



AFRL-RH-WP-TR-2021-0108

**IDENTIFICATION OF HYPERCAPNIA THROUGH VOICE
ANALYSIS AND ASSOCIATED NEUROLOGICAL AND
PERFORMANCE EFFECTS**

**Jeffrey B. Phillips, Arash Mahyari, Ian Perera, Adrien Moucheboeuf, Madison C. McInnis, Anil
Raj, Allison Bew, Andrew Dorsey**

The Florida Institute for Human and Machine Cognition

**SEPTEMBER 2021
Final Report**

Distribution Statement A: Approved for public release.

See additional restrictions described on inside pages

**AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
AIRMAN SYSTEMS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the Air Force Research Laboratory Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RH-WP-TR-2021-0108 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signature//

AMY M. MCKLEROY
Medical and Operational Biosciences

//signature//

R. ANDY MCKINLEY, DR-III, PhD
Core Research Area Lead
Cognitive and Physical Performance
Performance Optimization Branch
Airman Biosciences Division

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 29-09-2021		2. REPORT TYPE Final		3. DATES COVERED (From – To) September 2019 – June 2021	
4. TITLE AND SUBTITLE Identification of Hypercapnia through Voice Analysis and Associated Neurological and Performance Effects				5a. CONTRACT NUMBER FA8650-19-C-6131	
				5b. GRANT NUMBER 2019_027_01	
				5c. PROGRAM ELEMENT NUMBER 62202F	
6. AUTHOR(S) Jeffery B. Phillips, Arash Mahyari, Ian Perera, Adrien Moucheboeuf, Madison C. McInnis, Anil Raj, Allison Bew, Andrew Dorsey				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER H10E	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Florida Institute for Human and Machine Cognition 40 S. Alcaniz Street Pensacola, FL, 32502				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Air Force Research Laboratory 711 th Human Performance Wing Airman Systems Directorate Airman Biosciences Division Performance Optimization Branch Wright-Patterson AFB, OH 45433				10. SPONSORING/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-RH-WP-TR-2021-0108	
12. DISTRIBUTION / AVAILABILITY STATEMENT Distribution Statement A: Approved for public release. AFRL-2022-2662, cleared 6 June 2022					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT This investigation seeks to develop a non-invasive method to identify hypercapnia using a machine learning algorithm that can detect changes in speech and breath features specific to an individual with excess arterial carbon dioxide (CO ₂). Method: Thirty-five active-duty student Naval aviators ($M = 24.11$ years, $SD = 2.14$) performed two iterations of the Performance Assessment Tool (PAT) and two reading tasks while breathing four normobaric CO ₂ concentrations, 1.0%, 2.5%, 4.0%, 5.5%, delivered via an MBU-20/P flight mask, for 15-minutes each. Results: A repeated measures Analysis of Variance conducted on tracking performance during the PAT revealed significant differences between conditions, $f(1.69, 28.78) = 4.39$, $p = 0.027$, <i>partial eta square</i> = 0.21. Pairwise comparisons revealed significantly better tracking performance in the 2.5% and 4.0% exposure conditions when compared to the 1.0% exposure condition. Both the speech and breath machine learning approaches accurately identified the CO ₂ concentrations that participants were exposed to. The approaches based on modeling each individual's speech and breath features performed best with accuracy at 89% and 84.5% respectively. The speech and breath machine learning approaches also resulted in Area Under the Receiver Operating Characteristic values in the 80s and 90s indicating excellent diagnostic capabilities for hypercapnia. Conclusion: The results of the current investigation suggest that speech and breath contain hypercapnia specific signals during CO ₂ exposures. Machine learning approaches should be further pursued to identify changes in physiological status and mitigate operational stressors in tactical aviation.					
15. SUBJECT TERMS Analysis of Variance (ANOVA), Area Under the Receiver Operating Characteristic (AUROC)					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 42	19a. NAME OF RESPONSIBLE PERSON Amy McKleroy
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code) N/A

TABLE OF CONTENTS

LIST OF FIGURES	ii
LIST OF TABLES	iii
1.0 SUMMARY	1
2.0 INTRODUCTION	1
3.0 METHODS	3
3.1 Participants	3
3.2 Instrumentation.....	4
3.3 Procedure.....	5
4.0 ANALYSIS	6
4.1 Performance Analysis	6
4.2 Speech Analysis	6
4.3 Breath Analysis	7
4.3.1. Feature Extraction	7
4.3.2. Deviation from the Baseline	8
4.3.3. Breath-Based CO ₂ Level Detection and Classification	8
4.4 Receiver Operating Characteristics (ROC) Analysis of Models.....	9
5.0 RESULTS	9
5.1 Descriptive Statistics of Physiological Parameters Across Conditions	9
5.1.1. PaCO ₂ across Conditions during PAT Tasks.....	9
5.1.2. PaCO ₂ across Conditions during Reading Tasks.....	10
5.1.3. Heart Rate Across Conditions during PAT tasks.....	11
5.1.5. Breath Rate Across Conditions.....	12
5.1.6. Tidal Volume Across Conditions.....	13
5.1.7. Minute Ventilation Across Conditions	14
5.2 Performance Assessment Tool (PAT) Results	14
5.3 Speech Results.....	16
5.4 Breath Analysis Results	20
5.4.1. Expert User Approach (Trained on one participant and evaluated on the same participant).....	20
6.0 DISCUSSION	27
7.0 CONCLUSION.....	31
8.0 REFERENCES	32
LIST OF ABBREVIATIONS AND ACRONYMS	36

LIST OF FIGURES

Figure 1 a. Graphical Display of Increase in PaCO ₂ from Baseline Across Conditions During PAT Tasks	10
Figure 1 b. Graphical Display of Increase in PaCO ₂ from Baseline Across Conditions During Reading Tasks	11
Figure 2 a. Graphical Display of Heart Rate Across Conditions During PAT Tasks	11
Figure 2 b. Graphical Display of Heart Rate Across Conditions During Reading Tasks	12
Figure 3. Graphical Display of Breath Rate Across Conditions During PAT Tasks	13
Figure 4. Graphical Display of Tidal Volume Across Conditions During PAT Tasks	13
Figure 5. Graphical Display of Minute Ventilation Across Conditions During PAT Tasks	14
Figure 6. ANOVA results for overall PAT performance across CO ₂ conditions	15
Figure 7. ANOVA results for PAT tracking performance across CO ₂ condition	16
Figure 8. Pearson Correlation for Each Participant (and Average) in the Unseen Experiment – Where the Individual was not Present in the Training Data	17
Figure 9. Pearson Correlation for each Participant (and Average) in the Combined Experiment – Where the Tested Individual was Present in the Training Data Along with the Other Individuals	18
Figure 10. Pearson Correlation for Each Participant (and Average) in the Divergence Experiment – Where only the Tested Individual was Present in the Training Data to Evaluate Individual-Specific Hypercapnia Detection	19
Figure 11. ROC for Predicting Each Participant’s CO ₂ level in the Unseen Experiment – where the Tested Individual was not Present in the Training Data	20
Figure 12. The ROC of Detecting Changes in CO ₂ Level from Breath Sounds for Each Participant Using Unsupervised Approaches (Expert User Unsupervised Approach)	21
Figure 13. The Accuracy of Detecting Changes in CO ₂ Level from Breath Sounds for Each Participant Using Unsupervised Approaches.....	22
Figure 14. The ROC of Detecting CO ₂ Level from Breath Sounds for Each Participant when the Model is Trained on the Same Participant (Expert User Supervised Approach)	23
Figure 15. The Accuracy of Detecting CO ₂ Level from Breath Sound for Each Participant when the Model is Trained on the Same Participant (Expert User Supervised Approach)	23
Figure 16. The ROC of Classifying CO ₂ Levels from Breath Sound for New Participants	24
Figure 17. The Accuracy of Classifying CO ₂ Levels from Breath Sounds for New Participants Using K-folding	25
Figure 18. The Accuracy of Classifying Pairwise CO ₂ Levels (1.0 vs 2.5, 1.0 vs 4.0, 1.0 vs 5.5) from Breath Sounds for New Users K-folding	26
Figure 19. The Accuracy of Classifying Pairwise CO ₂ Levels (2.5 vs 4.0, 2.5 vs 5.5, 4.0 vs 5.5) from Breath Sounds for New Users K-folding	26
Figure 20. Similarities of breathing segments for 3 different participants in the unsupervised approach	27

LIST OF TABLES

Table 1 a. Increase in PaCO ₂ from Baseline Across Conditions During PAT Tasks	9
Table 1 b. Increase in PaCO ₂ from Baseline Across Conditions During Reading Tasks	10
Table 2 a. Heart Rate Across Conditions During PAT Tasks.....	11
Table 2 b. Heart Rate Across Conditions During Reading Tasks.....	12
Table 3. Breath Rate Across Conditions During PAT Tasks.....	12
Table 4. Tidal Volume Across Conditions During PAT Tasks	13
Table 5. Minute Ventilation Across Conditions During PAT Tasks	14
Table 6. Mean PAT Performance Across CO ₂ Exposure Concentrations	15
Table 7. The Area Under the Curve (AUC) of Detecting CO ₂ Level from Speech Using the Unseen Experiment.....	19
Table 8. The Area Under the Curve (AUC) of Detecting CO ₂ Level from Breath Sounds for Each Participant Using the Expert User Unsupervised Approach.....	21
Table 9. The Area Under the Curve (AUC) of Detecting CO ₂ Level from Breath Sounds for Each Participant When the Model is Trained on Each Participant (Expert User Supervised Approach)	22
Table 10. The Area Under the Curve (AUC) of Detecting CO ₂ Level from Breath Sounds for New Participants	24
Table 11. The Accuracy of Detecting Pairwise CO ₂ Levels.....	25

1.0 SUMMARY

As the number of unidentified physiologic episodes (UPEs) in tactical aviation continue to increase, a need exists for early detection of these occurrences in flight to ensure operator safety. Hypercapnia is believed to be a significant contributor to UPEs as operators encounter breathing resistance from oxygen mask valves, breathing hoses, and regulators, which can lead to hypoventilation and carbon dioxide (CO₂) retention. While current methods to detect hypercapnia exist, the ability to use this technology in the cockpit remains difficult due to environmental issues and interference from and with flight equipment. This investigation seeks to develop a non-invasive method to identify hypercapnia using a machine learning algorithm to detect changes in speech and breath features specific to an individual with excess arterial CO₂.

Method: Thirty-five active-duty student Naval aviators ($M = 24.11$ years, $SD = 2.14$) performed two iterations of the Performance Assessment Tool (PAT) and two reading tasks while breathing four normobaric CO₂ concentrations, 1.0 percent (%), 2.5%, 4.0%, 5.5%, delivered via an MBU-20/P flight mask, for 15-minutes each.

Results: Transcutaneous arterial partial pressures of CO₂ (PaCO₂) increased in participants across the four CO₂ exposures with an average increase of 15.30 millimeters of mercury (mmHg) from baseline PaCO₂. A repeated measures Analysis of Variance (ANOVA) conducted on tracking performance during PAT revealed significant differences between conditions, $f(1.69, 28.78) = 4.39$, $p = 0.027$, *partial eta square* = 0.21. Pairwise comparisons revealed significantly better tracking performance in the 2.5% and 4.0% exposure conditions when compared to the 1.0% exposure condition. There was not an improvement in tracking performance during the 5.5% exposure condition. Both the speech and breath machine learning approaches accurately identified the CO₂ that participants were being exposed to. The approaches based on modeling each individual's speech and breath features performed best with accuracy at 89% and 84.5% respectively. The speech and breath machine learning approaches also resulted in Area Under the Receiver Operating Characteristic (AUROC) values in the 80s and 90s indicating that they are excellent as diagnostic instruments to identify CO₂ concentrations.

Conclusion: The results of the current investigation suggest that speech and breath contain hypercapnic specific signals during CO₂ exposures. Machine learning approaches should be further pursued to identify changes in physiological status and mitigate operational stressors in tactical aviation.

2.0 INTRODUCTION

Fifth-generation high-performance tactical aircraft challenge a fighter pilot's normal physiologic state through exposure to higher altitudes, speeds, and accelerations (g-forces) than previous aircraft. Therefore, the capabilities and reliability of the cockpit life support systems are evermore critical for management of the pilot's partial pressure of oxygen, breathing gas flow rates, cabin pressurization, and in normal flight and in response to combat g-forces.¹ Unfortunately, due to the complexities of the human respiratory system and the aircraft life support system, unforeseen physiologic issues can arise as these systems interact and seek to maintain operator homeostasis

in flight. Operators have reported symptoms like those they experienced during hypoxia recognition training.² However, incident investigations performed by the U.S. Navy and Air Force have revealed little evidence of equipment malfunctions that would explain these hypoxia-like occurrences in the F/A-18, F-22, F-35, T-45, EA-6Bs, and T-6 platforms.^{3,1,4,5,6} As a result, the military has designated these incidents as UPEs. Since 2007, over 600 UPEs have occurred in the F/A-18 alone, significantly eroding operator confidence in life support systems across platforms and in operational readiness.¹ Investigators now believe a significant percentage of UPEs are likely associated with a buildup of CO₂ in the operator's arterial blood, known as hypercarbia or hypercapnia.¹ Hypercapnia in aircraft is thought to be caused by malfunctions in life support systems that result in restricted breathing and the inability to adequately off-gas, through exhalation, the CO₂ created by metabolic activity.⁷ Hypercapnia causes increased minute ventilation, heart rate, and vasodilation in the cerebral cortex that results in tissue hypoxia in neurons and, at extreme levels, can lead to operator incapacitation.⁸ The sudden rise in heart rate and breath rate induced by hypercapnia can often be indistinguishable from a sudden autonomic nervous response or panic attack. When this occurs, in-flight operators can easily confuse hypercapnia with other aeromedical risks such as hypoxia or contaminated breathing gas.

To mitigate hypercapnia risk in tactical aviation, methods to identify the early signs of the condition in-flight are required. Hypercapnia is typically measured through capnography, a measure of the amount of CO₂ expired in each breath. Efforts to develop accurate capnography systems for in-flight use have failed due to multiple issues that include mixing of expired gas and supply gas in the flight mask, humidity of expired breath, and rapidly changing pressures inside the flight mask, all of which can generate unreliable measurements of expired CO₂. Because aircraft operators communicate verbally throughout a flight, analyzing operator voice dynamics may provide a powerful and unobtrusive method to monitor hypercapnia and other operator neurophysiologic status in-flight. The Florida Institute for Human and Machine Cognition (IHMC) has previously developed speech analysis techniques that have shown success in staging neurodegenerative disorders diseases. Such techniques may also provide a real-time, reliable, unobtrusive, and flight-worthy indicator of operator cognitive status.

Prior research on speech in extreme environments, such as high-altitude operations, has demonstrated detectable changes in the speech signal that occur during hypoxia, as well as various other physiological and/or psychological stressors and conditions. For example, studies have shown that acoustic and clinical features in speech correlate and track with the progression of Amyotrophic Lateral Sclerosis (ALS), a progressive neurodegenerative condition that affects muscle function.^{9,10,11} Other conditions, such as Traumatic Brain Injury (TBI), can induce impairments in speech that indicate the stages of the condition.¹² Additionally, automatic identification of features in speech can detect changes in psychological states, such as stressed versus neutral,^{13,14,15} and emotional states, including negative versus positive¹⁶ and agitated versus calm,¹⁷ as well as conditions such as depression.^{18,19,20,21,22,23} In the present investigation, operator speech (10ms segments of a standard passage read by the participants) and specific breathing gestures (sniffing, normal breathing, deep breathing) were analyzed through audio analysis software including Praat²⁴ and openSMILE²⁵ to identify a large number of acoustic (energy, voice, and spectrum related) and duration-related features. Such features have proven useful for detection of various operational stressors. For example, as mentioned above, acoustic, and duration-related features of vocalization such as root mean square (RMS) signal frame energy, jitter, Mel-frequency cepstral coefficient 2 (MFCC 2) mean peak distance, and degree of voice breaks have been used

to identify the early signs of ALS.¹⁰ Duration related features such as total recording durations, total pause time, and variability of pause durations strongly correlate with Hamilton Depression rating scores (HAMD) in depressed individuals,²⁶ who also show impact in acoustic features such as pitch variability.¹⁹ Additionally, frequency and duration-based features can help identify abnormal breath sounds from normal breath sounds. Normal breathing sound frequency range (for vesicular breath sounds) ranges between 60-600 Hertz (Hz).²⁷ Abnormal breath sounds from wheezing (lower airway obstruction) or stridor (upper airway obstruction) both fall in higher frequency bands, e.g., above 2000 Hz.²⁷ Rhonchi (whistling or snoring due to fluid in the large airways) have a duration greater than 100ms and a frequency of less than 300 Hz.²⁸ Spectral analysis can also classify lung sounds associated with conditions such as pneumonia.²⁹

In the present work, acoustic and duration-related features of individual speech and breath signals produced during hypercapnia exposures were extracted from voice and breath and analyzed with two approaches: (a) a *feature-based* approach and (b) a *divergence-based* approach. The feature-based approach used acoustic and duration-based features directly in predicting four normobaric, mildly elevated inspired CO₂ states whereas the divergence-based approach computed the degree of divergence between current speech samples and baseline speech samples for each speaker; divergence scores were in turn used to predict the four CO₂ states. We hypothesized that a larger divergence from the baseline could represent the degree or intensity of physiological stress and correspondingly, the level of inspired CO₂ exposure. As a secondary aim, the current investigation sought to measure the psychomotor performance effects of hypercapnia using the PAT previously developed by the Air Force Research Laboratory (AFRL).

3.0 METHODS

3.1 Participants

Thirty-five active-duty student aviators from the Marine Aviation Training Support Group-21 (MATSG-21) attached to Naval Air Station Pensacola volunteered for the study. A total of 32 males and three female participants between the ages of 20-45 years-old ($M = 24.11$ years, $SD = 2.14$) with current aviation duty up-chits (Medical Recommendation for Flying or Special Operational Duty, Department of Defense Form 2992) participated. In addition, all participants self-reported verification that they had no history of asthma, emphysema, chronic obstructive pulmonary disease (COPD), anemia, hypertension, epilepsy, sickle cell trait, or a diagnosis of pneumonia within the past twelve months. Participants who reported a positive diagnosis of COVID-19 within twelve months preceding their test day and before their up-chit was issued, were included only if they denied pulmonary symptoms during and after the illness. All participants provided informed consent prior to participation in the human research participants protocol, which was approved by the Florida IHMC Institutional Review Board (IRB) prior to the start of data collection.

3.2 Instrumentation

Cognitive performance of each participant was evaluated using the PAT. The PAT is designed to be sensitive to the performance effects of extreme environments such as high altitude, high workload, fatigue, and extreme cold and heat. The PAT is composed of a manual tracking task, a working memory task, a math task that requires you to add numbers on the screen and compare their total to a number displayed in the MFD, and a mannequin (spatial processing) task previously described by Phillips 2019.³⁰ The PAT can be administered with one, two, three or all four of the tasks simultaneously. For this study, participants completed all the PAT subcomponent tasks simultaneously in two-minute iterations. The PAT and reading tasks were presented on a monitor with an integrated high speed video eye tracker designed for behavior and eye movements research use (Spectrum, Tobii Pro AB, Stockholm, Sweden). The system captures eye movement data while allowing for natural head movement. Two cameras capture stereo images of both eyes for accurate measurement of eye gaze and eye position in 3D space, as well as pupil diameter.

Participants wore an MBU-20/P Flight Mask (Gentex Corp., Carbondale, PA) during the investigation. A mainstay in tactical aviation, the MBU-20/P was developed for Pressure Breathing for G (PBG) breathing schedules up to 60,000 feet. The mask features a side entry breathing gas supply hose and inhalation valve designed to lower flow resistance. A bespoke computer controlled blended gas on demand delivery system (BGoD3000, Environics, Inc., Tolland, CT) provided the appropriate CO₂ and room air gas concentrations to the participant via the mask.

The BGoD3000 consists of a microcontroller unit (MCU)-based blended gas mixing/delivery system designed specifically to support this project. In addition to MCU controlled closed-loop gas mixing, the BGoD3000 receives control commands over the serial port from a simulation workstation running IHMC's Adaptive Multiagent Integration (AMI) software framework.³¹ An AMI software agent communicated with the BGoD3000 at 1 Hz to check device status, read CO₂ and oxygen (O₂) gas levels, and control the timing and concentrations of the CO₂ level changes. A recovery switch is mounted on both the front panel of the BGoD3000 and remotely within reach of the participant. This allows both the researcher or the participant to override the current gas mix and trigger either 100% oxygen or room air. The BGoD3000 is not approved by the Food and Drug Association (FDA). An inline flow meter (model 4040, TSI Incorporated, Shoreview, MN) provided continuous monitoring of the rate participants inhaled the mixed gas in liters per minute (minute volume). A metabolic cart (CardioCoach-CO₂ Korr Medical Technologies, Inc., Salt Lake City, UT) continuously provided heart rate using a chest strap heart rate monitor (model H1, Polar Electro, INC., Bethpage, NY) and end-tidal CO₂ measurements. Expired breath from the participant was captured at the exhaust valve of the MBU-20/P and passed through a hose to the CardioCoach-CO₂ which analyzed the breath to provide the CO₂ measurements of expired breath.

A sensor placed on the upper left chest provided continuous transcutaneous measurements of arterial CO₂ and O₂ partial pressure (Transcutaneous Monitor (TCM) CombiM, Radiometer, America Inc., Brea, CA) to monitor participant hypercapnia level throughout their experimental sessions. A pulse co-oximeter (Rainbow SET, Masimo Corp, Irvine, CA) provided a real time measure of percent oxygen saturation in the blood and pulse rate using a sensor worn on the left fifth digit. The measure of Oxygen Saturation (SpO₂) provided by the pulse oximeter was continuously monitored throughout the experiment to ensure the blood oxygen saturation of each

participant remained stable. In addition to the BGoD3000, all instrumentation was linked through device specific AMI agents for control, time synchronization and data logging.

3.3 Procedure

The IRB-approved protocol consisted of two days of participation from each participant. On Day 1, the participant received a study in-briefing from the researchers, provided informed consent and completed a medical history and activity questionnaire to confirm no deviation from their current aviation up-chit. A physician reviewed the medical history questionnaire to determine if the participant met any exclusion criteria. Upon enrollment in the study, the participant executed twelve, two-minute iterations of PAT on the practice mode setting. This was completed to establish an individual baseline difficulty level for each of the four PAT tasks.

On Day 2, before the beginning of data collection, the participant donned the heart rate sensor and strap, the TCM CombiM sensor attached to their upper left pectoral region, a pulse-oximeter placed on the fifth digit of the left hand, and an appropriately sized MBU-20/P flight mask, held in place with an elastic sleep apnea head gear with custom mounts for flight helmet bayonet clips. The inhalation valve on the MBU-20/P was connected to the BGoD3000 to deliver the appropriate gas mixtures to the participant and the exhalation valve was connected to the CardioCoach to calculate end-tidal CO₂ from the participant's expired breath. The lab space was light-controlled with all windows covered and the only light source being provided from ceiling lights. The room was not climate controlled. However, the temperature of the room was consistently set at 71°F. The only distractor present was a high-pitched sound emitted from the carbon monoxide detector that would periodically sound off a short 'beep' lasting approximately one second.

The participants would perform the PAT using a joystick located on the right side of their chair, regardless of the hand dominance of the participant. The monitor on which the PAT was performed was located at eye level approximately 22" to 30" from the participant's face. The monitor size was 23.8 inches.

To calibrate the Tobii eye tracker, the participant stared at calibration dots on a screen which allowed the Tobii to capture and analyze images of their eyes. If Tobii was unable to calibrate to the participant (usually due to interference from the MBU-20/P mask), eye tracking data was not collected for that participant. Eye tracking data was collected on 20 out of 35 participants.

To calibrate the audio signal from the Gentex M169A/AIC microphone in the MBU-20/P flight mask each participant was asked to read a sentence twice; once loudly and once using their regular speaking voice to ensure the audio signal was not clipped but still audible for analysis. We used the free and open-source audio software Audacity® (The Audacity Team, Pittsburgh, PA) to visualize the audio signal waveform. An AMI agent recorded the audio digitized by the experiment workstation's audio hardware as individual waveform audio file format (.wav) files.

The participant completed two PAT tasks and two reading tasks in an alternating order during CO₂ exposures. The BGoD3000 delivered four levels of CO₂ (1.0%, 2.5%, 4.0%, and 5.5%) at a constant positive pressure to the participants via the mask at an average flow rate of 70 liters per minute, each for a fifteen-minute period. During experimental conditions, participants completed

PAT trials in assessment mode. The reading tasks consisted of three passages commonly used in the clinical and research speech literature: (1) “My Grandfather”³²; (2) “The Caterpillar”³³; and (3) “The Rainbow”³⁴. The data collected under the 1.0% CO₂ condition were used as a baseline sample of speech and breath for each participant.

Between each CO₂ exposure level, the BGoD3000 provided room air to the participant through the MBU-20/P for at least five-minutes while the participant remained seated with the sensors in place. The participant’s baseline arterial CO₂ was established at the end of the five-minute break after the 1.0% CO₂ condition. The researchers empirically determined this was sufficient time for the TCM sensor heating element to reach operating temperature in order to warm the skin surface sufficiently to induce vasodilation, which increases the permeability of the skin to O₂ and CO₂. The participant’s vitals were continuously monitored by the researchers using the described instrumentation. If a participant’s partial pressure of CO₂ rose more than 20 mmHg, or if their arterial CO₂ reached 65 mmHg, the session was terminated, and the BGoD3000 participant provided 100% room air.

Participants were debriefed after the completion of the experiment and monitored for hypercapnic symptomology including dizziness, shortness of breath, headache, and anxiety. Either the Principal Investigator or Co-Investigator cleared the participant to leave if he or she was well-appearing and did not self-report any significant symptoms.

4.0 ANALYSIS

4.1 Performance Analysis

The current investigation employed a repeated measures cross-over design. Repeated measures ANOVAs were used to determine if there were significant changes in performance as indicated by PAT performance across the four CO₂ conditions.

4.2 Speech Analysis

A divergence score was computed for each of the individual features corresponding to the participant’s baseline speech sample at 1.0% CO₂. An absolute value of the difference between individual feature values in the experimental sample and the corresponding feature values in the baseline sample was used as the divergence score for each individual feature.

Speech analysis was performed on audio samples from the reading portions of the recorded audio, including both breathing and reading speech. We used the OpenSmile audio analysis toolkit to extract auditory features using the same methodology that was implemented in the Interspeech 2011 Speaker State challenge, which was previously used to determine sleepiness and blood alcohol level from speech.³⁵ The roughly 4,000 features extracted represent the result of combining low-level descriptors with functionals across small windows of audio to determine speech characteristics in both absolute and dynamic terms. Two methods were explored for extracting features: full-passage and breath-segmented. In the full-passage extraction, each reading of a

passage was used to generate a single feature vector. In this case, both breath and speech were averaged together to form a feature vector representing the speaker's state across the reading of the passage. For breath-segmented extraction, we used breath detection to determine speech boundaries, and extracted breath and speech independently. Each speech and breath segment was then labeled and considered independently, allowing us to determine the relative contributions of breath analysis versus speech analysis (using the Interspeech feature set).

To automatically determine the CO₂ concentration from the participant's speech sample, we experimented with multiple different regression models that took the feature vector as input and labels for each CO₂ condition. In-line with previous research and verified with initial experiments, a random forest regressor (provided by the scikit-learn library)³⁶ provided the best performance. It automatically chooses from the many features and cross validates against the training set to find those features that best predict an individual's CO₂ level. The initial hyper-parameters for the random forest regressor (feature selection rate and number of estimators) were determined using a development set consisting of about 10% of the total data. The remaining data was split for multi-fold evaluation for three situations – one in which the speaker to be identified was not in the training data (to determine generalizable detection of hypercapnia based on speech features), one in which the speaker to be identified was included in the training data (although no specific reading appeared in both the training and test set) to determine whether a general model can be augmented with individual-specific information for improved recognition performance when available, and finally, a model that is trained solely on a single individual to make a determination based on individual-specific divergence of speech quality. This split allows us to gauge how specific the hypercapnia features were to the individual and their expected assessment capabilities for a previously seen or unseen individual.

In addition to initial tests of models, we also experimented with existing speaker-state data sets to determine if features from those data sets could be adapted for the hypercapnia recognition. We separately used both the Sleepy Language Corpus³⁷ and the Alcohol Language Corpus³⁸ to train initial models and that were tested on each individual with levels of CO₂ scaled to be proportional to the sleepiness ratings used in the corpus.

4.3 Breath Analysis

Breath analysis was conducted during the PAT performance segments across exposure conditions. Breath analysis was not conducted during completion of the reading task. Each audio file was divided into 10-second-long audio samples, meaning each 10-second sample was used for training or evaluation of the method. Let x_i represent the i th 10-second sample.

4.3.1. Feature Extraction

The Fourier transform method was used to extract features from the breath sounds.

$$X_i(f) = \frac{1}{N} \sum_{k=1}^{N-1} x_i(k) e^{-\frac{j2\pi f k}{N}}$$

Equation 1. The Fourier transform of x_i defined

where N is the dimension of the x_i and $X_i(f)$; $f = 0, 1, \dots, N - 1$ are the Fourier coefficients. We use $|X_i(f)|$; $f = 0, 1, \dots, N - 1$ as the features representing the i th 10-second-long samples.

4.3.2. Deviation from the Baseline

To capture any deviation in the breathing pattern and strategy of participants with respect to their own normal breathing patterns (baseline), we calculated the Euclidean distance between X_i and the baseline X_0 . Because we did not collect the breath sounds at 0% CO₂ in this study, we used the first 10-seconds of the 1.0% CO₂ condition as a participant's baseline.

$$d_{i0} = \sqrt{\sum_{f=0}^{N-1} (|X_i(f)| - |X_0(f)|)^2}$$

Equation 2. The Euclidean distance defined

The distance was calculated for each participant separately and used for CO₂ level detection and classification using an unsupervised approach that represents the divergence of the breathing pattern from normal breathing patterns in the presence of CO₂.

4.3.3. Breath-Based CO₂ Level Detection and Classification

In the current investigation, we utilized two different procedures to detect and classify CO₂ levels from breath sounds:

1. Expert User Approach (Trained on one participant and evaluated on the same participant)

In this procedure, we trained a model on breath sound samples of a participant. Then, the model was used to detect CO₂ levels of the same participant.

1.a. Unsupervised approach: In this method, we used the *k-means* clustering algorithm to put the deviations from the baseline values (d_{i0}) into distinct groups. This method has the advantage of not needing any training data, allowing immediate use with new participants. This approach has a drawback, though, as it requires *a priori* knowledge of the CO₂ exposure levels targeted for identification.

1.b. Supervised approach: In this method, we fitted normal distributions to the deviations from the baseline values (d_{i0}) for each CO₂ level separately, meaning we fitted four different normal distributions for each participant. This approach requires the levels of CO₂ exposure during the training so that we assign each distribution to one known CO₂ level. During the test, we compare the deviations with the mean and variance of these four distributions and determine what the level of CO₂ exposure is.

2. New Participant Approach (Trained on multiple participants and evaluated on a new participant)

This procedure proved the most challenging because participant breathing patterns and strategies differ significantly between individuals (e.g., faster, louder, etc.) To achieve the

best performance, we used six different random forests to detect pairwise CO₂ levels, namely 1.0% vs 2.5%, 1.0% vs 4.0%, 1.0% vs 5.5%, 2.5% vs 4.0%, 2.5% vs 5.5%, 4.0% vs 5.5% CO₂, and then voted among the results to determine the final CO₂ level.

4.4 Receiver Operating Characteristics (ROC) Analysis of Models

We calculated both the area under the curve (AUC) and the ROC curve to determine the utility of each of these approaches as a diagnostic for detection of hypercarbia through the auditory features of breath and speech.

5.0 RESULTS

5.1 Descriptive Statistics of Physiological Parameters Across Conditions

5.1.1. PaCO₂ across Conditions during PAT Tasks

Participants' arterial partial pressure of CO₂ (PaCO₂) as measured by the TCM increased steadily across the four CO₂ exposures while participants performed the PAT (Table 1a, Figure 1a). The average increase from baseline in PaCO₂ across conditions is as follows, the 1.0% CO₂ condition resulted in an average PaCO₂ increase of 5.54 mmHg, the 2.5% condition resulted in an average PaCO₂ increase of 8.92 mmHg, the 4.0% condition resulted in an average PaCO₂ increase of 13.20 mmHg, and the 5.5% condition resulted in an average PaCO₂ increase of 14.46 mmHg (Table 1a, Figure 1a).

Table 1 a. Increase in PaCO₂ from Baseline Across Conditions During PAT Tasks

Condition	<u>1% CO₂ PAT</u>		<u>2.5% CO₂ PAT</u>		<u>4.0% CO₂ PAT</u>		<u>5.5% CO₂ PAT</u>	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Mean (mmHg)	3.91	5.54	4.87	8.92	7.20	13.20	9.24	14.46
SD	4.32	4.24	5.28	4.39	6.86	4.98	9.32	8.62

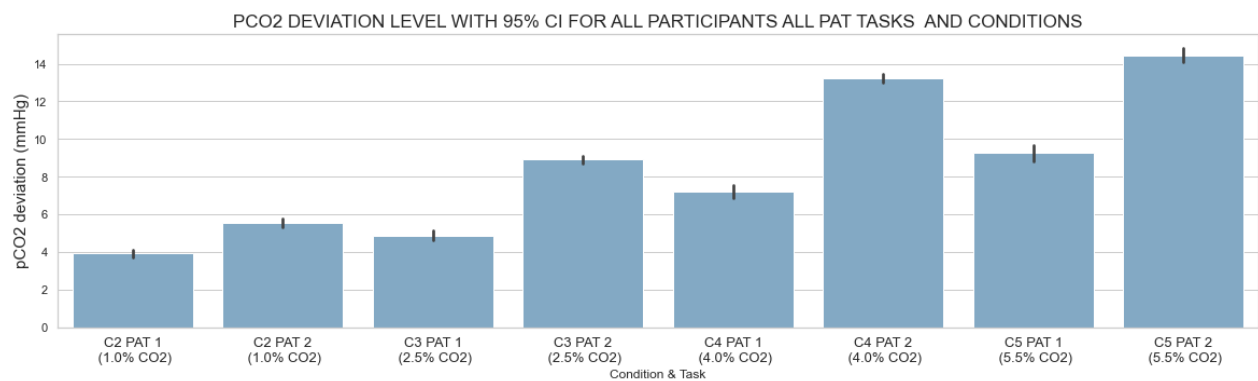


Figure 1 a. Graphical Display of Increase in PaCO₂ from Baseline Across Conditions During PAT Tasks

5.1.2. PaCO₂ across Conditions during Reading Tasks

Participants' PaCO₂ as measured by the TCM during the reading tasks also increased steadily from baseline across the four CO₂ exposure levels (Table 1.b, Figure 1.b). The average increase in PaCO₂ across conditions is as follows: the 1.0% CO₂ condition resulted in an average increase in PaCO₂ of 4.71 mmHg, the 2.5% condition resulted in an average increase in PaCO₂ of 8.35 mmHg, the 4.0% condition resulted in an average increase in PaCO₂ of 12.26, and the 5.5% condition resulted in an average increase in PaCO₂ of 16.29 mmHg (Table 1b, Figure 1b).

Table 1 b. Increase in PaCO₂ from Baseline Across Conditions During Reading Tasks

Condition	<u>1.0% CO₂</u>		<u>2.5% CO₂</u>		<u>4.0% CO₂</u>		<u>5.5% CO₂</u>	
	<u>Reading</u>		<u>Reading</u>		<u>Reading</u>		<u>Reading</u>	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Mean (mmHg)	4.16	4.71	5.43	8.35	8.19	12.26	10.86	16.29
SD	4.26	3.92	5.26	4.68	6.19	5.03	7.96	6.44

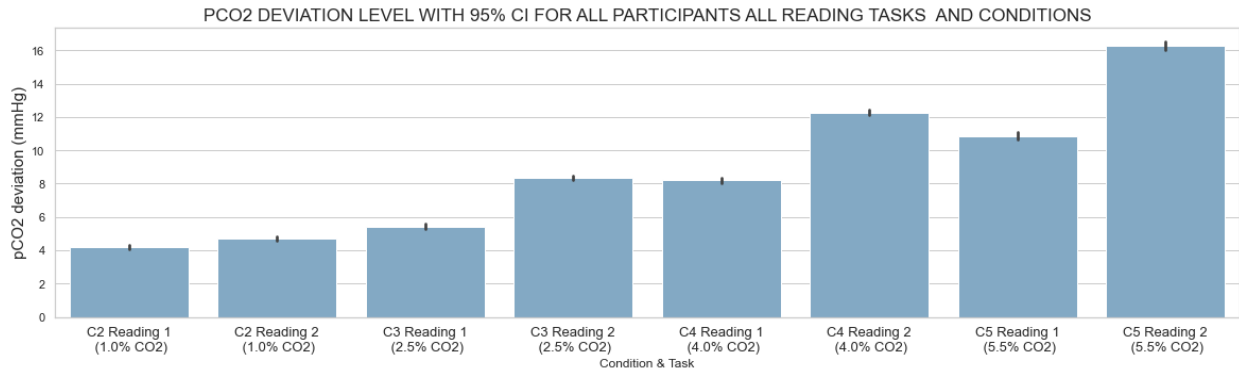


Figure 1 b. Graphical Display of Increase in PaCO₂ from Baseline Across Conditions During Reading Tasks

5.1.3. Heart Rate Across Conditions during PAT tasks

Heart rate rose slightly higher in the 1.0% condition when compared to the 2.5% and 4.0% CO₂ conditions. However, beginning with the 2.5% condition, heart rate did appear to increase in a stepwise fashion with increasing CO₂ levels (Table 2a, Figure 2a). The 5.5% CO₂ condition resulted in the highest average heart rate of 88.56 beats per minute (bpm). On average, heart rate increased 8.1 beats per minute from the 2.5% condition to the 5.5% condition (Table 2a, Figure 2a).

Table 2 a. Heart Rate Across Conditions During PAT Tasks

Condition	1% CO ₂ PAT		2.5% CO ₂ PAT		4.0% CO ₂ PAT		5.5% CO ₂ PAT	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Mean (bpm)	82.97	80.46	78.97	79.23	79.61	82.18	82.78	88.56
SD	12.62	11.30	11.76	12.15	11.69	12.94	11.68	14.92

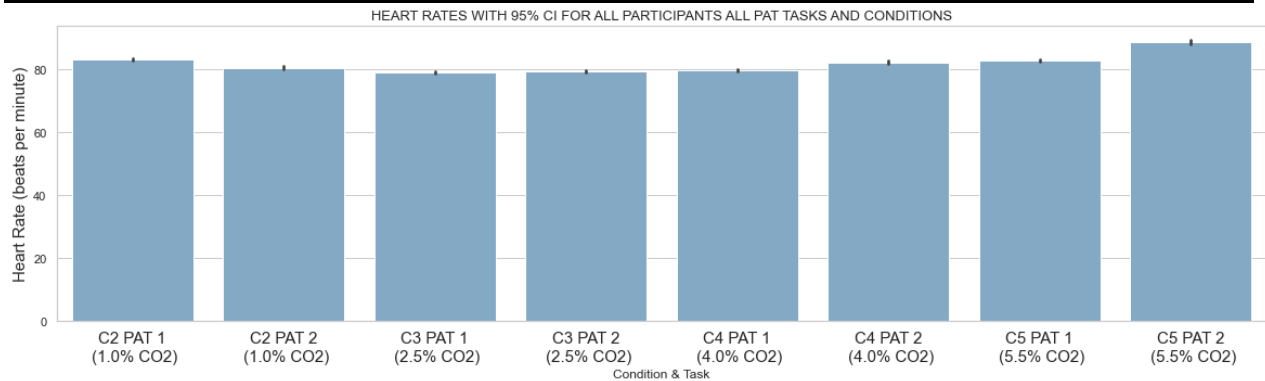


Figure 2 a. Graphical Display of Heart Rate Across Conditions During PAT Tasks

5.1.4. Heart Rate Across Conditions during Reading tasks

During the reading tasks, heart rates fluctuated in the 1.0% and the 2.5% CO₂ condition. However, heart rate consistently increased in the 4.0% and 5.5% CO₂ condition (Table 2b, Figure 2b). The 5.5% CO₂ exposure resulted in the highest average heart rate of 85.31 bpm. On average, heart rate increased 6.95 beats per minute from the 4.0% condition to the 5.5% condition (Table 2b, Figure 2b).

Table 2 b. Heart Rate Across Conditions During Reading Tasks

Condition	<u>1% CO₂</u>		<u>2.5% CO₂</u>		<u>4.0% CO₂</u>		<u>5.5% CO₂</u>	
	<u>Reading</u>		<u>Reading</u>		<u>Reading</u>		<u>Reading</u>	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Mean (bpm)	79.34	78.25	77.22	78.40	78.36	80.32	81.23	85.31
SD	11.62	11.16	11.12	11.85	12.53	13.23	14.07	15.23

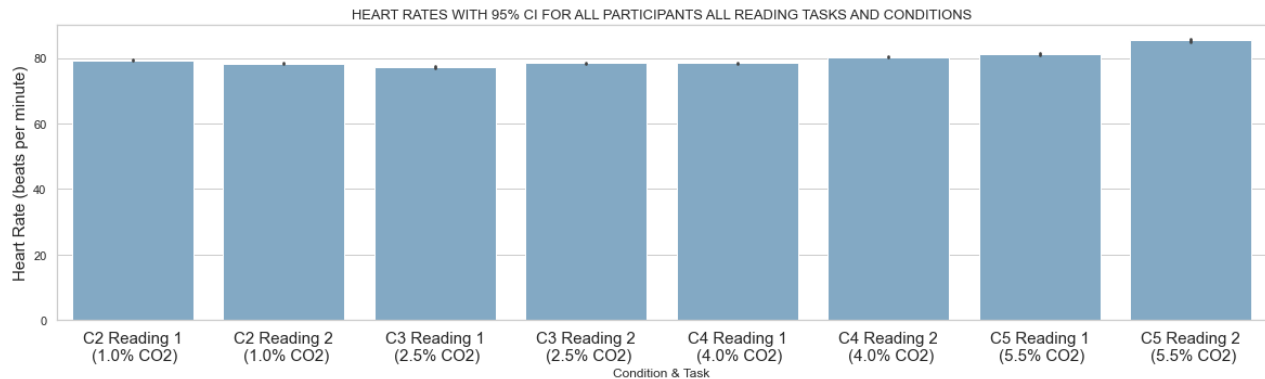


Figure 2 b. Graphical Display of Heart Rate Across Conditions During Reading Tasks

5.1.5. Breath Rate Across Conditions

Average breath rate increased with CO₂ exposure concentration in a stepwise fashion like the average PaCO₂ measurements (Table 3, Figure 3). Overall, average breath rate increased by 13.34 breaths per minute from the 1.0% CO₂ condition to the 5.5% CO₂ condition.

Table 3. Breath Rate Across Conditions During PAT Tasks

Condition	<u>1% CO₂ PAT</u>		<u>2.5% CO₂ PAT</u>		<u>4.0% CO₂ PAT</u>		<u>5.5% CO₂ PAT</u>	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Mean (breaths/min)	7.95	8.89	9.57	12.19	14.07	16.56	19.06	21.29
SD	6.95	7.25	7.75	8.33	10.07	10.37	13.65	12.48

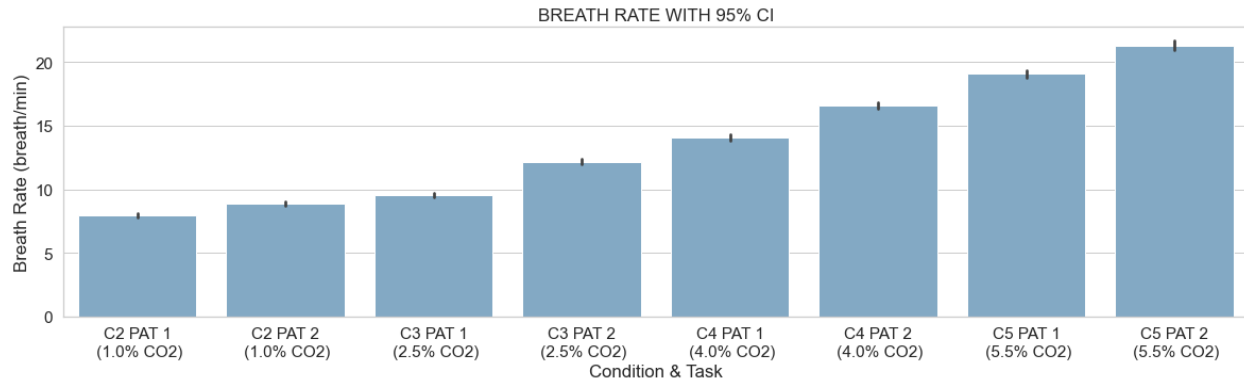


Figure 3. Graphical Display of Breath Rate Across Conditions During PAT Tasks

5.1.6. Tidal Volume Across Conditions

Much like respiration rate and PaCO₂, tidal volume correlated positively with CO₂ exposure concentration. On average, tidal volume increased by 1.314 liters from the 1.0% CO₂ condition to the 5.5% CO₂ condition (Table 4, Figure 4).

Table 4. Tidal Volume Across Conditions During PAT Tasks

Condition	<u>1% CO₂ PAT</u>		<u>2.5% CO₂ PAT</u>		<u>4.0% CO₂ PAT</u>		<u>5.5% CO₂ PAT</u>	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Mean (L/min)	0.8861	0.9656	1.064	1.259	1.452	1.775	1.827	2.276
SD	0.4204	0.3811	0.4018	0.3704	0.5631	0.4503	0.7164	0.6051

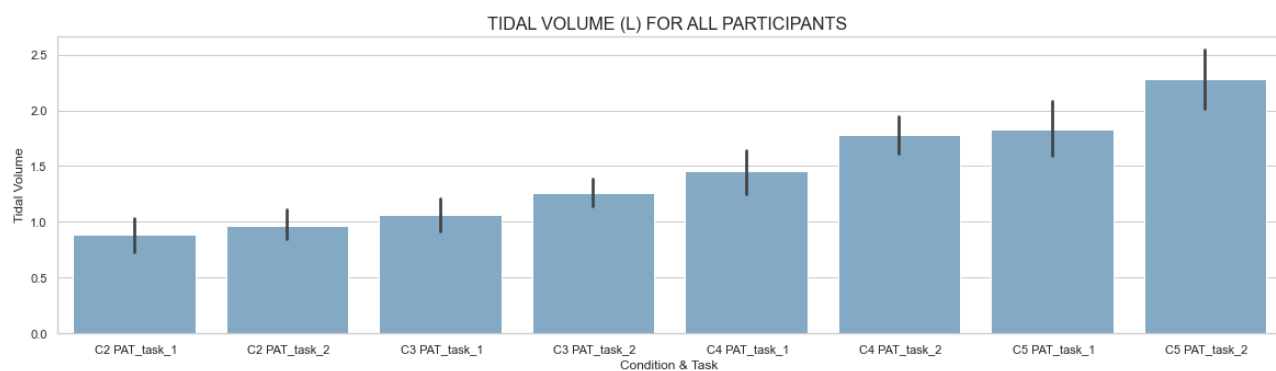


Figure 4. Graphical Display of Tidal Volume Across Conditions During PAT Tasks

5.1.7. Minute Ventilation Across Conditions

Minute ventilation increased from 3.37 liters per minute in the 1% CO₂ condition to 10.80 liters per minute in the 5.5% CO₂ condition, an average increase of 7.02 liters per minute (Table 5, Figure 5).

Table 5. Minute Ventilation Across Conditions During PAT Tasks

Condition	<u>1% CO₂ PAT</u>		<u>2.5% CO₂ PAT</u>		<u>4.0% CO₂ PAT</u>		<u>5.5% CO₂ PAT</u>	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Mean (L/min)	3.37	3.78	3.88	5.33	5.84	8.07	7.99	10.80
SD	1.63	1.45	1.90	1.62	2.79	2.02	3.97	2.54

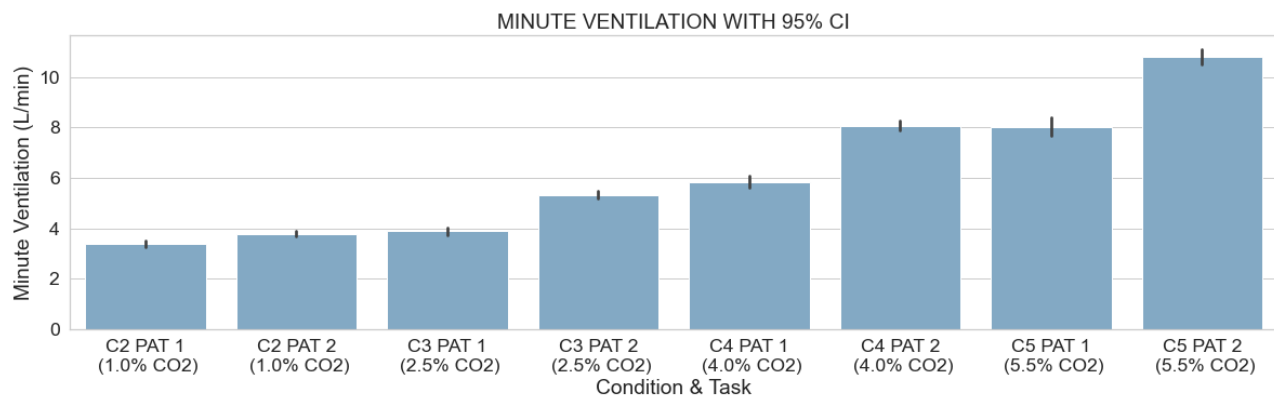


Figure 5. Graphical Display of Minute Ventilation Across Conditions During PAT Tasks

5.2 Performance Assessment Tool (PAT) Results

A repeated measures Analysis of Variance (ANOVA) was conducted on overall PAT performance and PAT Tracking Scores to establish whether there were significant performance effects associated with exposure to higher levels of CO₂. The analysis compared the second PAT assessment across each exposure condition. The second PAT session was used because it was conducted later in the exposures during the participants' maxim hypercarbia state associated with the respective condition. Table 6 shows the mean performance for overall PAT Scores and PAT Tracking Scores per CO₂ exposure.

Table 6. Mean PAT Performance Across CO₂ Exposure Concentrations

CO ₂ Exposure	1.0%	2.5%	4.0%	5.5%
PAT Scores Mean	848.42	852.20	867.64	866.32
SE Mean	14.05	15.20	13.82	11.01
Tracking Scores Mean	508.03	527.62	536.40	535.32
SE Mean	16.50	15.50	15.96	18.23

The ANOVA conducted on overall PAT performance did not show any significant differences across conditions, $f(1.9, 60.95) = 1.16$, $p=0.32$, *partial eta square* = 0.035. The ANOVA conducted on the tracking component of PAT without the subtasks showed significant differences between conditions, $f(1.69, 28.78) = 4.39$, $p = 0.027$, *partial eta square* = 0.21. Pairwise comparisons revealed significantly better tracking performance in the 2.5% and 4.0% exposure conditions when compared to the 1.0% exposure condition. The improvement in tracking performance ceased to exist in the 5.5% condition. (figure 6 and 7)

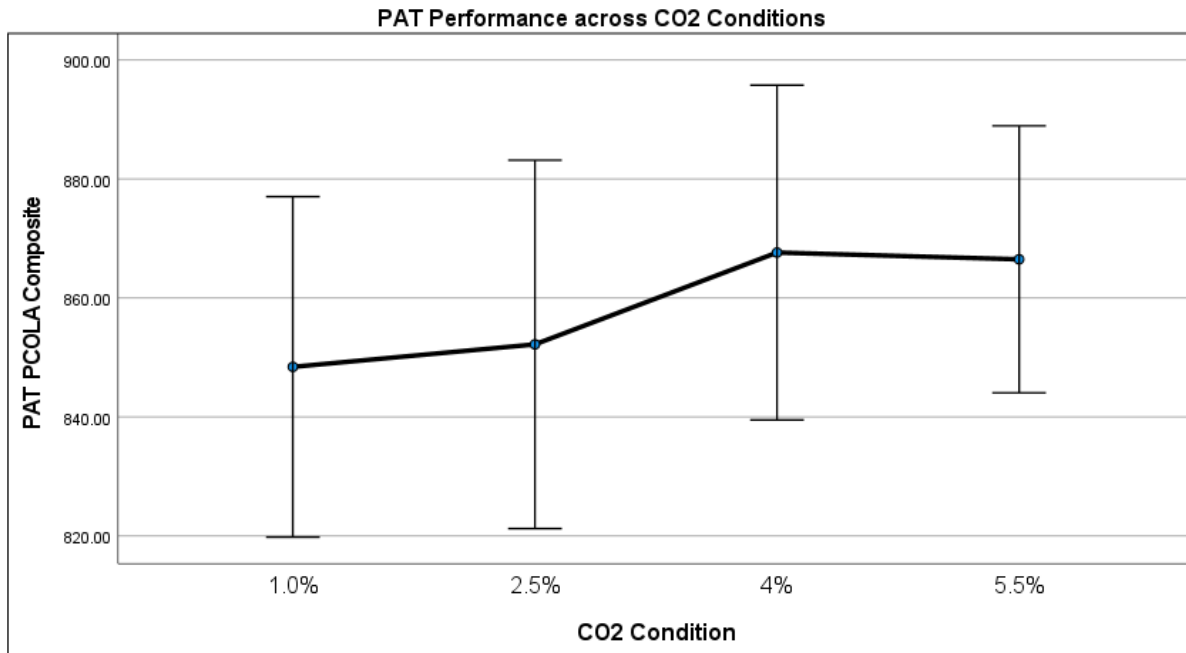


Figure 6. ANOVA results for overall PAT performance across CO₂ conditions

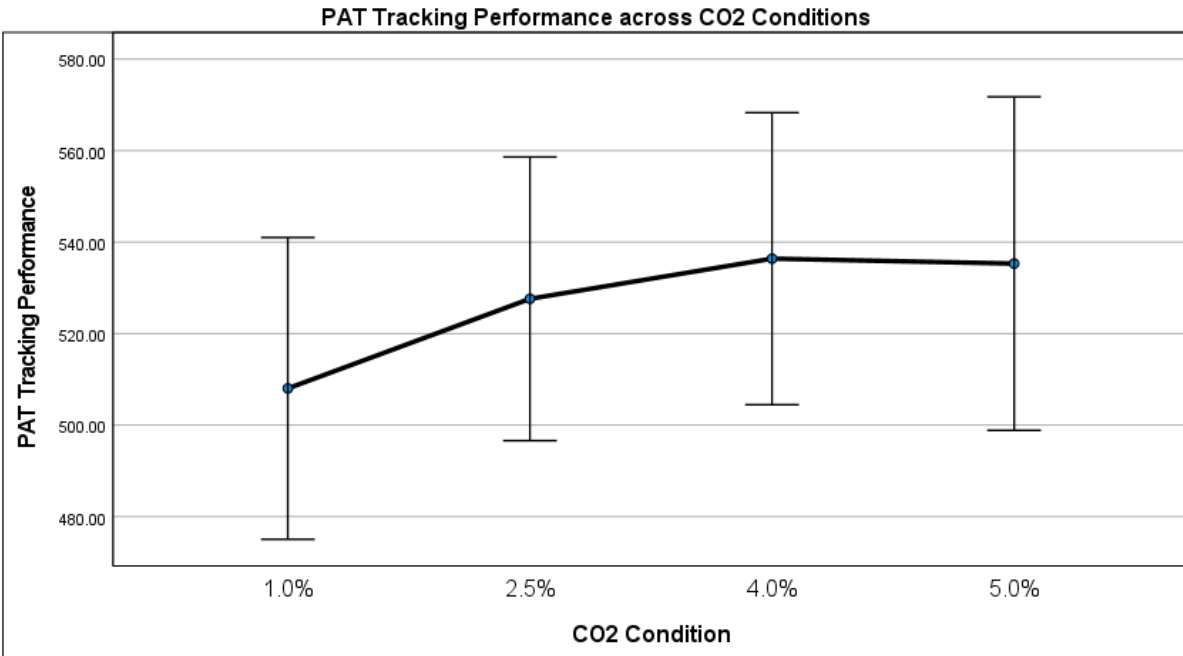


Figure 7. ANOVA results for PAT tracking performance across CO₂ condition

5.3 Speech Results

We extracted a single vector of features for each reading passage using both breathing and speaking together and also by using the breath segmentation method to separate speech and breathing from the recordings. The speech was more predictive than breath using this latter feature set, but the former feature set, which incorporated the two variables, provided better results. We, therefore, focused on experiments combining features extracted from both speech and breath taken together.

The models trained to detect sleepiness and blood alcohol levels using the Large Sleepy Language Corpus and the Alcohol Language Corpus were ineffective for determining CO₂ levels, indicating that hypercapnia affects speech in a unique way.

Initial hyper-parameters were trained with a development set consisting of about 10% of the total data. For the tested speaker not included in the training set, we achieved an average R^2 of .29 (an internal metric of the goodness of fit of the model), with a Pearson $R = .69$ ($p < .001$) calculated through the correlation of the ground truth CO₂ supply values with those generated by the model for the test set (Figure 8).

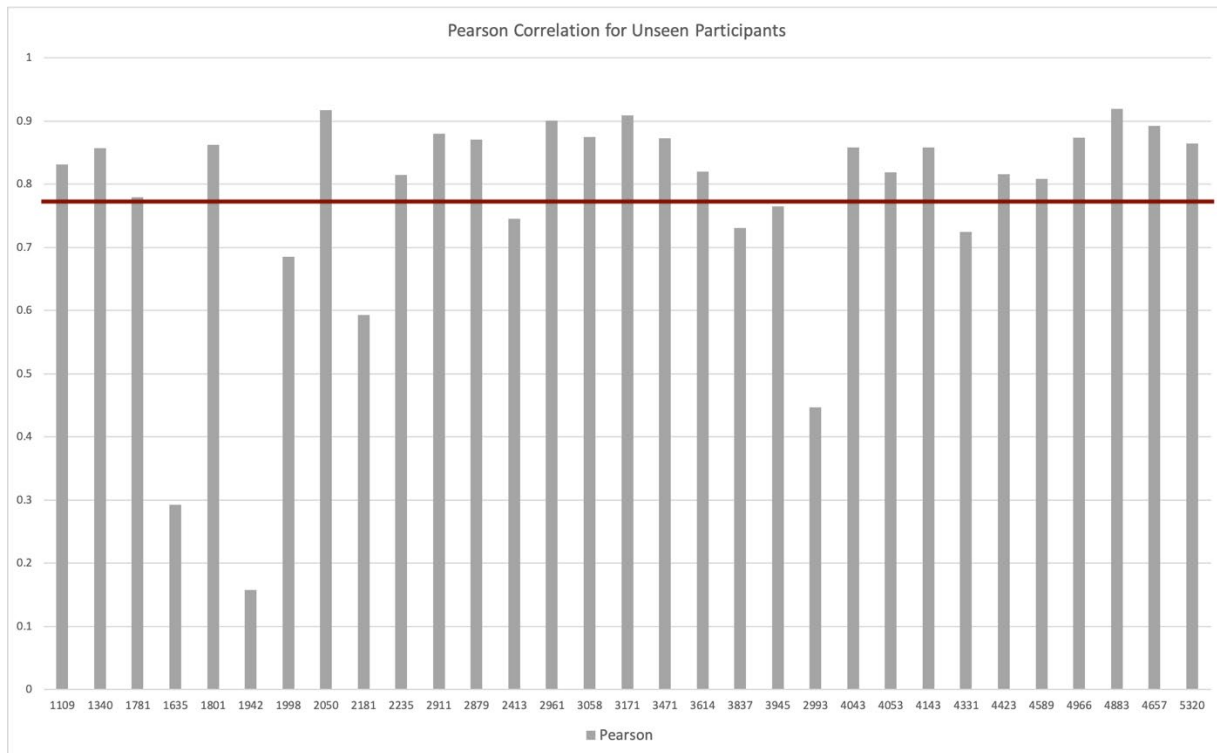


Figure 8. Pearson Correlation for Each Participant (and Average) in the Unseen Experiment – Where the Individual was not Present in the Training Data

For the prediction models where the tested individual was present in the training data (Combined), we achieved an average R^2 of .56, with a Pearson $R = .89$ ($p < .001$) (Figure 9).



Figure 9. Pearson Correlation for each Participant (and Average) in the Combined Experiment – Where the Tested Individual was Present in the Training Data Along with the Other Individuals

Finally, for the divergence evaluation (model trained on half of the speaker's data and tested on the other half), we achieve an average R^2 of .58, with a Pearson $R = .89$ ($p < .001$) (Figure 10).

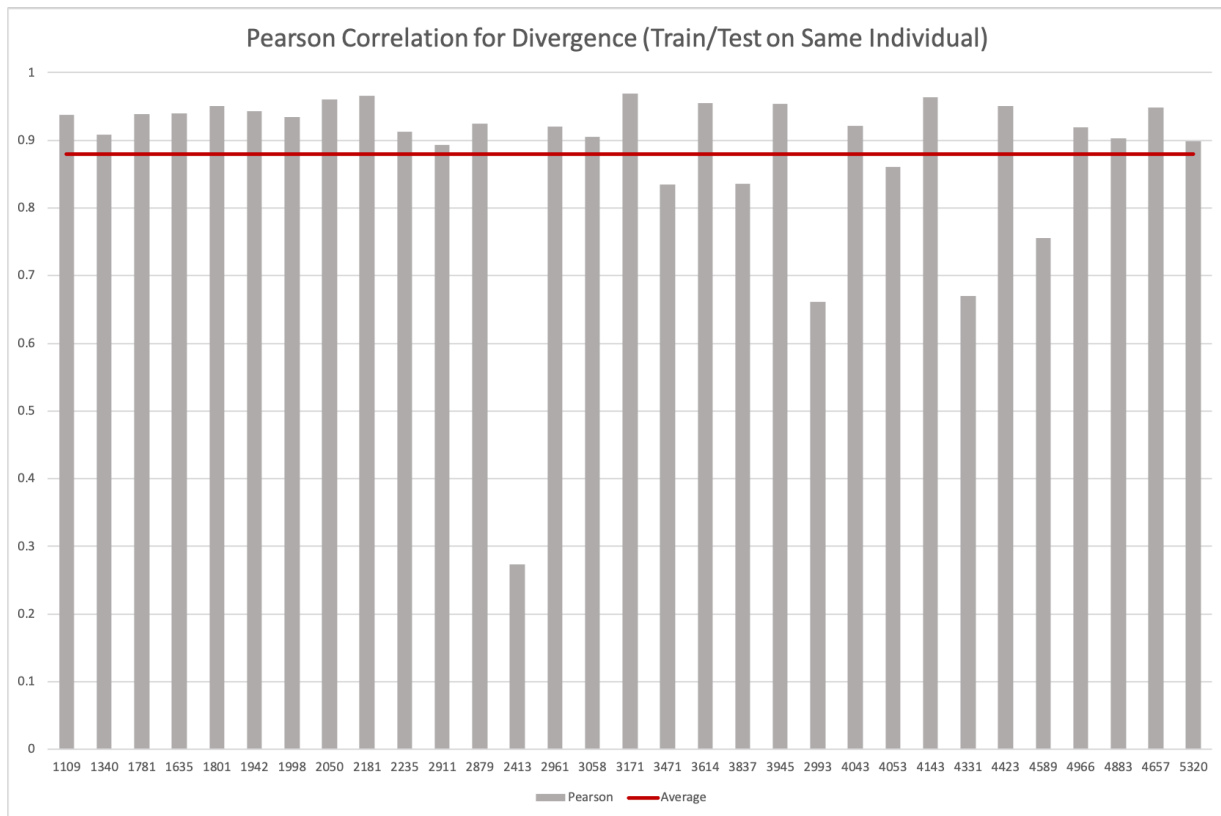


Figure 10. Pearson Correlation for Each Participant (and Average) in the Divergence Experiment – Where only the Tested Individual was Present in the Training Data to Evaluate Individual-Specific Hypercapnia Detection

In addition to applying regression models to the speech feature analysis, we also developed a classification model for distinguishing between various levels of CO₂. This classifier was trained using the same multi-fold approach used for the new user evaluation. Results are shown in Table 7 and Figure 11.

Table 7. The Area Under the Curve (AUC) of Detecting CO₂ Level from Speech Using the Unseen Experiment

CO ₂ Exposure	1.0%	2.5%	4.0%	5.5%
AUC	0.82	0.71	0.69	0.88

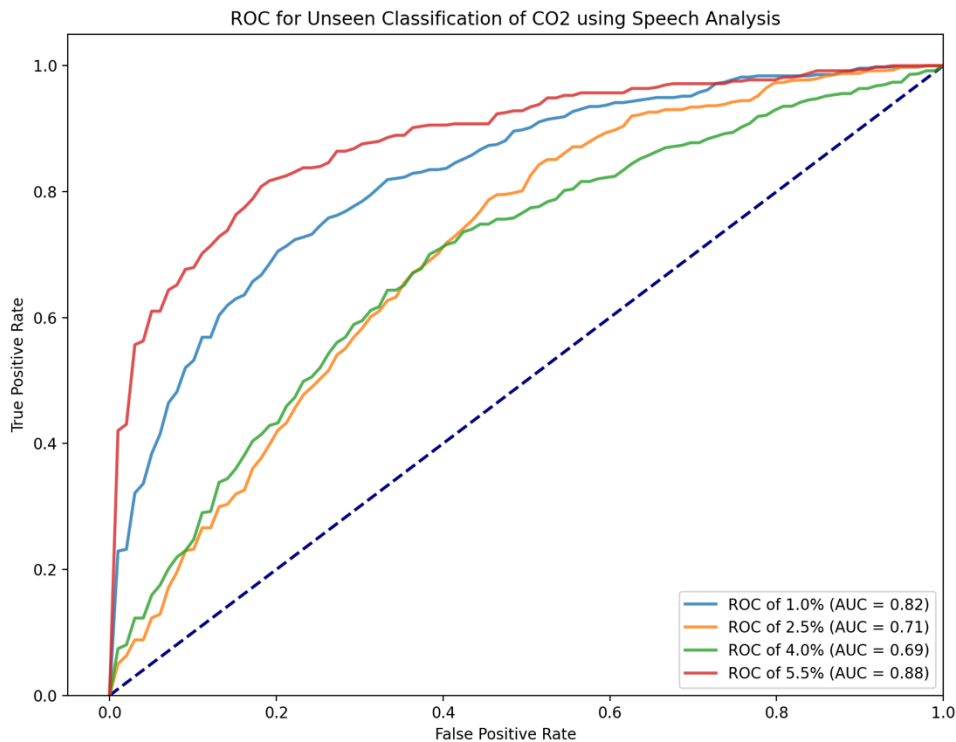


Figure 11. ROC for Predicting Each Participant’s CO₂ level in the Unseen Experiment – where the Tested Individual was not Present in the Training Data

5.4 Breath Analysis Results

5.4.1. Expert User Approach (Trained on one participant and evaluated on the same participant)

1.a. Unsupervised Approach

In the unsupervised approach described in Section 4.3.3, we compared each 10-second segment of each participant with their own 10-second baseline. The deviation of samples from the baseline was clustered into four separate groups. The clustered members were then compared with the ground truth to calculate the accuracy of clustering for each participant. Table 8 and Figure 12 shows the diagnostic utility of the unsupervised approach through corresponding AUC and ROC analyses. Figure 13 shows the accuracy of the test data using the unsupervised approach for each participant separately.

Table 8. The Area Under the Curve (AUC) of Detecting CO₂ Level from Breath Sounds for Each Participant Using the Expert User Unsupervised Approach

CO ₂ Exposure	1.0%	2.5%	4.0%	5.5%
AUC	0.91	0.92	0.89	0.85

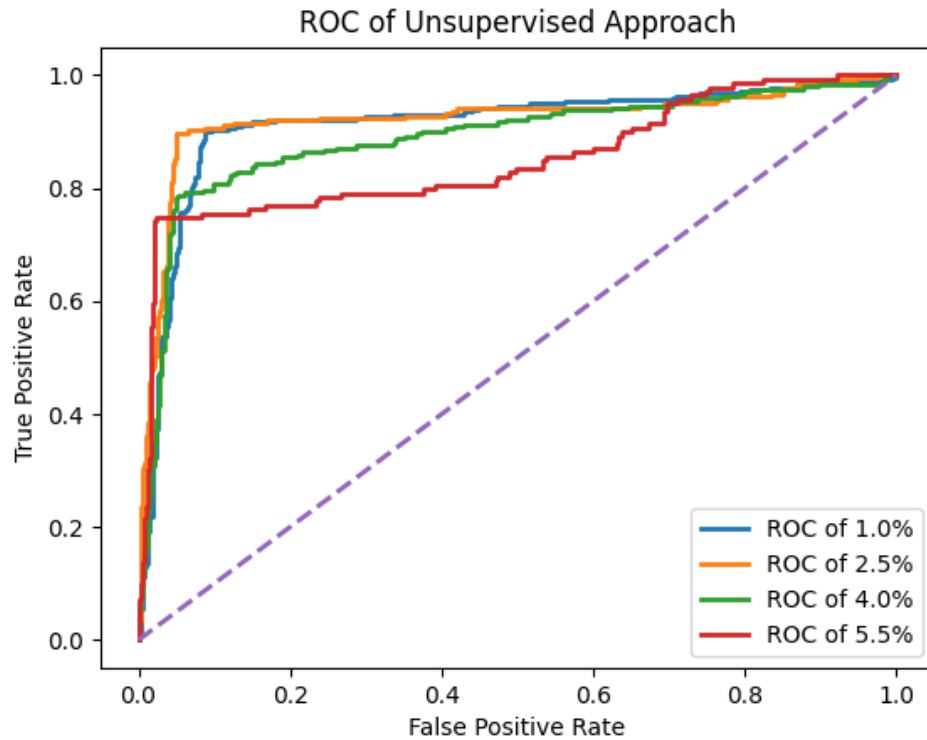


Figure 12. The ROC of Detecting Changes in CO₂ Level from Breath Sounds for Each Participant Using Unsupervised Approaches (Expert User Unsupervised Approach)

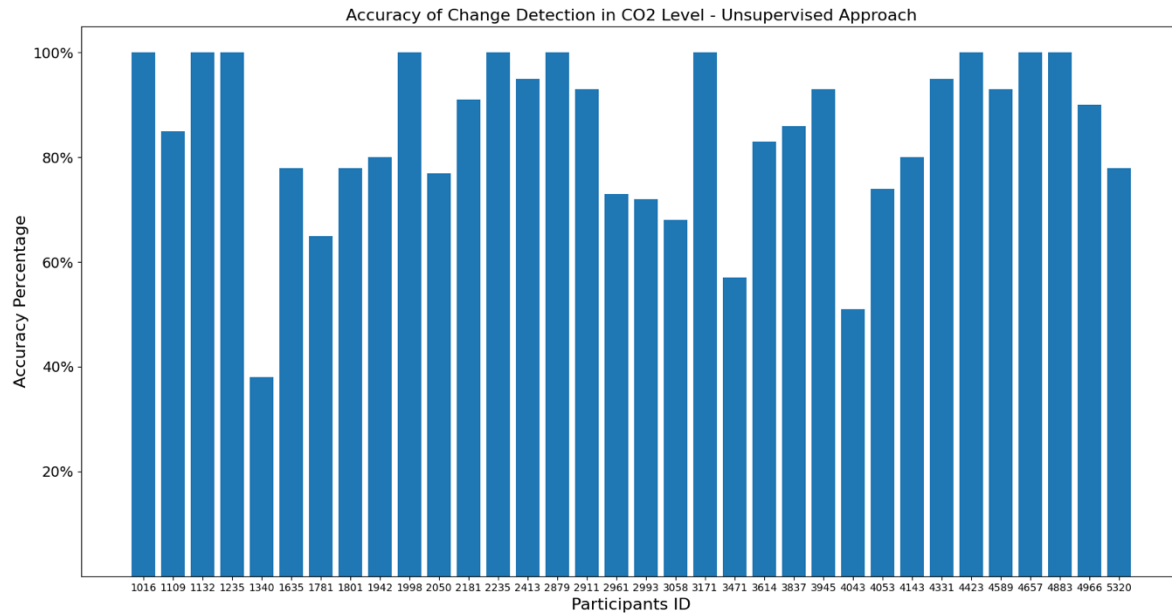


Figure 13. The Accuracy of Detecting Changes in CO₂ Level from Breath Sounds for Each Participant Using Unsupervised Approaches

Total Accuracy is 84.5%. The Total Accuracy After Taking Out the Outlier (id: 1340) is 85.91%

1.b. Supervised Approach

Within a participant's sample, 80% of the data in each CO₂ condition was allocated for training and 20% of the data in each CO₂ condition was allocated for testing. We trained the method described in Section 4.3.3 on the training dataset of each participant, and then used the trained model to predict the CO₂ level of the test dataset. The accuracy of the trained model was calculated based on the predicted CO₂ level for the test dataset of each participant separately when compared with the ground truth. Table 9 and Figure 14 shows the diagnostic utility of the supervised approach through corresponding AUC and ROC analyses. Figure 15 shows the accuracy of the test data for each participant separately.

Table 9. The Area Under the Curve (AUC) of Detecting CO₂ Level from Breath Sounds for Each Participant When the Model is Trained on Each Participant (Expert User Supervised Approach)

CO ₂ Exposure	1.0%	2.5%	4.0%	5.5%
AUC	0.94	0.83	0.83	0.91

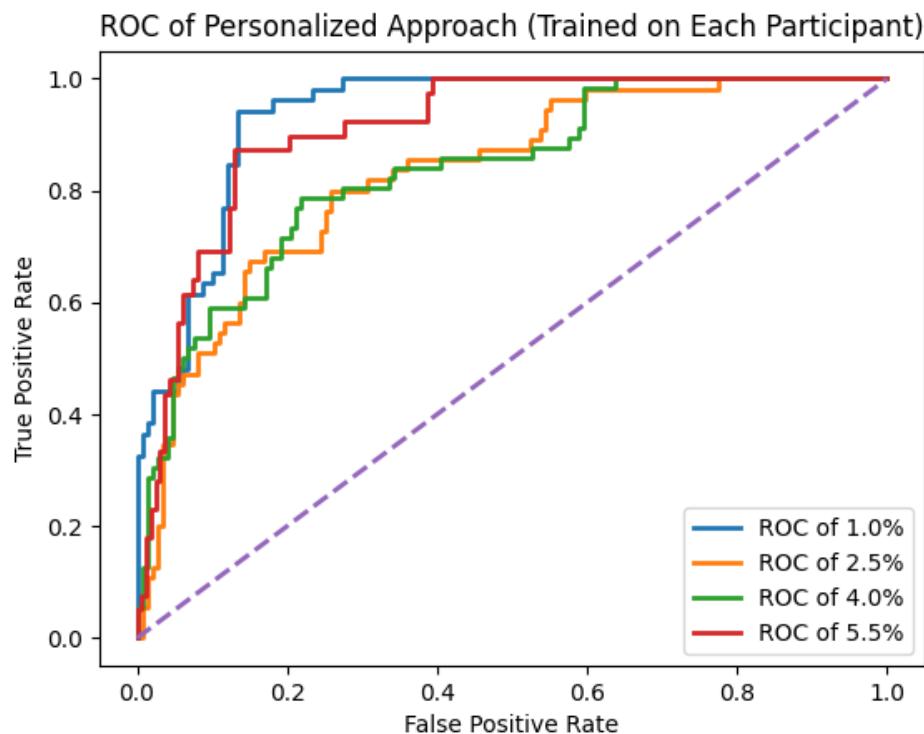


Figure 14. The ROC of Detecting CO₂ Level from Breath Sounds for Each Participant when the Model is Trained on the Same Participant (Expert User Supervised Approach)

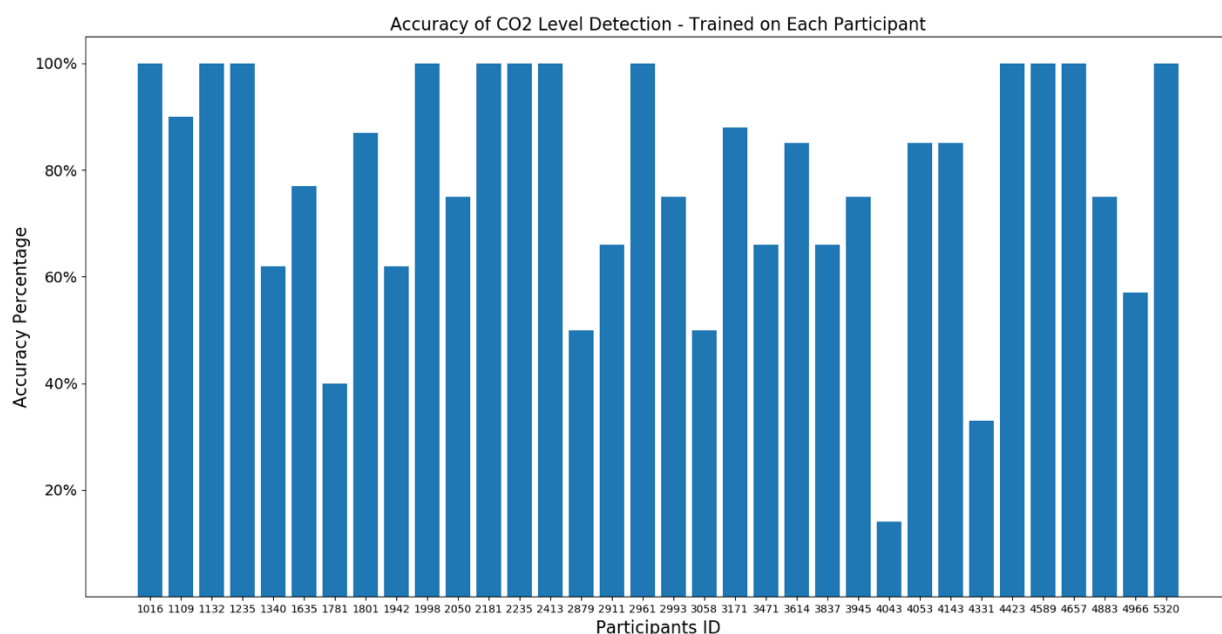


Figure 15. The Accuracy of Detecting CO₂ Level from Breath Sound for Each Participant when the Model is Trained on the Same Participant (Expert User Supervised Approach)
The Averaged Accuracy is 78.32%. When Outliers (4043, 4331, 1781) are Taken Out, the Averaged Accuracy Increases to 83.10%

2. Trained on Multiple participants and evaluated on a new participant

We also used a k-folding approach to train a model on a subset of participants and evaluated it on new participants as described in Section 4.3.3. Six different random forest models were used to detect differences between each pair of CO₂ conditions (e.g., 1.0% vs 2.5%, 1.0% vs 4.0%, 1.0% vs 5.5%, 2.5% vs 4.0%, 2.5% vs 5.5%, 4.0% vs 5.5%). Figure 16 shows the ROC of the experiment, and Table 10 shows the AUC for each CO₂ level. Figure 17 shows the accuracy of the test data for each participant separately.

Table 10. The Area Under the Curve (AUC) of Detecting CO₂ Level from Breath Sounds for New Participants

CO ₂	1.0% vs.	1.0% vs.	1.0% vs	2.5% vs.	2.5% vs.	4.0% vs.
Exposure	2.5%	4.0%	5.5%	4.0%	5.5%	5.5%
AUC	0.73	0.85	0.82	0.64	0.71	0.57

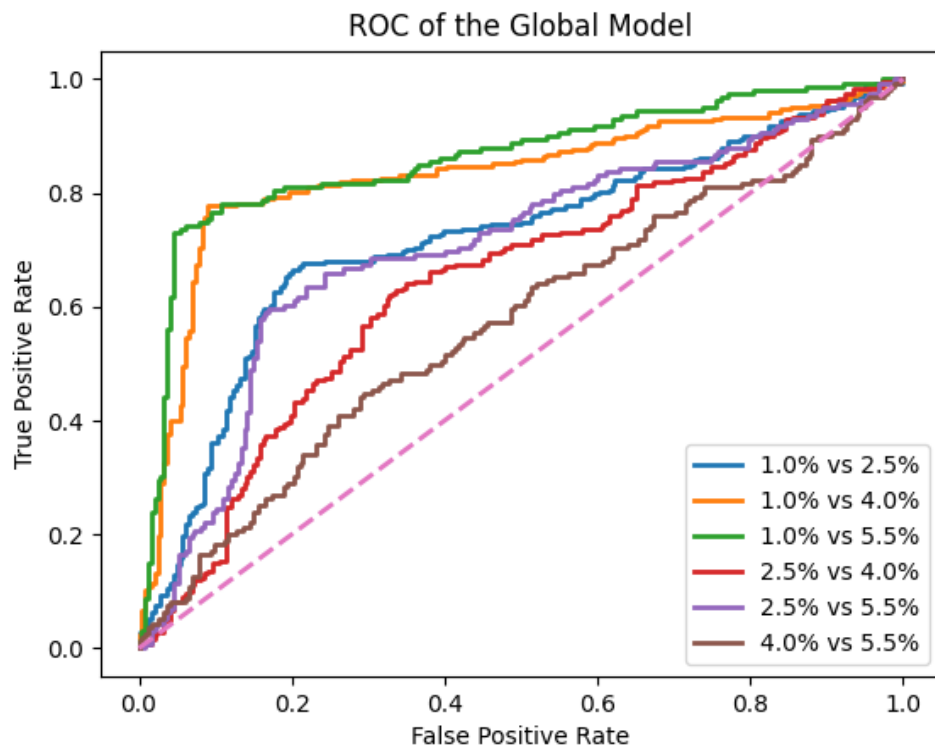


Figure 16. The ROC of Classifying CO₂ Levels from Breath Sound for New Participants

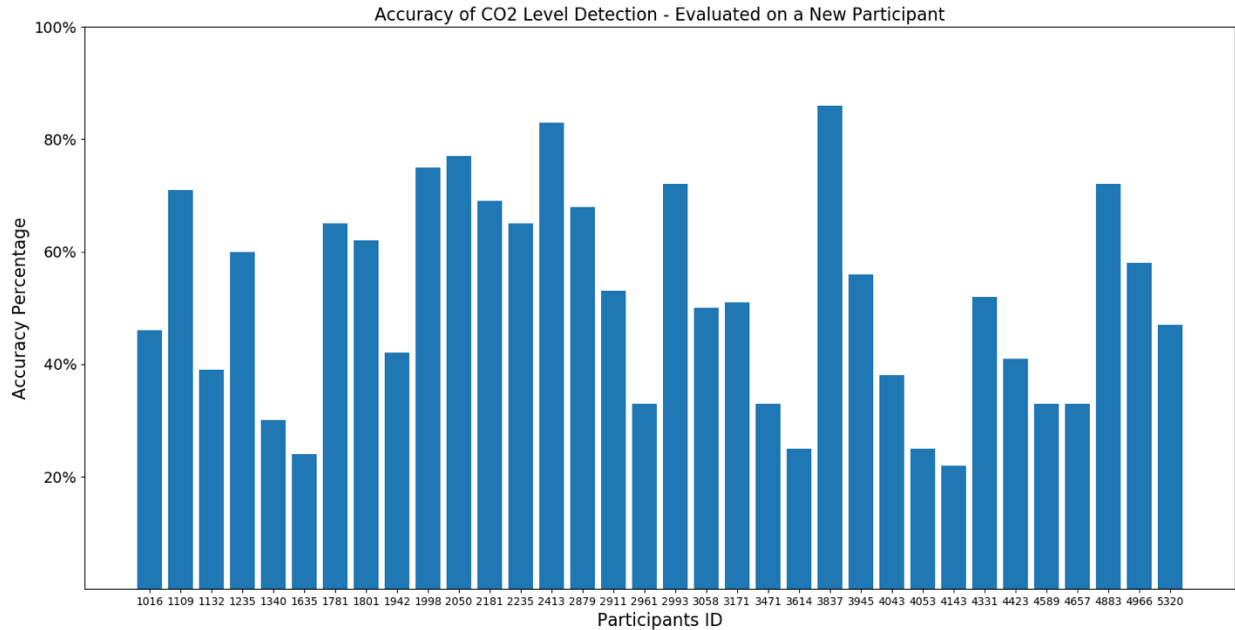


Figure 17. The Accuracy of Classifying CO₂ Levels from Breath Sounds for New Participants Using K-folding
The Averaged Accuracy is 51.65%

As a component of this study, we also evaluated the performance of the proposed method on new participants when assuming only two CO₂ levels to determine if the algorithm can distinguish two different levels of CO₂ accurately. Figure 18 and Figure 19 show the accuracy of this evaluation for each new participant when the algorithm is trained on the remaining participants. The accuracy of the algorithm averaged across all participants for each category is given in Table 11. The best accuracy manifested when distinguishing between the 1.0% and 5.5% CO₂ levels while the lowest accuracy occurred when distinguishing between CO₂ levels close to each other, e.g., 2.5% vs 4.0% and 4.0% vs 5.5%. This was likely because the breathing patterns and strategies are very similar to each other in these similar CO₂ levels.

Table 11. The Accuracy of Detecting Pairwise CO₂ Levels

CO ₂ levels	1.0 vs 2.5	1.0 vs 4.0	1.0 vs 5.5	2.5 vs 4.0	2.5 vs 5.5	4.0 vs 5.5
Accuracy	72.62%	84.90%	88.68%	64.68%	74.30%	62.22%

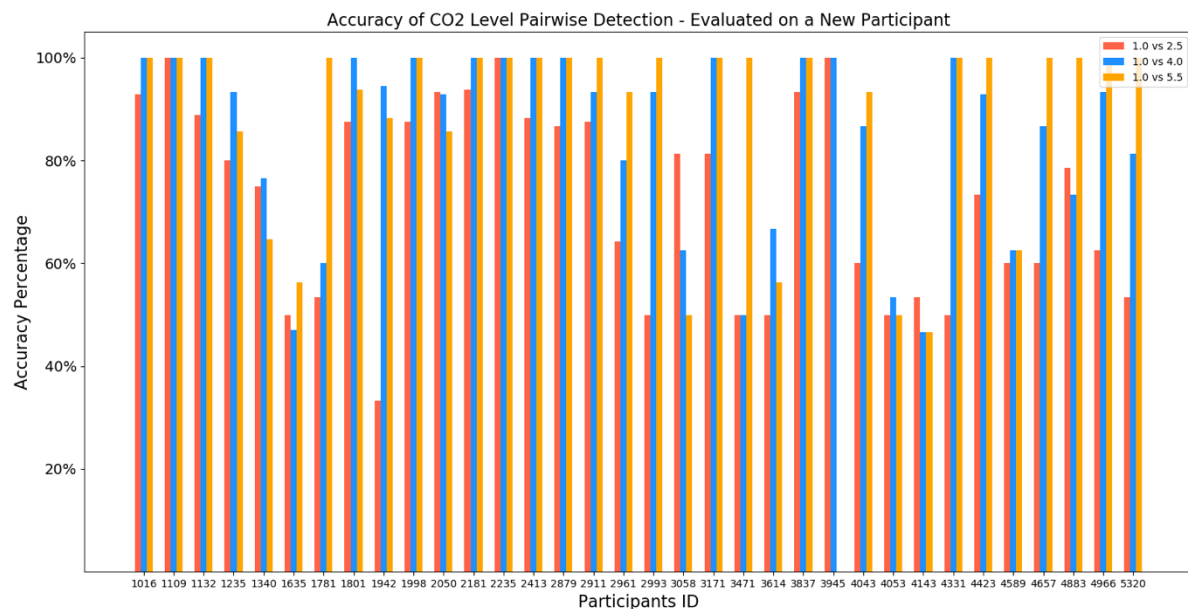


Figure 18. The Accuracy of Classifying Pairwise CO₂ Levels (1.0 vs 2.5, 1.0 vs 4.0, 1.0 vs 5.5) from Breath Sounds for New Users K-folding

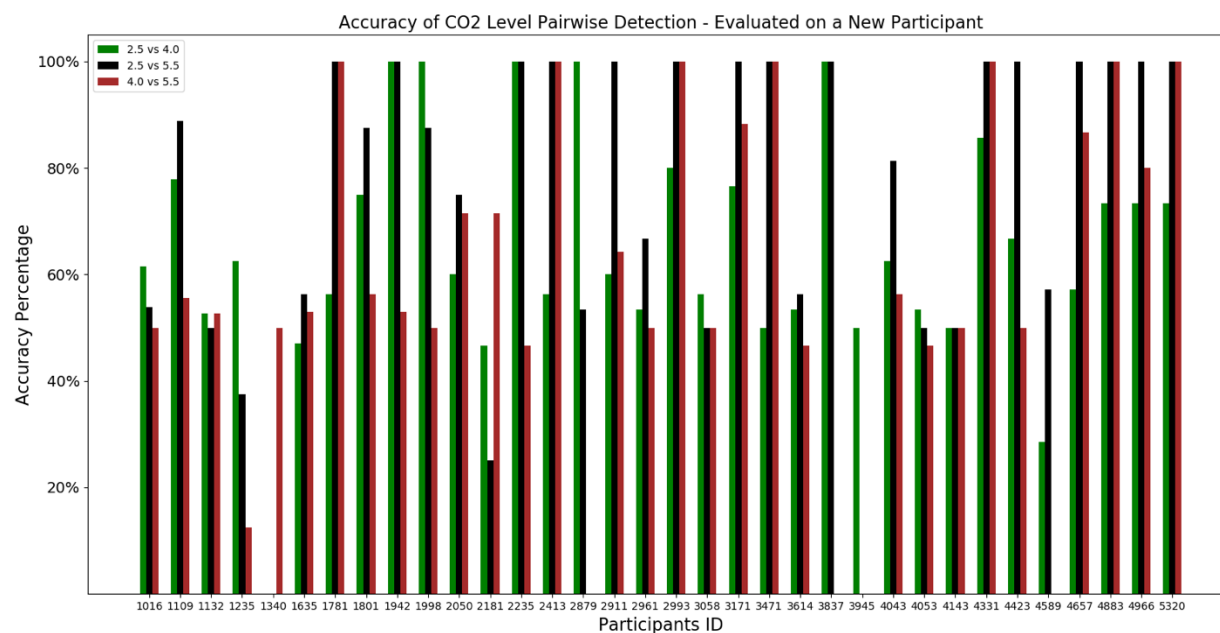


Figure 19. The Accuracy of Classifying Pairwise CO₂ Levels (2.5 vs 4.0, 2.5 vs 5.5, 4.0 vs 5.5) from Breath Sounds for New Users K-folding

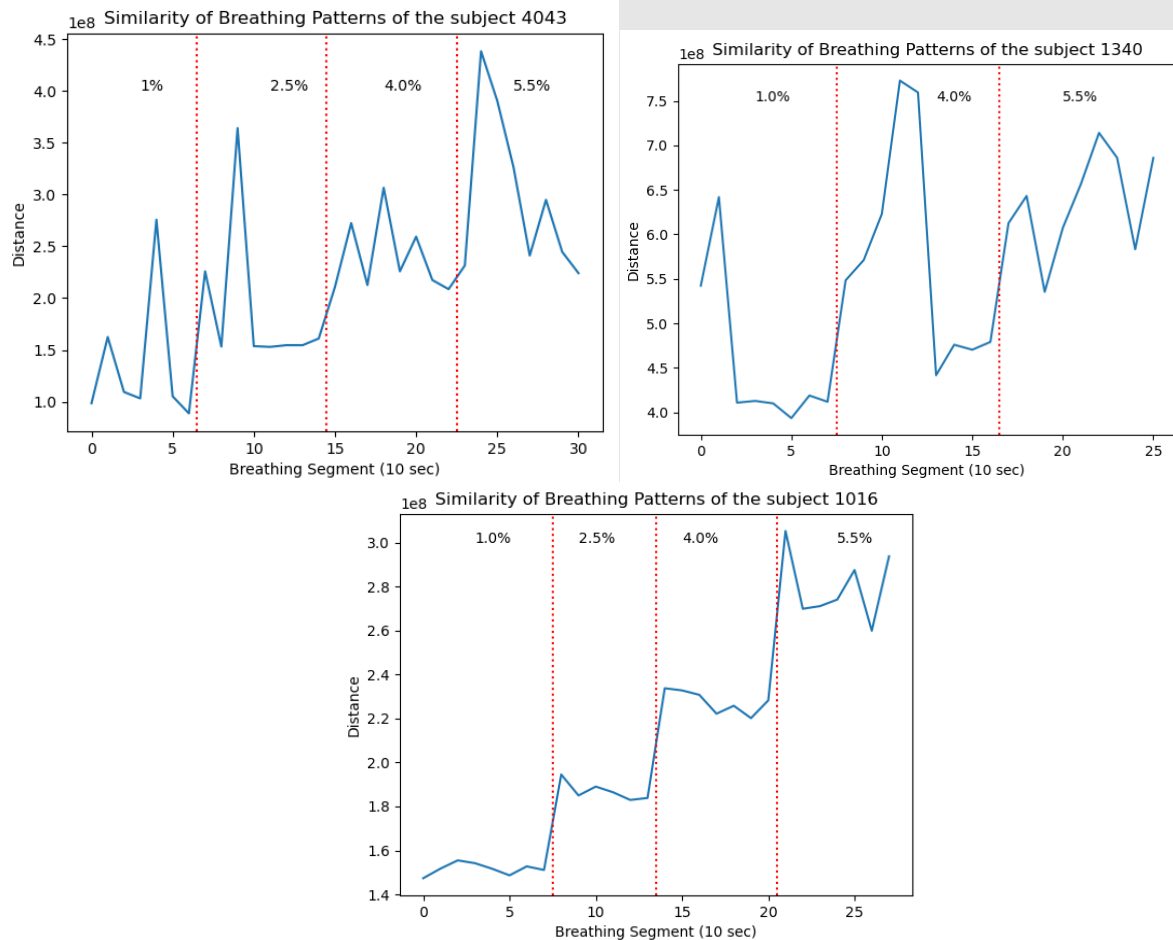


Figure 20. Similarities of breathing segments for 3 different participants in the unsupervised approach

6.0 DISCUSSION

The present study successfully developed and evaluated a suite of machine learning tools and algorithms to identify hypercapnic states through the analysis of operator acoustic features in speech and breathing using the standard microphone mounted inside the MBU-20/P flight mask. This proof-of-concept study could provide the foundation for establishing more robust and real-time methods for real-world monitoring of hypercapnia using existing microphones mounted in flight masks, dive masks and space suits.

The experimental manipulation of CO₂ concentration resulted in a steep increase in PaCO₂ during PAT tasks as participants were exposed to increasing levels of CO₂. On average participants experienced a 5.54 mmHg increase in the 1.0% condition, an 8.92 mmHg increase in the 2.5% condition, a 13.20 mmHg increase in the 4.0% condition, and a 14.46 mmHg increase in the 5.5% condition from their baseline PaCO₂. Likewise, there was also a strong positive trend of increasing

PaCO₂ across CO₂ exposure conditions during the reading task. On average participants experienced a 4.71 mmHg increase in the 1.0% condition, an 8.35 mmHg increase in the 2.5% condition, a 12.26 mmHg increase in the 4.0% condition, and a 16.29 mmHg increase in the 5.5% condition from their baseline PaCO₂. The participants experienced a slightly larger increase in PaCO₂ during the reading task when compared to the PAT task. The observed increases in PaCO₂ indicate that the experimental manipulation was adequate to evoke the variability of hypercapnia trends. Therefore, we were able to successfully test for performance differences across conditions and develop machine learning techniques to identify hypercapnia using speech and breath analysis.

Altering the CO₂ concentrations given to participants did not have a significant effect on average heart rate across conditions during reading and PAT tasks. There was an increase of 8.1 bpm between the 1.0% and the 5.5% CO₂ exposures during PAT tasks. Heart rate was higher during PAT tasks, on average, in the 1.0% condition than in the 2.5% and the 4.0% conditions. Elevated participant heart rate during the 1.0% exposure condition, which always occurred first, might have resulted from apprehension upon beginning the experiment and adapting to the breathing resistance imposed by the flight mask and BGoD3000. Following the 1.0% exposure condition during the PAT task heart rate trends began to rise with increasing CO₂ exposures as expected. During the reading task heart rate increased from 78.25 bpm in the 1.0% CO₂ exposure condition to 85.3 bpm in the 5.5% CO₂ exposure condition with a 6.98 bpm increase across conditions. The observed increase in heart rate across conditions was expected in association with the increase in inspired CO₂.

The pattern between breath rate and tidal volume observed in the 1.0% condition, a lower-than-average breath rate and a higher-than-average tidal volume, are consistent with observations taken in studies of breathing during deep sea diving where participants encounter increased ventilation resistance.^{39,40,41} In the current study breathing resistance was associated with exhalation resistance from the high flow rate of 70 liters per minute which makes it difficult for participants to exhale through the pressure balanced exhalation port of the MBU-20/P. This was done to simulate mask over pressurization events thought to occur in-flight due to significant regulator or mask valve malfunctions. Hypercarbia is known to increase breath rate, tidal volume, and minute ventilation, so it was expected that these variables would increase across the increasing CO₂ exposure conditions. Average breath rate increased by 13.34 breaths per minute from the 1.0% CO₂ condition to the 5.5% CO₂ condition. Average tidal volume increased by 1.314 liters from the 1.0% CO₂ condition to the 5.5% CO₂ condition. Minute ventilation increased an average of 7.02 liters per minute. The increasing breath rate and volume observed indicate that the experimental manipulation used here resulted in significant increases consistent with an increasing hypercarbia state due to higher CO₂ exposures.

The CO₂ exposures of 1.0%, 2.5%, 4.0%, and 5.5% used in the study did not negatively affect the participants' average PAT performance. The stable effect on performance during low CO₂ exposures has been cited in previous studies^{42,43,44,45} suggesting that a threshold exists for CO₂ related cognitive effects. Upon analysis, PAT performance showed improvement in association with higher CO₂ exposures. When PAT Tracking was analyzed without the other subtasks, performance showed a significant increase in higher CO₂ conditions when compared to the 1.0% condition. Pairwise comparisons revealed that performance in the 2.5% and 4% conditions was significantly better than performance during the 1.0% condition. This pattern of results suggests

that the stress evoked by the CO₂ exposures used in this study resulted in a slight improvement in performance as would be predicted by the Yerkes-Dodson curve when the magnitude of stress is on the right side of the peak.⁴⁶ If participants had been exposed to higher levels of CO₂ ($\geq 7.0\%$) performance would have been significantly degraded.

The speech analysis completed using machine learning provides a strong divergence measure to identify CO₂ exposure conditions when the model is trained on the user's own speech patterns $r = .89$, and a strong generalized measure for a new user $r = .69$, when the model is trained on speech patterns from other users. The generalized model resulted in good area under the receiver operating characteristic (AUC) values in the 2.5% and 4.0% CO₂ exposure conditions and excellent AUC values for correctly categorizing the 1.0% and the 5.5% CO₂ exposure conditions Figure 11.

Performance in the classification task was reduced in comparison to the regression approach, as edge cases may be counted as false positives, and there was no distinction made between a close classification (e.g., 5.5% vs. a true 4.5%) and a high error classification (e.g., 1.0% vs. a true 5.5%). We also noted the expected result that conditions on either end of the spectrum were the easiest to classify when compared to the middle two conditions (2.0% and 4.5%). We did not have sufficient data to perform the divergence or combined evaluation using this classification method, as this method requires the data available typically used for testing to be allocated to training data. In future work, this could be addressed by changing the size of the splits of data between training and test.

Even with a large dataset, occasional situations can arise where the model would fail to predict an individual's speech dynamics. With a limited amount of training data, however, the approach can quickly calibrate itself to an individual's speech patterns to achieve very good performance. Finally, we show that a combined training method can provide excellent generalized results while incorporating individual speech qualities in cases where there may be little opportunity to collect additional training data. This study demonstrates that the machine learning approach developed in this effort to identify hypercarbia through the analysis of operator speech is a viable option for unobtrusive monitoring of operator physiologic status.

The expert user breath analysis employing both unsupervised and supervised approaches, trained on each individual participant's features and evaluated on the same participant's holdout data, demonstrated excellent predictability and diagnostic capabilities. The unsupervised approach demonstrated the most predictive ability when compared to the ground truth with a significant correlation coefficient of 84.5. When one outlier was removed (participant number 1340) the average accuracy of the model improved to 85.91%. This approach also resulted in a high AUC for a diagnostic instrument. The unsupervised approach is trained for each participant separately, low accuracy for a participant, e.g., 1340, indicates that the participant did not have consistent breathing patterns and breathing techniques when they became hypercapnic (Figure 13). When we use the known CO₂ levels, 1.0%, 2.5%, 4.0%, and 5.5%, of a participant for the training model, the model learns the participant breathing patterns and techniques at these CO₂ exposure conditions and can identify and classify these patterns into different CO₂ exposure levels. However, if the participant uses different breathing patterns during the same CO₂ exposure level (e.g., at 1.0%) the ability of the model to discriminate the breathing patterns for that same CO₂ exposure level decreases. For example, if the participant holds his or her breath for 30 seconds when exposed to

5.5% CO₂, and then the participant begins to breath, the algorithm sees two different patterns (not breathing and breathing in 5.5% exposure). For example, Figure 20 shows the similarities of breathing patterns across different segments for different CO₂ levels of 3 participants (1340, 4043, 1016). The similarity of breathing patterns of the participant 1016 is very well separated, but not the same for the other two participants. Thus, if we train the model on the breathing pattern of the same participant exposed 5.5%, but the person holds their breath, the algorithm will not accurately identify the CO₂ level (because the person is not breathing). The unsupervised expert user approach, trained on each individual's personal breathing features, resulted in excellent AUC results supporting its accuracy as a diagnostic tool to discriminate between CO₂ exposure conditions Figure 12 shows the ROC of the experiment, and Table 8 shows the AUC for each CO₂ level.

The expert user supervised approach, which also trained on each participant's individual data, was associated with the second most predictive ability when compared to ground truth with a significant correlation coefficient of 78.32 which improved to 83.10 when three outliers were removed. Like the unsupervised approach, participants with low accuracy using this approach have similar breathing patterns during different CO₂ exposure conditions. For example, if the participant holds their breath during 1.0% and 4.0% CO₂ exposure conditions, the unsupervised algorithm will put the samples from 1.0% and 4.0% into one cluster, lowering the accuracy. Such issues would eventually be reconciled as the model incorporated more data. The supervised approach also demonstrated excellent AUC values demonstrating its accuracy as a diagnostic instrument. Figure 14 shows the ROC of the experiment, and Table 9 shows the AUC for each CO₂ level.

The K-folding approach that trained a model using average data to predict the ground truth of a new user not included in the training data yielded the lowest predictive power with an average correlation of 51.65. It did, however, result in good to excellent AUC values suggesting that it too may hold some promise as a diagnostic tool Figure 16. Applying a trained model to a new participant is the most challenging task because the trained model is not tailored to accommodate the new participant's individual features. This is a well-known issue in machine learning and data analytics and exists across different application domains such as recommendation systems like Netflix. As we expected, the overall accuracy of the system is reduced because the trained model struggles to detect CO₂ levels of a new participant with a new breathing pattern and strategy. This issue will be resolved as more training samples from different individuals are collected. More advanced analytical approaches will also increase the accuracy of the model. However, these advanced methods require a lot of training samples. Figure 17 shows the accuracy of the CO₂ level detection and classification for each participant. In this figure, the accuracy reported for each participant is the accuracy of the test on that participant when the model is trained on the remaining participants. Participants with low test accuracy (e.g., 4143) had significantly different breathing patterns and techniques while under the CO₂ exposure conditions when compared to the other participants. When a new user has a breathing pattern similar to the other participants, the trained model can identify the correct CO₂ level breathing patterns, resulting in higher accuracy. The K-folding approach may be useful to provide predictions based on average users while the expert user approach models each individual's feature set.

7.0 CONCLUSION

The current study demonstrates that analysis of speech and breath features through a machine learning approach shows great promise as a tool to diagnose hypercarbia using the standard microphone mounted in the MBU-20/P. This was especially true for distinguishing between the lowest (1.0%) and highest (5.5%) exposure levels. The current results provide a proof-of concept to justify further research and development of machine learning approaches of acoustic features of speech and breath to detect significant changes in operator state due to CO₂. The ability to monitor sudden changes in physiological status through analysis of features present in speech and breath would have far reaching implications beyond hypercapnia detection. This technique could easily be adapted to identify and mitigate many other operational stressors present in tactical aviation such as, hypocapnia, hypoxia, and breathing resistance associated with malfunctioning life support equipment. For example, recent studies have identified multiple issues with the standard aviation mask most used by the US military, the MBU-20/P. The current approach could be used to identify auditory features that are associated with fluttering valves mask leaks, and exhalation resistance. The approach may also provide an indication of regulator function that would be useful to diagnose and root causes for in-flight physiological episodes. More studies should be conducted to further refine and validate these techniques on additional research participants

The real-time analysis of breath and speech and other sounds that can be captured through communication equipment already present in the MBU-20/P represents a significant addition to the suite of tools being developed to track operator status in-cockpit across military tactical aviation. Systems could be calibrated to each operator during aviation training exposures like hypoxia familiarization and then continue to follow the individual through primary flight training into their operational squadrons to develop expert user models based on each operator's individual features.

The machine learning tools developed through this effort also hold promise for real-time monitoring in extreme environments where individuals are breathing on artificial life support systems. An example of an artificial life support system is the self-contained breathing apparatus used by divers, first responders, and astronauts. Machine learning could provide the ability to track physiological status of these operators and the function of the life support system in-real-time during real world scenarios using standard communication systems. The current approach may also be useful for tracking patient status during en route care operations or in other emergency medicine settings when patients are receiving respiratory therapy.

8.0 REFERENCES

1. Elliot, J.J. & Schmitt, D.R. (2019). Unexplained physiological episodes: A pilot's perspective. *Air & Space Power Journal*. 1-18.
2. Chiang, K.T., Tu, M.Y., Cheng, C.C., Chen, H.H., Huang, W.W., Chiu, Y.L., Wang, Y.Y., et al. (2021). Contributions of hypoxia-awareness training to the familiarization of personal symptoms for occupational safety in the flight environment. *International Journal of Environmental Research and Public Health*, 18(6), 2904.
3. Carey, B. (2017). U.S. military tackles vexing issue of physiologic episodes. *Aviation International News*. <https://www.ainonline.com/october-2017>.
4. Leinonen, A., Varis, N., Kokki, H., & Leino, T. K. (2021). Normobaric hypoxia training in military aviation and subsequent hypoxia symptom recognition. *Ergonomics*, 64(4), 545–552.
5. Phillips, J.B., Geyer, D.J., McEwen, K.A., Olson, T.M., & Lee, R.L. (2018). Addressing hypoxia-like physiologic episodes in modern tactical aviation: A review of research from 2011 to 2016. Naval Medical Research Unit Technical Report Number 17-143. 1-42.
6. Rice, G.M., Snider, D., Drollinger, S., Greil, C., Bogni, F., Phillips, J., Raj, A., Marco, K., & Linnville, S. (2019). Dry-EEG manifestations of acute and insidious hypoxia during simulated flight. *Aerospace Medicine and Human Performance*, 90(2), 92-100.
7. Gill, M., Natoli, M. J., Vacchiano, C., MacLeod, D. B., Ikeda, K., Qin, M. Pollock, N. W., Moon, R. E., Pieper, C., & Vann, R. D. (2014). Effects of elevated oxygen and carbon dioxide partial pressures on respiratory function and cognitive performance. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, 406-412.
8. Goldberg, S., Olila, H. M., Lin, L., Sharifi, H., Rico, T., Andlauer, O., Aran, A. Bloomrosen, E., Faraco, J., Fang, H., & Mignot, E. (2017). Analysis of hypoxic and hypercapnic ventilatory response in healthy volunteers. *PloS One*, 12(1), e0168930.
9. Bhatia, A., Dorr, B.J., Hollingshead, K., Phillips, S.L. and McKenzie, B. (2017a). Characterization of divergence in impaired speech of ALS patients. *The Proceedings of ACL BioNLP 2017 Workshop*, 149-158.
10. Bhatia, A., Hollingshead, K., Dorr, B.J., Perera, I., McKenzie, B, and Phillips, S.L. (2017b). Acoustic analysis of read speech for detection and monitoring of ALS. *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference*, 776.
11. Norel R., Pietrowicz M., Agurto C., Rishoni S., Cecchi G. (2018). Detection of amyotrophic lateral sclerosis (ALS) via acoustic analysis. *bioRxiv [preprint]*. doi: 10.21437/Interspeech.2018-2389
12. McHenry M. (2000). Acoustic characteristics of voice after severe traumatic brain injury. *The Laryngoscope*, 110(7), 1157–1161. <https://doi.org/10.1097/00005537-200007000-00017>
13. Kirchhübel, C., & Howard, D. M. (2013). Detecting suspicious behaviour using speech: acoustic correlates of deceptive speech -- an exploratory investigation. *Applied ergonomics*, 44(5), 694–702. <https://doi.org/10.1016/j.apergo.2012.04.016>

14. Tavi, L. (2018). Classifying females' stressed and neutral voices using acoustic-phonetic analysis of vowels: an exploratory investigation with emergency calls. *Int J Speech Technol* 22, 511–520 (2019). <https://doi.org/10.1007/s10772-018-09574-6>
15. Raj, A. K., Roberts, B., Hollingshead, K., McDonald, N., Poquette, M. & Soussou, W. (2018, July) A Wearable Multisensory, Multiagent Approach for Detection and Mitigation of Acute Cognitive Strain: Phase I- Vocalization analysis. In *Proceedings of HCI International 2018*. Las Vegas, NV (pp. 180-200).
16. Lee, C. M., Narayanan, S. S., & Pieraccini, R., (2002). Classifying emotions in human-machine spoken dialogs. *Proceedings. IEEE International Conference on Multimedia and Expo*, pp. 737-740 vol.1, doi: 10.1109/ICME.2002.1035887.
17. Petrushin, V.A. (1999). Emotion in speech recognition and application to call centers. In: *Proc. Artificial Neural Networks in Engineering (ANNIE '99)*, Vol. 1, pp. 7–10
18. Alghowinem, S., Goecke, R., Wagner, M., and Epps, J. (2013). A comparative of different classifiers for detecting depression from spontaneous speech. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, 8022-8026.
19. Cannizzaro M., Harel B., Reilly N., Chappell P., Snyder P.J. (2004). Voice acoustical measurement of the severity of major depression. *Brain and Cognition* 56, 30-35.
20. Höning, D., Hansen-Goos, H., Airo, A., & Spohn, T. (2014). Biotic vs. abiotic earth: a model for mantle hydration and continental coverage. *Planetary and Space Science*, 98, 5–13. <https://doi.org/10.1016/j.pss.2013.10.004>
21. Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological psychiatry*, 72:7, 580–587. doi:10.1016/j.biopsych.2012.03.015.
22. Scherer, S., Stratou, G., Gratch, J., & Morency, L. P. (2013). Investigating voice quality as a speaker-independent indicator of depression and PTSD. *Proc. INTERSPEECH*, Lyon, France, 847-851.
23. Stassen, H. H., Kuny, S., & Hell, D. (1998). The speech analysis approach to determining onset of improvement under antidepressants. *European Neuropsychopharmacology* 8, 303-310.
24. Boersma, P., van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glott International*, 5, pp. 341-347.
25. Eyben, F., Wöllmer, M., Schuller, B. (2010). openSMILE – the Munich versatile and fast open-source audio feature Extractor. In: *Proc. ACM Multimedia*, Florence, Italy, pp. 1459–1462.
26. Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K., & Geralts, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of neurolinguistics*, 20(1), 50–64. <https://doi.org/10.1016/j.jneuroling.2006.04.001>
27. Gross, V., Dittmar, A., Penzel, T., Schüttler, F., & von Wichert, P. (2000). The relationship between normal lung sounds, age, and gender. *American journal of respiratory and critical care medicine*, 162(3 Pt 1), 905–909. <https://doi.org/10.1164/ajrccm.162.3.9905104>
28. Reichert, S., Gass, R., Brandt, C., & Andrès, E. (2008). Analysis of respiratory sounds: state of the art. *Clinical medicine. Circulatory, respiratory and pulmonary medicine*, 2, 45–58. <https://doi.org/10.4137/ccrpm.s530>

29. Murphy, J. M., Sexton, D. M., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., & Stainforth, D. A. (2004). Quantification of modeling uncertainties in a large ensemble of climate change simulations. *Nature*, 430(7001), 768–772. <https://doi.org/10.1038/nature02771>
30. Phillips, J.B. (2019). Evaluation of the reliability and criterion validity of the performance assessment tool. Report to Sponsor (Air Force Research Laboratory).
31. Johnson, M. J., Kulkarni, S. P., Raj, A. K., Carff, R. W. & Bradshaw, J. M. (2005). AMI: An Adaptive Multi-Agent Framework for Augmented Cognition. In: *Proceedings of the 1st International Conference on Augmented Cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.
32. Van Riper, C. (1963). *Speech correction: Principles and Methods* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall
33. Patel, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., Ramage, L., Tyler, E., Russell, S. (2013). “The Caterpillar”: A novel reading passage for assessment of motor speech disorders. *American Journal of Speech-Language Pathology* 22, 1-9.
34. Fairbanks, G. (1940). *Voice and Articulation Drillbook* (2nd ed.). New York: Harper & Row.
35. Schuller, B., Steidl, S., Batliner, A., Schiel, F., Krajewski, J., Weninger, F., & Eyben, F. (2014). Medium-term speaker states—A review on intoxication, sleepiness and the first challenge. *Computer Speech & Language*, 28(2), 346-374.
36. Montiel, J., Read, J., Bifet, A., & Abdessalem, T. (2018). Scikit-multiflow: A multi-output streaming framework. *The Journal of Machine Learning Research*, 19(1), 2915-2914
37. Krajewski, J., Schnieder, S., Monschau, C., Titt, R., & Sommer, D. (2016). Large sleepy reading corpus (LSRC): applying read and speech for detecting sleepiness. *Speech Communication*. 12. ITG Symposium, pp. 1–4.
38. Schiel, F., Heinrich, Ch., Barfüßer, S., & Gilg, T. (2008). ALCAIcohol language corpus, In *Proceedings of Language Resources and Evaluation Conference*, Marrakesh, 1-5.
39. Anthony, TG. & Mitchell SJ. (2016). Respiratory physiology of rebreather diving. Pollock NW, Sellers SH, Godfrey JM, eds. *Rebreathers and Scientific Diving*. Proceedings of NPS/NOAA/DAN/AAUS June 16-19, 2015 Workshop. Wrigley Marine Science Center, Catalina Island, CA; p. 66-79.
40. Moon, R. E., Cherry, A. D., Stolp, B. W., & Camporesi, E. M. (2009). Pulmonary gas exchange in diving. *Journal of applied physiology (Bethesda, MD.: 1985)*, 106(2), 668–677. <https://doi.org/10.1152/japplphysiol.91104.200>
41. Segadal, K., Gulsvik, A., & Nicolaysen, G. (1990). Respiratory changes with deep diving. *The European respiratory journal*, 3(1), 101–108.
42. Dunworth, S. A., Natoli, M. J., Cooter, M., Cherry, A. D., Peachner, D. F., Potter, J. F., Wester, T. E., Freiburger, J. J., & Moon, R. E. (2017). Hypercapnia in diving: a review of CO₂ retention in submersed exercise at depth. *Undersea & hyperbaric medicine: journal of the Undersea and Hyperbaric Medical Society, Inc*, 44(3), 191–209. <https://doi.org/10.22462/5.6.2017.1>
43. Freiburger, J. J., Derrick, B. J., Natoli, M. J., Akushevich, I., Schinazi, E. A., Parker, C., Stolp, B. W., Bennett, P. B., Vann, R. D., Dunworth, S. A., & Moon, R. E. (2016). Assessment of the interaction of hyperbaric N₂, CO₂, and O₂ on psychomotor

- performance in divers. *J. Appl. Physiol.* 121, 953–964. doi: 10.1152/japplphysiol.00534.2016
44. Inouye, K.G., Nagaya, T., & Shigematsu, Y. (1955). Effect of high carbon dioxide concentration on mental performance and some sensory functions. *Jap J Physiol.* 5:109-121.
 45. Sayers, J. A., Smith, R. E., Holland, R. L., & Keatinge, W. R. (1987). Effects of carbon dioxide on mental performance. *Journal of applied physiology* (Bethesda, Md.: 1985), 63(1), 25–30. <https://doi.org/10.1152/jappl.1987.63.1.25>
 46. Yerkes, R.M., & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.

LIST OF ABBREVIATIONS AND ACRONYMS

%: percent
AFRL: Air Force Research Laboratory
ALS: Amyotrophic Lateral Sclerosis
AMI: Adaptive Multi-agent Integration
ANOVA: Analysis of Variance
AUC: Area Under the Curve
AUROC: Area Under the Receiver Operating Characteristic
BGoD3000: Blended Gas on Demand
bpm: beats per minute
CO₂: Carbon Dioxide
COPD: Chronic Obstructive Pulmonary Disease
EA-6Bs: Prowler
F-22: Raptor
F-35: Lightning
F/A-18: Hornet
FDA: Food and Drug Association
HAMD: Hamilton Depression Rating Scale
Hz: Hertz
IHMC: Institute for Human and Machine Cognition
IRB: Institutional Review Board
MATSG-21: Marine Aviation Training Support Group-21
MCU: microcontroller unit
MFCC: Mel-frequency cepstral coefficient
PAT: Performance Assessment Tool
PaCO₂: Partial Pressure of Carbon Dioxide
MBU-20/P: Combat Edge Oxygen Mask
mmHg: millimeters of mercury
O₂: Oxygen
ROC: Receiver Operating Characteristics
RMS: Root Mean Square
SpO₂: Oxygen Saturation
T-45: Boeing T-45 Goshawk two-seat Naval jet training aircraft
T-6: Beechcraft T-6 Texan II two-seat Naval turboprop training aircraft
TBI: Traumatic Brain Injury
TCM: Transcutaneous Monitor
UPEs: Unexplained Physiologic Episodes