasks 2019 Dataset

The primary dataset we used to showcase our framework was *Office Tasks 2019*, which is publicly available in the Open Science Framework (OSF)[2]. We reported the descriptor of this dataset in [7], where one can find detailed information. Here we outline the experimental design that gave rise to this dataset, in order to help the reader contextualize our analytic results.

The dataset contains multimodal data from $n = 62$ university student participants, who carried out a series of knowledge work tasks under an institutionally approved experimental protocol. The protocol featured the following experimental sessions:

Resting baseline (RB). Participants were asked to close their eyes and think of relaxing nature scenes for ~ 4 min. The aim was to minimize behavioral effects on breathing and other peripheral physiological channels, leaving only basal metabolic needs at play. As such, RB meant to serve as reference point for arousal normalization.

Short task (ST). This was a mixed mental-emotional task. Participants were given 5 min to write a short essay expressing their opinions on the subject of competition vs. collaboration. The aim was to simulate short bursts of mental activity, dominated by time pressure. A timer on the participants' screen was amplifying the intended sense of urgency.

Long task (LT). This was a mental task. Participants were given 50 min to compose an essay on the topic of technological singularity, that is, when machines overtake human intelligence. The aim was to simulate prototypical knowledge work of non-trivial duration that induces mental load.

Presentation (PR). This was a mixed mental-social task. Participants were asked to deliver a short speech (~ 4 min) about their LT essay, in front of a three-judge panel who

2. https://osf.io/zd2tn/

attended via Skype. The aim was to simulate a well-known socio-cognitive stressor in knowledge work environments. The fact that the participants were speaking, complicated breathing analysis.

Altogether, the protocol associated with the *Office Tasks 2019* dataset features speaking and nonspeaking sessions of various knowledge work tasks. Hence, the dataset fits the aims of our research because it allows: a) to examine the discriminating power of our method among different types of mental activities, and b) to test our method's ability to account for the breathing effect of speech. Furthermore, the dataset has been peer-reviewed, and contains data that underwent quality control and validity checks [7].

During the experimental sessions, several imaging and wearable sensors were continuously recording the participants' observational and physiological data. In this paper, we focus on the breathing data channel, which was collected via a Zephyr Bio-Harness 3.0 device (Zephyr Technology, Annapolis, MD) the participants worn on the chest. We used from Zephyr the raw breathing signals, captured at full resolution (25 Hz). These signals are made up of the instantaneous values of the Bio-Harness's length, which track the evolving rib cage circumference. Out of the 62 participants included in the *Office Tasks 2019* dataset, full resolution breathing data for five participants (T021, T032, T035, T037, T066) were either missing altogether or at least the RB session was missing. For one participant (T113) the RB session was not properly done. Hence, we carried out our analysis on the usable dataset of $n = 56$ participants (age 23 ± 7.37), representing nearly 58 hours of breathing recordings. The mean $\pm$ standard deviation duration of each experimental session was as follows: for RB, 3.26 ± 0.94 min; for ST, 4.68 ± 0.00 min; for LT, 49.23 ± 3.24 min; and for PR, 2.56 ± 0.53 min. With one experimental session nearly an hour-long (LT), and another experimental session featuring natural speech (PR), *Office Tasks 2019* is unique among the currently available affective datasets, and highly appropriate to test the expressed aims of our proposed breathing feature set.

2.2 CASE Dataset

To demonstrate the scalability of our framework, we opted to apply it to a second dataset. To fulfill this goal, however, we ran into significant practical difficulties. Breathing is the slowest peripheral physiological channel. An average sitting person exhibits each minute about 60 heartbeats, but only 12 breaths [27], [28]. Just for this reason alone, proper testing of sets of breathing features requires long observation sessions. Furthermore, our breathing feature set was designed to differentiate arousal responses generated from naturally evolving stressful conditions, where speech is often involved. Accordingly, it fits best naturalistic experiments. Unfortunately, there is a dearth of datasets with breathing function recordings from naturalistic experiments with long sessions. The *Office Tasks 2019* dataset [7] is an exception, representing a new generation of studies. Popular conventional datasets, like DEAP [2], are from highly stylized experiments, where subjects were recorded while watching video clips of minuscule duration (e.g., one minute). Moreover, a significant portion of the feature set we propose, are normalized features meant to ameliorate inter-individual differences. Such features require individual baseline measurements for their computation. Unfortunately, many open datasets either do not have baseline measurements or they have trivial baseline measurements. For instance, DEAP's baseline measurements last just 10 seconds, which is the time needed for only 1-2 breaths - hardly a solid statistical basis for normalization.

Given the aforementioned considerations, we chose to work with the Continuously Annotated Signals of Emotion (*CASE*) dataset [8], as the best available option. This is still the product of a highly stylized experimental design, where participants are recorded while watching specific movie clips. Nevertheless, the average duration of the *CASE* movie clips is close to 3 minutes, which is an improvement over the one-minute clips featured in other datasets. Furthermore, the movie clips were drawn from lists reported in reputable psychological studies [29], [30], offering some assurances about their arousal effect. Importantly, the *CASE* dataset features clips that can serve as baseline sessions with mean duration 1.84 min, thus offering a credible basis for feature normalization.

The experimental protocol included eight movie clips, which were watched in random order by $n = 30$ participants. The emotional ratings of the movie clips are as follows: two are boring, two are amusing, two are relaxing, and two are scary. Since our framework is unimodal, encompassing only breathing features, we focus on arousal and not valence classification, to avoid ambiguity [31]. As Fig. 5 in the *CASE* paper suggests [8], the two scary movies generated arousal that was clearly higher than the largely overlapping arousal of the amusing and relaxing movies. The arousal associated with the boring movies tended to be the lowest and was set as baseline. Accordingly, we constructed two classes for our testing: the Scary class associated with breathing signals produced by the two scary movies, and the Non-Scary class associated with breathing signals from the amusing and relaxing movies. The *CASE* dataset includes a little over 12 hours of breathing recordings, compared to the 58 hours of the *Office Tasks 2019* dataset. The mean $\pm$ standard deviation duration of Non-Scary vs. Scary movies is as follows: for Non-Scary, 2.58 ± 0.39 min; for Scary, 2.84 ± 0.62 min.

3 METHODS

3.1 Breathing Signal Processing

To prepare breathing signals for analytic work, first we reduce noise through filtering, and then we capture fundamental functional information by identifying breathing cycles. After these signal processing steps, we extract features for analysis (Fig. 1).

3.1.1 Breathing Signal Filtering & Normalization

Breathing rates for adults take values in the range 0.1Hz (6 breathing cycles/min) to 1Hz (60 breathing cycles/min) [32]. Accordingly, we apply two Gaussian filters, as per Equation (1) [33], on the raw breathing signals S:

$$f_c = \frac{F_s}{2\pi\sigma}, \tag{1}$$

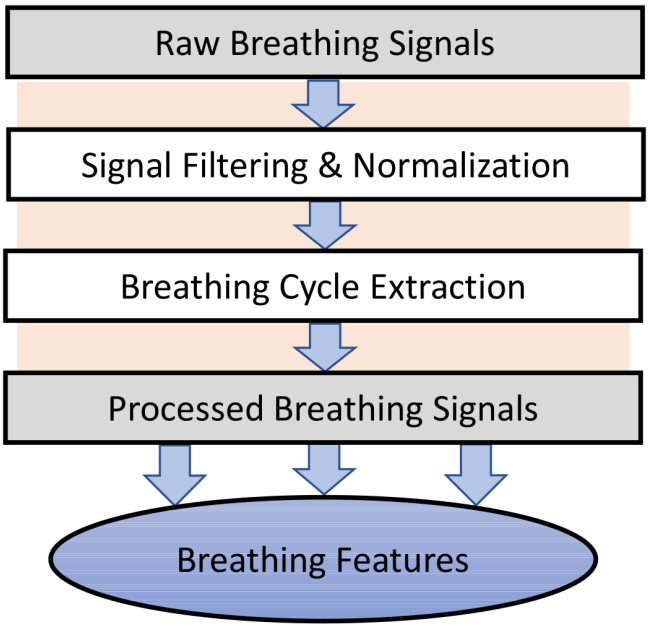

Fig. 1: **Methodological flow that precedes our analysis.** Breathing signal filtering is followed by breathing cycle extraction, before the computation of features takes place.

where f_c is the desired cut-off frequency, σ is the filter's standard deviation, and $F_s = 25$ Hz is the breathing signal sample rate used in Zephyr Bio-Harness. To eliminate frequencies $>$ 1Hz from a raw breathing signal S, we set $f_c \equiv f_{high} = 1$; from Eq. (1), the standard deviation σ_{high} of this high-pass filter, which is a measure of its size, is computed to be:

$$\sigma_{high} = \frac{F_s}{2\pi f_{high}} = \frac{25}{2\pi * 1} \approx 4 \text{ samples.} \quad (2)$$

To eliminate frequencies $>$ 0.1Hz from a raw breathing signal S, we set $f_c \equiv f_{low} = 0.1$; from Eq. (1), the standard deviation σ_{low} of this low-pass filter, which is a measure of its size, is computed to be:

$$\sigma_{low} = \frac{F_s}{2\pi f_{low}} = \frac{25}{2\pi * 0.1} \approx 40 \text{ samples.} \quad (3)$$

Hence, we correspondingly obtain two filtered breathing signals S'_{high} and S'_{low}. The normalized breathing signal S'', band-passed between 0.1Hz and 1Hz, is computed to be:

$$S'' = S'_{high} - S'_{low}. \quad (4)$$

Figure 2 shows results of the filtering process applied on the breathing signals of participant T003 in the *Office Tasks 2019* dataset. Note that Eq. (4) detrends S'', centering it on the zero line. Such a signal transformation facilitates tidal volume checking during the extraction of breathing cycles.

3.1.2 Breathing Cycle Extraction

Figure 3 illustrates the landmark points (valleys and peak) that delineate a breathing cycle, complete with their anatomical correspondence. We use the following algorithm to identify breathing cycles in the normalized signal $S''(i, j, t)$ of participant i, for session j, over time t.

Step 1 - Estimate start/end of breathing cycle (valleys): Find two consecutive local minima S''_{in} and S''_{ex}; potentially, these are the breathing cycle's inspiratory start and expiratory end points. To locate such minima, scan the signal in time windows that correspond to the minimum possible duration of a breathing cycle. Given that our high-pass filter has been set to 1 Hz, the scan window must be set to 1 second. In such a window, find the minimum value S''_{in}. Then, test if S''_{in} conforms to a standard valley detection template: $S''(t-2) > S''(t-1) > S''_{in}(t) < S''(t+1) < S''(t+2)$. In essence, this template is the discrete definition of the derivative test for locating curve minima. We validated such a valley detection template for physiological signals in [34]. If S''_{in} conforms to the valley template, then it is marked as the breathing cycle's potential inspiratory start point. If not, the same procedure is applied in the next scan window and so on, until a legitimate breathing cycle start point is identified. Following that, the same procedure is applied in subsequent scan windows, in search this time of a second valley point S''_{ex}, which will be the breathing cycle's potential expiratory end point.

Step 2 - Demarcate inspiration from expiration (peaks): Find the maximum signal value between S''_{in} and S''_{ex}; this is the peak point S''_p that likely demarcates the inspiration from the expiration phase in the breathing cycle. Figure 3a shows the detected valleys (S''_{in}, S''_{ex}) and peak (S''_p) in a breathing cycle of participant T003 in the *Office Tasks 2019* dataset; Fig. 3b shows the anatomical correspondences of these landmarks.

Step 3 - Proceed: Repeat Step 1 - Step 2 until the end of signal $S''(i, j, t)$ is reached. Figure 4 shows examples of the algorithm's performance on T003's breathing signals.

Step 4 - Eliminate spurious valleys and peaks: Although the band-pass filter eliminates a significant amount of noise in the breathing signals, some residual noise remains, which may give rise to spurious valleys and peaks. We eliminate such spurious valleys and peaks by leveraging physiological information in the form of tidal volume (TV). The participants' TV is the amount of air they inhale and exhale in a standard breath [35]. The best session to estimate a participant's TV in the *Office Tasks 2019* dataset is the RB session, where the participants are relaxed and free of any behavioral intervention. The difference between the median peak and valley in the RB session of participant i provides a reasonable approximation to individual TV. Due to behavioral interventions in the other sessions of the experiment, the participants are likely to hyperventilate at times in sessions ST, LT, and PR. Hyperventilation can reduce TV by over 50% [36]. Accordingly, we take 40% of TV as a safe threshold, symmetrically arranged ($\pm 20\%$ TV) around the zero line of the normalized breathing signal (Fig. 3a). The valleys and peaks that fall within the threshold band indicate minimal breathing cycles that are physiologically unlikely even under hyperventilating conditions; thus, they are eliminated from consideration.

Step 5 - Eliminate redundant valleys and peaks: By the definition of breathing cycle, there should be only one peak between two valleys and only one valley between two

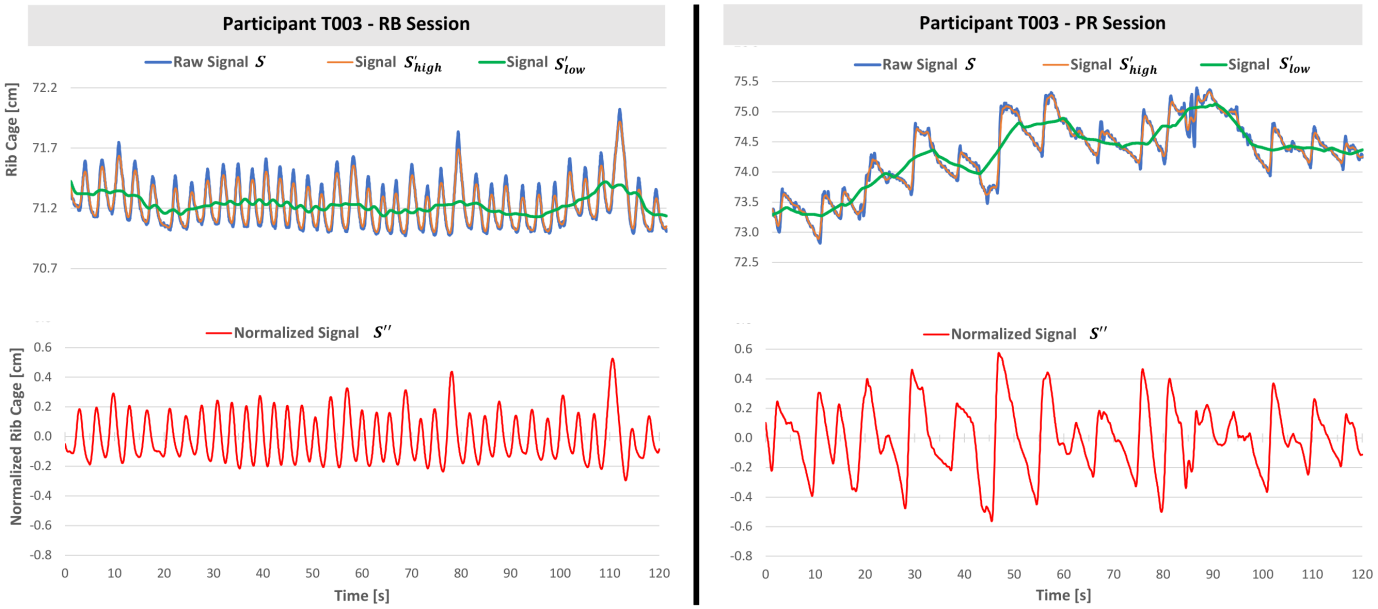

Fig. 2: **Breathing signal filtering and normalization. Left panels:** Raw, filtered, and normalized breathing signals of participant's T003 RB session in the *Office Tasks 2019* dataset. Superimposed on the raw signal S are the high- and low-pass filtered signals, S'_{high} and S'_{low}, respectively, whose difference produces the normalized signal S''. **Right panels:** Same information as in the left panels, but for the PR session of participant T003.

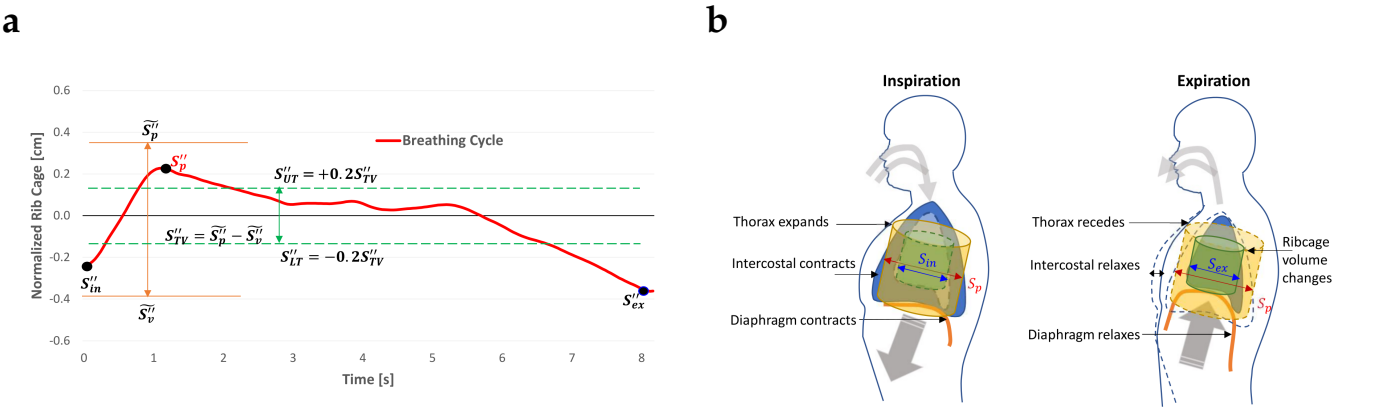

Fig. 3: **Normalized signal and anatomical illustration of breathing cycle. a.** In a breathing cycle of participant T003 in session RB of *Office Tasks 2019*, the detected valleys S''_{in}, S''_{ex} and peak S''_p are shown in black and red font, respectively, on the normalized signal. The threshold band, computed as 40% of the participant's T003 tidal volume S''_{TV}, is demarcated by the hyphenated green lines. Within this band, any valleys and peaks are considered physiologically unlikely and ignored. **b.** Correspondence between rib cage evolution and breathing cycle landmark points. The inspiratory start S_{in}, expiratory end S_{ex}, and peak S_p states are expressed in their actual (i.e., non-normalized) form.

peaks. A quality control (QT) script checks for this condition across the $S''(i, j, t)$ signal; where the condition is violated, QT keeps only the lowest valley (or highest peak), while discarding the redundant valleys (peaks). These redundant landmarks in the breathing signals are often the results of asynchronous breathing patterns, when the chest and abdomen are not moving in unison [37]. Figure 4d gives an example of the algorithm's handling of spurious and redundant valleys/peaks. Figure 5 shows how the algorithm worked step by step to produce the result shown in Fig. 4d.

Figures 6a and 6b show the breathing cycles identified by this five-step algorithmic process for participants T003 and 1 in the *Office Tasks 2019* and *CASE* datasets, respectively. To facilitate cycle viewing in each experimental session, we translate cycles so that are separated by a few points from each other on the vertical axis. Accordingly, the vertical axis shows relative and not absolute rib cage displacement. The LT session in the *Office Tasks 2019* dataset has a lot more cycles than the other sessions due to its much longer length. In fact, only 120 out of the 379 LT cycles are shown in Fig. 6a, as it was not practical to fit the rest in the panel. In each session displayed in Fig. 6, the breathing cycles appear congruent - a sign of accurate extraction. One can observe the large depth and asymmetric form of the cycles in the

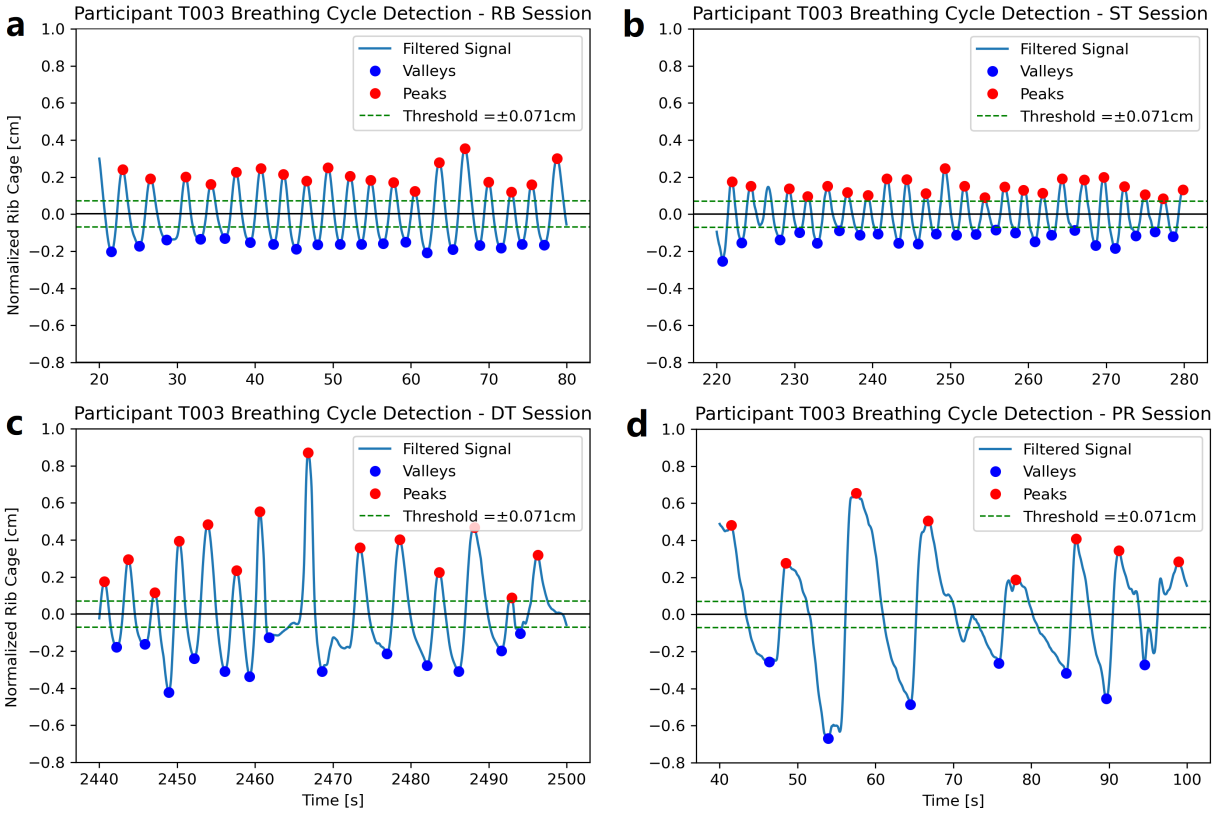

Fig. 4: **Breathing cycle detection for participant T003 in the *Office Tasks 2019* dataset.** The algorithmically detected valleys and peaks are marked in blue and red, respectively. The threshold band, identifying the portion of the signal where the appearance of valleys and peaks is physiologically unlikely, is demarcated between hyphenated green lines. Any valleys/peaks identified within this band are dropped from consideration. **a.** RB session sample. **b.** ST session sample. **c.** LT session sample. **d.** PR session sample, where between 110 s and 120 s there is a peak and a valley that fall within the threshold band; thus, they are ignored.

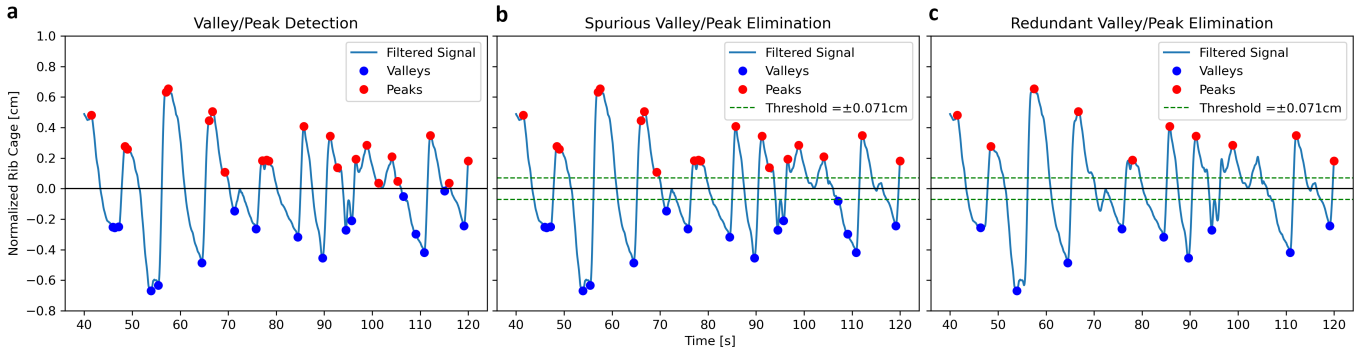

Fig. 5: **Three-stage estimation of valleys and peaks for participant T003 in session PR of the *Office Tasks* 2019 dataset.** **a.** Detection of valleys and peaks. **b.** Elimination of spurious valleys and peaks that fall within the threshold band. **c** Elimination of redundant valleys and peaks.

PR session, where participants speak. Deep cycles are also abundant in the Scary session, with a fear-related sighing effect apparent in the blue waveform.

3.2 Breathing Feature Computation

After having extracted all cycles in breathing signal $S''(i, j, t)$, we compute three groups of features that tend to capture different behavioral aspects. First, we compute breathing depth features for capturing increased metabolic demands beyond the basal level. In sitting subjects, such demands are typically associated with mental activities. Occasionally, breathing depth features capture sighing effects associated with fear responses Second, we compute phase shift features for capturing any rearrangement between inspiration and expiration. We expect phase shift features to contribute in the differentiation between speaking and

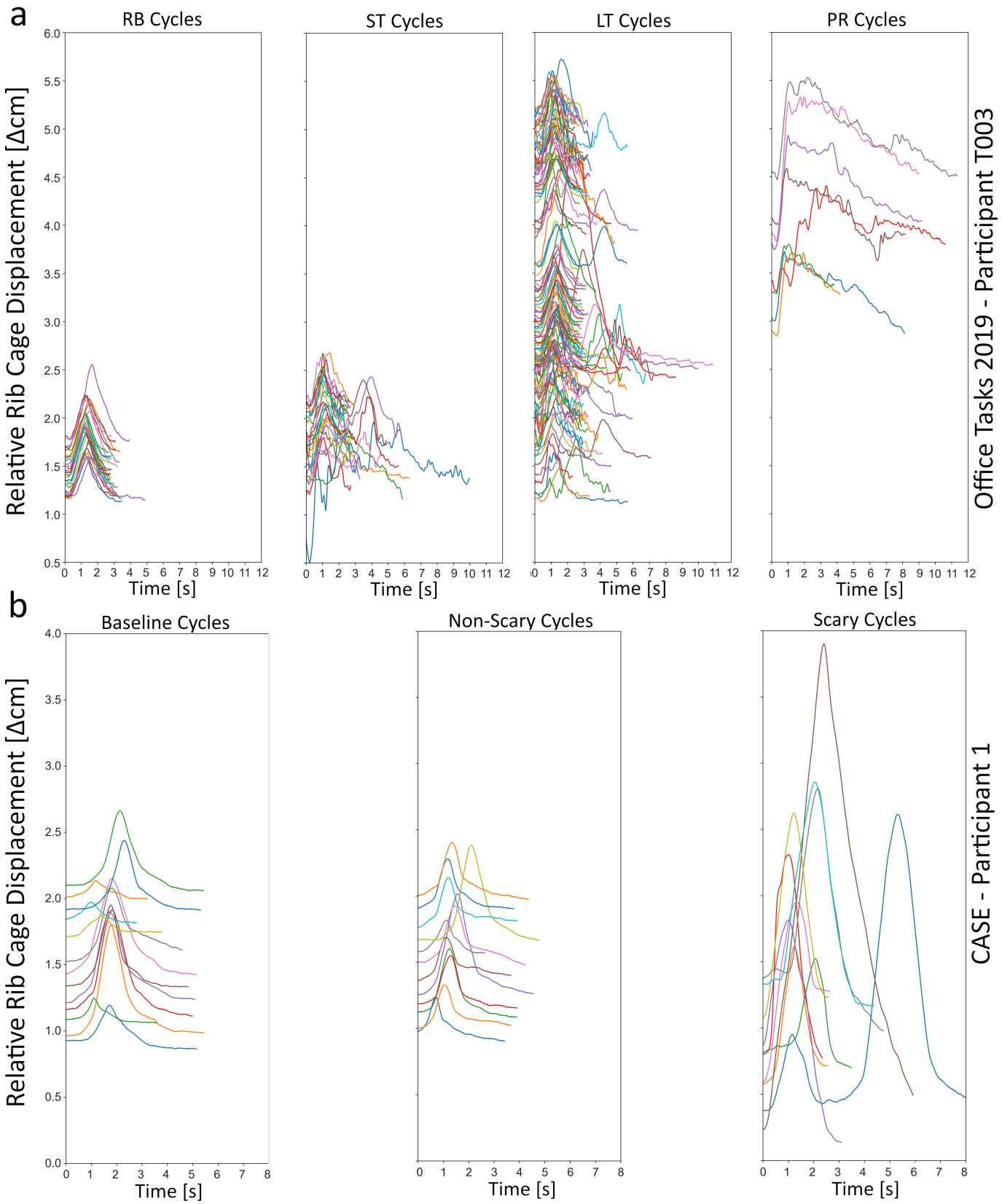

Fig. 6: **Algorithmically extracted breathing cycles. a.** For participant T003 in the *Office Tasks 2019* dataset: all cycles of sessions RB, ST, and PR, as well as 120 out of the 379 cycles in the long LT session. **b.** For participant 1 in the *CASE* dataset: all cycles of sessions Baseline, Non-Scary, and Scary. We applied a 'jittering' function on the vertical axis to avoid cycle overlapping. For this reason, the vertical axis shows relative (not absolute) rib cage displacement. Within each experimental session, cycles appear congruent - a sign of accurate extraction.

nonspeaking states. Third, we compute breathing speed features for capturing 'fight or flight' effects, associated with emotional states. For all features, we compute normalized versions that are 'free' of the basal metabolic component of breathing. This comprehensive set of features aims to endow our ML models with strong discriminating capabilities, by providing them with breathing information relevant to mental loading, fear responses, speaking/nonspeaking state, and emotional influences.

3.2.1 Breathing Depth Features

The breathing depth feature group includes breathing volume and breathing amplitude features. We calculate a correlate of the breathing volume per cycle, and then we derive a correlate of the breathing volume per second from it. With the understanding that these measures are correlates and not exact estimates, we will call them breathing volume per cycle and breathing volume per second, for brevity.

Our calculation is rooted in the respiratory inductive plethysmography (RIP), which was introduced by Kono and Mead in 1967 and has since been used for monitoring patients in clinical settings [38]. In RIP, the breathing volume is derived by measuring the changes brought about by respiration on two inductive belts placed around the abdomen and rib cage of a subject. In the *Office Tasks 2019* experiment, however, there was only one inductive belt, set up to measure rib cage changes. The *CASE* experiment, also featured a single inductive belt, as is the case with most affective computing experiments involving breathing measurements. Hence, we needed to simplify the RIP model. Per Kono and Mead's study, the rib cage accounts for about half of the tidal volume when the total volume change is constrained. Since the participants in the studies of interest to us remain seated and our data processing eliminated spurious breathing cycles, rib cage volume changes are proportional to the total lung volume changes (Fig. 3b).

We model the rib cage shape as a cylinder, following a methodology reported by Augusti in [39]. Accordingly, the rib cage volume V can be calculated as $V = \pi r^2 h$, where r and h are the radius and height of the rib cage, respectively. Taking the approximation $h = 2r$ [39], simplifies the rib cage volume equation to $V = 2\pi r^3$. Replacing the radius r with the perimeter P ($r = P/2\pi$), yields the final rib cage volume equation $V = P^3/4\pi^2$. Each breathing cycle features two phases, the inspiration phase in and the expiration phase ex. Per our standard symbolism, S_{in}, S_p, and S_{ex} denote the inspiration valley, the peak, and the expiration valley, respectively, in a breathing cycle; they express the corresponding length of the inductive belt in cm, which is tantamount to the rib cage. Note that we do not use the normalized values S''_{in}, S''_p, and S''_{ex}, but the corresponding filtered values S'_{in}, S'_p, and S'_{ex}, exactly because they are associated with the actual rib cage measurements.

We calculate the inspiration volume δV_{in} by summing up infinitesimal volumes computed every second $t \in [1, \ldots, n]$ of the inspiration:

$$\delta V_{in} = \frac{S'^3_p(t) - S'^3(t-1)}{4\pi^2} + \ldots + \frac{S'^3(t-n-1) - S'^3_{in}}{4\pi^2} \quad (5)$$

We calculate the expiration volume δV_{ex} by summing up infinitesimal volumes computed every second $t \in [1, \ldots, m]$ of the expiration:

$$\delta V_{ex} = \frac{S'^3_p - S'^3(t+1)}{4\pi^2} + \ldots + \frac{S'^3(t+m-1) - S'^3_{ex}}{4\pi^2} \quad (6)$$

We estimate the total volume δV_c for breathing cycle c by adding δV_{in} and δV_{ex}:

$$\delta V_c = \delta V_{in} + \delta V_{ex}. \quad (7)$$

Using Eq. (7), we estimate the volume of each breathing cycle and compute the following features: $\overline{\delta V}_c(i,j) \equiv$ mean breathing volume per cycle for each treatment j of each participant i; SD$[\delta V_c(j)] \equiv$ the standard deviation of the participants' breathing volumes per cycle for each treatment j; $\|\overline{\delta V}_c(i,j)\|_{\mathcal{N}} \equiv$ the normalized mean breathing volume per cycle for each treatment j of each participant i, where normalization is realized by dividing with the participant's mean breathing volume per cycle in the RB session. The same breathing volume features are also computed on a per second basis. We also compute the mean, normalized mean, and standard deviation for the breathing waveform amplitude, which is a direct measure of breathing depth. The statistical formulas for the breathing volume and amplitude features are given in Table 1.

3.2.2 Breathing Phase Shift Features

To track shifts between the inspiration and expiration phases of breathing cycles, we use the respiratory time quotient (RTQ). RTQ is a concept first introduced by Conrad et al. [40]. It expresses the relation of inspiration and expiration as the duration of inspiration divided by that of expiration. Here we calculate the RTQ value for each breathing cycle through Eq. (8):

$$RTQ = \frac{T_p - T_{in}}{T_{ex} - T_p}, \quad (8)$$

where T_{in} is the timestamp at point S'_{in} and T_{ex} is the timestamp at point S'_{ex}, T_p is the timestamp at point S'_p (Fig. 3). Then, we compute the mean and normalized mean of RTQ for each treatment j of each participant i; we also compute for each treatment j the standard deviation of the RTQs of the participants in that treatment (Table 1).

3.2.3 Breathing Speed Features

We compute breathing speed features, first by computing the mean and normalized mean of breathing rate for each treatment j of each participant i. For each treatment j, we also compute the standard deviation of the breathing rates of the participants in that treatment. Similarly, we compute the mean, normalized mean, and standard deviation for the breathing waveform length, which indicates breathing duration and is related to breathing frequency. Table 1 shows the statistical formulas for these and all the other features used in our analytical framework.

4 Analytic Results

In the *Office Tasks 2019* dataset, the knowledge work tasks ST, LT, and PR are challenging and are expected to increase arousal [7]. The nature of the challenge, however, differs

TABLE 1: **Breathing function features used in our analytical framework.** The breathing cycle index is c and the total number of breathing cycles in a session is denoted as C. Features listed on a white background belong to the breathing depth group. Features listed on a light gray background belong to the breathing phase shift group. Features listed on a dark gray background belong to the breathing speed group.

C	Features	Description	Symbol	Domain*
1	$\overline{\delta V}_c(i,j) = \frac{\sum_{c=1}^{\mathcal{C}} \delta V(i,j,c)}{\mathcal{C}}$	Mean breathing volume/cycle	BVC_AVG	S'
2	$\text{SD}[\delta V_c(i,j)] = \sqrt{\frac{\sum_{c=1}^{C}[\delta V(i,j,c) - \overline{\delta V}_c(i,j)]^2}{C}}$	SD of breathing volumes/cycle	BVC_SD	S'
3	$\|\overline{\delta V}_c(i,j)\|_{\mathcal{N}} = \frac{\overline{\delta V}_c(i,j)}{\overline{\delta V}_c(i,\text{RB})}$	Normalized $\overline{\delta V}_c(i,j)$	BVC_NORM	S'
4	$\overline{\delta V}_t(i,j) = \frac{\sum_{c=1}^{\mathcal{C}}[\delta V(i,j,c)/t_c]}{C} = \frac{\sum_{c=1}^{\mathcal{C}} \delta V_t(i,j,c)}{C}$	Mean breathing volume/second	BVT_AVG	S'
5	$\text{SD}[\delta V_t(i,j)] = \sqrt{\frac{\sum_{c=1}^{C}[\delta V_t(i,j,c) - \overline{\delta V}_t(i,j)]^2}{C}}$	SD of breathing volumes/second	BVT_SD	S'
6	$\|\overline{\delta V}_t(i,j)\|_{\mathcal{N}} = \frac{\overline{\delta V}_t(i,j)}{\overline{\delta V}_t(i,\text{RB})}$	Normalized $\overline{\delta V}_t(i,j)$	BVT_NORM	S'
7	$\overline{WA}(i,j) = \frac{\sum_{c=1}^{\mathcal{C}} WA(i,j,c)}{\mathcal{C}}$	Mean waveform amplitude	WA_AVG	S'
8	$\text{SD}[WA(i,j)] = \sqrt{\frac{\sum_{c=1}^{C}[WA(i,j,c) - \overline{WA}(i,j)]^2}{C}}$	SD of waveform amplitudes	WA_SD	S'
9	$\|\overline{WA}(i,j)\|_{\mathcal{N}} = \frac{\overline{WA}(i,j)}{\overline{WA}(i,\text{RB})}$	Normalized $\overline{WA}(i,j)$	WA_NORM	S'
10	$\overline{RTQ}(i,j) = \frac{\sum_{c=1}^{\mathcal{C}} RTQ(i,j,c)}{\mathcal{C}}$	Mean RTQ	RTQ_AVG	S'
11	$\text{SD}[RTQ(i,j)] = \sqrt{\frac{\sum_{c=1}^{C}[RTQ(i,j,c) - \overline{RTQ}(i,j)]^2}{C}}$	SD of RTQs	RTQ_SD	S'
12	$\|\overline{RTQ}(i,j)\|_{\mathcal{N}} = \frac{\overline{RTQ}(i,j)}{\overline{RTQ}(i,\text{RB})}$	Normalized $\overline{RTQ}(i,j)$	RTQ_NORM	S'
13	$\overline{BR}(i,j) = \frac{\sum_{c=1}^{\mathcal{C}} BR(i,j,c)}{\mathcal{C}}$	Mean breathing rate	BR_AVG	S'
14	$\text{SD}[BR(i,j)] = \sqrt{\frac{\sum_{c=1}^{C}[BR(i,j,c) - \overline{BR}(i,j)]^2}{C}}$	SD of breathing rates	BR_SD	S'
15	$\|\overline{BR}(i,j)\|_{\mathcal{N}} = \frac{\overline{BR}(i,j)}{\overline{BR}(i,\text{RB})}$	Normalized $\overline{BR}(i,j)$	BR_NORM	S'
16	$\overline{WL}(i,j) = \frac{\sum_{c=1}^{\mathcal{C}} WL(i,j,c)}{\mathcal{C}}$	Mean waveform length	WL_AVG	S'
17	$\text{SD}[WL(i,j)] = \sqrt{\frac{\sum_{c=1}^{C}[WL(i,j,c) - \overline{WL}_c(i,j)]^2}{C}}$	SD of waveform lengths	WL_SD	S'
18	$\|\overline{WL}(i,j)\|_{\mathcal{N}} = \frac{\overline{WL}(i,j)}{\overline{WL}(i,\text{RB})}$	Normalized $\overline{WL}(i,j)$	WL_NORM	S'

among tasks, and the question is if the comprehensive set of breathing features in Table 1 can capture these differences while also accounting for speech effects. There is also a question as to whether our feature set can capture the fear responses in the *CASE* dataset. The ensuing exploratory and model analysis addresses these questions.

4.1 Exploratory Analysis

4.1.1 Exploratory Analysis on Office Tasks 2019

Exploratory analysis brings to the fore a consequential interplay between breathing rate and breathing volume in the various tasks of the knowledge work experiment associated with *Office Tasks 2019*. Normalized mean breathing rate (BR_NORM) increases in ST, while recedes in LT, and recedes even further in PR, dropping below baseline levels (Fig. 7-E1). Accordingly, if one were to use exclusively BR_NORM to track arousal, as is often the case in the literature [7], s/he would have concluded that ST is the task that produces the strongest arousal responses, while PR is a non-stressful task! We observe, however, that the normalized mean breathing volume per second (BVT_NORM) increases in LT with respect to ST, and increases even further

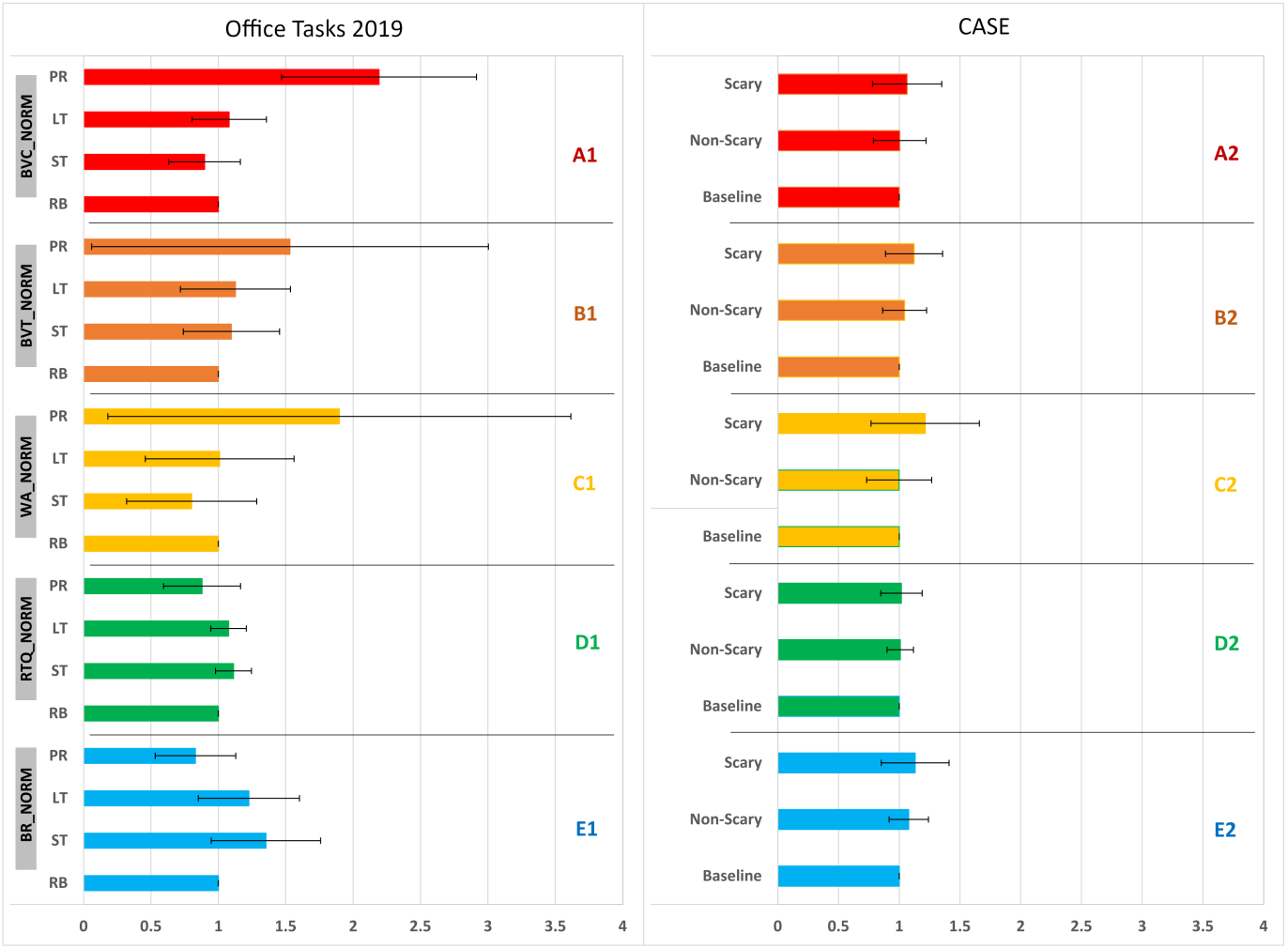

Fig. 7: **Bar plots of participant means with error bars for normalized breathing features in each experimental session of the Office Tasks 2019 and CASE datasets. A1-A2.** BVC_NORM ≡ Normalized mean breathing volume/cycle. **B1-B2.** BVT_NORM ≡ Normalized mean breathing volume/second. **C1-C2.** WA_NORM ≡ Normalized mean waveform amplitude. **D1-D2.** RTQ_NORM ≡ Normalized mean respiratory time quotient. **E1-E2.** BR_NORM ≡ Normalized mean breathing rate.

in PR (Fig. 7-B1), revealing an antithetical trend to the one observed for normalized breathing rate. Although at face value these two results appear to contradict each other, on closer examination offer complementary insights.

Since all the experimental tasks are of cognitive nature to one degree or another, arousal is expected to facilitate the increased metabolic needs of the brain, which is hard at work. In this respect, the normalized breathing volume per second (BVT_NORM) reliably captures this physiological necessity, indicating that tasks ST, LT, and PR led to an aroused state (Fig. 7-B1). What, however, BVT_NORM does not clearly show is how these aroused states were effected. To elucidate this information, we need to look into the normalized mean breathing volume per cycle (BVC_NORM) and normalized mean waveform amplitude (WA_NORM) vis a vis the normalized mean breathing rate (BR_NORM).

In ST, as can be observed in Fig. 7-A1 and Fig. 7-C1, the measures of breathing depth BVC_NORM and WA_NORM are even lower than the baseline. This is compensated by a significant increase of BR_NORM in ST - the highest among all treatments (Fig. 7-E1). The combined result is that the mean breathing volume per second (BVT_NORM) remains above the baseline, indicating an aroused state (Fig. 7-B1). The participants' physiology manages to furnish the brain with increased amounts of oxygen to help it cope with ST's elevated mental load. The process through which this is achieved involves fast but shallow breathing. Such a decomposition of the breathing phenomenon beautifully explains the additional stressor present in ST. ST was a short essay writing session, lasting only 5 minutes and with a ticking clock in constant reminder of that. Thus, time pressure loomed large in ST, triggering a flight response. This flight response is the likely culprit behind the significant increase in breathing rate during ST (Fig. 7-E1), which was accompanied by a reduction in the amplitude of the breathing waveform (Fig. 7-C1).

In task LT, the participants engage in an essay writing task for nearly an hour. Hence, time pressure dissipates and the flight response recedes in LT, bringing about a decrease in breathing rate with respect to ST. Nevertheless, the need

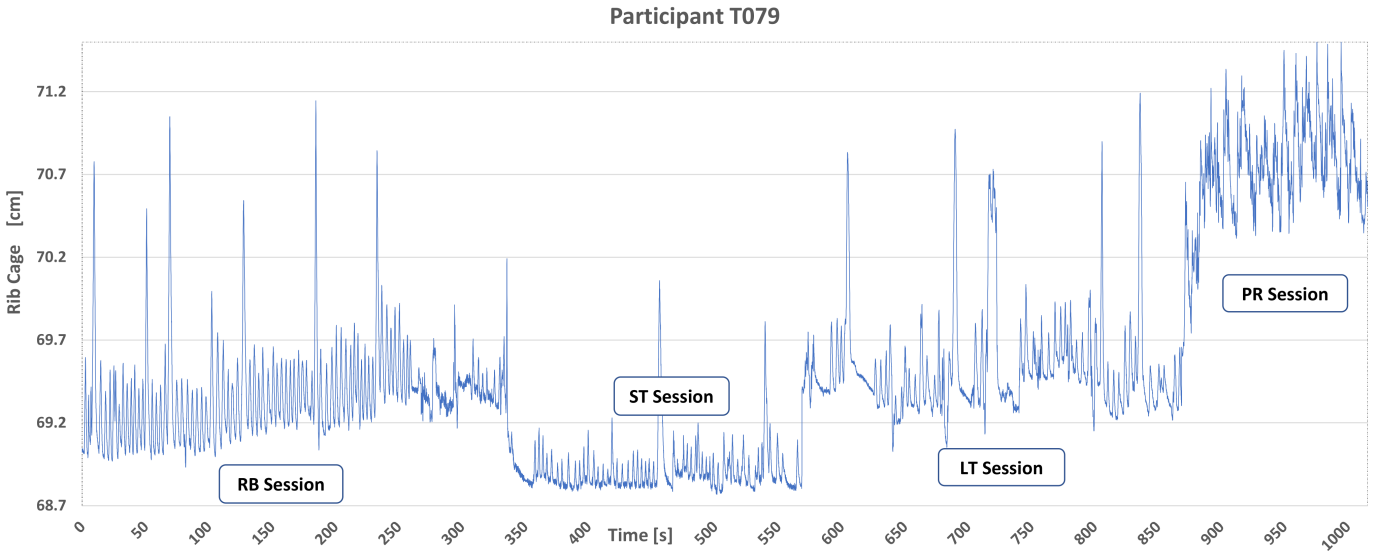

Fig. 8: **Example of changing breathing signal profiles in response to behavioral interventions.** Breathing signals during the RB, ST, LT, and PR sessions for participant T079 in the *Office Tasks 2019*. Since the signals were captured through the BioHarness, they express the evolving rib cage perimeter. It is clearly shown how shallow breaths in ST, change to deeper breaths in LT, and even deeper breaths in PR.

for providing the brain with increased levels of oxygen in LT is even stronger than the ST task. For this reason, the breathing rate decrease (Fig. 7-E1) is more than compensated by an increase in the breathing volume per cycle (Fig. 7-A1), resulting in an increase of the breathing volume per second, which surpasses ST (Fig. 7-B1). Hence, arousal addresses the significantly increased metabolic needs of the brain in LT by instituting deep breathing patterns instead of the shallow ones it effects in ST.

In task PR, the breathing rate is reduced with respect to task LT, dropping even lower than the baseline (Fig. 7-E1). However, this breathing rate diminution is compensated by a dramatic increase in the breathing volume per cycle (Fig. 7-A1). The combined result is that the breathing volume per second in PR emerges as the highest among all the tasks, suggesting the strongest arousal (Fig. 7-B1). This outcome is in agreement with the facial electrodermal (EDA) result reported by Zaman et al. in the dataset's data descriptor [7]. It appears that in PR, arousal accommodates the increased metabolic needs of the brain, during the demanding task of articulating and delivering a talk on the fly, by deepening the breathing patterns even further. In the context of these deeper PR breathing patterns, the drop of the mean RTQ value suggests a relative increase of the expiration over the inspiration period - a within cycle rearrangement meant to facilitate speech. This result too is in agreement with other reports in the literature [40]. In fact, one can observe the lengthier expiration phase in PR in the right-skewed waveform cycles of participant T003 in Fig. 4d. Figure 8 shows the breathing signals of participant T079 in ST, LT, and PR, providing a comprehensive example of the distinct breathing patterns arising from the three behavioral interventions.

4.1.2 *Exploratory Analysis on CASE*

For the CASE dataset, where there is no speaking session, things are simpler. Participants are more aroused when they watch scary movie clips. This is evident in Fig. 7-C2, where the breathing depth measure WA_NORM for the Scary treatment stands well above the baseline, while this is not the case for the Non-Scary treatment. Hence, fear conditioning appears to be associated with deeper breaths, something that is also supported by the psyhcophysiological literature [41]. Unlike in the *Office Tasks 2019* dataset, here the RTQ measure is non-discriminating, because neither the Scary nor the Non-Scary treatment involves speaking participants (Fig. 7-D2).

4.2 Regression Analysis

4.2.1 *Multinomial Regression on Office Tasks 2019*

The key question in our research is if the physiologically motivated groups of breathing features in Table 1 can be used to identify the different arousal profiles manifested in knowledge work tasks exemplified by ST, LT, and PR. We opt to use multinomial regression because not only is an effective multi-class classification model for moderate size datasets, but also provides highly explainable results.

Prior to applying the model, we check if any of the proposed features are highly correlated. In the cross-correlation matrix, we identify six sets of highly correlated features; four sets are centered on breathing depth statistics while the other two are centered on breathing speed statistics (Fig. 9a). Only one feature per cross-correlated set should be kept in the model, while the rest should be eliminated. To find the optimal feature in each case, we test exhaustively all the combinations that arise from candidate feature eliminations in the cross-correlated sets. The following six features survive this optimization process: BVT_AVG,

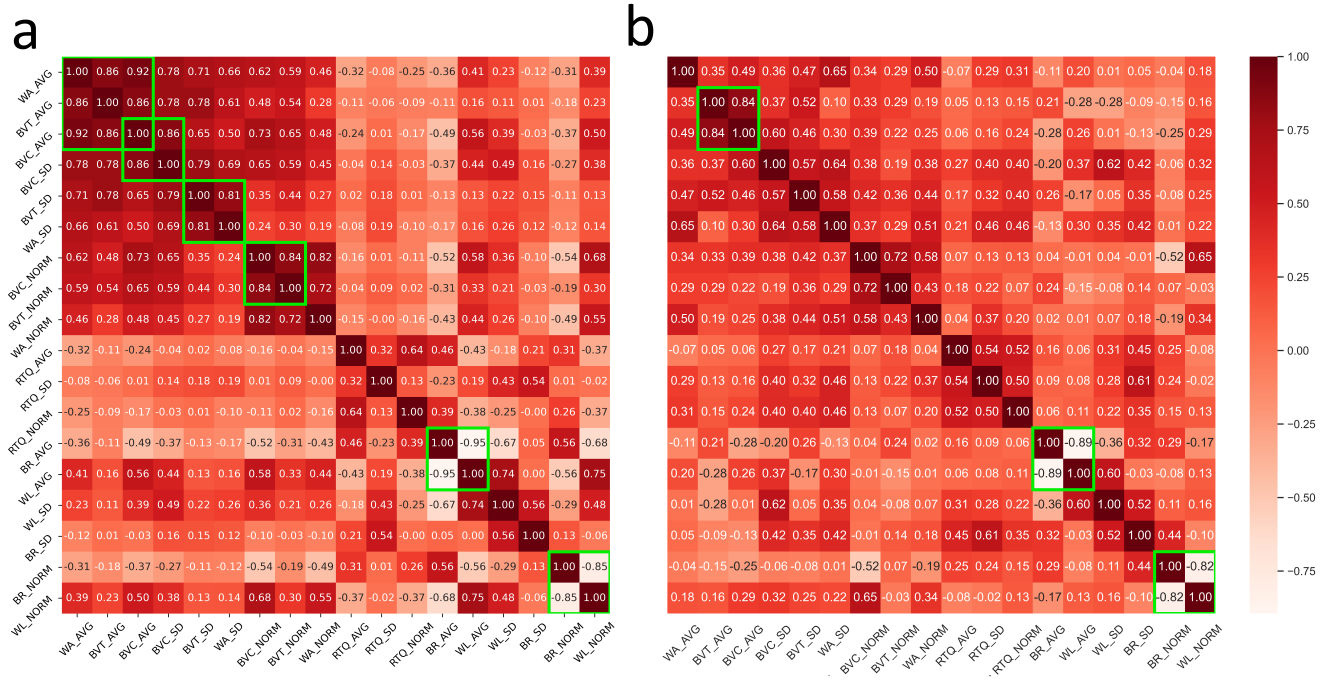

Fig. 9: **Cross-correlation matrices of all features for the a. *Office Tasks 2019* and b. *CASE* datasets.** The matrices reveal sets of highly correlated features that need to be optimally reduced. Highly correlated features are defined as features that have correlation coefficient $|r| \geq 0.80$; they correspond to dark (positive) and white (negative) cells in the cross-correlation matrices. Outlined in green, are six blocks of highly correlated features in *Office Tasks 2019* and three blocks in *CASE*.

BVC_SD, WA_SD, BVT_NORM, WL_AVG, and BR_NORM (Table 2 - Office Tasks 2019 section).

TABLE 2: **Sets of highly correlated features gleaned from the cross-correlation matrices in Fig. 9**. The highlighted features denote the optimal selections from each set, after an exhaustive combinatorial search.

	CORRELATED SETS	SET MEMBERS
OFFICE TASKS 2019	**Depth AVG Set**	{ WA_AVG, BVT_AVG, BVC_AVG }
	Depth AVG/SD Set	{BVC_AVG, BVC_SD }
	Depth SD Set	{BVT_SD, WA_SD }
	Depth NORM Set	{BVC_NORM, BVT_NORM }
	Speed AVG Set	{ BR_AVG, WL_AVG }
	Speed NORM Set	{ BR_NORM, WL_NORM }
CASE	**Depth AVG Set**	{ BVT_AVG, BVC_AVG }
	Speed AVG Set	{ BR_AVG, WL_AVG }
	Speed NORM Set	{ BR_NORM, WL_NORM }

After this first round of feature elimination, we construct a multinomial regression model with the remaining 12 out of the original 18 features (Eq. (9)). The response variable in this model is session-related arousal AR with three levels corresponding to the ST, LT, and PR sessions of the experiment:

$$\begin{aligned} AR \sim & \beta_I + \beta_a \text{BVC\_SD} + \\ & \beta_b \text{BVT\_AVG} + \beta_c \text{BVT\_NORM} + \\ & \beta_d \text{WA\_SD} + \beta_e \text{WA\_NORM} + \\ & \beta_f \text{RTQ\_AVG} + \beta_g \text{RTQ\_SD} + \beta_h \text{RTQ\_NORM} + \\ & \beta_i \text{BR\_SD} + \beta_j \text{BR\_NORM} + \\ & \beta_k \text{WL\_AVG} + \beta_l \text{WL\_SD}. \end{aligned} \tag{9}$$

We apply the backward elimination method to optimize the model described in Eq. (9). The result is a reduced model with six predictors, which is described in Eq. (10).

$$\begin{aligned} AR \sim & \beta_0 + \beta_1 \text{BVT\_NORM} + \beta_2 \text{WA\_SD} + \\ & \beta_3 \text{RTQ\_AVG} + \beta_4 \text{RTQ\_NORM} + \\ & \beta_5 \text{WL\_AVG} + \beta_6 \text{WL\_SD} \end{aligned} \tag{10}$$

For this ultimate model, the coefficients of the predictors, along with their standard errors and significance levels are given in Table 3a. We observe that normalized depth and phase shift features play a significant predictive role across the ST, LT, and PR classes - a result that justifies the accounting of the basal metabolic component in our framework. Indeed, the normalized breathing volume per second (BVT_NORM) and the normalized RTQ (RTQ_NORM) are significant predictors across the board ($p < 0.01$ for all cases). These results establish the progressive increase (decrease) of normalized breathing volume per second (normalized RTQ) as participants move from ST to LT and eventually to PR. The phase shift feature

RTQ in particular, is a significant differentiator between PR ($p < 0.001$) and the other two sessions. The strong negative coefficient ($\beta = -2.118$) indicates a dramatic RTQ decrease in PR - a sure sign of longer expiration, which is typically associated with speech production.

One-vs-Rest Classification w/ Regression: In this and the other classifications presented in this paper, we classify arousal through breathing signals, using our proposed feature set. The classification labels for the breathing signals are PR, LT, and ST, symbolizing the three types of hyperarousal produced in the respective sessions of the *Office Tasks 2019* experiment. Using the optimized set of predictors suggested in Eq. (10), we perform one-vs-rest classification. One-vs-rest is a standard method to perform multinomial classification, by performing a set of binomial classifications. In this case, we perform three binomial tests: a) PR vs. [ST $\cup$ LT]. b) LT vs. [ST $\cup$ PR]. c) ST vs. [LT $\cup$ PR]. Each of these three tests is performed on every piece of data, and the test that yields the strongest score becomes the ultimate classification result per occasion.

To assess robustness, we operationalize classification via k-fold cross-validation. We choose k=3. Accordingly, we split the data in each interventional session in three equitable subsets $D_i^{\text{ST}}, D_i^{\text{LT}}, D_i^{\text{PR}}\ i = 1 \dots 3$. Then, we produce representative subsets $D_i, i = 1 \dots 3$ for the entire experiment by joining the corresponding equitable subsets of the different interventions: $D_i = D_i^{\text{ST}} \cup D_i^{\text{LT}} \cup D_i^{\text{PR}}$. We keep one D_i for testing, while we use the other two $D_j,\ j \neq i$ for training. We repeat the process k=3 times. Figure 10a shows the 3-fold ROCs for the three different treatments PR, LT, and ST. One-vs-rest classification remains robust in all three treatments. The mean AUC in PR is in the mid 90s, while in the ST and LT stands around 90 and in the mid 70s, respectively. In all three cases the standard deviation is small. Specifically, for PR, ST, and LT, the one-vs-rest 3-fold AUC distributions are $0.96 \pm 0.02, 0.90 \pm 0.00,$ and 0.75 ± 0.05, respectively.

The summative confusion matrices and statistics of key classification metrics are reported in Table 3b and 3c, respectively. The confusion matrices indicate that in terms of true positives vs false negatives one vs. rest classification performs very well in all three tasks. In terms of true negatives vs. false positives, however, the performance is superior in the PR task followed by the ST task; the LT task performs the worst. This performance ranking is also reflected in the accuracy, recall, and F1 scores. For instance, for PR the F1 score distribution is in the upper 80s, for ST in the low 80s, and for LT in the low 70s.

4.2.2 Logistic Regression on CASE

The *CASE* dataset is of a much simpler study than the *Office Tasks 2019* study. In the context of testing our breathing feature set (Table 1), the key objective is to differentiate between arousal responses from watching Scary vs. Non-Scary movie clips. As we are confronted with a binary classification problem, we construct a logistic regression model, where the response variable is the odds $P(AR)$ of arousal associated with Scary movie clip watching.

Prior to applying the said model, we check if any of the features in Table 1 are highly correlated. In the cross-correlation matrix shown in Fig. 9b, we identify three sets of highly correlated features; one set is centered on breathing depth statistics, while the other two are centered on breathing speed statistics. Only one feature per cross-correlated set should be kept in the model, while the rest should be eliminated. To find the optimal feature in each case, we test exhaustively all the combinations that arise from candidate feature eliminations in the cross-correlated sets. The following three features survive this optimization process: BVT_AVG, WL_AVG, and BR_NORM (Table 2 - CASE section). After this first round of feature elimination, we construct the logistic regression model with the remaining 15 out of the original 18 features (Eq. (11)).

$$\begin{aligned} P(AR) \sim\ & \beta_I' + \\ & \beta_a' \text{BVC\_SD} + \beta_b' \text{BVC\_NORM} + \\ & \beta_c' \text{BVT\_AVG} + \beta_d' \text{BVT\_SD} + \beta_e' \text{BVT\_NORM} + \\ & \beta_f' \text{WA\_AVG} + \beta_g' \text{WA\_SD} + \beta_h' \text{WA\_NORM} + \\ & \beta_i' \text{RTQ\_AVG} + \beta_j' \text{RTQ\_SD} + \beta_k' \text{RTQ\_NORM} + \\ & \beta_l' \text{BR\_SD} + \beta_m' \text{BR\_NORM} + \\ & \beta_n' \text{WL\_AVG} + \beta_o' \text{WL\_SD}. \end{aligned} \tag{11}$$

We apply the backward elimination method to optimize the model described in Eq. (11). The result is a reduced model with two predictors, which is described in Eq. (12).

$$P(AR) \sim \beta_0' + \beta_1' \text{BVT\_AVG} + \beta_2' \text{WA\_NORM}. \tag{12}$$

For this ultimate model, the coefficients of the predictors, along with their standard errors and significance levels are given at the bottom of Table 3a. Like in *Office Tasks 2019*, there is at least one normalized depth feature (i.e., WA_NORM) that plays a significant predictive role - a result that further justifies the accounting of the basal metabolic component in our framework. Unlike in *Office Tasks 2019*, no normalized phase shift feature (i.e., RTQ_x) plays any significant predictive role. This is to be expected, as participants are silent in *CASE*.

Binary Classification w/ Regression: The classification labels for the breathing signals are Scary and Non-Scary, symbolizing the two types of arousal produced in the respective movie viewing sessions of the *CASE* experiment. Using the optimized set of predictors suggested in Eq. (12), we perform binomial classification in the *CASE* dataset.

To assess robustness, we operationalize classification via 3-fold cross-validation. Figure 10c shows the resulting 3-fold ROC. Regression classification performance is good, with the 3-fold AUC distribution standing at 0.71 ± 0.04. The summative confusion matrix and statistics of key classification metrics are reported in Table 3b and 3c, respectively. The confusion matrix indicates that in terms of true positives vs false negatives regression classification performs very well. In terms of true negatives vs. false positives, however, the performance is relatively lower. This performance signature is reflected in the accuracy, recall, and F1 scores. Characteristically, recall stands at 0.86 ± 0.09 while accuracy is 0.72 ± 0.03.

TABLE 3: **Regression parameters and classification metrics. a.** Parameter estimates for the multinomial and logistic regression models in Eqs. (10) and (12), corresponding to the *Office Tasks 2019* and *CASE* datasets. $\beta.$ are the estimated regression coefficients. SE are the standard errors of the individual regression coefficients. **Pr(**$>|z|$**)** are the p-values of the coefficients. Significance levels have been set as follows: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. **b.** Summative confusion matrices for regression and random forest operating upon the *Office Tasks 2019* and *CASE* datasets. The summation was done over the results of 3-fold cross-validation in each task. **c.** Summary statistics of key classification metrics for regression and random forest operating upon the *Office Tasks 2019* and *CASE* datasets. The statistics reflect the performance results of 3-fold cross-validation in each task. **d.** Derivation of total records shown in the summative confusion matrices. Resampling is performed by standard Python routines to ameliorate imbalances in the data.

a

		Office Tasks 2019 Dataset									CASE Dataset		
		ST			LT			PR					
		$\beta.$	SE	**Pr(**$>\|z\|$**)**	$\beta.$	SE	**Pr(**$>\|z\|$**)**	$\beta.$	SE	**Pr(**$>\|z\|$**)**	$\beta.$	SE	**Pr(**$>\|z\|$**)**
	BVT_NORM	-1.741	0.41	< **0.001** ***	1.900	0.518	< **0.001** ***	2.292	0.756	**0.002** **			
	WA_SD	-1.936	0.518	< **0.001** ***	2.174	0.625	**0.001** **	2.471	0.776	**0.001** **			
	RTQ_AVG	0.311	0.181	0.087	0.246	0.183	0.179	-0.775	0.311	**0.013** *			
	RTQ_NORM	0.996	0.238	< **0.001** ***	-0.813	0.239	**0.001** **	-2.118	0.546	< **0.001** ***			
	WL_AVG	-0.004	0.524	0.994	-2.728	0.554	< **0.001** ***	2.536	0.663	< **0.001** ***			
	WL_SD	0.973	0.393	**0.013** *	1.999	0.571	< **0.001** ***	-5.864	1.282	< **0.001** ***			
	BVT_AVG										0.367	0.166	**0.028** ·
	WA_NORM										0.694	0.171	< **0.001** ***

b

		ST			LT			PR					
		Actual YES	Actual NO	Predicted Total	Actual YES	Actual NO	Predicted Total	Actual YES	Actual NO	Predicted Total	Actual YES	Actual NO	Predicted Total
Regression	Predicted Yes	TP = 39	FP = 11	50	TP = 36	FP = 26	62	TP = 35	FP = 1	36	TP = 51	FP = 27	78
	Predicted No	FN = 6	TN = 76	82	FN = 8	TN = 62	70	FN = 8	TN = 88	96	FN = 8	TN = 34	42
	Actual Total	45	87	132	44	88	132	43	89	132	59	61	120
Random Forest	Predicted Yes	TP = 32	FP = 8	40	TP = 36	FP = 23	59	TP = 40	FP = 7	47	TP = 51	FP = 15	66
	Predicted No	FN = 10	TN = 82	92	FN = 7	TN = 66	73	FN = 7	TN = 78	85	FN = 9	TN = 45	54
	Actual Total	42	90	132	43	89	132	47	85	132	60	60	120

c

	ST			LT			PR					
	Accuracy	Recall	F1	Accuracy	Recall	F1	Accuracy	Recall	F1	Accuracy	Recall	F1
Regression	0.87 ± 0.01	0.87 ± 0.00	0.82 ± 0.02	0.76 ± 0.08	0.84 ± 0.17	0.70 ± 0.03	0.93 ± 0.02	0.81 ± 0.04	0.89 ± 0.04	0.72 ± 0.03	0.86 ± 0.09	0.75 ± 0.06
Random Forest	0.86 ± 0.06	0.77 ± 0.06	0.79 ± 0.06	0.77 ± 0.06	0.84 ± 0.04	0.71 ± 0.04	0.89 ± 0.06	0.85 ± 0.04	0.84 ± 0.07	0.80 ± 0.05	0.85 ± 0.06	0.81 ± 0.06

d

Originally (ST = 56) + (LT = 53) + (PR = 51) = 160 records →	Originally 180 records →
after resampling = 132 records =	after resampling = 198 records =
88 training + 44 testing records →	158 training + 40 testing records →
w/ 3-fold = 44 × 3 = **132 records**	w/ 3-fold = 40 × 3 = **120 records**

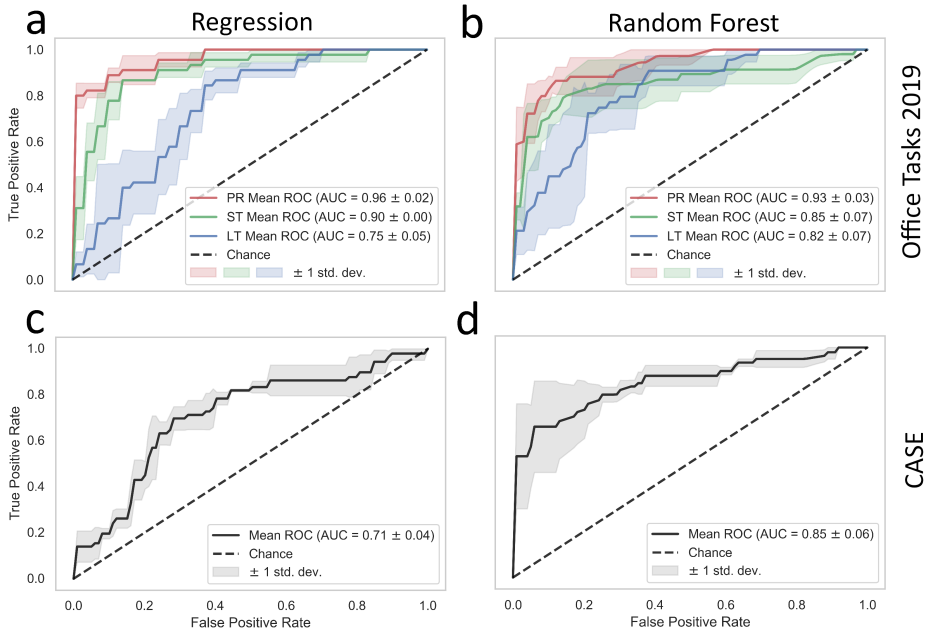

Fig. 10: **ROC curves for three-fold cross-validation in a.** Multinomial regression in *Office Tasks 2019*. **b.** Random forest in *Office Tasks 2019*. **c.** Logistic regression in *CASE*. **d.** Random forest in *CASE*. Solid lines represent the 3-fold mean curves and shaded regions the standard deviation.

4.3 Random Forest Analysis

4.3.1 Random Forest on Office Tasks 2019

To appreciate how the proposed framework of breathing features performs in the context of different ML algorithms, we also carry out classification of treatments in the *Office Tasks 2019* dataset using random forest. Unlike decision tress, random forest runs a reduced risk of overfitting and for this reason is a popular ML algorithm.

To assess robustness, we operationalize classification via 3-fold cross-validation. Figure 10b shows the 3-fold ROCs for treatments PR, LT, and ST. Random forest remains robust in all three treatments. Comparing Fig. 10b with Fig. 10a, we observe that the random forest's performance is on par with the regression performance in treatments PR and ST, while it exceeds regression in treatment LT. In more detail, for PR, ST, and LT, the random forest 3-fold AUC distributions are $0.93 \pm 0.03, 0.85 \pm 0.07,$ and 0.82 ± 0.07, respectively.

The summative confusion matrices and statistics of key classification metrics are reported in Table 3b and 3c, respectively. The confusion matrices indicate that in terms of true positives vs false negatives random forest classification performs very well in all three tasks. In terms of true negatives vs. false positives, however, the performance is superior in the PR and ST tasks; the LT task performs the worst. This performance ranking is also reflected in the accuracy, recall, and F1 scores. For instance, accuracy for PR stands at 0.89 ± 0.06, for ST at 0.86 ± 0.06, and for LT at 0.77 ± 0.06.

4.3.2 Random Forest on CASE

Figure 10d shows the 3-fold ROC for *CASE*, when we use random forest. The performance appears to improve over that of regression for the same dataset (Fig. 10c) - AUC = 0.85 ± 0.06 vs. 0.71 ± 0.04. The summative confusion matrix and statistics of key classification metrics are reported in Table 3b and 3c, respectively. The confusion matrix indicates that in terms of true positives vs false negatives random forest classification performs very well. In terms of true negatives vs. false positives, the performance is relatively lower. This performance signature is reflected in the accuracy, recall, and F1 scores. Characteristically, recall stands at 0.85 ± 0.06 while accuracy is 0.80 ± 0.05.

4.4 Feature Importance Across Datasets-Algorithms

The cross-comparison of feature importance among the datasets and algorithms used in this research offers valuable insights into the operationalization of our feature set. Figure 11 shows the feature importance graphs regarding classification via regression and random forest in the *Office Tasks 2019* and *CASE* datasets. In regression classification, the feature importance is determined by the coefficients of the applicable regression model, while in random forest classification the feature importance is determined by the mean decrease in impurity.

Figures 11b - 11d show features that account for more than 5% of reduction in impurity in random forest classification for the *Office Tasks 2019* and *CASE* datasets, respectively. In *Office Tasks 2019*, in particular, there are four features that exceed 10% contribution in random forest classification. These features include both breathing depth (BVC_NORM) and speed (WL_NORM, WL_AVG, BR_AVG) features in normalized and non-normalized form. There is also an RTQ feature that immediately follows them. This diverse set of important features covers all three categories of breathing features that constitute the essence of our research design, as laid out in Table 1. Hence, the said results affirm the goodness of our proposition for depth, speed, and RTQ breathing features to be used in naturalistic knowledge work studies with speech sessions. Furthermore, in the *Office Tasks 2019* the important features in random forest (Fig. 11b) largely match the important features in regression (Fig. 11a), establishing the relative insensitivity of our feature set to different ML algorithms, when used in applications for which was designed for.

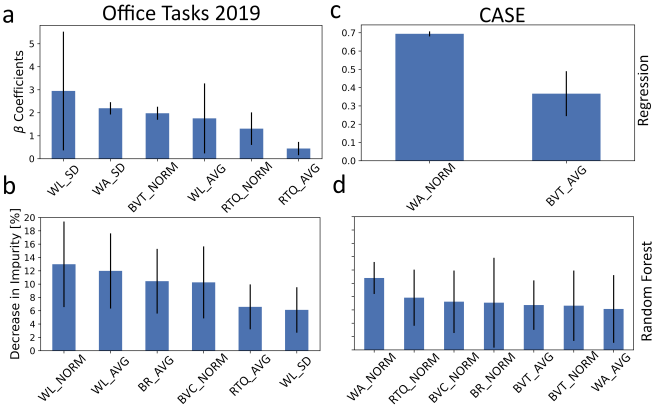

Fig. 11: **Feature importance graphs for: a.** Regression classification in the *Office Tasks 2019* dataset. **b.** Random forest classification in the *Office Tasks 2019* dataset. **c.** Regression classification in the *CASE* dataset. **d.** Random forest classification in the *CASE* dataset.

In *CASE*, there is only one feature that exceeds 10% contribution in random forest classification. This is the breathing depth feature WA_NORM (Fig. 11d), which captures fear responses during scary movie watching, as we explained earlier in the text. Some speed and RTQ features follow, but they are all below 10%. The regression feature importance graph in Fig. 11c agrees with the random forest graph in Fig. 11d regarding the importance of the WA_NORM feature. The regression graph largely ignores all the other low scoring features that show in the random forest graph. Hence, in the *CASE* dataset our proposed feature set, as laid out in Table 1, is underutilized. This is simply because the dataset is void of relevant content, such as various mental tasks in the presence and absence of speech. Nevertheless, the one depth feature that is relevant here, captures the one-dimensional emotional content of the dataset (i.e., fear), helping the algorithms to deliver decent classification performance. The random forest method manages to scrape information from a few other features, which likely explains its relatively higher performance.

5 Discussion

The research presented in this paper aims at establishing a new analytical framework for breathing signals, which

can live up to the challenges of naturalistic affective studies for sitting subjects. Such studies are reflective of the knowledge economy era, where subjects engage in mental activities either sustained or time-pressured, interspersed with speech delivery during video calls. These types of activity are likely to cause hyperarousal and thus are of interest to affective investigations. Existing breathing analysis methods have difficulty dealing comprehensively with such activities, either because the said methods confound speech effects or because can capture certain behavioral effects (e.g., emotional) but not others (e.g., mental).

The key characteristic of the framework we introduce is a comprehensive set of psychophysiologically informed breathing features. Specifically, the framework is capable of accounting for the basal metabolic, involuntary behavioral, and volitional behavioral components of breathing in individuals. The method attains this performance thanks to two provisions: First, it estimates the basal metabolic component of breathing from a resting baseline session, which the method considers to be integral to affective computing studies. By removing the basal component, the method can focus more effectively on the involuntary and voluntary behavioral components of breathing. Second, the method uses a multitude of features that capture breathing characteristics laden with behavioral information. These features include breathing rate (i.e., speed), which tracks involuntary behavioral effects of emotional nature. The features also include breathing volume, which tracks increased metabolic demands due to mental stress, or the instantaneous effect of sighing associated with fear responses. Finally, the features also include the RTQ ratio, which captures shifts between the inspiratory and expiratory phases of breathing, indicating the presence or absence of speech - a volitional behavior that interferes with breathing.

By testing our breathing feature set on the *Office Tasks 2019* dataset, we found that it can reliably classify hyperarousal instigated by mental stressor vs. mixed mental-emotional stressor vs. mixed mental-social stressor in the presence of speech. The first class is an example of an involuntary side effect caused by an internal stimulus (i.e., mental effort). The second class is an example of an involuntary side effect caused by a combination of internal (i.e., mental effort) and external stimulus (i.e., strict time limit). The third class is an example of mixed involuntary-voluntary effects. The involuntary effect stems from the socio-cognitive stressor of articulating a public speech on the fly, while the voluntary effect stems from delivering this speech, and thus volitionally altering one's breathing patterns.

There are two issues that merit further discussion. First, what happens if the affective study does not feature a resting baseline session. Second, can this framework being extended beyond the realm of naturalistic knowledge work studies? With respect to the first question, if the study is a controlled experiment, then the absence of a baseline session is a problem, because it will render impossible the computation of normalized features, which constitute one third of our set (Table 1). If the study is in the wild, however, featuring lengthy daily observations, then the basal metabolic component of breathing can be estimated with careful analysis. For instance, one could look for consistent low breathing values on Sunday mornings, when most people tend to relax, as a reasonable approximation to a basal metabolic state.

To start addressing the second question, we tested our feature set on *CASE* - a second dataset derived from a stylized experiment, where participants are exposed to short scary and non-scary movie clips. The dataset has few of the characteristics our feature set has been designed for. For instance, participants are not speaking in *CASE*, while the experimental sessions are short and do not involve naturalistic knowledge work tasks. Despite all these, our feature set attains respectable performance in classifying scary from non-scary arousal responses. In doing so, our feature set demonstrates its capacity for application beyond our original design aims. Moreover, the feature set's performance appears to be independent of the choice of classification algorithms in the *Office Tasks 2019* dataset, where both regression and random forest perform equally well. For the *CASE* dataset performance remains on solid grounds across classification methods but the feature set attains better performance through random forest rather than regression. In conclusion, the proposed breathing feature set demonstrates:

1) Excellent stand-alone classification performance of affect when used within its designed application envelope, with naturalistic tasks featuring mental and emotional stressors plus speech. Full feature utilization renders classification performance independent of algorithmic choice; cross-validated AUC mostly in the 90s for PR and ST in *Office Tasks 2019*.
2) Very good stand-alone classification performance of affect when used in naturalistic tasks that feature strong mental stressors without speech; cross-validated AUC in the 80s and 70s for LT in *Office Tasks 2019*.
3) Solid stand-alone classification performance, even when used outside its designed application envelope, in contrived experimental tasks with moderate stressors and no speech. Random forest better utilizes sparse features in this instance, gaining a performance edge; cross-validated AUC in the 70s through regression and in the 80s through random forest in *CASE*.

REFERENCES

[1] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang *et al.*, "A systematic review on affective computing: Emotion models, databases, and recent advances," *Information Fusion*, vol. 83-84, pp. 19–52, 2022.

[2] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2011.

[3] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[4] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, 2005.

[5] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A dataset for affect, personality and mood research on individuals and groups," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 479–493, 2018.

[6] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 2018, pp. 400–408.
[7] S. Zaman, A. Wesley, D. R. D. C. Silva, P. Buddharaju, F. Akbar, G. Gao, G. Mark, R. Gutierrez-Osuna, and I. T. Pavlidis, "Stress and productivity patterns of interrupted, synergistic, and antagonistic office activities," *Scientific Data*, vol. 6, 2019.
[8] K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, and F. Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," *Scientific Data*, vol. 6, no. 1, p. 196, 2019.
[9] S. A. Shea, "Behavioural and arousal-related influences on breathing in humans," *Experimental Physiology: Translation and Integration*, vol. 81, no. 1, pp. 1–26, 1996.
[10] H.-D. Park, C. Barnoud, H. Trang, O. A. Kannape, K. Schaller, and O. Blanke, "Breathing is coupled with voluntary action and the cortical readiness potential," *Nature Communications*, vol. 11, no. 1, pp. 1–8, 2020.
[11] G. W. Fincham, C. Strauss, J. Montero-Marin, and K. Cavanagh, "Effect of breathwork on stress and mental health: A meta-analysis of randomised-controlled trials," *Scientific Reports*, vol. 13, no. 1, p. 432, 2023.
[12] L. M. Schleifer, R. Ley, and T. W. Spalding, "A hyperventilation theory of job stress and musculoskeletal disorders," *American Journal of Industrial Medicine*, vol. 41, no. 5, pp. 420–432, 2002.
[13] M. Kusserow, O. Amft, and G. Tröster, "Monitoring stress arousal in the wild," *IEEE Pervasive Computing*, vol. 12, no. 2, pp. 28–37, 2013.
[14] M. I. Y. Arafath K. and A. Routray, "Automatic measurement of speech breathing rate," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5.
[15] B. A. Kelsen, "Exploring public speaking anxiety and personal disposition in EFL presentations," *Learning and Individual Differences*, vol. 73, pp. 92–101, 2019.
[16] E. Heim, P. H. Knapp, L. Vachon, G. G. Globus, and S. J. Nemetz, "Emotion, breathing and speech," *Journal of Psychosomatic Research*, vol. 12, no. 4, pp. 261–274, 1968.
[17] Q. Zhang, X. Chen, Q. Zhan, T. Yang, and S. Xia, "Respiration-based emotion recognition with deep learning," *Computers in Industry*, vol. 92, pp. 84–90, 2017.
[18] N. Jadhav and R. Sugandhi, "Survey on human behavior recognition using affective computing," in *2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN)*. IEEE, 2018, pp. 98–103.
[19] D. Wu, C. G. Courtney, B. J. Lance, S. S. Narayanan, M. E. Dawson, K. S. Oie, and T. D. Parsons, "Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality Stroop task," *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 109–118, 2010.
[20] G. Valenza, A. Lanata, and E. P. Scilingo, "The role of nonlinear dynamics in affective valence and arousal recognition," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 237–249, 2012.
[21] P. Sanches, K. Höök, E. Vaara, C. Weymann, M. Bylund, P. Ferreira, N. Peira, and M. Sjölinder, "Mind the body! Designing a mobile stress management application encouraging personal reflection," in *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, 2010, pp. 47–56.
[22] E. Asmussen, "Regulation of respiration: "the black box"," *Acta Physiologica Scandinavica*, vol. 99, no. 1, pp. 85–90, 1977.
[23] D. Carroll, J. Rick Turner, and J. C. Hellawell, "Heart rate and oxygen consumption during active psychological challenge: The effects of level of difficulty," *Psychophysiology*, vol. 23, no. 2, pp. 174–181, 1986.
[24] F. A. Boiten, "The effects of emotional behaviour on components of the respiratory cycle," *Biological Psychology*, vol. 49, no. 1-2, pp. 29–51, 1998.
[25] H. K. Brumback, "Investigation of breath counting, abdominal breathing and physiological responses in relation to cognitive load," in *International Conference on Augmented Cognition*. Springer, 2017, pp. 275–286.
[26] D. J. McDuff, J. Hernandez, S. Gontarek, and R. W. Picard, "COGCAM: Contact-free measurement of cognitive stress during computer tasks with a digital camera," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 4000–4004.
[27] R. Avram, G. H. Tison, K. Aschbacher, P. Kuhar, E. Vittinghoff, M. Butzner, R. Runge, N. Wu, M. J. Pletcher, G. M. Marcus *et al.*, "Real-world heart rate norms in the Health eHeart study," *NPJ Digital Medicine*, vol. 2, no. 1, p. 58, 2019.
[28] A. Rowden, "What is a normal respiratory rate based on your age?" 2023, accessed = 2023-12-21. [Online]. Available: https://www.medicalnewstoday.com/articles/324409
[29] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," *Cognition & Emotion*, vol. 9, no. 1, pp. 87–108, 1995.
[30] J. Hewig, D. Hagemann, J. Seifert, M. Gollwitzer, E. Naumann, and D. Bartussek, "A revised film set for the induction of basic emotions," *Cognition and Emotion*, vol. 19, no. 7, p. 1095, 2005.
[31] A. Wesley, P. Lindner, and I. Pavlidis, "Eustressed or distressed? Combining physiology with observation in user studies," in *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, 2012, pp. 327–330.
[32] R. Bartlett Jr and H. Specht, "Maximum breathing capacity with various expiratory and inspiratory resistances (single and combined) at various breathing rates," *Journal of Applied Physiology*, vol. 11, no. 1, pp. 79–83, 1957.
[33] M. Nixon and A. Aguado, *Feature Extraction and Image Processing for Computer Vision*. Academic Press, 2019.
[34] P. Tsiamyrtzis, M. Dcosta, D. Shastri, E. Prasad, and I. Pavlidis, "Delineating the operational envelope of mobile and conventional EDA sensing on key body locations," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 5665–5674.
[35] M. Lutfi, "The physiological basis and clinical significance of lung volume measurements," *Multidisciplinary Respiratory Medicine*, vol. 12, 02 2017.
[36] J. Ohlsson and B. Wranne, "The influence of hyperventilation on the measurement of stroke volume using a CO2 rebreathing method," *European Journal of Applied Physiology and Occupational Physiology*, vol. 55, no. 1, pp. 19–23, 1986.
[37] M. Tobin, "Breathing pattern analysis," *Intensive Care Medicine*, vol. 18, no. 4, pp. 193–201, 1992.
[38] K. Konno and J. Mead, "Measurement of the separate volume changes of rib cage and abdomen during breathing," *Journal of Applied Physiology*, vol. 22, no. 3, pp. 407–22, 1967.
[39] A. Augousti, "A theoretical study of the robustness of the iso-volume calibration method for a two-compartment model of breathing, based on an analysis of the connected cylinders model," *Physics in Medicine & Biology*, vol. 42, no. 2, p. 283, 1997.
[40] B. Conrad and P. Schönle, "Speech and respiration," *Archiv fur Psychiatrie und Nervenkrankheiten*, vol. 226, no. 4, p. 251—268, April 1979. [Online]. Available: https://doi.org/10.1007/bf00342238
[41] I. Van Diest, M. M. Bradley, P. Guerra, O. Van den Bergh, and P. J. Lang, "Fear-conditioned respiration and its association to cardiac reactivity," *Biological Psychology*, vol. 80, no. 2, pp. 212–217, 2009.

Nanfei Sun Dr. Sun is Assistant Professor in the Management Information System Department at the University of Houston-Clear Lake. He received his PhD from the Computer Science Department at the University of Houston in 2006. He previously worked for the IBM T.J. Watson Research Centre and Hubwoo. His research interests are in computational physiology, computer vision, and machine learning. Dr. Sun is known for his pioneering work on unobtrusive measurement methods for heart function, which he published in a series of articels in IEEE Transactions and Proceedings. He is currently working on developing the full potential of the breathing channel for affective computing applications.

Ioannis Pavlidis Dr. Pavlidis is the Eckhard-Pfeiffer Distinguished Professor of Computer Science and Director of the Affective and Data Computing Laboratory at the University of Houston. His research is funded by multiple sources including the National Science Foundation, transportation agencies, and medical institutions. He has published extensively in the areas of affective computing, data science, and science of science. He was the first to conceive and develop contact-free methods for measuring physiological variables, including electrodermal activity, breathing, and heart function, which he used to study stress in the wild.